Proof of concept of CRISPR/Cas9 to treat corneal dystrophies

Kathleen A. Christie

BSc (Hons) Biomedical Genetics

Research conducted within

Genomic Medicine Research Group

School of Biomedical Sciences

Faculty of Life and Health Sciences

Ulster University

A thesis submitted for the degree of

Doctor of Philosophy

October 2018

I confirm that the word count of this thesis is less than 100,000 word

TABLE OF CONTENTS

Declaration

I hereby declare that with effect from the date on which the thesis is deposited in the Research Office of the University of Ulster, I permit

1. The librarian of the University to allow the thesis to be copied in whole or in part without reference to me on the understanding that such authority applies to the provision of single copies made for study purposes or for inclusion within the stock of another library.
2. The thesis to be made available through the Ulster Institutional Repository and/or EThOS under the terms of the Ulster e Theses Deposit Agreement which I have signed

This thesis is the sole work of the author and has not been submitted for any previous application for a higher degree.

Kathleen Christie

Acknowledgements

First and foremost I would like to thank my supervisors Tara and Andrew, without your continued support and advice this thesis and all of the fantastic experiences I've had throughout my PhD would not have been possible. Tara, thank you for your unwavering belief in me and for challenging me to aim for the top. Andrew, you have taught me so much, thank you for being such an exceptional mentor. I would like to thank my external supervisor David, for teaching me the fundamentals of PhD life and for only being a skype call away. Thank you to everyone in Avellino, with a special mention to Larry, Connie and Kevin, for all of your support over the past few years. I would like to thank my surgical nurse Colin for all of your advice and your aptitude for creating nicknames. I would like thank everyone in my group, past and present, with a special mention to David, Eleanora, Davide, Paulina, Marie, Laura, Soma, Gideon, and Sarah. Thanks for the laughs and the mutual support, I couldn't have done it without you all. Thank you to my office buddy Laura, we have shared a lot and have many great memories. Thanks to Kevin, my conference bestie, for your constant encouragement and advice. Thanks to Colum for your support, advice and encouragement over the past three years. Thanks to Declan, for your reassurance and guidance throughout this overwhelming process. I would like to thank absolutely everyone in the CMB for making it such a friendly place to be, I will miss you all. A very special mention to Caroline, Stores (Rob, Lisa and Davey), Robert Rainey, Sharon, Nigel, Rhonda, Glenn and James in the BBRU, Keith, Danny and everyone that has helped me in any way throughout the last three years. Thank you for putting up with me and for everything you have done for me, I really appreciate it. Thank you to Rachel for your amazing ability to turn a scribble into a publication worthy image. A big thank you to Barry for being my thesis guru and keeping me motivated. Thanks to all the Cromlech gals (Mother Mary, Laura, Colleen and Emer) for all the parties and the lols. Thanks to the Cromlech yogis (Darren and Emma) for attempting to keep me calm. Thanks to everyone at the gym, with a special mention to Helen, Adrian and Hugh (my personal gluten free baker), circuits is the best therapy and I will miss all the laughs…life motto – only give 100%. To Zoga, for your constant positivity and encouragement. To SJ for your witty banter and for looking after me. To my post-doc Rachelle, thanks for always being there for me and remember 'life can always get worse'. Thanks to all my fabulous Newcastle pals, I definitely couldn't have survived without our bi-annual reunions. With a special mention to Eleanor, thank you for always supporting me, for our eventful Northern Ireland reunions and for giving me full custody of Simba. To George, for your abstract ideas for my project, unfortunately they didn't

make the final thesis. To Shazza Bee and all the Foyle girls, with a special mention to AC, Hev and Holly, for always checking in and reminding me there is life outside the lab. To all my eccentric and colourful family who have unconditionally supported me throughout this journey, with a special mention to Auntie Kate, Auntie Mary and Uncle Ben (Rosie has played an integral part!). To Bengelina, for providing constant entertainment in the story of 'What Ben did next'. To Marco/Jan, for being an inspiration and teaching me that the only way to succeed is to never give up #supersonic. Finally, a massive and whole-hearted thank you to my Mum & Dad. Dad, you have sacrificed so much to give me every opportunity possible, and I am so grateful. Gnome, you have always believed I was capable of anything and have supported me through every high and low, I love you.

A bit of madness is key,

To give us new colours to see,

Who knows where it might lead us?

And that's why they need us.

Abstract

Introduction

The corneal dystrophies are a group of inherited disorders that affect the shape or transparency of the cornea. The unique qualities of the cornea coupled with their mendelian traits make the them an ideal target for gene therapy. The main aim of this thesis is to investigate the potential to use CRISPR/Cas9 to treat the autosomal dominant *TGFBI* corneal dystrophies.

Methods

Paper II: Plasmid expressing CRISPR/Cas9 targeted to *Luc2* was delivered to the cornea via an intrastromal pressure injection. Knockdown of luciferase was measured using an IVIS *in vivo* imager. The allele-specificity of CRISPR/Cas9 was assessed *in vitro* using a dual-luciferase assay and an *in vitro* digestion.
Paper III: A morpholino targeted to *tgfbi* was electroporated into the regenerating adult zebrafish tail fin, and the effect of Tgfbi knockdown was measured by the quantification of regenerated tissue. 10X phased sequencing and EBV-transformation of isolated lymphocytes was performed on a patient harbouring a R124H granular corneal dystrophy type II (GCD2) mutation. The phasing information was used to design allele-specific sgRNAs which were then tested using an *in vitro* digest and targeted resequencing across the target locus in *TGFBI*. qPCR was used to infer the efficiency of a dual-cut using a combination of sgRNAs.
Paper IV: The transduction efficiency of 3 AAV-GFP serotypes in immortalised human corneal cells was assessed using flow cytometry. *In vivo* transduction of AAV-GFP following intracameral injection was assessed using the IVIS *in vivo* imager and fluorescent microscopy. AAV-Cas9 and AAV-sgRNA-GFP were co-injected, DNA was extracted from the whole cornea and TIDE analysis was performed to determine efficiency of indels.

Results

Paper II: The failure of a mutation-dependent approach to target 20% of *TGFBI* missense mutations was demonstrated. Comparison of two widely used allele-specific strategies revealed a PAM-specific approach conferred superior specificity than that of a guide-specific approach. The inability of *S.pyogenes* Cas9 to distinguish between single base pair changes in the guide sequence was confirmed.
Paper III: Knockdown of Tgfbi in the regenerating zebrafish tail fin was shown to impair wound healing. The ability to selectively target the mutant allele by means of non-disease-causing SNPs, which are associated with a PAM on the same allele as the disease-causing mutation, was demonstrated. The addition of a 50:50 ssODN with the ribonucleoprotein complex was shown to significantly increases the frequency of a dual-cut event.
Paper IV: AAV-2/9 was shown to transduce all corneal layers *in vivo* following a single intracameral injection. A dual-AAV-2/9 CRISPR/Cas9 system was shown to generate 25.7% within the whole cornea.

Conclusion

Cas9 lacks the specificity to discriminate between single base pair mismatches within the 20bp guide sequence, however mutations within the 2bp PAM are much less tolerated. As such, using currently available nucleases a PAM-specific approach is necessary to discriminate between alleles. AAV was demonstrated as a robust vehicle to deliver gene-editing reagents to the corneal layers.

Abbreviations

AAV	Adeno associated virus
*As*Cas12a	*Acidaminococcus* Cas12a
ASNIP	Allele-specific, SNP-derived PAM, *in cis*, personalised
Cas12a	CRISPR associated protein 12 a
Cas9	CRISPR associated protein 9
CHB	Hans Chinese in Beijing, China
COL8	Type VIII collagen
COL8A1	Collagen type VIII isoform alpha 1
COL8A2	Collagen type VIII isoform alpha 2
CRISPR	Clustered regularly interspersed palindromic repeats
CROPT	cysteine-rich domain of periostin and TGFBIp
crRNA	CRISPR-RNA
ddNTP	Dideoxynucleotides
DNA	Deoxyribonucleic acid
dpa	Days post amputation
DSB	Double-strand break
dsODN	double-stranded oligonucleotide
EAS	East Asian
ECM	Extracellular matrix
FA	Fanconi anaemia
FECD	Fuchs' endothelial corneal dystrophy
GCD2	Granular corneal dystrophy type II
GWAS	Genome wide association study
HDR	Homology directed repair
HR	Homologous recombination
HTT	Huntington
IC3D	International Classification of Corneal Dystrophies
indels	Insertions or deletions
JPT	Japanese in Tokyo, Japan
LASIK	Laser assisted in-situ keratomileusis
*Lb*Cas12a	*Lachnospiraceae bacterium* Cas12a
LCD1	Lattice corneal dystrophy classic
LCL	Lymphocyte cell line
LC-MS/MS	Liquid chromatography mass-spectrometry
LD	Linkage disequilibrium
LESCs	Limbal epithelial stem cells
MAF	Minor allele frequency
MUT	mutant
NHEJ	Non-homologous end-joining
NLS	Nuclear localisation signals
NMD	Nonsense mediated decay

NUC	Nuclease
PAM	Protospacer adjacent motif
PBMCs	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PI	PAM-interacting
pre-crRNA	precursor-CRISPR-RNA
RBCD	Reis-Bücklers corneal dystrophy
REC	recognition
REP	reporter
RGD	integrin-recognition domain
RNP	Ribonucleoprotein
S.aureus	*Staphylococcus aureus*
S.pyogenes	*Streptococcus pyogenes*
S.thermophilus	*Streptococcus thermophilus*
sgRNA	single-guide RNA
siRNA	small-interfering RNA
SNPs	Single nucleotide polymorphisms
ssODN	single-stranded oligonucleotide
SSTR	single-stranded template repair
TAC	Transient amplifying cells
TBCD	Thiel-Behnke corneal dystrophy
TCF4	Transcription factor 4
TGFB	Transforming growth factor β
TGFBI	Transforming growth factor β-induced
TGFBIp	Transforming growth factor β-induced protein
tracrRNA	trans-activating CRISPR RNA
UTRs	Untranslated regions
UV	Ultraviolet
WGS	Whole genome sequencing
WT	Wild-type

General Introduction

1 General Introduction

The prospect of genome engineering as a therapeutic for genetic disease has finally come to light. This new era is possible due to over a century of research in which genes where discovered as units of hereditary, the helical structure of DNA was revealed, the central dogma of life was unraveled and the age of recombinant DNA was born. As the link between genetic variation and disease became apparent, so too did the possibility of curing disease by correcting or eliminating the genetic cause. The emergence of gene editing tools has enabled the progression of personalised medicine to the clinic and the cornea offers the ideal tissue to pioneer these novel treatments.

1.1 Cornea

The cornea is an avascular, transparent tissue found in the anterior segment of the eye. The main functions of the cornea are to act as a structural barrier to the outside world and to provide the majority of the eye's refractive power[1]. The first barrier that exists between the outside world and the internal structures of the eye is the tear-film, which coats the cornea. The cornea itself can be subdivided into 3 distinct regions; an outer anterior region which comprises of multiple layers of proliferating epithelial cells built upon a basement membrane, a middle region which is a connective tissue stroma which contributes to 90% of the cornea's thickness, composed of collagen lamellae interspersed with keratocytes and lastly a posterior region which contains a monolayer of endothelial cells attached to Descement's membrane (Figure 1a)[2]. Surrounding the cornea is the limbus which is where the limbal epithelial stem cells (LESCs) reside, within the limbal region the basement membrane disappears and Pallisades of Vogt form. The limbus partitions the avascular cornea from the vascular conjunctiva. (Figure 1b)

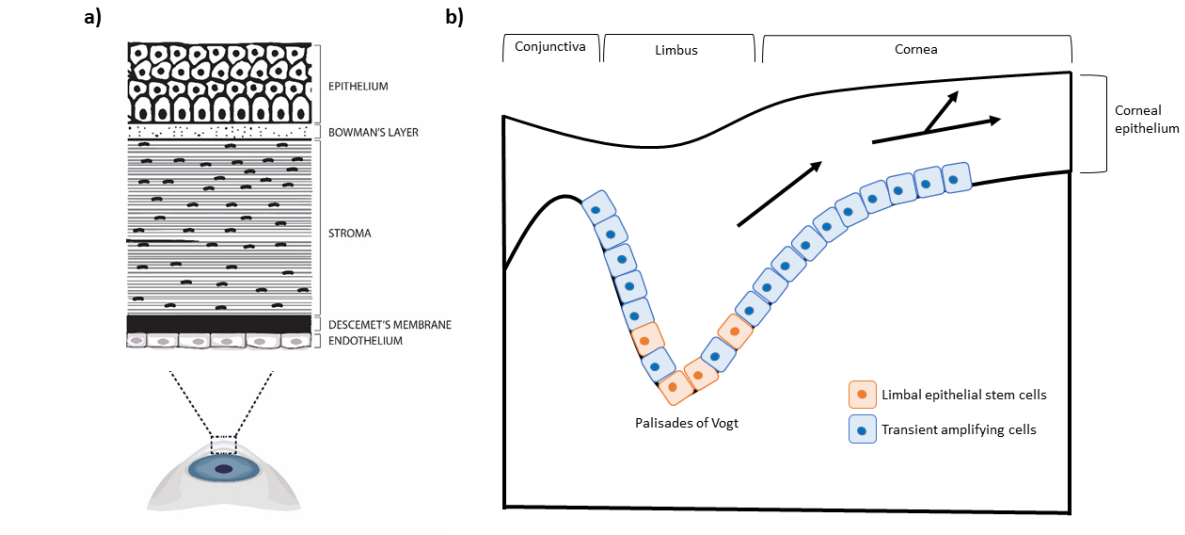


Figure 1: a) Illustration to show the 5 different layers of the cornea; the epithelium, Bowman's layer, stroma, Descemet's membrane and endothelium. Each plays in intergral part in maintaining the structure and transparency of the cornea. **b)** Schematic of the Palisades of Vogt in which the LESCs are resident. They are capable of self-renewal to produce new LESCs and also divide asymmetrically to produce TACs that migrate centripetally and repopulate the corneal epithelium.

The corneal epithelium is a non-keratinising stratified squamous epithelium and it functions as a physical barrier preventing foreign bodies from entering the eye. It an approximately 7 cell layer thick structure consisting of 3 layers of superficial flat polygonal cells which form intracellular tight junctions and are covered in microvilli that allow close interaction with the tear film, 3 layers of suprabasal wing cells whose cell membranes are linked with numerous desmosomes and a basal columnar cell layer which links to the basement membrane via hemi desmosomes[2]. The corneal epithelium is continually turned over every 1-2 weeks as deeper epithelial cells replace the superficial epithelial cells[3]. The LESCs are important in this process, LESCs are found the Palisades of Vogt in the basal region of the limbus at the corneoscleral junction. LESCs are capable of dividing symmetrically to self-renew and asymmetrically to produce daughter transit amplifying cells (TAC) that migrate centripetally to populate the basal layer of the corneal epithelium (Figure 1b)[4,5].

The basement membrane, known as Bowman's layer, is made up of randomly orientated collagen fibrils, it provides structure and strength to the cornea and allows epithelial attachment to the stroma[6]. The stroma consists of regularly arranged bundles of collagen with keratocytes located between them.[2] The collagen fibrils are made up of collagen type I and type V and are formed into very uniform lamellae; the uniformity of this arrangement is crucial to corneal function, the organisation allows light transmission and maintains cornea curvature and strength[7]. The collagen lamellae are uniformly spaced from each other due to interactions with proteoglycans[8]. Keratocytes are sparsely populated between these collagen lamellae, they produce crystallins which contribute to corneal transparency[9]. Any disturbance of this highly ordered layer can potentially lead to corneal opacity.

Descemet's membrane acts as a specialised basement membrane for the corneal endothelium, it provides an anchor for corneal endothelial cells, in addition to assisting in the corneal

endothelium function[10]. The corneal endothelium is a monolayer of cells that play an integral role in the hydration of the cornea[11]. They are considered to have a "leaky" barrier, allowing passage of nutrients from the aqueous humour into the corneal stroma; contrary to this the endothelium then simultaneously uses Na+/K+ ATPase pumps to remove excess fluid in the stroma to maintain corneal transparency. Corneal endothelial cells decrease with age, injury, trauma or disease, such as Fuchs' endothelial corneal dystrophy (FECD)[12]. They have a very low proliferative capacity and as the number of cells progressively decrease the remaining cells will spread and enlarge to compensate. As more cells are lost the remote barrier function in addition to the pump function of the cornea endothelium is affected, resulting in the excess fluid from the aqueous humour not being adequately removed causing a loss in corneal transparency.

1.2 Corneal dystrophies

Corneal dystrophies are a group of inherited, heterogeneous, bilateral disorders that affect the transparency or shape of the cornea[13]. To-date there are currently 23 characterised corneal dystrophies (Table 1), associated with different corneal layers[14,15]. The genetics behind corneal dystrophies is well documented; the inheritance pattern, gene locus and causative gene is known in ten of these characterised corneal dystrophies, with an additional four partially understood. Current treatment strategies are extremely limited, at present the only curative measure that can be taken is a partial or complete corneal transplant, however this is associated with its own risks such as donor shortage, graft rejection and recurrence of disease[16].

1.2.1 *TGFBI* corneal dystrophies

As described corneal dystrophies are a group of diseases that affect the transparency or shape of the cornea. Missense mutations within *TGFBI* result in the accumulation of mutant TGFBI protein within the cornea that impair its function to maintain transparency. *TGFBI* was first

Table 1: List of known corneal dystrophies, including; associated inheritance pattern, gene locus and causative genes.

		Inheritance Pattern	Genetic Locus	Gene known	Gene/s Affected	IC3D Category
	EBMD	Minority of cases, mostly sporadic	5q13	Some cases	*TGFBI*	Some C1
Epithelial and Sub-Epithelial Dystrophies	ERED	Autosomal Dominant	Unknown	Unknown	N/A	C3
	SMCD	Likely autosomal Dominant	Unknown	Unknown	Unknown	C4
	MECD	Autosomal Dominant	12q13 and 17q12	Yes	*KRT3* and *KRT12* (Stocker–Holt variant)	C1
	LECD	X-chromosomal dominant	Xp 22.3	Unknown	Unknown	C2
	GDCD	Autosomal Recessive	1p32.	Yes	*TACSTD2*, previously *M1S1*	C1
Epithelial Stromal Dystrophies	RBCD	Autosomal Dominant	5q13	Yes	*TGFBI*	C1
	TBCD	Autosomal Dominant	5q13	Yes	*TGFBI*	C1
	LCD1	Autosomal Dominant	5q13	Yes	*TGFBI*	C1
	GCD1	Autosomal Dominant	5q13	Yes	*TGFBI*	C1
	GCD2	Autosomal Dominant	5q13	Unknown	*TGFBI*	C1
Stromal Dystrophies	MCD	Autosomal Recessive	16q22	Yes	*CHST6*	C1
	SCD	Autosomal Dominant	1p36	Yes	*UBIAD1*	C1
	CSCD	Autosomal Dominant	12q21.33	Yes	*DCN*	C1
	FCD	Autosomal Dominant	2q34	Yes	*PIKFYVE*, previously *PIP5K3*	C1
	PACD	Autosomal Dominant	12q21.33	Yes	*KERA, LUM,DCN,EPYC*	C1
	CCDF	Unknown	Unknown	Unknown	Unknown	C4
	PDCD	Reported AD, similar deposits seen with X-linked ichthyosis	Isolated PDCD = Unknown	Yes	Unknown	C4
			X-linked ichthyosis = Xp22.31	Unknown	*STS*	C1
Descemet's Membrane and Endothelial Dystrophies	FECD	Unknown, reported autosomal dominant	Early onset = 1q34.3–p32 (FECD 1)	Yes	*COL8A2*	C1
			Late onset –13pter q12.3 (FECD 2), 18q21.2 q21.3 (FECD 3),20p13 q12 (FECD 4), 5q33.1 q35.2 (FECD 5), 10p11.2 (FECD 6), 9p24.1 p22.1 (FECD 7), 15q25 (FECD 8)	Some cases	Unknown, *TCF4*, *SLC4A11*, Unkown, *ZEB1*, Unkown, *AGBL1*	C2 = identified genetic loci, C3 = without known inheritance
	PPCD	Autosomal Dominant	PPCD 1 = 20p11.2–q11.2	Unknown	Unknown	C2
			PPCD 2 = 1p34.3–p32.3	Yes	*COL8A2*	C1
			PPCD 3 = 10p11.2	Yes	*ZEB1*	C1
	CHED	Autosomal recessive	20p13	Yes	*SLC4A11*	C1 (some cases C3)
	XECD	X-chromosomal dominant	Xq25.	Unknown	Unknown	C2
Total:	22	Known = 17	Known = 18	Known = 13 Partially known = 3 Unknown = 6		

reported to be expressed in the cornea in 1994[17], within the same year *TGFBI* and three separate corneal dystrophies were both linked to chromosome 5. In light of this *TGFBI* became a strong candidate for the causative gene behind these dystrophies. Subsequently, Munier *et al* reported the association of 3 different missense mutations with *TGFBI* to distinct corneal dystrophies. To-date there are 62 missense mutations within *TGFBI* which are clustered in hotspots in exons 4,11,12 and 14 (Figure 2 and supplementary table 1 in paper II)[14,15]. R124C, R124H, R124L, R555Q and R555W are the most common missense mutations reported, across all ethnic groups[18,19]. *TGFBI* corneal dystrophies affect several layers in the cornea and collectively are known as the epithelial-stromal *TGFBI* corneal dystrophies[14,15]. They exhibit strong allelic and phenotypic heterogeneity; these missense mutations result in strikingly different corneal deposits and the *TGFBI c*orneal dystrophies are divided into sub-types based on the clinical appearance of these corneal deposits. Broadly they can be classified as lattice corneal dystrophy (LCD) or granular corneal dystrophy (GCD). Prior to genetic analysis theses sub-groups of the *TGFBI* corneal dystrophies were all considered to be distinct clinical forms, the introduction of molecular techniques elucidated that all of these dystrophies were caused by mutations within *TGFBI*. Missense mutations within *TGFBI* have not been shown to cause adverse phenotypes in other tissues in the body. The fundamental function of the cornea is to maintain transparency, *TGFBI* corneal dystrophies result in a disease phenotype due to the accumulation of mutant proteins in the corneal stroma that impair its function to remain transparent. The pathomechanism of how mutant TGFBIp results in corneal deposits is incompletely understood. Extensive work has been performed to uncover the mechanisms behind the formation of these deposits. Potential mechanisms include susceptibility of the mutant protein to oxidative stress, which is induced by continued exposure to ultraviolet light (UV)[20–23]. In addition, the abnormal proteolytic processing of mutant protein has been hypothesised to result in an accumulation of mutant protein resulting in corneal deposits [24–27].

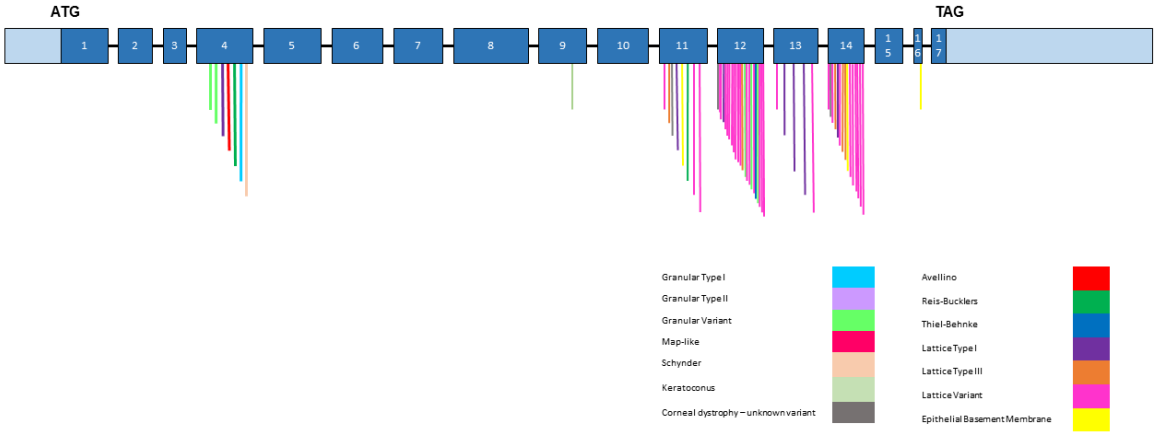


Figure 2: The *TGFBI* gene, including untranslated regions (UTRs) (shown by light blue shading) and introns (shown by black interlinking lines), covers ~35kb, and there are 17 coding exons (shown by dark blue shading). To-date 62 missense mutations within *TGFBI* have been associated with corneal dystrophies, each mutation is depicted by a single drop-down line and the colours correspond to the dystrophy the mutation is associated with, described in the colour coded key. These missense mutations are found in exons 4 to 16 of the gene; however, the majority of mutations are clustered in hot-spots in exons 4, 11, 12 and 14 of the gene.

1.2.1.1 Manifestations of *TGFBI* corneal dystrophies

Classic lattice corneal dystrophy Type I (LCDI) presents as lines of amyloid deposits, arranged radially (Figure 3 A,B)[14]. Typically, it presents in the first or second decade and is associated with recurrent erosions and eventually causes vision loss. The most common mutation associated with LCDI is R124C[28]. Several variations of LCDI have been described and IC3D grouped these collectively as 'Variant LCD'[14]. They are also due to other missense mutations within *TGFBI* however do not always present with characteristic thin branching refractile lines and in some cases erosions do not occur.

Granular corneal dystrophy type I (GCDI) presents early in life and is associated with white granular opacities in the cornea, which typically increase in number and size over time (Figure 3C)[14]. Corneal erosions often occur, which over time result in a loss in visual acuity. GCDI is predominantly associated with the R555W mutation within *TGFBI*[28].

Granular corneal dystrophy type II (GCDII) commonly referred to as Avellino corneal dystrophy, presents in the second decade of life. It is associated with granular deposits that form a snowflake pattern, recurrent corneal erosions and leads to a progressive vision loss (Figure 4)[14]. Avellino corneal dystrophy is caused by a R124H mutation within *TGFBI*[28]. LCDI deposits consist of amyloid deposits while GCDI has been shown to contain hyaline deposits, GCDII has both amyloid and hyaline deposits present, thus is somewhat a combination of lattice and granular dystrophies[29].

Granular corneal dystrophy type III (GCDIII), also known as Reis-Bücklers corneal dystrophy (RBCD), presents in early childhood (Figure 3D). Initial symptoms include formation of irregular opacities in a geographical pattern, this is followed by corneal erosions and scarring which lead to progressive visual impairment.[14] R124L is the predominant mutation associated with RBCD[28].

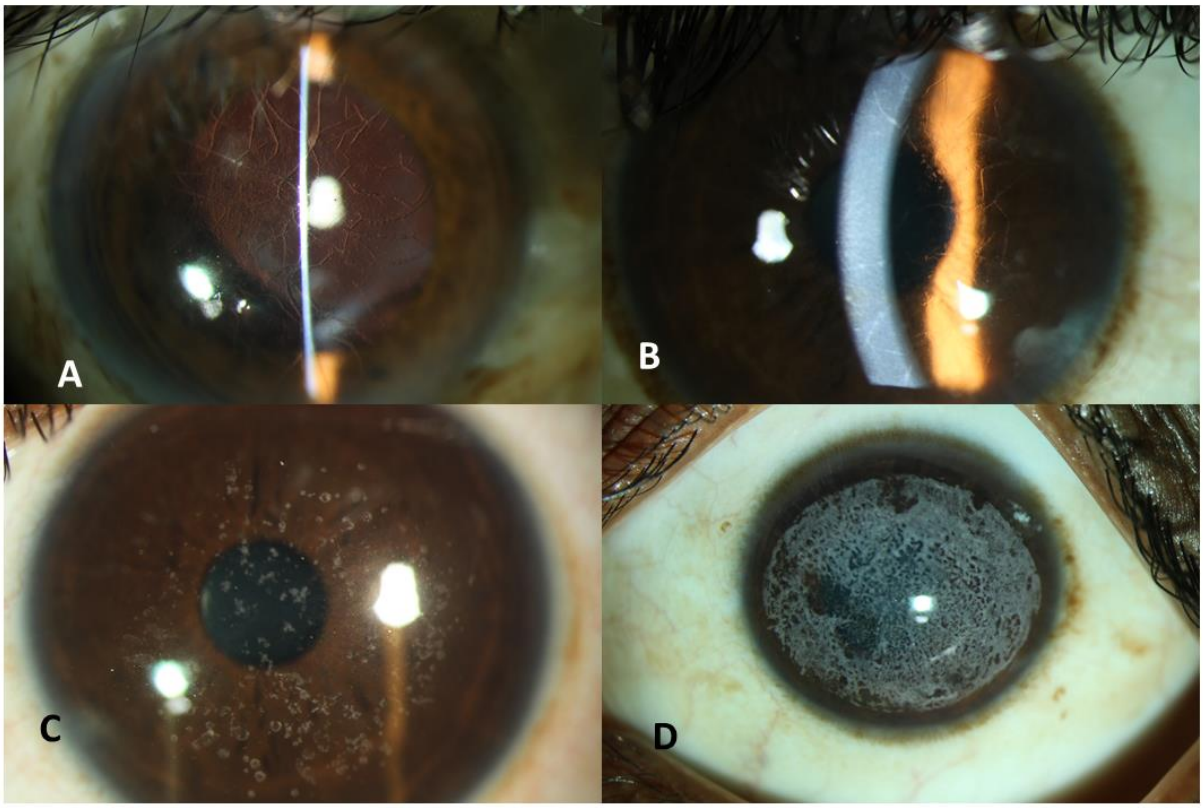


Figure 3: A and B) Lattice corneal dystrophy type I (LCD1) **C)** Granular corneal dystrophy type I (GCD1) **D)** Reis-Bücklers corneal dystrophy

Source: Dr N Venkatesh Prajna, Aravind Eye Hospital, Madurai, Tamil Nadu, India.

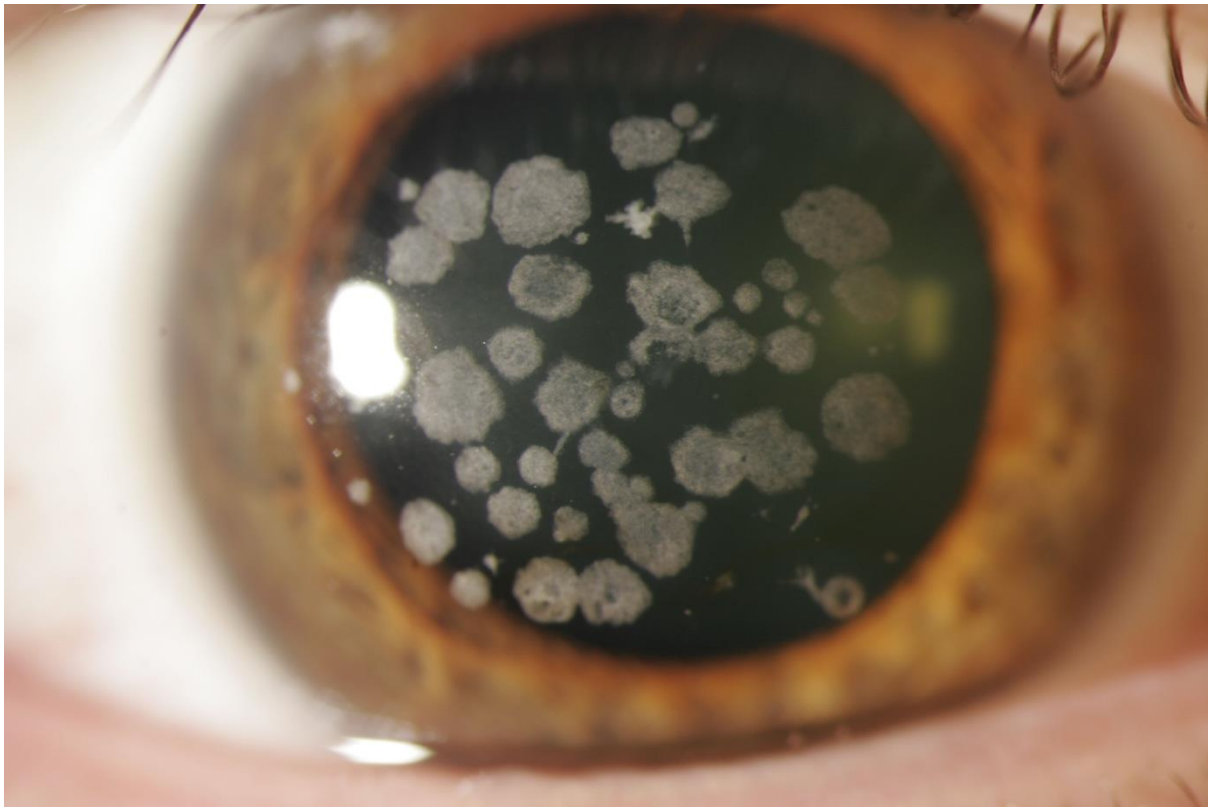


Figure 4: Avellino corneal dystrophy

Source: Prof Colin Willoughby, Belfast Health and Social Care Trust

Thiel-Behnke corneal dystrophy (TBCD) manifests in a similar way to RBCD, however is less aggressive; it presents early in life, results in corneal erosions and an eventual loss of vision. However, it presents as distinctive honeycomb-shaped opacities[14]. TBCD can be distinguished from RBCD by the presence of curly fibres beneath the corneal epithelium, these fibres can only be detected by transmission electron microscopy[30]. R555Q is the predominant mutation associated with TBCD[28].

1.2.1.2 *TGFBI*

The *TGFBI* gene, including untranslated regions (UTRs) and introns, covers ~35kb, and there are 17 coding exons. *TGFBI* produces a 683-residue extracellular matrix (ECM) protein. The crystal structure of TGFBIp reveals it consists of a 23-residue signal peptide for secretion, a flexible linker, an N-terminal cysteine rich domain, four FAS1 domains and a 46-residue C-terminal segment containing an integrin-recognition motif (RGD) [31]. As previously reported by Lukassen *et al*, the N-terminal cysteine rich domain represents a novel domain[32], as such it has been referred to as a cysteine-rich domain of periostin and TGFBIp (CROPT). Thus, TGFBIp consists of a CROPT domain, 4 FAS domains and a RGD enabling integrin binding (Figure 5). The majority of the missense mutations known to date are found in either the FAS1-1 or FAS1-4 domains[31].

TGFBI was first identified in a human lung adenocarcinoma cell line (A549) which had been treated with TGFβ[33], *TGFBI* is widely expressed in most tissues of the body[34,35]. The TGFβ signalling pathway affects many biological processes such as cell proliferation, differentiation morphogenesis, tissue homeostasis and regeneration[36]. However, elucidating the role of TGFβ signalling in these processes is very challenging as, depending on the cell type and conditions in question, the effects of this signalling pathway can have different outcomes. TGFβ is a large superfamily of proteins with over 30 members, including TGFβ, BMP, Activin and Inhibin.

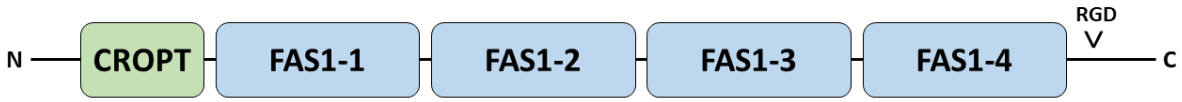


Figure 5: Schematic to show the different domains present in TGFBIp. The majority of reported missense mutations occur in FAS1-1 or FAS1-4

Upon ligand binding a heteromeric complex of type I and type II serine/threonine kinase receptors forms. The type II receptor phosphorylates the type I receptor which then enables the phosphorylation of receptor-regulated SMAD proteins (R-SMAD). SMAD signalling has a central role in all downstream TGFβ effects and they act to regulate transcriptional output in addition to opening repressive chromatin. SMAD 2 and SMAD 3 are activated by TGFβ, Activin or Nodal signals, whereas SMAD 1, SMAD 5 and SMAD 8 are activated by BMP or GDP signals. Phosphorylation of these SMAD proteins then enables binding to a common mediator SMAD protein known as SMAD4. These activated SMAD4-R-SAMD complexes bind other DNA-binding transcription factors allowing the regulation of transcription. Reports have indicated that activation of Smad2/3-Smad 4 via TGFβR1 results in TGFβI expression[37].

1.2.2 Fuchs' endothelial corneal dystrophy

Fuchs' Endothelial corneal dystrophy (FECD) is a common, age-related, inherited degenerative disease of the corneal endothelium which in advanced disease affects all layers of the cornea (Figure 6). It is identified by the presence of corneal guttae, which are excrescences of Descemet's membrane. These corneal guttae are associated with a progressive loss of corneal endothelial cells, the loss in endothelial cells below a critical threshold results in the inability of the corneal endothelium to successfully dehydrate the stroma; causing fluid accumulation in the stroma and the development of painful epithelial bullae leading to corneal clouding and a reduction in visual acuity.[38] In the US, approximately 5% of Caucasians over 40 years of age exhibit corneal guttae which may develop to corneal decompensation.[39] There are two categories of FECD; early-onset which presents with symptoms in the first decade and late-onset which presents with symptoms in the sixth decade.[40] Within these two categories there are 8 subtypes, in all subtypes of FECD the genetic locus has been identified, however the causative gene has not been elucidated in all cases.[15] (Table 1)

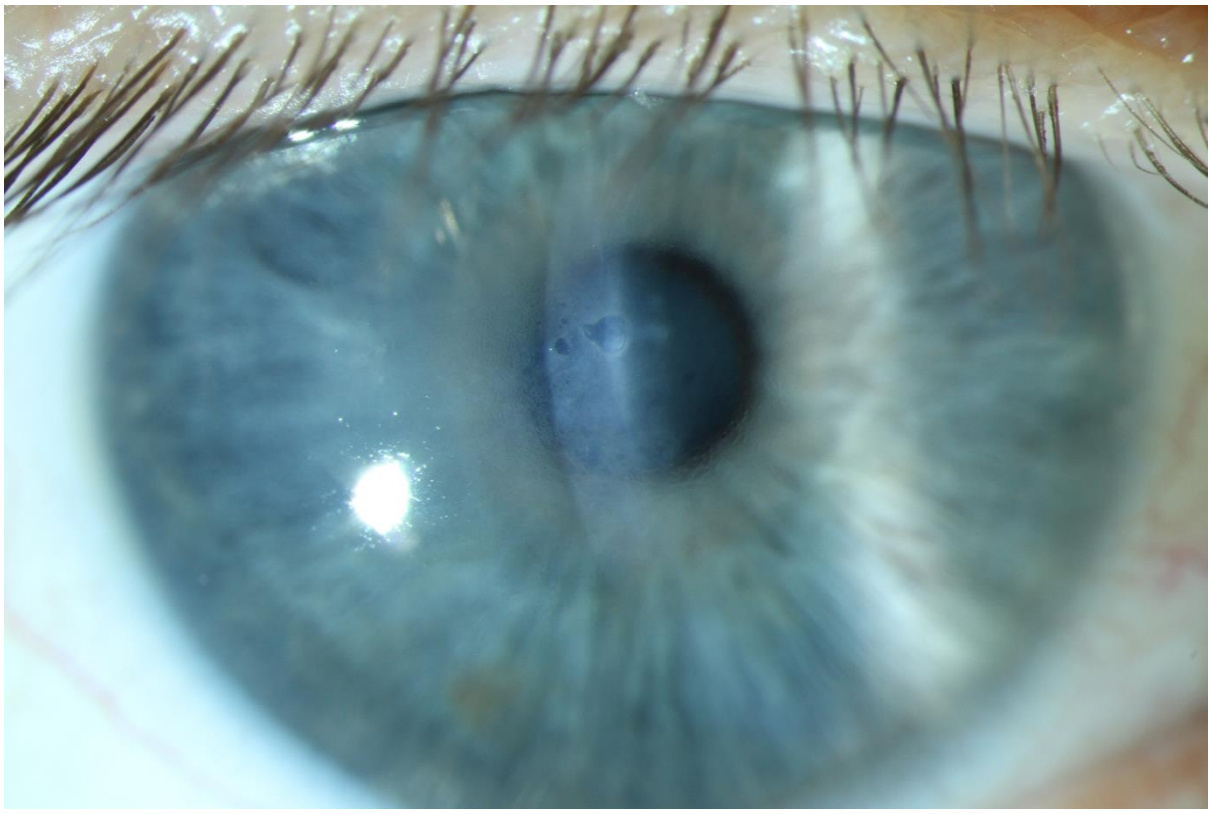


Figure 6: Fuchs' endothelial corneal dystrophy.
Source: Dr B Steger, Department of Ophthalmology, Medical University of Innsbruck, Innsbruck, Austria.

The endothelium secretes type VIII collagen (*COL8*) which is a major constituent of Descemet's membrane. *COL8* has two isoforms, collagen type VIII isoform αlpha 1 (*COL8A1*) and collagen type VIII isoform αlpha 2 (*COL8A2),* together they form a heterotrimer that has a highly organised structure.[41] Mutations in *COL8A2* can result in instability of the heterotrimer and have been associated with early-onset FECD. To date 3 autosomal dominant missense mutations in *COL8A2* (Q455K, L450W and Q455V), have been linked to early-onset FECD.[42–45] These mutations are rare, when considered alongside their late-onset FECD counterparts.

Several genes have been associated with late-onset FECD, the most prevalent cause is linked to transcription factor 4 (*TCF4*). 70% of reported FECD cases are associated with *TCF4*. A genome wide association study (GWAS) strongly associated the rs613872 variant in *TCF4* with FECD.[46] It was later discovered that this SNP was in linkage disequilibrium with CTG18.1, a CTG repeat expansion in the third intron of *TCF4*.[47] CTG repeat expansions cause formation of nuclear RNA foci, these RNA foci act to sequester splicing factors such as MBNL1 and MBNL2; causing a deficiency of these proteins and missplicing of the *TCF4* gene leading to FECD. [48,49]

1.3 The cornea as a target for genome engineering

The cornea serves as an ideal tissue for genome engineering strategies. As previously described the corneal dystrophies are monogenic and present with a mendelian pattern of inheritance, allowing for the design of a gene-editing approach targeted to the single causative gene. In addition, it has a small surface area, therefore gene-editing can be localised reducing the number of cells in which editing must occur. Importantly, it is readily accessible and easily visualised, therefore both treatment and phenotypic readout can be performed in a routine manner. Finally, as the cornea is avascular it holds a unique immune privileged status. While there are undoubtedly benefits to gene-editing based therapies for the corneal layers there are

also major restrictions. One of the primary functions of the cornea is to provide a structural barrier between the outside world and the eye. As such the cornea has many adaptations preventing entry of foreign substances, such as characteristically impermeable junctions, making delivery of gene therapy reagents to the corneal layers very challenging. Considerations for the delivery to the cornea are extensively discussed in Section 5 of Paper I.

1.4 Overview of the CRISPR/Cas system

Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/CRISPR associated (Cas) systems provide a tool to achieve targeted double strand breaks (DSB) within the genome, upon creation of a DSB the cell will initiate endogenous repair pathways to correct this damage and it is by these cellular responses that genome modifications can been achieved. CRISPR-Cas systems can be divided into two classes; Class I which require multiple effectors and Class II which only require a single effector protein[50]. These classes are then further divided into five different types based on the effector proteins present (Table 2). As Class II CRISPR systems only require a single effector protein they have been heavily utilised in genome engineering; Class II systems consist of type II in which the effector protein is Cas9 and type V in which the effector protein is Cas12a (formerly known as Cpf1). Employing these Class II systems to generate DSBs in mammalian cells only requires two-components, a Cas nuclease which recognises a specific DNA motif known as a protospacer adjacent motif (PAM) and a single guide RNA (sgRNA), the Cas nuclease and sgRNA will form a complex and scan the genome for suitable targets (Figure 7). Cas9 and Cas12a have become widespread tools across biomedical science both for basic research and gene therapy.

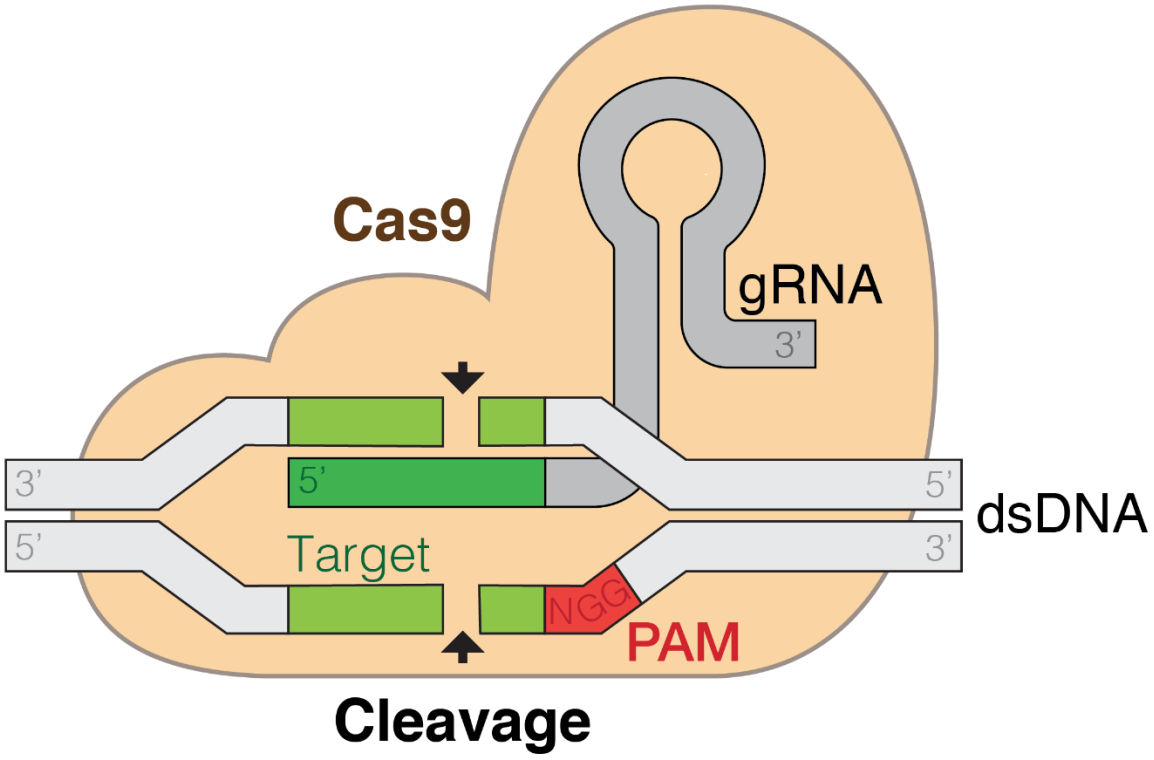


Figure 7: Cas9 (orange) scans the genome in search of a PAM (red), once a PAM is encountered it will determine if the 20bp guide sequence (dark green) is complementary to the sequence immediately adjacent to the PAM (light green), if there is adequate specificity between guide and target DNA Cas9 will induce a DSB 3bp upstream of the PAM.

Source: https://commons.wikimedia.org/wiki/File:GRNA-Cas9.png

Table 2: Current classification of known CRISPR systems

Class	Type	Multi/single protein effector
Class I	Type I	Multisubunit effector complex
	Type III	
	Type IV	
Class 2	Type II	Single effector protein - Cas9
	Type V	Single effector protein - Cas12a
	Type VI	Single effector protein - Cas13b (RNA targeting)

1.5 Discovery of the CRISPR/Cas system as an adaptive immune response

CRISPR repeats were first reported by a Japanese group in 1987[51], they had observed these repeat sequences in *E.coli* but had not considered their presence to be remarkable. While studying gene expression of *Haloferax mediterranei* in response to extreme salt concentration Francisco Mojica again found these repeat sequences[52], which he reported to be multiple copies of palindromic, repeated sequence of 30 base pairs, separated by spacers of approximately 36 base pairs. Interestingly, the spacer sequences were unlike any family of repeats previously characterised in microbes. He then began to discover these repeats in additional microbes [53], at this point Mojica recognised the significance of these repeats. Speaking in an interview with the CRISPR journal[54], Mojica recalled "they are found in many distantly related prokaryotes. Some Archaea have 2% of their genome made up of these repeats, so that tells me this is really important." He believed that in order to understand the function of the repeats, he had to understand the origin of the intervening spacers, as the spacers differ greatly between species. By 2000 Mojica had identified these repeat sequences in 20 different organisms[55], which he had coined Short Regularly Spaced Repeats (SRSRs), SRSRs were later renamed to Clustered Regularly Interspaced Palindromic Repeats (CRISPR). Jansen *et al* reported the existence of CRISPR associated (*cas*) genes[56], which were found to be invariably located adjacent to the CRISPR locus, indicating a common function. In an attempt to understand the identity of the spacer sequences Mojica BLASTed them one by one. In 2003, after many failed attempts, a spacer isolated from an *Escherichia coli (E.coli)* strain matched the sequence of a P1 phage. Critically, the *E.coli* strain that spacer was derived from is known to be resistant to P1 infection. Further investigation revealed that 2/3 of spacers with sequence matches corresponded to extrachromosomal plasmids and bacteriophages[57], indicating CRISPR has a role as an adaptive immune system that can protect the bacteria against these elements. In parallel, Gilles Vergnaud found that the spacers sequenced from *Yersinia pestis (Y.pestis)* were homologous

to an inactive prophage within the *Y.pestis* genome[58], leading to the same conclusion that CRISPR acted an adaptive immune system. At the same time Alexander Bolotin found that the resistance of *Streptococcus thermophilus* (*S.thermophilus)* to phage correlated with the number of spacers that were present and he also came to the conclusion that CRISPR acted as an adaptive immune system[59]. He also proposed that CRISPR suppressed these extrachromosomal elements by coding for anti-sense RNA. Three groups had now independently shown homology between the spacer sequences and foreign elements and were all proposing the CRISPR system to act as an adaptive immune system. Barrangou *et al* were the first to substantiate this hypothesis[60]; demonstrating that not only is the sequence similarity between the spacer and target used for recognition but that when polymorphisms between the spacer and target were present the bacteria would lose resistance. Furthermore, inactivation of *cas7* did not affect resistance while inactivation of *cas5* (now referred to *cas*9) resulted in a loss of resistance, hypothesising that *cas5* (*cas9*) acts as a critical nuclease, supported by its previously reported RuvC and HNH nuclease domains [59,61].

1.5.1 Reprogramming of CRISPR/Cas systems

The first demonstration of reprogramming of a CRISPR/Cas system was using the Class I, type I CRISPR system from *E.coli*; Brouns *et al* reported that the CRISPR loci was transcribed into a pre-crRNA (precursor-crisprRNA), which was then cleaved to form a mature crRNA. Moreover, the crRNA together with the effector complex was responsible for cleaving the target sequence and finally utilising synthetic spacers targeted to lambda (λ) phage, resistance to λ phage was induced. By achieving resistance with both coding and non-coding spacers they were the first to hypothesise that CRISPR targeted DNA, challenging previous hypotheses that CRISPR worked by an anti-sense RNA mechanism. Further work again showed target similarity directs CRISPR activity but additionally provided confirmation that CRISPR targeted DNA[62]; utilising an isolate of *Staphylococcus epidermidis* which contained a spacer

complementary to the *nickase (nes)* gene, which is present in the majority of staphylococcal conjugative plasmids, CRISPR interference was used to prevent conjugation. Moreover, disruption of the *nes* spacer target site by the introduction a self-splicing intron, which after a splicing event would reform the target sequence, revealed that conjugation was not inhibited. This proved that when the target exists in the mRNA but not in the DNA, CRISPR does not function, therefore the system must act on DNA.

1.5.2 Target recognition and cleavage

Two papers identified an additional CRISPR element that indicated spacer acquisition was not random. An exact sequence motif was always located downstream of the protospacer[63,64], now referred to as a protospacer adjacent motif (PAM). Furthermore, different Cas nucleases were shown to recognise unique PAMs. PAM recognition has since been extensively studied confirming that Cas nucleases isolated from different strains have distinct PAMs; such as *Streptococcus pyogenes* Cas9 (*Sp*Cas9) that recognises a NGG motif, *Staphylococcus aureus* Cas9 (*Sa*Cas9) that recognises NNGRRT, *Acidaminococcus* Cas12a (*As*Cas12a) and *Lachnospiraceae bacterium* Cas12a (*Lb*Cas12a) which both recognise a TTTN motif[65–69]. PAM sequences are critical recognition domains because they provide a mechanism for the Cas nuclease to discriminate between self and non-self DNA[70]. By sequencing the immediate products of Cas9 cleavage Garneau *et al* demonstrated that Cas9 creates a blunt double-strand break (DSB) in the target DNA and that this event always occurs 3bp upstream of the PAM[71], demonstrating the precision of this tool at making targeted DSBs. In contrast Cas12a, the other Class II system that has been repurposed for genome engineering, was found to generate a staggered DSB with a 5' overhang[69]. Deltcheva *et al* uncovered a novel small RNA that was transcribed from the region adjacent to the CRISPR locus, critically this transcript has a 24-nucleotide complementarity to the repeat regions of pre-crRNA[72]. This trans-activating CRISPR RNA (tracrRNA) was shown to hybridise with the pre-crRNA and upon cleavage by

RNase III was processed into mature crRNA. Sapranaukas *et al* showed that the CRISPR system from *S.thermophilus* could be transferred to *E.coli* were it retained its function to cleave both plasmid and viral DNA[73]. Importantly, they also demonstrated that the only protein necessary to maintain function was Cas9.

1.5.3 Repurposing CRISPR/Cas9 for use in mammalian cells

The fundamental components for the CRISPR system were now known, using purified Cas9 and *in vitro* transcribed crRNA and tracrRNA Jinek *et al* showed that it was possible to reprogram the CRISPR system to cleave a target *in vitro*[65]. In addition, by mutating the catalytic domains of RuvC and HNH they confirmed RuvC cuts the strand complementary to the crRNA and HNH cut the strand non-complementary to the crRNA. Finally, in what would become a key aspect of genome editing by CRISPR systems, they demonstrated that the crRNA and tracrRNA could be fused to produce a single-guide RNA (sgRNA), negating the need for processing by RNase III. Gasiunas *et al* published work very parallel to this just weeks later[74]. Cong *et al* were the first to achieve gene-editing in mammalian cells using the CRISPR-Cas9 system[66]. This was achieved by codon-optimising *Sp*Cas9 and attaching nuclear localization signals (NLS), this optimised *Sp*Cas9 together with an adopted version of the sgRNA described by Jinek *et al,* which again was a chimera of crRNA and tracrRNA but of different lengths, was both necessary and sufficient to achieve efficient gene editing. They also created a mutant nickase *Sp*Cas9 containing a previously described alanine substitution (D10A)[65,73,75] in the RuvC domain. As this nickase would only result in a single strand DNA nick different DNA repair mechanisms will process it and thus different genomic modifications can be achieved. Finally, they demonstrated that it was possible to target multiple genes simultaneously by using a CRISPR array encoding several spacers. Published in the same issue of Science, Mali *et al* also showed a two-component CRISPR system achieved efficient edits in mammalian cells and also used multiplex gene-editing to target several sites. Several Cas9 orthologues and Cas12a

nucleases have since been repurposed for use in mammalian cells[68,69]. These landmark papers by the labs of Doundna & Charpentier, Zhang and Church established CRISPR/Cas systems as a tool for mammalian gene-editing, however CRISPR was first reported in 1987 and it is only by the instrumental efforts of a large number of people that this feat has been achieved.

1.6 DNA Repair related to a DSB

The Cas nuclease employed generates a DSB within the cell, however it is by the DNA repair mechanisms initiated in response to the DSB by which genome editing can be achieved. It has been estimated that ~10^5 DNA lesions occur in the mammalian genome each day[76], including DSBs, as such mammalian cells have various DNA repair mechanisms to deal with this damage. DSBs are considered the most lethal lesion in DNA, it is widely accepted that they are repaired by either homology directed repair (HDR) or non-homologous end joining (NHEJ) repair (Figure 8). Homologous recombination (HR) is only active during S and G2 phases of the cell cycle, therefore can only occur within dividing cells. Conversely, NHEJ occurs in both dividing and non-dividing cells.

The exact mechanism behind repair of the blunt-ended DSB induced by Cas9 is unknown. Work is ongoing to elucidate the exact pathways implemented and thus how to manipulate them to make the genome modifications more favorable for the desired outcome. Generally, it is accepted that HR is a precise mode of repair, requiring a template in the form of a homologous chromosome or synthetic donor, that can be utilised to introduce specified changes. Synthetic donors typically include either a double-stranded oligonucleotide (dsODN) or single-stranded oligonucleotide (ssODN). Conversely, NHEJ is an error-prone repair pathway, in which ends of the DSB are joined together in an imperfect manner that results in the introduction of small insertions or deletions (indels). If the indels are frameshifting they

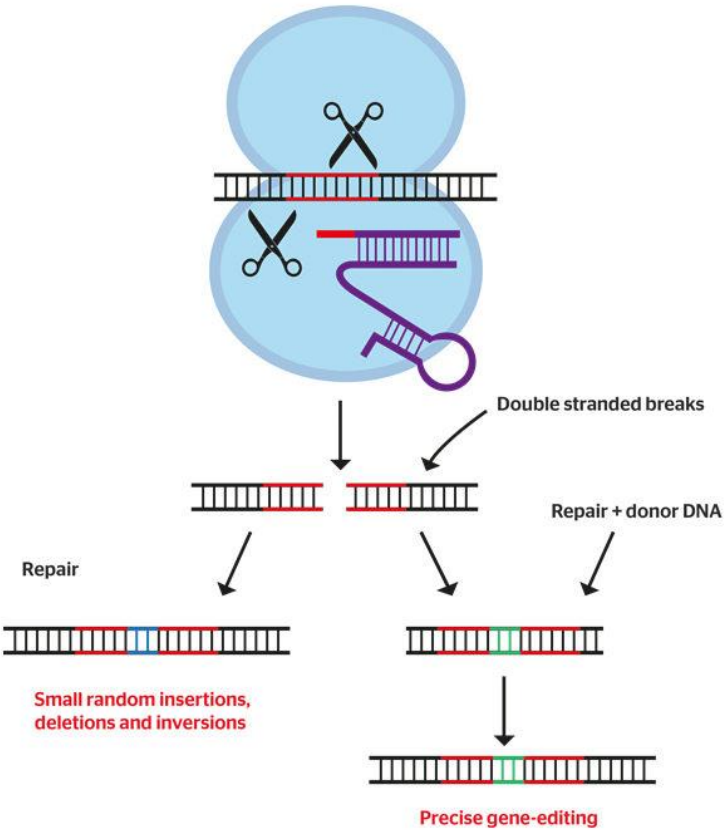


Figure 8: Different modes of repair of a Cas9 induced DSB. DSBs are either repaired by NHEJ which induced indels that may lead to frameshifting mutations resulting in a premature stop codon, resulting in gene disruption. Alternatively the DSB is repaired by HR in which a template, either the homologous chromosome or a donor template, is used for repair. HR can be used for precise editing, for example inserting or correcting mutations.

Source: https://www.flickr.com/photos/b4fa/27401360522

can result in a premature stop codon, which provided it occurs ≥50-55 nucleotides upstream of the 3' most exon-exon junction [77,78], can cause gene disruption.

Richardson *et al* recently published an interesting finding into HR in relation to a ssODN, which is referred to as single-stranded template repair (SSTR)[79]. Using a high-throughput screen they identified that the Fanconi Anaemia (FA) pathway, known to recognise and repair interstrand cross-links, was critical to SSTR but had no effect on NHEJ repair. However, using chromatin immunoprecipitation sequencing (ChIP-seq) they found that FANCD2, involved in the FA pathway, localizes to Cas9-induced DSBs, regardless of the presence of ssODN. They hypothesised that the FA pathway acts as a mediator to direct the repair pathway of the DSB. In the absence of a ssODN NHEJ will occur as normal, however in the presence of a ssODN HR by either standard HR, SSTR or single-strand annealing (SSA) can occur.

Predominantly DSBs will be repaired by NHEJ, allowing efficient gene disruption. HDR provides a tool to correct a genetic defect or introduce new genetic information. However, HR is extremely inefficient therefore efforts have been focused on increasing the efficiency of this repair outcome to maximise its potential. As more insights into the exact mechanisms of Cas9 DSB repair are made, potential strategies to increase HR may become apparent.

1.7 Structure of sgRNA:Cas9 complex

The crystal structures for inactive Cas9, a sgRNA:Cas9 complex and a sgRNA:Cas9 complex bound to target DNA have now been published[80–83]. The crystal structure of a sgRNA:Cas9 complex bound to target DNA revealed that there are two lobes; a recognition lobe (REC) and a nuclease lobe (NUC)[82]. The REC lobe can be sub-divided into seperate domains including REC1, REC2 REC3 and an α bridging helix. While the NUC lobe can be sub-divided to contain a RuvC domain, HNH domain and PAM-interacting (PI) domains. The sgRNA:DNA

heteroduplex that forms when the sgRNA binds to its target sequence is positioned in a positive groove between the REC and NUC lobes.

1.8 Mechanism of cleavage by Cas9

Broadly, target recognition by the sgRNA-Cas9 complex is thought to involve two main stages, the recognition of a suitable PAM and the formation of a RNA-DNA heteroduplex[84]. Initially *S.pyogenes* Cas9 will sample PAMs via the PAM interacting domain, following identification of a suitable NGG PAM the *Sp*Cas9 nuclease will then test the PAM-proximal sequence for complementarity with the guide sequence provided. If high sequence similarity exists *Sp*Cas9 will then test the PAM-distal portion of the guide and if further complementarity exists *Sp*Cas9 will then generate a DSB. Sternberg *et al* demonstrated that the conformational change of the HNH domain acts as an additional proofreading mechanism[85]. Indicating that a mechanism exists whereby Cas9 recognises a PAM and tests the adjacent sequence for complementarity, if an on-target sequence is identified this drives a conformational change within the HNH domain, which acts as a allosteric switch to trigger the RuvC domain and results in a concerted cleavage of both strands. Off-target sites are known to bind Cas9, however this explains why although mismatches allow Cas9 binding without a conformational change in the HNH domain, cleavage will not occur[86]. Further studies into the mechanism of Cas9 cleavage have revealed that HNH transitions between several conformations before docking into its active state, demonstrating that there is an intermediate state that governs transition of DNA binding and DNA cleavage by Cas9 [87].

1.9 Limitations of CRISPR/Cas9

While CRISPR-Cas9 holds immense potential for the field of gene editing there are also a number of issues impeding the translation of this technology to the clinic. Firstly, *Sp*Cas9 is quite large in size thus the efficient delivery of the CRISPR-Cas9 components presents quite a

challenge. Delivery is a challenge faced throughout the field of gene therapy and is not exclusive to ocular gene therapy. However, each target tissue has their own unique barriers to be overcome in order to achieve potent and specific delivery to the cell of interest. The cornea exists as a structural barrier to prevent entry of foreign substances into the eye. As such, it has evolved to be difficult to penetrate. Adeno associated virus (AAV) has the ability to infect a range of cell types due to the receptors present on the AAV capsid. Different AAV serotypes have tropisms for different tissues due to the capsid proteins present on their surface. The payload of this widely used gene therapy vehicle is <5kb; the coding sequence of *Sp*Cas9 alone is 4.2kb, inclusion of additional components including sgRNA and promoters exceed the 5kb limit. In order to overcome this limitation several solutions have been explored, including, splitting the CRISPR components across two vectors and utilising *Sp*Cas9 orthologues smaller in size. Another issue is that AAV results in the sustained expression of the transgene, dual AAV systems that carry an additional guide designed to disrupt the transgene have been explored[88–90]. Furthermore non-viral delivery solutions that result in reduced expression of Cas9 within the cell have been investigated[91,92]. The Cas9 protein in complex with the sgRNA is highly anionic, thus the carriage of this large negatively complex across a cell membrane can be quite inefficient. The specific challenges associated with both delivery to the eye and delivery of CRISPR/Cas9 components are extensively discussed in Paper I section 5.

Critically, it has been widely shown that *Sp*Cas9 can tolerate mismatches within the guide sequence leading to the generation of DSBs at sites with high sequence similarity to the guide sequence employed [93–95]. These off-target events can occur at sites with several mismatches relative to the guide sequence anywhere within the genome. The inadequate genome-wide specificity of *Sp*Cas9 has fueled extensive research into ways to reduce off-target cleavage. Such as selecting guides that do not have high homology to other sites in the genome, thus reducing the likelihood of off-target events. In addition, truncation of the guide sequence has

been shown to confer greater specificity [96]. Finally, several high fidelity *Sp*Cas9 variants have been engineered using rational design [97–99], these variants exhibit improved global specificity, however off-target events have not been completely eliminated.

As the majority of the corneal dystrophies are due to missense mutations that present with an autosomal dominant inheritance pattern an allele-specific knockout of only the mutant allele would be a viable treatment approach for these dystrophies. When considering global specificity, it is extremely rare to encounter an off-target site that differs by only a single base pair. However, for allele-specific editing this off-target site will inevitably exist and will act as the most perilous off-target event, as unintended cleavage at the non-target allele will result in a loss of the functional protein and may exacerbate the disease. Thus, in order to develop a gene therapy for the corneal dystrophies the ability to distinguish between single base pair changes is paramount.

1.10 Conclusion

The unique qualities of the cornea and genetics of the corneal dystrophies make it an amenable target for gene therapy. The CRISPR technology holds vast potential to treat the corneal dystrophies. However, in order to realise this potential a number of issues must be overcome, specifically, efficient delivery to the cornea and both improved discrimination between wild-type and disease alleles and genome-wide specificity.

1.11 References

1. Meek, K. M., Dennis, S. & Khan, S. Changes in the refractive index of the stroma and its extrafibrillar matrix when the cornea swells. *Biophys. J.* **85,** 2205–2212 (2003).

2. Derek W. DelMonte, T. K. Anatomy and physiology of the cornea and related structures. *J Cataract Refract Surg* **37,** 588–598 (2011).

3. Hanna, C., Bicknell, D. & O'Brien, J. Cell turnover in the adult human eye. *Arch. Ophthalmol.* **65,** 695–698 (1961).

4. Kinoshita, S., Friend, J. & Thoft, R. A. Sex chromatin of donor corneal epithelium in rabbits. *Investig. Ophthalmol. Vis. Sci.* **21,** 434–441 (1981).

5. Tseng, S. C. Concept and application of limbal stem cells. *Eye (Lond).* **3 (Pt 2),** 141–157 (1989).

6. Wilson, S. E. & Hong, J. W. Bowman's layer structure and function: Critical or dispensable to corneal function? A Hypothesis. *Cornea* **19,** 417–420 (2000).

7. Hassell, J. R. & Birk, D. E. The molecular basis of corneal transparency. *Experimental Eye Research* **91,** 326–335 (2010).

8. Müller, L. J., Pels, E., Schurmans, L. R. H. M. & Vrensen, G. F. J. M. A new three-dimensional model of the organization of proteoglycans and collagen fibrils in the human corneal stroma. *Exp. Eye Res.* **78,** 493–501 (2004).

9. Jester, J. V *et al.* The cellular basis of corneal transparency: evidence for 'corneal crystallins'. *J. Cell Sci.* **112 (Pt 5,** 613–622 (1999).

10. Schlötzer-Schrehardt, U., Bachmann, B. O., Laaser, K., Cursiefen, C. & Kruse, F. E. Characterization of the cleavage plane in Descemet's membrane endothelial keratoplasty. *Ophthalmology* **118,** 1950–1957 (2011).

11. Geroski, D. H., Matsuda, M., Yee, R. W. & Edelhauser, H. F. Pump function of the human corneal endothelium. Effects of age and cornea guttata. *Ophthalmology* **92,** 759–763 (1985).

12. Bahn, C. F. *et al.* Postnatal development of corneal endothelium. *Invest. Ophthalmol. Vis. Sci.* **27,** 44–51 (1986).

13. Klintworth, G. K. Corneal dystrophies. *Orphanet J. rare Dis. Cornea Dystrophy Foundaion* **4,** 7 (2009).

14. Weiss, J. S. *et al.* The IC3D Classification of the Corneal Dystrophies. *Cornea* **27,** S1–S42 (2008).

15. Weiss, J. S. *et al.* IC3D Classification of Corneal Dystrophies—Edition 2. *Cornea* **34,** 117–159 (2015).

16. Moore, C. B. T., Christie, K. A., Marshall, J. & Nesbit, M. A. Personalised genome editing - The future for corneal dystrophies. *Progress in Retinal and Eye Research* (2018). doi:10.1016/j.preteyeres.2018.01.004

17. Escribano, J., Hernando, N., Ghosh, S., Coca- Prados, M. & Crabb, J. cDNA from human ocular ciliary epithelium homologous to βig- h3 is preferentially expressed as an extracellular protein in the corneal epithelium. *J. Cell. Physiol.* **160,** 511–521 (1994).

18. Kannabiran, C. & Klintworth, G. K. TGFBI gene mutations in corneal dystrophies. *Hum. Mutat.* **27,** 615–625 (2006).

19. Korvatska, E. *et al.* Mutation hot spots in 5q31-linked corneal dystrophies. *Am. J. Hum. Genet.* **62,** 320–4 (1998).

20. Choi, S. Il *et al.* Decreased catalase expression and increased susceptibility to oxidative stress in primary cultured corneal fibroblasts from patients with granular corneal dystrophy type II. *Am. J. Pathol.* **175,** 248–261 (2009).

21. Kim, T. *et al.* Altered mitochondrial function in type 2 granular corneal dystrophy. *Am. J. Pathol.* **179,** 684–692 (2011).

22. Grothe, H. L. *et al.* Altered protein conformation and lower stability of the dystrophic transforming growth factor beta-induced protein mutants. *Mol. Vis.* **19,** 593–603 (2013).

23. Kolozsvári, L., Nógrádi, A., Hopp, B. & Bor, Z. UV absorbance of the human cornea in the 240- to 400-nm range. *Investig. Ophthalmol. Vis. Sci.* **43,** 2165–2168 (2002).

24. Choi, S. Il, Maeng, Y. S., Kim, K. S., Kim, T. I. & Kim, E. K. Autophagy is induced by raptor degradation via the ubiquitin/proteasome system in granular corneal dystrophy type 2. *Biochem. Biophys. Res. Commun.* **450,** 1505–1511 (2014).

25. Korvatska, E. *et al.* On the role of kerato-epithelin in the pathogenesis of 5q31-linked corneal dystrophies. *Investig. Ophthalmol. Vis. Sci.* **40,** 2213–2219 (1999).

26. Korvatska, E. *et al.* Amyloid and non-amyloid forms of 5q31-linked corneal dystrophy resulting from kerato-epithelin mutations at Arg-124 are associated with abnormal turnover of the protein. *J. Biol. Chem.* **275,** 11465–11469 (2000).

27. Courtney, D. G. *et al.* Protein composition of TGFBI-R124C- and TGFBI-R555W-associated aggregates suggests multiple mechanisms leading to lattice and granular corneal dystrophy. *Investig. Ophthalmol. Vis. Sci.* **56,** 4653–4661 (2015).

28. Munier, F. L. *et al.* BIGH3 mutation spectrum in corneal dystrophies. *Investig. Ophthalmol. Vis. Sci.* **43,** 949–954 (2002).

29. Folberg, R. *et al.* Clinically Atypical Granular Corneal Dystrophy with Pathologic Features of Lattice-like Amyloid Deposits: A Study of Three Families. *Ophthalmology* **95,** 46–51 (1988).

30. Kuchle, M., Green, W. R., Volcker, H. E. & Barraquer, J. Reevaluation of corneal dystrophies of Bowman's layer and the anterior stroma (Reis-Bucklers and Thiel-Behnke types): A light and electron microscopic study of eight corneas and a review of the literature. *Cornea* **14,** 333–354 (1995).

31. García-Castellanos, R. *et al.* Structural and Functional Implications of Human Transforming Growth Factor β-Induced Protein, TGFBIp, in Corneal Dystrophies. *Structure* **25,** 1740–1750.e2 (2017).

32. Lukassen, M. V., Scavenius, C., Thøgersen, I. B. & Enghild, J. J. Disulfide Bond Pattern of Transforming Growth Factor β-Induced Protein. *Biochemistry* **55,** 5610–5621 (2016).

33. Skonier, J. *et al.* cDNA cloning and sequence analysis of beta ig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-beta. *DNA Cell Biol.* **11,** 511–522 (1992).

34. Hashimoto, K. *et al.* Characterization of a cartilage-derived 66-kDa protein (RCD-CAP/βig-h3) that binds to collagen. *Biochim. Biophys. Acta - Mol. Cell Res.* **1355,** 303–314 (1997).

35. Kim, J. E. *et al.* Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-β-induced gene, βig-h3. *J. Biol. Chem.* **275,** 30907–30915 (2000).

36. Massagué, J. TGFβ signalling in context. *Nature Reviews Molecular Cell Biology* **13,** 616–630 (2012).

37. Jeon, E. S., Kim, J. H., Ryu, H. & Kim, E. K. Lysophosphatidic acid activates TGFBIp expression in human corneal fibroblasts through a TGF-β1-dependent pathway. *Cell. Signal.* **24,** 1241–1250 (2012).

38. Borboli, S. & Colby, K. Mechanisms of disease: Fuchs' endothelial dystrophy. *Ophthalmology Clinics of North America* **15,** 17–25 (2002).

39. Lorenzetti, D. W. C., Uotila, M. H., Parikh, N. & Kaufman, H. E. Central Cornea Guttata. *Am. J. Ophthalmol.* **64,** 1155–1158 (1967).

40. Vedana, G., Villarreal, G. & Jun, A. S. Fuchs endothelial corneal dystrophy: Current perspectives. *Clinical Ophthalmology* **10,** 321–330 (2016).

41. Shuttleworth, C. A. Type VIII collagen. *Int. J. Biochem. Cell Biol.* **29,** 1145–1148 (1997).

42. Biswas, S. Missense mutations in COL8A2, the gene encoding the alpha2 chain of type VIII collagen, cause two forms of corneal endothelial dystrophy. *Hum. Mol. Genet.* **10,** 2415–2423 (2001).

43. Magovern, M., Beauchamp, G. R., McTigue, J. W., Fine, B. S. & Baumiller, R. C. Inheritance of Fuchs' Combined Dystrophy. *Ophthalmology* **86,** 1897–1920 (1979).

44. Gottsch, J. D. *et al.* Inheritance of a novel COL8A2 mutation defines a distinct early-onset subtype of fuchs corneal dystrophy. *Investig. Ophthalmol. Vis. Sci.* **46,** 1934–1939 (2005).

45. Mok, J. W., Kim, H. S. & Joo, C. K. Q455V mutation in COL8A2 is associated with Fuchs' corneal dystrophy in Korean patients. *Eye* **23,** 895–903 (2009).

46. Baratz, K. H. *et al.* E2-2 Protein and Fuchs's Corneal Dystrophy. *N. Engl. J. Med.* **363,** 1016–1024 (2010).

47. Wieben, E. D. *et al.* A Common Trinucleotide Repeat Expansion within the Transcription Factor 4 (TCF4, E2-2) Gene Predicts Fuchs Corneal Dystrophy. *PLoS One* **7,** (2012).

48. Mootha, V. V. *et al.* TCF4 Triplet Repeat Expansion and Nuclear RNA Foci in Fuchs' Endothelial Corneal Dystrophy. *Invest. Ophthalmol. Vis. Sci.* **56,** 2003–2011 (2015).

49. Du, J. *et al.* RNA toxicity and missplicing in the common eye disease fuchs endothelial corneal dystrophy. *J. Biol. Chem.* **290,** 5979–5990 (2015).

50. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13,** 722–736 (2015).

51. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169,** 5429–5433 (1987).

52. Mojica, F. J. M., Juez, G. & Rodriguez- Valera, F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* **9,** 613–621 (1993).

53. Mojica, F. J. M., Ferrer, C., Juez, G. & Rodríguez- Valera, F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Mol. Microbiol.* **17,** 85–93 (1995).

54. Davies, K. & Mojica, F. Crazy About CRISPR: An Interview with Francisco Mojica. *Cris. J.* **1,** crispr.2017.28999.int (2018).

55. Mojica, F. J. M., Díez-Villaseñor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* **36,** 244–246 (2000).

56. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).

57. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening

sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182 (2005).

58. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151,** 653–663 (2005).

59. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).

60. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science (80-.).* **315,** 1709–1712 (2007).

61. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* **1,** (2006).

62. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science (80-.).* **322,** 1843–1845 (2008).

63. Horvath, P. *et al.* Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *J. Bacteriol.* **190,** 1401–1412 (2008).

64. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.* **190,** 1390–1400 (2008).

65. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity. *Science* **337,** 816–822 (2012).

66. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–23 (2013).

67. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (80-.).* **339,** 823–826 (2013).

68. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520,** 186–191 (2015).

69. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759–771 (2015).

70. Swarts, D. C. & Jinek, M. Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdisciplinary Reviews: RNA* (2018). doi:10.1002/wrna.1481

71. Garneau, J. E. *et al.* The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).

72. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).

73. Sapranauskas, R. *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39,** 9275–9282 (2011).

74. Gasiunas, G. & Barrangou, R. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. in *Proceedings of the …* (2012). doi:10.1073/pnas.1208507109

75. Gasiunas, G. & Barrangou, R. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. in *Proceedings of the …* (2012). doi:10.1073/pnas.1208507109

76. Iyama, T. & Wilson, D. M. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst).* **12,** 620–636 (2013).

77. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends in Biochemical Sciences* **23,** 198–199 (1998).

78. Popp, M. W. & Maquat, L. E. Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell* **165,** 1319–1332 (2016).

79. Richardson, C. D. *et al.* CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* **50,** 1132–1139 (2018).

80. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (80-.).* (2014). doi:10.1126/science.1247997

81. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science (80-.).* (2015). doi:10.1126/science.aab1452

82. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).

83. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573 (2014).

84. Cencic, R. *et al.* Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One* **9,** (2014).

85. Sternberg, S. H., Lafrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* (2015). doi:10.1038/nature15544

86. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32,** 670–676 (2014).

87. Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* (2017). doi:10.1126/sciadv.aao0027

88. Chen, Y. *et al.* A Self-restricted CRISPR System to Reduce Off-target Effects. *Mol. Ther.* **24,** 1508–1510 (2016).

89. Moore, R. *et al.* CRISPR-based self-cleaving mechanism for controllable gene delivery in human cells. *Nucleic Acids Res.* **43,** 1297–1303 (2015).

90. Petris, G. *et al.* Hit and go CAS9 delivered through a lentiviral based self-limiting circuit. *Nat. Commun.* **8,** (2017).

91. Lee, K. *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* (2017). doi:10.1038/s41551-017-0137-2

92. D'Astolfo, D. S. *et al.* Efficient intracellular delivery of native proteins. *Cell* **161,** 674–690 (2015).

93. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases

in human cells. *Nat. Biotechnol.* **31,** 822–826 (2013).

94. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31,** 839–843 (2013).

95. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).

96. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32,** 279–284 (2014).

97. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490–495 (2016).

98. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.).* **351,** 84–88 (2016).

99. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550,** 407–410 (2017).

1.12 Thesis aims

The overall aim of this thesis was to investigate the potential of treating autosomal dominant corneal dystrophies by CRISPR/Cas9, by inducing allele-specific gene disruption of only the mutant allele. This thesis aims to overcome the current limitations of the CRISPR/Cas9 system, to enable the treatment of the *TGFBI* corneal dystrophies.

In fulfilment of this aim, the objectives of the research presented in this thesis were to:

- Write a comprehensive review of the position of the corneal dystrophies in relation to genome editing
- Investigate possible mechanisms of how *TGFBI* corneal dystrophy patients develop an accumulation of corneal deposits following laser eye surgery
- Define the ability to target the *TGFBI* corneal dystrophy missense mutations utilising known allele-specific strategies
- Explore the possibility of exploiting natural variation within the genome to achieve allele-specific editing
- Assess the ability of adeno associated virus to transduce the corneal layers and deliver CRISPR/Cas9 components

Paper I

Paper I

2 Paper I - Personalised genome editing – the future for corneal dystrophies

C.B. Tara Moore[*], Kathleen A. Christie[*], John Marshall, M. Andrew Nesbit

[*] These authors are co-first authors

The main aim of this paper was to:

1. Generate an informative review article collating relevant literature relating to the field of gene editing and discuss how recent advancements can be applied to the treatment of the corneal dystrophies

CONTRIBUTION

I authored this review article and designed all figures and tables.

C. B. Tara Moore – proofread final manuscript

John Marshall – proofread final manuscript

M. Andrew Nesbit – proofread manuscript at all stages

Progress in Retinal and Eye Research 65 (2018) 147–165

Contents lists available at ScienceDirect

Progress in Retinal and Eye Research

journal homepage: www.elsevier.com/locate/preteyeres

Personalised genome editing – The future for corneal dystrophies

C.B. Tara Moore[a,b,*,1,2], Kathleen A. Christie[a,1,2], John Marshall[c,2], M. Andrew Nesbit[a,2]

[a] Biomedical Sciences Research Institute, Ulster University, Coleraine, Northern Ireland, BT52 1SA, UK
[b] Avellino Laboratories, Menlo Park, CA 94025, USA
[c] Department of Genetics University College London (UCL) Institute of Ophthalmology, University of London, London, UK

ARTICLE INFO

Keywords:
Genome editing
Corneal dystrophies
CRISPR/Cas9
Personalised medicine

ABSTRACT

The potential of personalised genome editing reaching the clinic has come to light due to advancements in the field of gene editing, namely the development of CRISPR/Cas9. The different mechanisms of repair used to resolve the double strand breaks (DSBs) mediated by Cas9 allow targeting of a wide range of disease causing mutations. Collectively, the corneal dystrophies offer an ideal platform for personalised genome editing; the majority of corneal dystrophies are monogenic, highly penetrant diseases with a known pattern of inheritance. This genetic background coupled with the accessibility, ease of visualisation and immune privilege status of the cornea make a gene editing strategy for the treatment of corneal dystrophies an attractive option. Off-target cleavage is a major concern for the therapeutic use of CRISPR/Cas9, thus current efforts in the gene editing field are focused on improving the genome-wide specificity of Cas9 to minimise the risk of off-target events. In addition, the delivery of CRISPR/Cas9 to different tissues is a key focus; various viral and non-viral platforms are being explored to develop a vehicle that is highly efficient, specific and non-toxic. The rapid pace and enthusiasm with which CRISPR/Cas9 has taken over biomedical research has ensured the personalised medicine revolution has been realised. CRISPR/Cas9 has recently been utilised in the first wave of clinical trials, and the potential for a genome editing therapy to treat corneal dystrophies looks promising. This review will discuss the current status of therapeutic gene editing in relation to the corneal dystrophies.

1. Introduction

Corneal dystrophies are a group of inherited, heterogeneous, bilateral disorders that affect the transparency or shape of the cornea (Klintworth, 2009). Historically, these dystrophies were sub-classified according to the corneal layer predominantly affected. Advances in genetic analysis and the completion of the human genome project gave researchers the capability to identify the causative genes (Shendure et al., 2017). These advances transformed our understanding of corneal dystrophies and revealed the extensive genetic heterogeneity that exists, leading to the necessity of a new classification system. In 2008, The International Committee for Classification of Corneal Dystrophies (IC3D) published a new classification system that aimed to preserve the traditional grouping while making way for the new era of genetic advancements; an updated version has since been published (Weiss et al., 2015).

The severity of the dystrophic phenotype can vary substantially, and therefore the treatment strategy required will need to be tailored to suit the individual patient accordingly (Klintworth, 2009). In some cases the corneal dystrophy can be asymptomatic and no treatment is required, while in other instances opacities which reduce visual acuity may result in complete loss of vision. Currently, corneal dystrophies are treated in a stage–related process (Seitz and Lisch, 2011). The decision of which treatment strategy will be most effective for the patient is made based on the current stage of the dystrophy. For milder cases, conservative therapies implemented include; gels/ointments, application of therapeutic contact lenses and/or conventional corneal abrasion (Seitz and Lisch, 2011). However, if these are not successful a surgical approach must be employed. The most effective surgical approach chosen will be based on the anatomical location of the opacities. Phototherapeutic keratectomy (PTK) can be considered for superficial dystrophies of the epithelium and, with less success, stromal dystrophies. However, it is not curative and in many cases the opacities may return (Chen and Xie, 2013; Dinh et al., 1999; Hafner et al., 2005). PTK is more often than not a temporary solution, it will likely require repeated treatments, with an ultimate goal of avoiding keratoplasty (NoRathi et al., 2016).

Unfortunately, in many instances sight deteriorates to the point

* Corresponding author. Biomedical Sciences Research Institute, University of Ulster, Coleraine, Northern Ireland BT52 1SA, United Kingdom.
E-mail address: t.moore@ulster.ac.uk (C.B.T. Moore).
[1] These authors contributed equally to this work.
[2] Percentage of work contributed by each author in the production of the manuscript is as follows: Prof Tara Moore = 35%; Miss Kathleen Christie = 35%; Prof John Marshall = 5%; Dr Andrew Nesbit = 25%.

https://doi.org/10.1016/j.preteyeres.2018.01.004
Received 23 November 2017; Received in revised form 19 January 2018; Accepted 22 January 2018
Available online 31 January 2018
1350-9462/

where a keratoplasty is required. Although corneal transplantation is well-established, drawbacks include the perpetual shortage of corneal donors and graft rejection following transplant. Despite advancements in ocular surgery no significant improvements in graft survival rates have been observed in the last 30 years (Bidaut-Garnier et al., 2016). The 5 year survival rate of grafts from 2004 to 2014 is documented to be 76.5%; however, when in the presence of risk factors, this value falls to 57.1% (Bidaut-Garnier et al., 2016). Predisposing factors for high-risk keratoplasty include a preoperative vascularised cornea caused by inflammation related to infection or chemical injury; this leads to a disruption in the immune privilege status of the cornea, allowing entry of immunologically competent cells (Arentsen, 1983; Hill, 1994). Another critical issue is that patients harbouring a causative *TGFBI* mutation see a re-emergence or in some cases a novel occurrence of mutant protein in the corneal graft (Aldave et al., 2007; Han et al., 2016; Jun et al., 2004; Kim et al., 2008). There will always be an element of risk in undergoing a surgical procedure, and as such, keratoplasty, whether lamellar or penetrating, is reserved as the absolute last resort (Seitz and Lisch, 2011). Unfortunately, in practice this means that nothing curative may be done for the patient until they are effectively blind. Due to the complete penetrance observed with many of the corneal dystrophies an approach that tackles the underlying genetic cause permanently, with a minimally invasive technique, seems a very attractive option.

Gene therapy seeks to treat genetic diseases by the introduction of foreign therapeutic DNA into a patient's cells. Recent advancements in the field of gene therapy, such as the development of new tools coupled with improvements in delivery, safety and efficiency; have accelerated the possibility of gene therapy reaching the clinic as a treatment, the eye is central to this genetic revolution. This movement has been pioneered by developments in retinal gene therapy; currently there are several on-going clinical trials for retinal diseases including; Leber congenital amaurosis (LCA), choroideremia, Usher's syndrome and Stargardt disease. These developments will be discussed in section 7.2 'Ocular clinical trials.'

The cornea offers the ideal candidate for targeted gene therapy due to its small surface area, accessibility and ease of visualisation. In addition, the cornea holds a unique immune privileged status which is critical for gene editing as it will minimise immune response to the gene-based therapeutics and delivery vehicles that are introduced (Charlesworth et al., 2018). Furthermore, as the cornea is avascular, any gene based therapeutics and delivery vehicle supplied to the cornea will not be able to reach other organs of the body minimising risk of off-target events in other tissues. Successful genome editing is reliant on i) strategic selection of a suitable gene therapy approach ii) efficient delivery to the targeted cell population iii) specific and efficient editing of the target gene in only the desired cell population. This review will discuss the current position of gene therapy in relation to the corneal dystrophies.

2. Genome engineering strategies

The concept of correcting disease status based on genetic information has fuelled decades of research. The most promising approaches for genetics based therapeutics that have emerged are; RNA interference (RNAi), gene augmentation and utilisation of genome engineering nucleases to achieve gene knockout or mutation correction. To-date the most utilised genome engineering strategy in ocular disease is that of gene augmentation, this will be discussed in sections 2.2 and 7.2.

2.1. RNAi

RNAi, first described in 1998 by Fire et al. (1998), utilises small interfering RNA (siRNA) molecules, 21 nucleotides in length, with complementarity to a specific gene's messenger RNA (mRNA) transcript. Upon target recognition the siRNA induces degradation of the complementary mRNA, preventing translation and protein expression (Fig. 2a). Currently, the database of clinical trials indicates that there are 40 ongoing clinical trials involving RNAi (RNAi clinical trials). At present 6 RNAi based therapeutic agents have progressed to phase 3 clinical trial stage (Sullenger and Nair, 2016). For example, vascular endothelial growth factor (VEGF) plays a well-established role in choroidal neo-vascularization (CNV), which leads to age-related macular degeneration (AMD). Bevasiranib, an siRNA targeted to the VEGF mRNA, reached phase 3 clinical trial, was administered by intravitreal injection every 12 weeks, almost 1/3 less frequently than current treatment options, such as Bevacizumab (Garba and Mousa, 2010). However, the trial was terminated as effects were not as potent as currently available therapies (Sullenger and Nair, 2016). It was shown that inhibition of Cas2, which is primarily activated by retinal ganglion cells, can prevent apoptosis in these retinal ganglion cells (Vigneswara et al., 2014). This interesting finding has now been translated to a potential therapy, QPI-1007, a siRNA targeted to Cas2, which is now in clinical trials to reduce retinal ganglion apoptisis in patients with Nonarteritic anterior ischemic optic neuropathy (NAION) (Sullenger and Nair, 2016).

2.2. Gene augmentation

Conventionally, gene therapy refers to the introduction of a functional copy of the gene to treat loss of function mutations, usually by viral transgene expression (discussed in detail in section 5.2 'Viral delivery to the eye'). The gene supplied by the virus allows the target cell to produce a functional protein in cases when the endogenous protein is defective (Fig. 2b). There are currently 2781 ongoing clinical trials for gene therapy listed on the clinical trial database, with 78 of these investigating gene therapy in eye diseases (Gene therapy clinical trials). Most ocular gene therapies tested to date target diseases of the retina, largely due to the fact that the majority of retinal diseases are caused by loss of function mutations (RetNet). For example, Leber's congenital amaurosis type II (LCA2), due to loss-of-function mutations in the *RPE65* gene, has been treated in three independent studies by delivery by single subretinal injection of the *RPE65* cDNA packaged in AAV2 (Bainbridge et al., 2008; Bennett et al., 2016; Hauswirth et al., 2008; Maguire et al., 2008). Improvement in vision that was stable for at least 3 years was observed in each study. The current stage of this therapeutic will be discussed in section 7.2 'Ocular clinical trials'.

2.3. Programmable nucleases

Programmable nucleases provide tools to manipulate the genome in a sequence specific manner, they consist of a nuclease that can be reprogrammed to cleave at a precise target sequence. They facilitate precise genome editing by inducing a double strand break (DSB) at a desired location. The cellular responses initiated to repair this damage are either non-homologous end-joining (NHEJ) or homology directed repair (HDR). Depending on which of the cellular responses that is employed, different modes of genome editing, such as gene knockout or gene correction, can be achieved. Which is discussed in detail in section 2.3.2 'Types of therapeutic genome modifications with CRISPR/Cas9'.

There are currently four classes of programmable nucleases that have been utilised: meganucleases (Belfort and Bonocora, 2014; Stoddard, 2011), zinc finger nucleases (ZFNs) (Urnov et al., 2010), transcription activator-like effector nucleases (TALENs) (Bogdanove and Voytas, 2011) and clustered regularly interspersed palindromic repeats (CRISPR) associated nuclease, Cas9 (Fig. 2 c–f). Although all of these programmable nucleases cause a DSB which mediates genome editing, the mode by which they achieve target recognition and their specific limitations differ, influencing which nuclease is most applicable for a given situation.

One of the major considerations is the ease with which a nuclease can be engineered for a specific target. For instance, meganucleases and ZFNs require extensive protein engineering, while CRISPR/Cas9 can be easily redirected with simple molecular cloning techniques. Due to this,

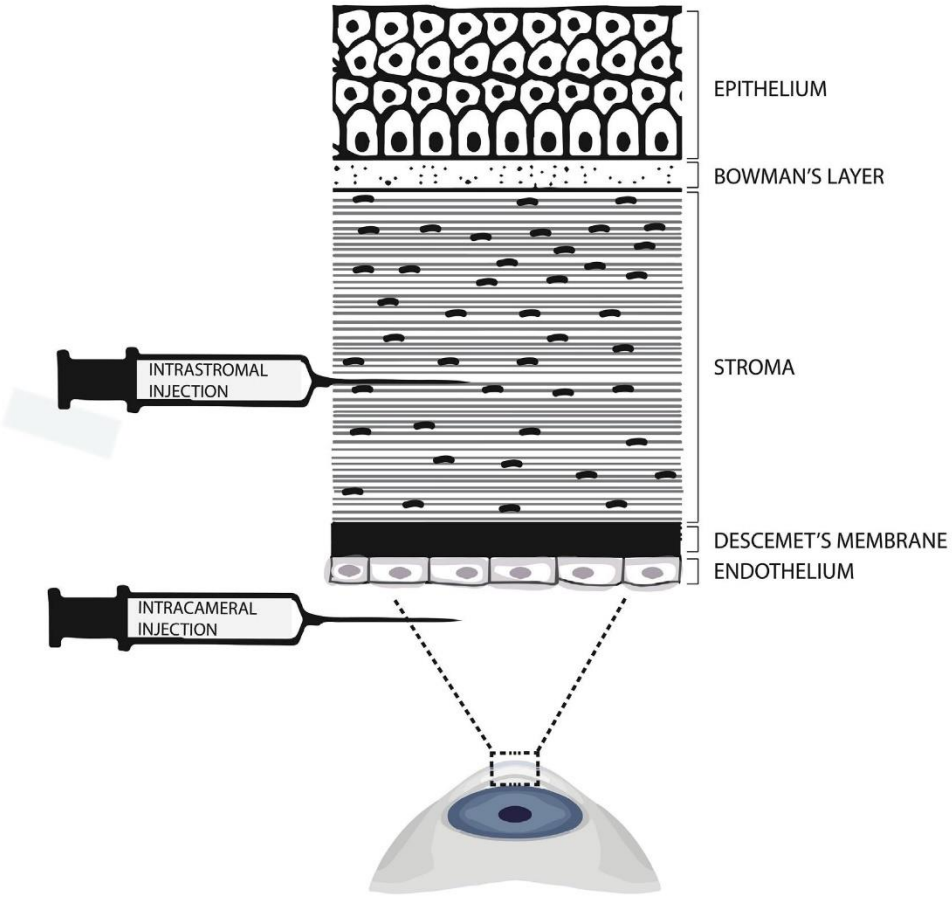


Fig. 1. Schematic of the corneal layers, indicating the location of an intrastromal injection into the stromal layer and a intracameral injection into the anterior chamber.

the CRISPR system has rapidly eclipsed the other programmable nucleases as the genome editing tool of choice. However, CRISPR/Cas9 has limitations also. For instance, due to the large size of the coding sequence of Cas9 (~4kbp), packaging into the most commonly used delivery vector, adeno-associated virus, poses substantial difficulties. Table 1 highlights the core differences between each nuclease. Cas9 has dominated the frontier of genome editing in recent years, and this review will focus mainly on Cas9 as the genome editing tool of choice.

2.3.1. CRISPR/Cas9

CRISPR/Cas systems were discovered in bacteria where they act as an adaptive immune response (Karginov and Hannon, 2010; Makarova et al., 2011). They can be divided into 2 classes and then further subdivided into 5 types (Makarova et al., 2011). Class 1 require several effector proteins whilst class 2, consisting of type II and type V, is characterised by having a single large effector protein. The large effector protein of type II systems is Cas9; whilst that of type V is Cas12a, formerly known as Cpf1. Due to the minimal components required class 2 systems have now been manipulated for use in mammalian cells (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013b; Zetsche et al., 2015).

Both Class 2 effector proteins, Cas9 and Cas12a, act as RNA-guided endonucleases (Jinek et al., 2012; Zetsche et al., 2015). The *S. pyogenes* type II CRISPR locus consists of 4 genes and 2 non-coding CRISPR RNAs (crRNAs); trans-activating crRNA (tracrRNA) and precursor crRNA (pre-crRNA). In bacteria it was shown that the essential components required to reprogramme Cas9 only consisted of a dual tracrRNA:crRNA (Jinek et al., 2012). To apply this mechanism in mammalian cells the necessary genes were codon-optimised and nuclear localisation signals were attached (Cong et al., 2013; Mali et al., 2013b). In addition, the dual tracrRNA:crRNA was optimised to include the mature crRNA fused to a truncated tracrRNA, now termed a single guide RNA (sgRNA) (Jinek et al., 2012; Mali et al., 2013b).

To achieve sequence specific edits in mammalian cells, all that is required is either a Cas9 or Cas12a nuclease and a sgRNA. The sgRNA for type II systems is a 100mer consisting of a pre-crRNA:tracrRNA fusion (F Ann Ran et al., 2013a,b), while a sgRNA for a type V system does not require a tracrRNA and is only a 42-44mer (Zetsche et al., 2015). The RNA-guided endonuclease can be redirected to a desired target sequence simply by altering the 20 nucleotide spacer sequence within the sgRNA, provided it lies directly upstream of a protospacer adjacent motif (PAM) recognised directly by the Cas9 protein (Jinek et al., 2012; Mojica et al., 2009); Cas9 or Cas12a nucleases from different bacterial species recognise different PAMs. The most characterised type II systems are *S. pyogenes* Cas9, which recognises a 5' – NGG – 3′ PAM and *S. aureus* Cas9, which recognises a 5' – NNGRRT – 3′ PAM (Jinek et al., 2012; Mojica et al., 2009), while the type V systems tend to have more AT rich PAMs, with the most characterised being *Acidaminococcus sp.*(AsCas12a) and *Lachnospiraceae bacterium* (LbCas12a) which both recognise a 5′- TTTN – 3′ PAM (Zetsche et al., 2015).

PAM recognition is the initial step in Cas9 binding to a target site, the PAM interacting domain samples the PAM and once the correct PAM has been identified the PAM proximal region of the guide will

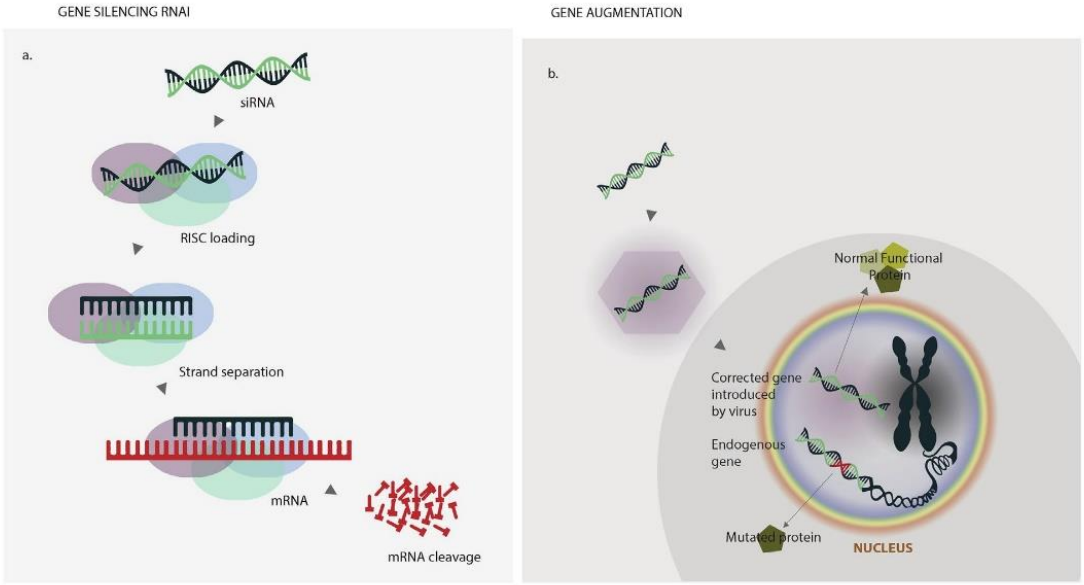


Fig. 2. a) RNA interference (RNAi): Upon entering the cell, small-interfering RNA (siRNA) becomes loaded onto the RNA-induced silencing complex (RISC), siRNA is separated into two strands (sense shown in light green and anti-sense shown in dark green), the sense strand is degraded and the anti-sense pairs with a complementary sequence in the target messenger RNA transcript (shown in red), initiating cleavage of the mRNA transcript. b) Gene augmentation: The wild-type copy of the cDNA sequence is packaged into a viral vector (shown by the purple hexagon), the viral vector containing the WT cDNA transduces the cell and is transported to the nucleus. The WT cDNA introduced by the virus is translated and produces a normal functional protein, in place of the mutated protein produced by the endogenous gene. c) Meganucleases: Achieve DNA recognition in different ways, the most common mechanism is to dock into the major grooves of their DNA target sites via antiparallel B sheets and using a series of sequence-specific and non-specific contacts. d) Zinc-finger nucleases (ZFNs): Each ZFN pair consists of 4 zinc finger binding motifs and each zinc finger binding motif recognises 3 base pairs. Each zinc finger binding motif recognises a different combination of base pairs, upon dimerization of the ZFN pairs the Fok1 endonuclease becomes active and generates a double strand break. e) Transcription activator like endonucleases (TALENs): Each TALEN pair typically recognises 16 base pairs; each DNA binding domain is designed using an array of single repeats based on their individual binding preferences. The TALEN pairs are fused to a FokI endonuclease, upon dimerization of the TALEN pairs the Fok1 endonuclease generates a DSB. f) CRISPR/Cas9: An RNA-guided endonuclease is composed of a Cas9 nuclease (show in grey) and a sgRNA (shown in purple). The Cas9 scans the genome in search of a protospacer adjacent motif (PAM) once a PAM is encountered the sgRNA will direct binding of a Cas9 and once bound Cas9 will generate a DSB.

direct target binding (Anders et al., 2015; Cencic et al., 2014; Sternberg et al., 2014). As the PAM proximal region dictates whether Cas9 binds to the target site, mismatches in this region are critical in determining specificity of Cas9 (discussed in detail in section 3.3 'Role of guide sequence in the specificity of Cas9'). If the PAM proximal region is complementary to the guide sequence, the PAM distal region will then bind. Although complementary between the PAM distal region and guide sequence is required, reports have shown mismatches here are better tolerated than in the PAM proximal region (Pattanayak et al., 2013; Sternberg et al., 2014; Wu et al., 2014b), again this will be discussed in section 3.3. As *S. pyogenes* Cas9 is the best characterised and most frequently used system in CRISPR research the remainder of the review will focus on this system unless stated otherwise.

2.3.2. *Types of therapeutic genome modifications with CRISPR/Cas9*

Once a double strand break is induced by a programmable nuclease, the cell will initiate one of two downstream repair pathways to resolve the break, Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR) (Jasin and Haber, 2016).

2.3.2.1. Gene disruption via NHEJ. NHEJ involves re-ligation of the two ends of the DSB in a non-precise way (F Ann Ran et al., 2013a,b). Although NHEJ often results in clean repair of the lesion, when a DSB is continually generated at the same region errors can be introduced (Richardson et al., 2016). These errors can consist of insertions or deletions (indels) of various sizes. Initially, it was thought these indels occurred in a random fashion, but it has recently been shown that each site generates a distinctive pattern of indels (van Overbeek et al., 2016). Interestingly, these 'indel fingerprints' are due to the guide sequence employed rather than the genomic context. These indels may result in a frameshifting mutation which can lead to mRNA degradation by nonsense mediated decay or they may generate a premature stop codon which may lead to a non-functional truncated protein (Hentze and Kulozik, 1999). In both cases, indel generating NHEJ can result in gene disruption and this can be used to knockout gain-of –function mutations (Fig. 3a). NHEJ has been shown to function at all stages in the cell cycle but is most active during G1, therefore it is possible to exploit gene disruption via NHEJ in both dividing and non-dividing cells, such as the corneal endothelium or retinal photoreceptors (Iyama and Wilson, 2013; Straub et al., 2014).

2.3.2.2. Gene correction via NHEJ. In addition, the multiplexing capability of CRISPR/Cas9 can be used to achieve NHEJ mediated gene correction (Cox et al., 2015). (Fig. 3b) Some diseases are due to a trinucleotide repeat expansions, for example Huntington's or Friedreich's Ataxia. In these cases, the copy number of trinucleotide repeats exceeds the threshold number and a disease phenotype manifests. It is possible to generate multiple DSBs flanking the trinucleotide repeat expansion, and remove it, restoring normal function (Cong et al., 2013; Tabebordbar et al., 2016).

2.3.2.3. Gene correction via HDR. HDR is a much more precise mechanism of repair, in which the DSB is repaired using a homologous repair template. This mode of repair can be used to achieve gene correction (Fig. 3c). In normal cellular repair, homologous recombination will utilise the homologous chromosome as the repair template. However, it is possible to introduce an exogenous DNA repair template, in large excess, to encourage repair mechanisms to introduce desired sequence changes. The repair template has sequence identity to the regions flanking the break site

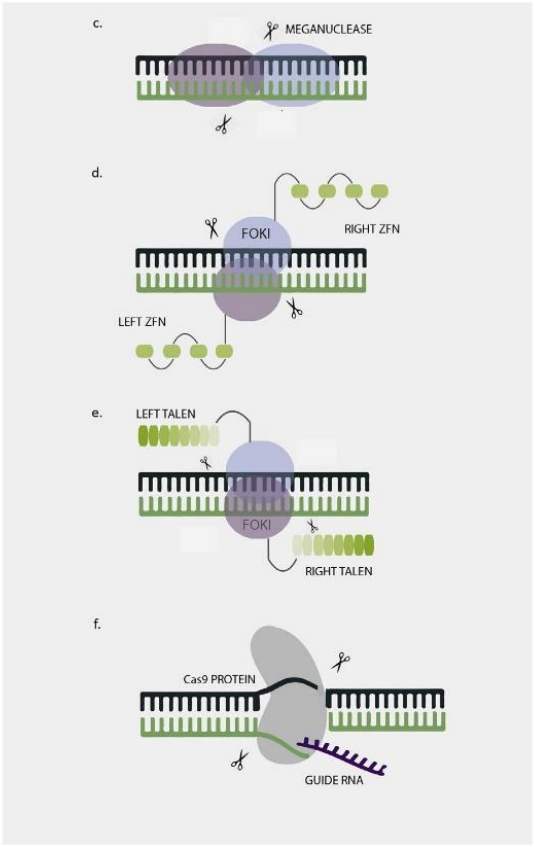


Fig. 2. (*continued*)

(Lombardo et al., 2011). Examples of genomic safe harbours are the adeno-associated virus site (AAVS1) (Kotin et al., 1992), the chemokine (CC motif) receptor 5 (*CCR5*) gene locus (Lombardo et al., 2011) and the human orthologue of the mouse ROSA2g locus (Irion et al., 2007). (Fig. 3d) Gene augmentation has previously been achieved using integrating viruses. It has been proposed that HDR be utilised to introduce the wild-type gene into a genomic safe-harbour locus following CRISPR/CAs9-mediated DSB, and thus this approach would provide a more controlled method of gene introduction.

2.3.2.5. Exciting additions to the CRISPR toolkit. The generation of a catalytically inactive Cas9 (dCas9) provided an invaluable tool to transport proteins regulating gene expression in a sequence specific manner by changing the guide RNA sequence (Jinek et al., 2012). dCas9 has been used to transport a host of proteins including both transcriptional activators and inhibitors (Gilbert et al., 2013; Konermann et al., 2014; Maeder et al., 2013; Mali et al., 2013a; Perez-Pinera et al., 2013), chromatin modifying enzymes (Hilton and Gersbach, 2015; Kearns et al., 2015; Liu et al., 2016) and fluorescent proteins (Chen et al., 2013; Ma et al., 2016). While all of these tools are of interest to scientific research the development of base editors and RNA editors by a similar approach offer potential benefit to genome engineering therapeutics. Komor et al. developed a fusion of dCas9 with a cytidine deaminase enzyme that facilitates the conversion of cysteine to uracil in the absence of a DSB. This allows base editing of C to U, or G to A. Provided the flanking sequence is suitable, this base editor offers a tool to correct missense mutations associated with human disease (Komor et al., 2016). Recently, the same group reported the development of Adenine Base Editors (ABE) which have the capability of converting A > G and T > C (Gaudelli et al., 2017). This was achieved by fusing an evolved tRNA adenosine deaminase (TadA) to dCas9, wild-type TadA is a adenine deaminase that converts adenine to inosine (I) in the single-stranded anticodon loop of tRNA. Wild-type TadA was evolved through mutagenesis to generate an adenine deaminase that acts on DNA. Finally, an exciting new development includes RNA editing for programmable A to I (G)

Table 1
Core differences between the four different programmable nucleases.

	Meganucleases	ZFNs	TALENs	Cas9
Mode of DNA recognition	Protein:DNA interactions	Protein:DNA interactions	Protein:DNA interactions	guideRNA:DNA and Protein:DNA interactions
Recognition Site	20-30bp	18-36bp/ZFN pair	30-402bp/TALEN pair	~ 20bp guide sequence and adjacent PAM
Ease of engineering	Substanial protein engineering required	Intricate molecular cloning methods required	Substantial protein engineering required	Simple molecular cloning techniques required
Size	1kb	~1kb X2	~3kb X2	4.2kb (S.pyogenes Cas9)
Specificity	Small number of positional mismatches tolerated	Small number of positional mismatches tolerated	Small number of positional mismatches tolerated	Positional and multiple consecutive mismatches tolerated

and is introduced into the cell along with the programmable nuclease. Once a DSB has been generated, the exogenous DNA repair template is utilised to direct repair of the lesion. Any nucleotide changes included in the exogenous repair template will then be incorporated in the repaired allele. This powerful mechanism can be used to correct point mutations or introduce protective mutations. However, as homologous recombination is restricted to the S and G2 phase of the cell cycle, gene correction via HDR is limited to dividing cells (Iyama and Wilson, 2013).

2.3.2.4. Gene addition via HDR. A genomic safe harbour locus is a region of the genome that can facilitate the predictable expression of integrated DNA without having any adverse effects on the host cell, such as malignant transformation or changing cellular functions

replacement (REPAIR) (Abudayyeh et al., 2017). They used a Type VI CRISPR/Cas system which, unlike Type II systems, has a single-effector RNA-guided RNase – Cas13. They fused catalytically inactive Cas13 (dCas13) to an inosine deaminase, as inosine is functionally equivalent to guanosine they were able to robustly achieve RNA editing in mammalian cells. These advances offer novel solutions for the treatment of genetic disease.

3. Maximising on-targets and minimising off-target

3.1. Guide selection

Early reports of CRISPR/Cas9 suggested that Cas9 can be redirected to cut any sequence in the genome provided it is directly upstream of

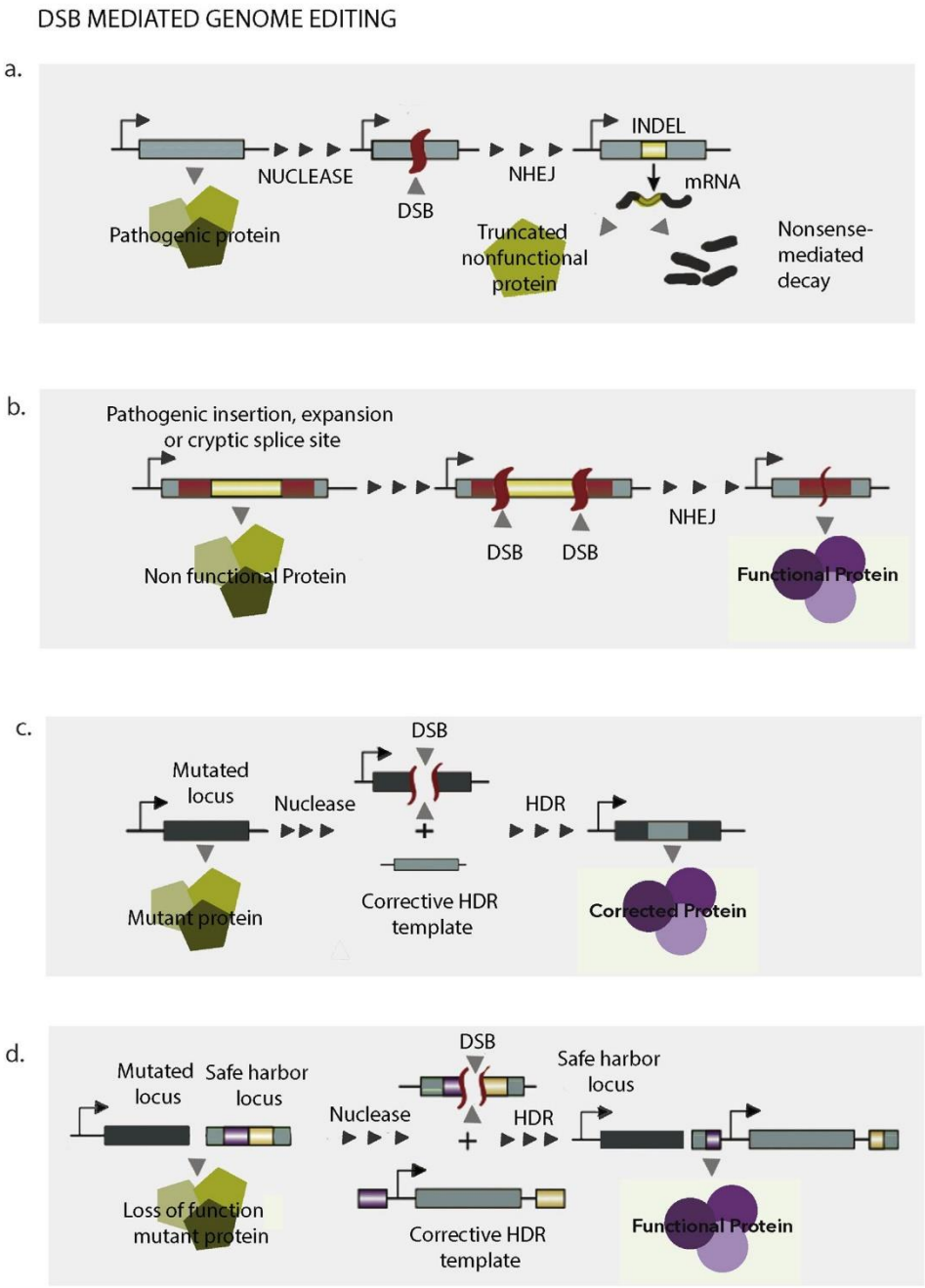


Fig. 3. a) Gene disruption via NHEJ: A pathogenic protein is silenced by targeting the gene locus and introducing indels, these indels can act to cause a frameshifting mutation that may cause a premature stop codon; this will lead to either a non-functional truncated protein or nonsense mediated decay. b) Gene correction by NHEJ: When a cryptic splice-site, pathogenic insertion or deletion is present in an intronic region (indicated in red) two DSBs can be generated either side of the region, which can result in a deletion of the sequence between the two DSBs, removing the pathogenic region. c) Gene correction by HDR: HDR can be used to correct disease causing mutations, a DSB is induced near the mutation and a repair template is provided with the correct sequence, HDR repair machinery can use the repair template rather than the homologous chromosome to repair the lesion. d) Gene addition via HDR: An alternative approach to gene augmentation is to introduce the therapeutic gene into a safe-harbour locus. A DSB is generated in the safe harbour locus and a repair template containing the therapeutic gene with flanking arms that have homology to the safe harbour locus is also introduced. Integration of the therapeutic gene into the safe harbour locus also allows a functional protein to be produced.

PAM. However, it rapidly became apparent that was not the case. Although a PAM site is obligatory, it does not guarantee gene editing can be achieved at that position. There are many considerations for choosing a target site, with a well-designed guide being fundamental to achieving a successful experimental outcome.

It is well documented that the activity of Cas9-sgRNA combinations can vary greatly between i) genomic loci and ii) cell type (Kim et al., 2014; Lin et al., 2014; F. Ann Ran et al., 2013a,b; Richardson et al.,

2016; Yang et al., 2013). It is absolutely critical that guides with high cleavage efficiency are selected for each application, especially if considering a clinical setting, as highly active guides will be required to achieve a therapeutic benefit. Recent work by our group highlights the necessity in meticulous guide design in determining both on-target efficiency and specificity (Christie et al., 2017), which will be discussed in detail in section 3.3.

3.2. Off-targets

The transition of CRISPR/Cas9 towards the clinic as a therapeutic is being made with caution. Unlike RNAi, which has a transient effect, programmable nucleases introduce permanent changes in the genome. This means that researchers need to be absolutely certain that the changes they are introducing are not causing any unwanted effects by off-target cleavage. Off-target is a broad term to describe binding or cleavage by Cas9 at a site other than the target site. The main area of concern is the specificity of Cas9 and much research has sought to elucidate both the off-target profile of Cas9 in addition to the most accurate way to determine off-target cleavage.

3.3. Role of guide sequence in the specificity of Cas9

Initial genome-wide analyses with CRISPR/Cas9 found that Cas9 had variable specificity and was capable of off-target cleavage at sequences with close homology to that of the target guide sequence (Cho et al., 2014; Cradick et al., 2013; Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013b; Pattanayak et al., 2013). These reports found that mismatches in the PAM proximal region were critical in determining specificity, while mismatches in the distal region had less impact on the specificity. The PAM proximal region has now been coined as the 'seed region'. While there is no uniformity on the exact length of the seed region, reports collectively suggesting it is the first 5-12bp of the guide sequence (Pattanayak et al., 2013; Sternberg et al., 2014; Wu et al., 2014b). These initial screens where carried out with limited knowledge of optimal guide design. Building on these original findings, Doench et al. generated an algorithm to define the 'rules' of guide design and reduce off-target cutting and maximise on-target activity (Doench et al., 2016, 2014). Several online tools have been generated based on these 'rules' that allow *in silico* assessment of on and off target cleavage of each individual guide sequence. These tools, such as Benchling, Synthego, MIT CRISPR, E-CRISP, DESKGEN, etc. generate a relative score of the chance that the chosen guide sequence will cleave at unintended sites. The programs search the genome for sites with homology to the guide sequence and provide an output score which takes into account the total number of mismatches in the similar sequences, whether the mismatches are consecutive and critically if the sites with homology to the guide sequence are in a coding or non-coding region. In addition to these basic parameters, these algorithms apply findings from large datasets to confer which guide sequences are likely to have low cleavage activity. With the new insight into optimal guide design several genome wide analyses have been conducted that have concluded minimal off-target cleavage occurs when care is taken to avoid promiscuous guides (Frock et al., 2014; Iyer et al., 2015; Kim et al., 2015; Tsai et al., 2017, 2014; Veres et al., 2014). When care in guide design is taken to prevent unwanted cleavage, CRISPR/Cas9 can be highly specific.

3.4. Determining off-target events

As an off-target event is likely to be an extremely rare event, whole genome sequencing of a population of targeted cells will not likely be able to detect it, as it will be present at such a low frequency not detectable at the read depths usually employed for WGS (Wu et al., 2014a). To identify off-target cleavage a method that is both sensitive and unbiased is required. To overcome this obstacle, several groups have developed innovative approaches to detect these rare events. These methods include; direct *in situ* breaks labelling, enrichment on streptavidin and next-generation sequencing (BLESS), genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-Seq), circularization for *in vitro* reporting of cleavage effects by sequencing (CIRCLE-Seq), *in vitro* Cas9-digested whole-genome sequencing (Digenome seq) and linear amplification–mediated high-throughput genome-wide sequencing (LAM-HTGTS) and selective enrichment and identification of tagged genomic DNA ends by sequencing (SITE-Seq) (Cameron et al., 2017; Crosetto et al., 2013; Frock et al., 2014; Hu et al., 2016; Kim et al., 2015; Tsai et al., 2017, 2014). A reflection from Zischewski et al. notes 'the few off-target mutations that occur in an experiment with a carefully selected gRNA are far less abundant than the spontaneous mutations that occur during the clonal expansion of cell cultures.' (Zischewski et al., 2017) While this may be the case as these infrequent off-targets will either be in living cells in a patient or transplanted into a patient, utmost care must be taken to ensure off-target events are minimised.

3.5. Rationally engineering Cas9

All of the findings described above modified specificity by guide selection and utilised wild-type *S. pyogenes* Cas9. Several groups hypothesised that the Cas9 protein itself could be engineered to improve the specificity of Cas9 while having no effect on on-target cleavage. Kleinstiver et al. proposed that the Cas9-sgRNA complex had more binding energy than it needed to carry out cleavage at on-target sites (Kleinstiver et al., 2016). Based on structural studies, they hypothesised if they disrupted the direct hydrogen bonds between the Cas9 and the phosphate backbone of the target DNA strand they could weaken the energetics of this complex to an extent that destabilises mismatched binding and eliminate off-target cleavage. Simultaneously, Slaymaker et al. proposed that they could weaken non-target strand binding by neutralising residues within the non-target strand groove and thus minimise off-target cleavage (Slaymaker et al., 2016). These hypotheses led to generation of high-fidelity Cas9 (SpCas9-HF1) and enhanced Cas9 (eSpCas9) respectively. In both cases SpCas9-HF1 and eSpCas9 saw a marked reduction in off-target cleavage when compared to wild-type *S. pyogenes*. A recent report by Chen et al. indicates that these original hypotheses were incorrect and that, in both cases, the improved specificity was unrelated to target binding (Chen et al., 2017). Remarkably, even though SpCas9-HF1 and eSpCas9 have alanine substitutions in different residues in distinct regions of the Cas9 protein it appears both variants reduce off-target cleavage by raising the threshold for conformational activation of the HNH nuclease domain, rather than affecting target binding. The continued research into the mechanism behind cleavage Cas9 will undoubtedly allow us to gain a much better appreciation for how off-target cleavage occurs and how to eradicate it completely.

3.6. Allele-specific genome editing

Our research focusses on developing an allele-specific genome editing strategy to treat corneal dystrophies caused by dominant negative missense mutations. Our primary aim is to cause gene disruption by NHEJ at only the mutant allele leaving the wild-type allele intact. This strategy is unique in that we can test both the efficiency and specificity of Cas9 by investigating the wild-type allele in comparison to the mutant allele.

Previously, we reported an innovative approach to achieve this by employing a PAM generated by the missense mutation itself. This approach exploits the novel PAM on the mutant allele and a guide RNA is designed utilising this novel PAM; there is no PAM present on the wild-type allele, therefore Cas9 should only be able to bind to and cleave the mutant allele, thus allele-specificity should be achieved. We have successfully shown this *in vitro* and *in vivo* in mice, using the L132P mutation in *Krt12* that causes Meesmann's corneal dystrophy (MECD)

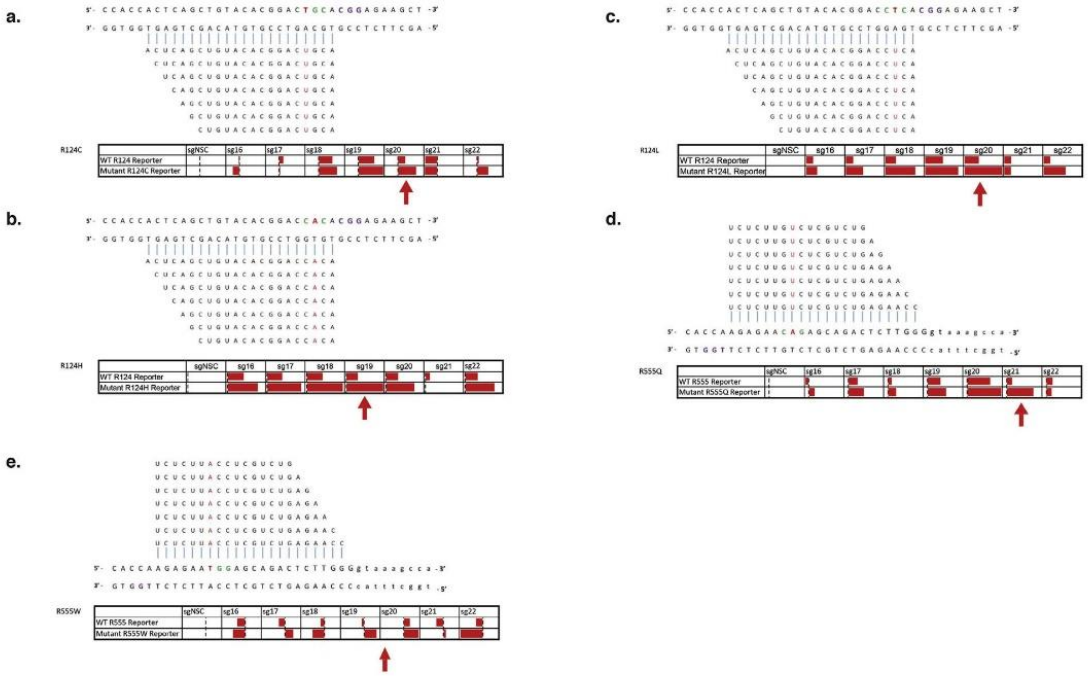


Fig. 4. Reports have indicated that truncating the length of the matching sequence within the guide from 20 to 18 nucleotides can reduce genome-wide off-target cutting, while maintaining on-target efficiencies. An assessment of the effect of guide-length upon specificity between the wild-type and mutant alleles, using a dual-luciferase assay, was conducted for the 5 most prevalent TGFBI mutations. Reports have shown that guide lengths < 16nt abolish cleavage activity. For each mutation a range of guide lengths from 16 to 22 nucleotides were tested, each guide was targeted to the wild-type and respective mutant sequence and the firefly luciferase activity was measured as an indicator of specificity.
For all mutations investigated, the truncated guides did not provide a marked improvement of specificity; for most cases maximal discrimination occurred with guides 20 or 19 nucleotides in length. For R124C, a 20 nt guide seemed to confer allele-specificity, however, no other guide length offered any adequate discrimination (Fig. 4a). In the case of the R555Q mutation, guides in the 18–20 nt range did not offer sufficient discrimination, although, interestingly, the 21 nt guide provided convincing allele-specificity(Fig. 4d). R555W did not offer any considerable allele-specificity for any length tested (Fig. 4e). R124H and R124L displayed clear allele-specific cleavage, especially in the 18–20 nt sgRNA range, with minimal cutting of the wild-type sequence (Fig. 4b and c). Interestingly for the R124 mutations guide lengths of 21 nt seemed to impair cleavage activity in all cases.

(Courtney et al., 2016).

Our recent research interests include investigating an allele-specific system to treat the *TGFBI* corneal dystrophies (Christie et al., 2017). To-date 60 missense mutations have been identified in *TGFBI* that cause a variety of corneal dystrophies. We focussed on developing an allele-specific system to treat the 5 most prevalent mutations, which account for the majority of cases of *TGFBI* corneal dystrophies. None of these 5 mutations generate a novel PAM, however the missense mutation is present in the seed region of the guide sequence in all cases. Therefore, we investigated if a guide-specific approach could confer stringent allele-specificity.

We found that S. *pyogenes* Cas9 was not able to adequately discriminate between wild-type and mutant alleles differing in sequence at only one base of the guide sequence. Reports which indicated that truncating the guide RNA can improve specificity, led us to perform a screen of guide-length from 16 to 22bp (Fig. 4). However, no significant improvement in specificity was observed. Interestingly, we found that mismatches closer to the PAM conferred more discrimination between alleles that PAM distal mismatches, confirming previous reports (Pattanayak et al., 2013; Sternberg et al., 2014; Wu et al., 2014b). However, even though PAM-proximal mismatches result in superior discrimination they not confer enough specificity to be suitable for gene therapy.

3.7. Improving efficiency of HDR

It has been shown that the format in which the Cas9 and sgRNA are introduced into cells can have a significant effect on efficiency (DeWitt et al., 2017; Kim et al., 2014). Ribonucleoprotein complexes (RNPs) of purified Cas9 protein and guide RNA have been shown to achieve higher efficiencies *in vitro* and *in vivo* (Kim et al., 2017; Liang et al., 2015). The mechanism of increased efficiency with RNPs is not well understood. It may be due to controlled complexing of the Cas9 protein and sgRNA *in vitro*, the Cas9 protein sheltering the sgRNA from degradation or the reduced cellular toxicity associated with introduction of DNA into cells (Liang et al., 2015). However, it is important to note that whether a plasmid expression system or RNP complex is utilised, a poorly designed guide sequence will perform with low efficiency.

Frequency of HDR mediated gene correction is characteristically low in comparison to indels generated by NHEJ (Wang et al., 2013). Many groups have sought to overcome this by either inhibiting the NHEJ pathway or enhancing HDR itself. The introduction of Scr7, an inhibitor of DNA Ligase IV which is required for NHEJ, has been shown to increase HDR *in vitro* and *in vivo* in mice (Maruyama et al., 2016). Similarly, groups have tried to inhibit NHEJ by introducing shRNAs targeted to DNA ligase IV and also KU60, another protein essential to NHEJ (Chu et al., 2015). In addition, the introduction of RS-1 an enhancer of HDR has been shown to result in a 2–5 fold increase of HDR *in vitro* depending on the locus and an increase from 14 to 33% HDR *in vivo* (Song et al., 2016). Finally, it was reported that using chemically modified guides to evade degradation by nucleases such as phosphodiesterase, to form the RNP complexes caused a marked increase in both NHEJ and HDR efficiencies (Hendel et al., 2015).

Dewitt et al. demonstrated an impressive 33% sequence correction when sickle cell disease (SCD) patient $CD34^+$ hematopoietic stem/progenitor cells (HSPCs) were treated *ex vivo* with RNPs and a single stranded oligonucleotide (ssODN) containing the corrected SNP, without applying selection (DeWitt et al., 2016). When these corrected

human HSPCs were engrafted into an immunocompromised mouse, gene corrected cells were still present after 16 weeks. This offers an encouraging result for gene therapy via HDR mediated gene correction. 13% gene correction in induced pluripotent stem cells (iPSCs) derived from a patient with a mutation in the retinitis pigmentosa GTPase regulator (RPGR) gene, which causes an X-linked variant of RP (XLRP), has also been demonstrated, using plasmid expression constructs (Bassuk et al., 2016). Autologous cell transplantation of corrected cells is a viable treatment option for retinal dystrophies (Wiley et al., 2015), which will be discussed in section 6 'Stem cell culture and transplantation'. It seems that improvements to the efficiency of HDR-mediated repair means that this technique will soon be more widely applicable to achieve therapeutic benefit.

4. Choice of therapeutic approach based on the genetics

The various types of genome modification that can be achieved via the CRISPR/Cas9 system mean that a host of genetic mutations will be targetable. The majority of corneal dystrophies are the result of a missense mutations that present with an autosomal dominant inheritance pattern - in which only one mutant allele is required for a disease to manifest. Others exhibit autosomal recessive inheritance, where two mutant alleles are required for the disease phenotype to develop, while a few corneal dystrophies present with X-linked inheritance, where transmission of the mutant gene is from the X chromosome.

To date, 23 distinct corneal dystrophies have been described. The inheritance pattern is known for 20, the genetic locus has been successfully mapped in 17 and the causative gene has been definitively identified in 10, while the causative gene has been elucidated in some cases in a further 4. Table 2. The Mendelian inheritance patterns of corneal dystrophies - namely their monogenic and highly penetrant nature, coupled with the known pattern of inheritance in most cases, make them an ideal target for genome editing therapeutics, particularly when we consider the accessibility of the cornea. Devising a gene-based therapeutic offers a promising treatment strategy for these diseases.

As discussed, depending on the type of repair employed, different forms of gene editing can be achieved. We can use this information to generate edits that are most appropriate for the underlying mutation, Fig. 5- roadmap to gene therapy, depicts the considerations one must make before choosing the most fitting genome editing strategy for one's mutation of interest.

5. Delivery to the eye

5.1. Considerations

Delivery of gene therapy components to the different corneal layers poses a substantial problem. As the outermost surface of the eye, the cornea has many protective adaptations that prevent entry of the gene therapy agent. The cornea is covered by a lachrymal film which hydrates, cleanses and lubricates the eye (de la Fuente et al., 2010). In addition to the tear film's functional roles, it acts as the first barrier to foreign substance entry into the eye. These substances mix with the lacrimal fluid and are drained into the lacrimal sac, resulting in rapid precorneal loss (Le Bourlais et al., 1998). The next barrier the foreign substances encounter is the cornea epithelium in which the cells are linked by characteristically impermeable tight junctions (Hämäläinen et al., 1997). This severely limits transcorneal entry of therapeutic agents, not only to the posterior segment of the eye, but also to each of the corneal layers. In addition to the biological barriers that exist, the entry of DNA itself into cells is challenging. DNA is hydrophilic, large in size and negatively charged, so penetration across the cell membrane is difficult (Oliveira et al., 2017). Substances can pass through the epithelium via the transcellular pathway or paracellular pathway. Hydrophilic molecules more readily pass along the paracellular pathway, and thus, DNA molecules will have to navigate through the exceptionally tight junctions of the corneal epithelium (Romanelli et al., 1994).

Due to the biological barriers discussed above topical application of therapeutic agents does not offer a promising mode of delivery, as it results in rapid precorneal loss and inefficient corneal penetration. However, it has been shown that topical application following debridement of the corneal epithelium, which removes the initial biological barriers, can allow passage of viral vectors (Spencer et al., 2000). Local administration via intrastromal or intracameral injection (Fig. 1) of gene therapy agents, such as naked DNA, siRNA, or AAV, circumvents the problem of biological barriers. Although effective in the small mouse eye, where the pressure achieved during injection can force DNA into cells (Matthaei et al., 2012), when scaled up to a human eye, entry of naked DNA into cells is not particularly effective - due to the reduced pressure, in addition to the large size and negative charge of DNA. Therefore, many alternative avenues of delivery have been explored, these different delivery vehicles can be broadly sub-divided into viral and non-viral approaches and are summarised in Table 3 (Deyle and Russell, 2009).

An ideal vehicle for gene therapy should have a cargo capacity large enough to packaging all necessary CRISPR/Cas9 components, cell tropism for the specific cell population of interest, high nucleus targeting efficiency, produce a negligible immune response, in addition to minimal genotoxicity and cytotoxicity. As we will show in this section, there is currently not a single mode that possesses all of the above qualities. Regardless of the vector utilised, it must be able to evade detection by the host's immune system and have efficient uptake by the target tissue, persisting long enough to elicit a therapeutic effect but not long enough to cause off-target effects.

5.2. Viral delivery to the eye

Viral vectors utilised for gene therapy include adenoviruses, adeno-associated viruses (AAV) and retroviruses, specifically lentivirus. Each vector has been extensively tested for use in corneal transduction (Mohan et al., 2012, 2005). No single vector offers all desirable criteria; adenovirus boasts a large cargo capacity but initiates a substantial immune response (Borrás et al., 2001; Tsubota et al., 1998), AAV provides efficient transduction with a minimal immune response but has a limiting packaging capability (Sharma et al., 2010; Zinn and Vandenberghe, 2014) and lentivirus has a considerable packaging capacity but can cause unwanted integration (Athanasopoulos et al., 2017). Table 3 summarises the immediate differences between each vector. As AAV has been the leading vector of choice for gene therapy, particularly ocular gene therapy, this review will focus on AAV as the viral vector of choice for delivery.

5.2.1. Adeno-associated virus (AAV)

AAV has a non-enveloped single-stranded DNA (ssDNA) viral genome 4.7kb in size. There are currently 12 naturally-occurring human serotypes, with AAV-2 being the most widely used as the vector backbone. AAV is a *Dependovirus*, meaning lytic infection by AAV requires a helper virus; this low level of pathogenicity makes it an attractive vehicle for gene therapy. The viral genome is relatively simple, consisting of two genes - *Rep* and *Cap*, flanked by two inverted terminal repeats (ITRs).

Current recombinant AAV (rAAV) vectors have both *Rep* and *Cap* genes removed from the backbone, leaving only the ITRs which are required for packaging (Choi et al., 2005). Different AAV serotypes have tropisms for different tissues due to the capsid proteins present on their surface. AAV-2 is the most characterised AAV vector; its mechanism of infection has been well studied which makes it an indispensable tool for ensuring safety in clinical settings (Choi et al., 2005; Daya and Berns, 2008). However, although AAV-2 has shown promise in several tissues, including the retina, it does not infect all cell types

Table 2
List of known corneal dystrophies, including; associated inheritance pattern, gene locus and causative genes.

		Inheritance Pattern	Genetic Locus	Gene known	Gene/s Affected	IC3D Category
Epithelial and Sub-Epithelial Dystrophies	**EBMD**	Minority of cases, mostly sporadic	5q13	Some cases	***TGFBI***	Some C1
	ERED	Autosomal Dominant	Unknown	Unknown	N/A	C3
	SMCD	Likely autosomal Dominant	Unknown	Unknown	Unknown	C4
	MECD	Autosomal Dominant	12q13 and 17q12	Yes	***KRT3*** and ***KRT12*** (Stocker–Holt variant)	C1
	LECD	X-chromosomal dominant	Xp 22.3	Unknown	Unknown	C2
	GDCD	Autosomal Recessive	1p32.	Yes	***TACSTD2***, previously *M1S1*	C1
Epithelial Stromal Dystrophies	**RBCD**	Autosomal Dominant	5q13	Yes	***TGFBI***	C1
	TBCD	Autosomal Dominant	5q13	Yes	***TGFBI***	C1
	LCD1	Autosomal Dominant	5q13	Yes	***TGFBI***	C1
	GCD1	Autosomal Dominant	5q13	Yes	***TGFBI***	C1
	GCD2	Autosomal Dominant	5q13	Unknown	***TGFBI***	C1
Stromal Dystrophies	**MCD**	Autosomal Recessive	16q22	Yes	***CHST6***	C1
	SCD	Autosomal Dominant	1p36	Yes	***UBIAD1***	C1
	CSCD	Autosomal Dominant	12q21.33	Yes	***DCN***	C1
	FCD	Autosomal Dominant	2q34	Yes	***PIKFYVE,*** previously *PIP5K3*	C1
	PACD	Autosomal Dominant	12q21.33	Yes	***KERA, LUM,DCN,EPYC***	C1
	CCDF	Unknown	Unknown	Unknown	Unknown	C4
	PDCD	Reported AD, similar deposits seen with X-linked ichthyosis	Isolated PDCD = Unknown X-linked ichthyosis = Xp22.31	Yes Unknown	Unknown ***STS***	C4 C1
Descemet's Membrane and Endotehlial Dystrophies	**FECD**	Unknown, reported autosomal dominant	Early onset = 1q34.3-p32 (FECD 1) **Late onset** =13pter-q12.3 (FECD 2), 18q21.2-q21.3 (FECD 3), 20p13-q12 (FECD 4), 5q33.1-q35.2 (FECD 5), 10p11.2 (FECD 6), 9p24.1-p22.1 (FECD 7), 15q25 (FECD 8)	Yes Some cases	***COL8A2*** Unknown, ***TCF4,*** ***SLC4A11***, Unkown, ***ZEB1***, Unkown, ***AGBL1***	C1 C2 = identified genetic loci, C3 = without known inheritance
	PPCD	Autosomal Dominant	PPCD 1 = 20p11.2–q11.2 PPCD 2 = 1p34.3–p32.3 PPCD 3 = 10p11.2	Unknown Yes Yes	Unknown **COL8A2** ***ZEB1***	C2 C1 C1
	CHED	Autosomal recessive	20p13	Yes	***SLC4A11***	C1 (some cases C3)
	XECD	X-chromosomal dominant	Xq25.	Unknown	Unknown	C2
Total:	22	Known = 17	Known = 18	Known = 12 Partially known = 4 Unknown = 5		

and indeed other serotypes may transduce specific tissues more efficiently (Srivastava, 2016). Hybrid serotypes were generated by transcapsidation, which involves packaging the ITR from one serotype into the capsid of a different serotype (Choi et al., 2005), has allowed the well characterised AAV-2 genome to be packaged into an array of serotypes, extending the range of cell types that can be targeted.

In addition to hybrid serotypes, chimeric capsids have been developed by recombination between capsid sequences from different serotypes (Hauck et al., 2003). Engineering the AAV capsid to produce vectors that are highly specific for a desired cell type (Büning et al., 2015) has been achieved by rational design strategies such as introducing sequence changes to target a specific cell surface receptor or directed evolution methodologies in which the virus undergoes multiple rounds of mutagenesis until variants with different infection capabilities emerge. Collectively, these efforts seek to generate a repertoire of AAV vectors with highly selective and efficient transduction of a wide range of cell types.

5.2.2. AAV in the eye

AAV has been the most widely used vector in ocular gene replacement therapy. The success of AAV in gene augmentation is largely due to its lack of immune response and persistence in the nucleus. The use of AAV-2 to deliver gene therapeutics has been extensively tested in clinical trials for the treatment of ocular diseases such as Leber's congenital amaurosis (Pierce and Bennett, 2015), all at different phases. In addition, AAV-2 is being used in a phase I trials for retinitis pigmentosa caused by mutations in *MERTK* and in phase I/II trials for the treatment of Choroideremia (Pierce and Bennett, 2015). The success and future of these therapies will be discussed in section 7.2 'Ocular clinical trials'.

The most utilised vector, and the most favourable vector for retinal transduction, AAV-2 does not show promise as a gene therapy vector for the anterior segment of the eye as it does not efficiently transduce the corneal layers (Borrás et al., 2002). However a multitude of AAV vectors of various serotypes have been shown to successfully transduce all layers of the cornea. Of a study testing AAV-2/1, AAV-2/2, AAV-2/5 and AAV-2/8 on both human corneas *ex vivo* and mouse *in vivo*, AAV-2/8 was found to be most efficient in both cases (Hippert et al., 2012). AAV-2 refers to wild-type AAV serotype 2, while AAV-2/1 etc. refers to a hybrid serotype in which the wild-type AAV-2 has been packaged by the capsid of the AAV-1 serotype. The different serotypes were delivered via an intrastromal injection and AAV-2/8 was found to achieve long-term transgene expression in the stroma keratocytes. AAV-2/9 and AAV 2/8 have shown successful transduction of the superficial cells of the stromal layer after epithelial debridement, with limited transduction of the same cells by AAV-2/6 (Sharma et al., 2010). Following an *ex vivo* intrastromal injection in human corneas, an AAV8 and AAV9 chimeric capsid vector (AAV8G9) was shown to successfully transduce the stromal layers in addition to some endothelial cells (Vance et al., 2016). AAV-2/9 has been shown to successfully transduce the cornea endothelium in mouse *in vivo*, via an intracameral injection

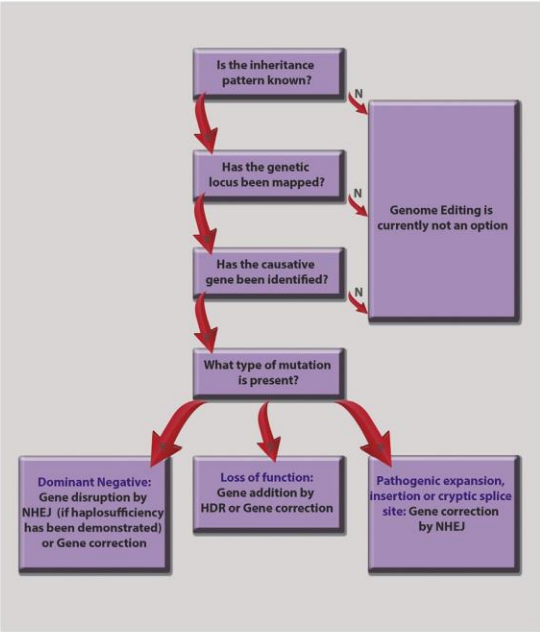


Fig. 5. Roadmap to gene therapy: When considering a genome engineering approach to treat disease status a series of questions must be considered. This flow chart offers an overview of how the most appropriate application could be chosen based on the genetics of the disease in question.

(O'Callaghan et al., 2017). A novel synthetic AAV, referred to as Anc80L65, is based on ancestral sequence reconstruction (Wang et al., 2017), was shown to efficiently transduce corneal stroma and endothelial cells following intrastromal injection in mouse.

5.2.3. AAV to deliver CRISPR/Cas9

The size of the transgene that can be packaged is limited by the size of the wild-type AAV genome (4.7 kb). The maximum payload that can be successfully packaged is 5kb, anything exceeding this has a negative effect on the viral titre produced (Zinn and Vandenberghe, 2014). When one considers the components that are needed to deliver a functional Cas9 nuclease, this maximum payload is very restrictive in relation to the therapeutic genes that can be delivered via AAV. The coding sequence of *S. pyogenes* Cas9 alone is ~4.2 kb, to which one must added a promoter (this can be general or tissue-specific) and terminator for Cas9 and a promoter and guide RNA sequence. Thus an entire *S. pyogenes* CRISPR/Cas9 system cannot be successfully packaged in AAV as the sum of these components exceeds the 5kb maximum payload. Cas9 nucleases from other species, such as *S. aureus* Cas9, have a smaller coding region allowing other components to be carried in a single capsid (Friedland et al., 2015); Successful delivery of an all-in-one *S. aureus* CRISPR/Cas9 AAV-8 system has been demonstrated in mouse *in vivo*, > 40% gene modification was observed when targeted to *Pcsk9* in the liver (Ran et al., 2015). In addition, *in vivo* editing and partial recovery of muscle function was demonstrated in a Duchenne muscular dystrophy (DMD) mouse, following delivery of an all-in-one *S. aureus* CRISPR/Cas9 AAV system (Tabebordbar et al., 2016). A limitation of *S. aureus* Cas9 is its more intricate PAM of 5' – NNGRRT – 3′, as this PAM will occurs less frequently than S. pyogenes NGG, reducing the number of potential targets significantly; however, studies have shown that it is possible to manipulate the PAM recognition by S. aureus to 5' – NNNRRT -3′ broadening the targeting capacity.

Alternative approaches have been explored to circumvent the payload limitations of AAV, which include: deleting the non-essential Cas9 REC2 lobe (Nishimasu et al., 2014), using orthologue Cas9 nucleases (Esvelt et al., 2013; Fonfara et al., 2014; Hou et al., 2013; Ran et al., 2015) and using a dual-AAV system – either in terms of splitting Cas9 and sgRNA into separate vectors (Swiech et al., 2015; Yang et al., 2016) or splitting Cas9 itself (Truong et al., 2015).

Once the crystal structure of *S. pyogenes* Cas9 was elucidated, insight into how to rationally engineer Cas9 to produce genetic variants, such as a smaller variant that could retain nuclease activity, was obtained. Although it was found the REC2 lobe was not essential for function, its removal reduced Cas9 activity in comparison to wild-type Cas9 by 50% (Nishimasu et al., 2014).

An alternative strategy is to split the CRISPR components into two separate AAV vectors. Although this provides spatiotemporal advantages it may also negatively impact gene editing efficiency, as each cell needs to take up both AAV vectors for cleavage to be achieved, however the dual vector system has been successfully applied in mouse brain, liver and eye (Hung et al., 2016; Swiech et al., 2015; Yang et al., 2016). This dual-AAV system has been demonstrated in mouse brain *in vivo* to study gene function of *Mecp2* and in mouse liver *in vivo* where gene correction of *Otc* was demonstrated (Swiech et al., 2015; Yang et al., 2016). Finally, splitting the Cas9 coding sequence across two vectors and fusing each half to a 'protein-intron', known as an intein, was demonstrated; upon co-expression, intein-mediated transsplicing occurs and the full Cas9 protein is reconstituted (Truong et al., 2015).

A dual-AAV system was used to deliver *S. pyogenes* Cas9 and sgRNA to retinal ganglion cells *in vivo,* to disrupt yellow fluorescent protein (YFP) in a Thy1-YFP transgenic mouse (Hung et al., 2016). The dual-AAV system was delivered via a single intravitreal injection and 5 weeks later there was a 84% reduction in YFP + cells. In addition, a dual-AAV system was also used to deliver *S. pyogenes* Cas9 and sgRNA to postmitotic photoreceptor cells *in vivo* in 3 separate retinal degeneration mouse models (Yu et al., 2017). In retinitis pigmentosa there is a progressive loss of rod cells which leads to the secondary death of cones. Neural retina leucine zipper (*Nrl*) has a role in determining rod cell fate, and knockout of this gene has been shown to lead to a loss of rod cell features and gain of cone characteristics; a recent study showed that ablation of *Nrl* by AAV-delivered CRISPR/Cas9 led to increased survival and preservation of cone function in 3 animal models of RP caused by rod-specific gene mutations.

A potential drawback of AAV in CRISPR/Cas9-mediated gene therapy is its persistent transgene expression, particularly when

Table 3
Characteristics of viral vectors utilised for gene therapy.

	Adenovirus	Retrovirus	Lentivirus	AAV
Genome Size	38-39kb	3-9kb	3-9kb	4.7kb
Genome Type	dsDNA	ssRNA	ssRNA	ssDNA
Host genome integration	No	Yes	Yes (integrase deficient versions available)	Reported at a very low frequency (Deyle and Russell, 2009)
Transgene expression	Days/weeks	Months/years	Months/years	Months/years
Immunogenicity	High	Moderate-high	Low-moderate	Low
Infection	Dividing cells	Dividing cells	Non-dividing and dividing cells	Non-dividing and dividing cells
Packaging capacity	< 7.5kb	< 8kb	< 8kb	< 5kb

considering off-target effects. A number of measures have been taken to minimise this risk, such as; the use of tissue-specific promoters or self-cleaving vectors. A tissue-specific reporter ensures expression is localised to only the desired target tissue. Two clinical trials for gene augmentation of *RPE65* to treat Leber's congenital amaurosis utilised an AAV-2 vector with a human *RPE65* promoter (hRPE65) (Bainbridge et al., 2015), this safeguards against unwanted expression in other tissues (NCT01496040, NCT02781480). One caveat of tissue-specific promoters is that they are generally considered to be weaker promoters (Nettelbeck et al., 1998), so although they ensure specific expression they may also restrict the level of expression. The clinical trials utilising hRPE65 employ gene augmentation to treat the disease, the reduced level of expression meant less wild-type cDNA was present in the cells resulting in inferior results compared to other vectors that utilised promoters such as CMV; however, in respect to gene-editing reduced expression of Cas9 may help to reduce off-target events. Of note, as tissue-specific promoters are generally larger in size this may pose difficulties of successful packaging into AAV vectors as previously discussed. Finally, in order to limit the time that the CRISPR components are expressed in the cell, and thus reduce off-targets effects, a self-cleaving system was developed which involves introducing both a sgRNA targeted to the gene of interest and a sgRNA targeted to the delivery vector (Chen et al., 2016; Moore et al., 2015; Petris et al., 2017).

5.3. Non-viral delivery

Non-viral delivery methods can be subdivided into physical and chemical approaches. Physical methods use force to increase the permeability of the cell membrane, whilst chemical methods utilise a carrier to carry the nucleic acid across the cell membrane. Physical methods of delivery include injection of naked DNA, electroporation, gene gun, ultrasound and magnetofection. These approaches have been reviewed extensively elsewhere (Mohan et al., 2005; Oliveira et al., 2017), this review will focus on the use of chemical non-viral delivery to the eye and of CRISPR/Cas9 components. Chemical non-viral delivery methods utilise electrostatic interactions between the negatively charged nucleic acid and either a natural or induced positive charge of the carrier, generating a complex with an overall positive charge (Al-Dosari and Gao, 2009). These positively complexes are attracted to the negatively charged cell membrane where they facilitate entry to the cell via endocytosis. However, these complexes must also be able to evade endosomal entrapment and degradation and must be transferred to the nucleus to have a therapeutic benefit (Riley and Vermerris, 2017). Substantial progress in nanotechnology has allowed the facile production of nanoparticles and a better understanding of nanoscale materials for gene delivery (Monopoli et al., 2012; Santos et al., 2014). Several carriers have been employed in non-viral delivery, including; lipoplexes, polyplexes and to a lesser extent inorganic compounds.

For viral delivery, the transgene is supplied as a DNA expression construct, which means it must be transcribed and translated by the transduced cell. For non-viral delivery, therapeutic reagents can be delivered either as mRNA or protein. In addition to challenges associated with RNA stability, the use of mRNA as a therapeutic has been hindered due its initiation of a strong immunogenic response (Van Tendeloo et al., 2007); these pitfalls have been partially overcome by the use of modified mRNA that suppresses the immune response (Hendel et al., 2015; Zangi et al., 2013). While intracellular delivery of proteins is limited by proteolytic instability coupled with insufficient membrane permeability (Fu et al., 2014), nevertheless, the use of mRNA or protein instead of DNA expression constructs immediately offers potential benefits for gene editing, as mRNA or protein Cas9 nuclease will not persist in the cells to cause off-target effects. Despite this obvious benefit, non-viral delivery solutions have been hampered by their low-efficiency (Putnam, 2006).

Several cationic polymers, either synthetic or naturally occurring, have been used for the delivery of nucleic acids. Examples include; polyethyleneimine (Lungwitz et al., 2005), poly-L-lysine (Kodama et al., 2014), polyamidoamine (Wu et al., 2011), chitosan (Mao et al., 2010), and cell penetrating peptides (CPPs) (Boisguérin et al., 2015). Chitosan-nanoparticles loaded with pEGFP plasmid were able to achieve up to 15% transfection efficiency in human corneal epithelial cells *in vitro* (De La Fuente et al., 2008b), and when topically applied to rabbit corneas *in vivo* (de La Fuente et al., 2008a) successfully transfected both corneal and conjunctival epithelial cells. In addition, an intrastromal injection of chitosan-nanoparticles loaded with a luciferase plasmid into rat cornea transduced the keratocytes (Klausner et al., 2010).

Gene therapy for LCA has been widely investigated, as described previously, three separate trials have successfully delivered the *RPE65* gene via AAV-2. However, replacement of *Rpe65* has also been demonstrated via peptide-modified lipid nanoparticles in a *Rpe65*-deficient mouse model following a single subretinal injection, at efficiencies comparable to AAV-2 (Rajala et al., 2014). The peptide-modified lipid nanoparticles were able to achieve efficient, lasting, cell-specific gene expression in a similar fashion to viral vectors.

Several groups are now harnessing the benefits of a non-viral system to deliver CRISPR/Cas9 RNP complexes. RNP complexes allow Cas9 to be active immediately upon entry into cells and degrade quickly minimising off-target effects. Catonic lipids have been utilised to successfully deliver CRISPR/Cas9 RNP to the inner ear cells in mice, where 20% genome modification of hair cells was observed (Zuris et al., 2014). A recent report described the development of CRISPR-Gold, which utilised gold nanoparticles for successful *in vivo* HDR in a DMD mouse (Lee et al., 2017). HDR gene modification requires the successful delivery of not only the Cas9 nuclease and sgRNA, but also the repair template. CRISPR-Gold is able to simultaneously deliver RNPs and donor DNA into cells. Following a single injection into the gastrocnemius and tibialis anterior muscle 5.4% of the expressed dystrophin was corrected to the wild-type sequence, while treatment with RNPs and donor DNA by themselves only achieved a 0.3% correction of the dystrophin gene. It has been shown that presence of 3–15% wild-type dystrophin is enough to relieve symptoms of DMD (van Putten et al., 2012; Van Putten et al., 2014, 2013). Of note, an AAV-mediated study previously described, utilising the same mouse model of DMD but attempting exon excision by NHEJ gene correction rather than HDR, saw 39% exon excision and partial recovery of muscle functional deficiencies (Tabebordbar et al., 2016).

Historically, non-viral vectors have not been widely used in ocular gene-editing research, likely due to the immense success of AAV in retinal gene therapy (Bainbridge et al., 2008; Bennett et al., 2016; Hauswirth et al., 2008; Maguire et al., 2008). A recent report has described the highest editing efficiency to-date *in vivo* with a non-viral delivery platform, so this could change current perceptions of the low efficiency of non-viral vectors (Yin et al., 2017). The report describes use of lipid nanoparticles to encapsulate the CRISPR components. However, the noteworthy difference is their use of enhanced sgRNAs (e-sgRNA) which had 70 out of 101 nt of sgRNA modified with a 2′ hydroxyl (OH) group in addition to a number of phosphorothioate bonds. These modifications did not disrupt the interaction between Cas9 and sgRNA but allowed the sgRNA to persist *in vivo* long enough to achieve cleavage of the desired target. A single intravenous injection of a lipid nanoparticle encapsulating Cas9 mRNA and 2 e-sgRNAs in mice resulted in a > 80% editing of *Pcsk9* in the liver. This remarkable result suggests a worthy place for non-viral delivery in gene editing and it could certainly offer a promising alternative to viruses for the eye.

5.4. Combined viral and non-viral methods

As no single approach meets every requirement a combination of viral and non-viral mediated approaches may be an alternative method to achieve efficient gene editing. In an initial report by Yin et al.,

separate plasmids, expressing CRISPR/Cas9 components targeted to the *Fah* gene and a donor oligo, were delivered systemically via a hydrodynamic injection into the tail vein in a *Fah* $^{-/-}$ mouse. In only 0.4% of hepatocytes did correction occur leading to expression of wild-type *Fah*. In a subsequent study using a combination of viral and non-viral delivery vehicles to introduce the CRISPR/Cas9 components > 6% of cells were corrected (Yin et al., 2016). A lipid nanoparticle was used to deliver Cas9 mRNA, whilst an AAV vector was used to express guide RNA and donor template.

5.5. Summary

Delivery is a challenge faced not only in ocular gene therapy research but all biomedical fields. Nucleic acids are large, negatively charged molecules and getting them across cell membranes with high efficiency and minimal detection by the host's immune system is no small task. Of the current array of potential vehicles available none meet all desirable criteria, the pros and cons of each vehicle must be considered for each application. A compromise may have to be made in order to identify a suitable vehicle with which to proceed. The field has already come so far in terms of addressing delivery issues it is likely the issues faced today such as small payload and low efficiency will be overcome by advances in engineering, understanding of essential components of Cas9 and improvements in non-viral complexes.

6. Stem cell culture and transplantation

All of the above examples describe the use of CRISPR/Cas9 *in vivo* in somatic cells. The ability to isolate and culture stem cells provides a promising approach to genetically manipulate cells *ex vivo* and transplant them back into the patient (Fig. 6). This approach has been adapted as a potential treatment for genetic skin disorders (Hainzl et al., 2017; Kocher et al., 2017; Mavilio et al., 2006). Mutations in laminin 5 (*LAM5*) cause junctional epidermolysis bullosa (JEB); epidermal stem cells from a patient with JEB caused by a mutation in *LAM5* were cultured and transduced with retrovirus encoding the cDNA of *LAM5* (Mavilio et al., 2006). Epidermal grafts were then prepared using these genetically modified cultured cells and transplanted onto regions on the patient's legs. These grafts resulted in firmly adherent epidermis that remained stable for the 1 year follow up.

Recently groups have focused on adapting this technique to genetically modify these epidermal stem cells using CRISPR/Cas9. Human recessive dystrophic epidermolysis bullosa (RDEB) keratinocytes, carrying a homozygous mutation (6527insC) in exon 80 of *COL7A1* were transfected with CRISPR/Cas9 nickase (D10A) (F. Ann Ran et al., 2013a,b) and a donor template used to correct the inherited mutation (Hainzl et al., 2017). Grafts of corrected keratinocytes were transplanted into immunocompromised mice and resulted in phenotypic regression. In addition, using the same targeting strategy to transfected human epidermolysis bullosa simplex (EBS) keratinocytes carrying a heterozygous mutation (c.1231G > A) in exon 6 of Kertain 14 (*KRT14*) (Kocher et al., 2017) achieved > 30% correction and the treated keratinocytes displayed phenotypic regression.

The corneal epithelium is a regenerating tissue, turning over every 1–2 weeks (Hanna et al., 1961). The limbal stem cells located on the conjunctiva-corneal boundary, known as the limbus, are responsible for regenerating the corneal epithelium (Schermer et al., 1986). Limbal stem cell deficiency (LSCD) is a pathology in which the cornea partially or completely loses its regenerative capability due to damage of the limbus region (Barut Selver et al., 2017). Transplantation of autologous limbal stem cell explants into a LSCD patients was first reported in 1997, producing remarkable results restoring epithelial function and vision in severely damaged corneas (Pellegrini et al., 1999; Rama et al., 2010, 2001).

We have previously reported the ability to culture limbal epithelial stem cells (LESC) from a limbal biopsy taken from a Meesmann's corneal dystrophy (MECD) patient with a L132P *KRT12* missense mutation (Courtney et al., 2014). We then demonstrated allele-specific silencing of the mutant SNP in these cultures using siRNA. This work is a proof of concept for genetically manipulating LESC in culture and CRISPR/Cas9 could be applied in a similar fashion to permanently correct these LESC before transplantation. However, transplantation of these explants would pose a series of challenges. Initially, a limbal biopsy would be taken from a patient and a culture established; LESC would be corrected by CRISPR/Cas9 via either gene disruption of the mutant allele by NHEJ or by corrective HDR. Single corrected clones with verified stemness would then be selected and used to grow clonal grafts. However, in order to permanently correct the underlying mutation in the patient, LSCD would have to be induced in the cornea before the transplant is engrafted, to remove the resident stem cells still harbouring the causative mutation. As LSCD is a severe condition the possibility that the graft would not take poses too much risk, as it would leave the patient in worse condition than prior to the treatment; for this reason, inducing LSCD in corneal dystrophy patients is not an accepted treatment strategy.

7. Current gene therapy clinical trials

7.1. Non-ocular clinical trials

Gene editing has progressed to clinical trials to treat diseases such as HIV-1, Haemophilia B and various types of cancer.

Various *ex vivo* approaches utilising ZFNs have been adapted to treat HIV-1. They are all targeted to the CCR5 receptor, which is a major co-receptor for HIV-1. Knocking out the CCR5 receptor makes cells resistant to HIV-1 infection. One study, currently in phase I clinical trial, delivers ZFNs mRNA targeted to the CCR5 receptor by *ex vivo* electroporation of Hematopoietic Stem/Progenitor Cells (HSPCs) (DiGiusto

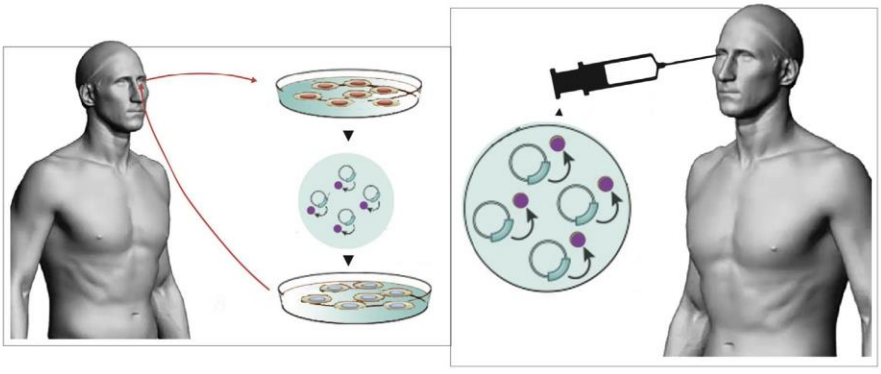


Fig. 6. Schematic to show *ex vivo* gene editing compared to *in vivo* gene editing. For corneal gene editing *ex vivo*, a limbal biopsy will be taken from the patient, limbal epithelial stem cells will then be cultured and edited *ex vivo*, the edited cells will then be selected, expanded and transplanted into the patient. For corneal gene editing *in vivo* the gene therapy agent and selected delivery vehicle will be injected into the eye, either by intrastromal or intracameral injection (Fig. 1).

et al., 2016). (NCT02500849) Following busulfan conditioning, modified autologous CCR5 HSPCs are infused into the patient bloodstream. 5 other trials using a similar approach are currently underway.

ZFNs have also been employed to treat severe Haemophilia B in a study that is currently in phase I clinical trial (NCT02702115). AAV2/6 vectors are being used to deliver ZFNs and template containing wild-type *F9* complementary DNA (cDNA). The aim of this study is the targeted integration of this template into the albumin-encoding locus of liver cells. A similar approach uses AAV vectors to introduction the correct cDNA copy of the α-L-iduronidase *IDUA* gene into patient hepatocytes to treat Mucopolysaccharidosis I (MPS I). (NCT02702115).

An emerging treatment for leukemia involves engineering T cells to express Chimeric antigen receptors (CAR) against leukemia antigens, such as CD19 (Qasim et al., 2017); this involves transducing T cells *ex vivo* with lentiviral vectors containing the desired transgene, in this case *CAR19*. However, these engineered T cells must also be able to evade host immunity and not induce graft-versus-host disease (GVHD) following infusion. A phase I clinical trial is now underway, using modified T cells to treat Pediatric Acute Lymphoblastic Leukemia (PALL). (NCT02808442). This study used TALENs to disrupt of CD52 expression, to allow infused cells to evade the depletion effects of alemtuzumab and prevent GVHD. In addition, it used lentivirus to integrate *CAR19*. The compassionate use of these modified T cells was granted to Great Ormond Street Hospital for Children in London for two infants with therapy-refractory, relapsed B cell leukemia; the infants are now reported to be in cytogenetic remission (Qasim et al., 2017).

Excitingly, CRISPR/Cas9 is now in phase I clinical trials using modified T cells to treat a range of solid tumours (NCT02793856, NCT02867345, NCT02863913, NCT02867332, NCT03081715). These include the treatment of; Metastatic Non-small Cell Lung Cancer, Castration Resistant Prostate Cancer, Muscle-invasive Bladder Cancer, Metastatic Renal Cell Carcinoma and Advanced Esophageal Cancer. The first use of CRISPR/Cas9 modified cells to treat human disease was carried out in 2016, when *ex vivo* modified T cells were used to treat metastatic non-small-cell lung cancer (Nature 539, 479, 2016) and it is now in phase I clinical trial (NCT02793856). This involved collecting peripheral blood lymphocytes and using CRISPR/Cas9 to knockout the programmed cell death protein 1 (*PDCD1*) gene *ex vivo*; modified cells were then selected and expanded before being infused back into patients who had been treated with cyclophosphamide for 3 days to deplete T regulatory cells. All currently active clinical trials described are using the exact approach to target a plethora of different cancers. In addition, there are a wealth of studies at preclinical stage to treat diseases such as multiple myeloma, melanoma, sarcoma, beta-thalassemia, sickle cell disease and transthyretin amyloidosis (Sheridan, 2017).

7.2. Ocular clinical trails

At present ocular clinical trials are underway for the treatment of inherited retinal diseases. They all employ gene augmentation for loss of functions diseases or optogenetics, rather than gene editing strategies. At present gene augmentation trials are on-going for Leber's congenital amaurosis, Choroideremia, Achromatopsia, X-linked Retinoschsis, Retinitis Pigmentosa (RP), RP in Usher Syndrome 1B, Stargardt disease and Leber's hereditary optic neuropathy (LHON).

Leber's congenital amaurosis gene replacement is a well-established gene therapy for inherited retinal eye disease. Three groups almost simultaneously initiated clinical trials utilising a AAV-2/8 vector encoding RPE65 cDNA (Bainbridge et al., 2008; Hauswirth et al., 2008; Maguire et al., 2008). The three studies differed in terms of dose, inclusion criteria, type of promoter, location of the injection and outcome measures. All three studies reported high levels of safety and critically demonstrated efficacy up until the 3-year follow-up. Subsequently, trials utilising AAV-4 and AAV-5 to deliver RPE65 cDNA (Bainbridge et al., 2015; Bennett et al., 2016; Jacobson et al., 2015; Weleber et al., 2016) also indicated excellent safety and initial large improvements in vision. However, one study reported a decline in visual improvements in 3/15 patients after 3 years(Jacobson et al., 2015). and another saw a decrease in visual improvements after 6–12 months (Bainbridge et al., 2015). It should be noted that the latter study employed a cell-specific (RPE) promoter which is weaker than promoters previously used and the initial response was not as strong. Finally, re-administration of gene therapy from one of the first studies demonstrated a further improvement in vision without an immune response (Bennett et al., 2016, 2012). Gene therapy for Leber's congenital amaurosis is the first ocular gene therapy to complete a phase III clinical trial. The success of this approach paves the way for additional gene therapy approaches in the eye.

At present there are no active clinical trials for gene editing in ocular diseases. However, the unique qualities of the eye such as immune-privilege status, small surface area, ease of visualisation and accessibility, coupled with the success of current retinal gene replacement studies, make gene editing in the eye a natural next step to target diseases for which gene replacement is not a viable option. Editas Medicine and Sanofi-Genzyme have a clinical trial planned for LCA in which CRISPR will be targeted to delete a cryptic splice site and restore normal splicing of the LCA10 gene. They have subsequently announced future plans for a similar trial targeted to Usher syndrome. It is likely these initial clinical trials will herald a new era for the treatment of ocular inherited diseases for which current disease prognosis is poor.

8. Translation to the clinic

After much anticipation, the era of personalised medicine is now upon us. For gene-editing to progress to the clinic there will have to be some form of collaboration being the regulatory bodies and the scientists developing these life-changing treatments. Although gene editing has come so far in recent years, applying these technologies still presents a substantial challenge. The long-term follow up of patients who have participated in genome editing clinical trials such as the two infants in the UCART19 clinical trial in PALL (NCT02808442), will likely provide invaluable insight into the *in vivo* activity and specificity of programmable nucleases.

Another issue moving forward will be the critical assessment of risk vs benefit. At present a key factor of entry into clinical trial includes the prerequisite that current treatments are lacking. How does one define what disease is serious enough to risk experimental gene therapy? For instance, inducing LSCD in patients with corneal dystrophy will almost certainly be considered too much of a risk especially if the patient in question still has a slight vision. However, a single injection of CRISPR/Cas9 components containing a well characterised guide sequence with a potential to permanently eradicate the patient's corneal dystrophy will likely confer more benefit than risk.

Another consideration, especially relevant for corneal dystrophies, is that entry into a clinical trial is extremely expensive. Conventionally this means a pharmaceutical company will likely invest and employ researchers to develop the therapy. However, in terms of very rare diseases it is likely that companies will not invest where they do not envisage revenue. So how can we ensure severe and sometimes life-threatening diseases with no current treatment options do not miss out on the personalised medicine revolution? Furthermore, if these therapies do reach the clinic who will cover costs? Will the cost of CRISPR gene therapy ever be met by the NHS or private healthcare? In order to justify these costs it would have to be shown that the cost of this one-off treatment was less than the lifetime costs of conventional treatment in addition to offering reduced morbidity or improved quality of life.

9. Conclusion

The rapid pace and enthusiasm with which CRISPR has engendered in the field of gene editing has seen advancements develop at an unprecedented rate. As CRISPR/Cas9 is now entering phase I clinical trials

it is likely the progression of gene editing will only accelerate. The genetics of corneal dystrophies, unique qualities of the eye and recent success of gene replacement present the ideal platform for gene editing therapies. The personalised medicine revolution offers a wealth of possibilities for both inherited and acquired diseases. In the future, it is entirely possible that every new-born child will have their genome sequenced. A time when gene-based therapeutics exist as standard treatment for ocular diseases is forthcoming. Individual genome sequences coupled with advances in gene editing tools will allow for preventive personalised medicine rather than treating a disease phenotype. It is an immensely exciting time for both the field of gene editing and ocular research, and the future for corneal dystrophies looks promising.

Declaration of interest

C.B.Tara Moore is a consultant for Avellino Laboratories.

Acknowledgments

This work was supported by Fight for Sight UK (Reference 1450/1451).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.preteyeres.2018.01.004.

References

Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev, A., Lander, E.S., Voytas, D.F., Ting, A.Y., Zhang, F., 2017. RNA targeting with CRISPR–Cas13. Nature. https://doi.org/10.1038/nature24049.

Al-Dosari, M.S., Gao, X., 2009. Nonviral gene delivery: principle, limitations, and recent progress. AAPS J. 11, 671–681. https://doi.org/10.1208/s12248-009-9143-y.

Aldave, A.J., Sonmez, B., Forstot, S.L., Rayner, S.A., Yellore, V.S., Glasgow, B.J., 2007. A clinical and histopathologic examination of accelerated TGFBIp deposition after LASIK in combined granular-lattice corneal dystrophy. Am. J. Ophthalmol. 143, 416–419. https://doi.org/10.1016/j.ajo.2006.11.056.

Anders, C., Niewoehner, O., Jinek, M., 2015. In vitro reconstitution and crystallization of Cas9 endonuclease bound to a guide RNA and a DNA target. Methods Enzymol. 558, 515–537. https://doi.org/10.1016/bs.mie.2015.02.008.

Arentsen, J.J., 1983. Corneal transplant allograft reaction: possible predisposing factors. Trans. Am. Ophthalmol. Soc. 81, 361–402.

Athanasopoulos, T., Munye, M.M., Yáñez-Muñoz, R.J., 2017. Nonintegrating gene therapy vectors. Hematol. Oncol. Clin. N. Am. 31, 753–770. https://doi.org/10.1016/j.hoc.2017.06.007.

Bainbridge, J.W.B., Mehat, M.S., Sundaram, V., Robbie, S.J., Barker, S.E., Ripamonti, C., Georgiadis, A., Mowat, F.M., Beattie, S.G., Gardner, P.J., Feathers, K.L., Luong, V. a., Yzer, S., Balaggan, K., Viswanathan, A., de Ravel, T.J.L., Casteels, I., Holder, G.E., Tyler, N., Fitzke, F.W., Weleber, R.G., Nardini, M., Moore, A.T., Thompson, D. a., Petersen-Jones, S.M., Michaelides, M., van den Born, L.I., Stockman, A., Smith, A.J., Rubin, G., Ali, R.R., 2015. Long-term effect of gene therapy on Leber's congenital amaurosis. N. Engl. J. Med. 372, 1887–1897. https://doi.org/10.1056/NEJMoa1414221.

Bainbridge, J.W.B., Smith, A.J., Barker, S.S., Robbie, S., Henderson, R., Balaggan, K., Viswanathan, A., Holder, G.E., Stockman, A., Tyler, N., Petersen-Jones, S., Bhattacharya, S.S., Thrasher, A.J., Fitzke, F.W., Carter, B.J., Rubin, G.S., Moore, A.T., Ali, R.R., 2008. Effect of gene therapy on visual function in Leber's congenital amaurosis. N. Engl. J. Med. 358, 2231–2239. https://doi.org/10.1056/NEJMoa0802268.

Barut Selver, Ö., Yağcı, A., Eğrilmez, S., Gürdal, M., Palamar, M., Çavuşoğlu, T., Ateş, U., Veral, A., Güven, Ç., Wolosin, J.M., 2017. Limbal stem cell deficiency and treatment with stem cell transplantation. Türk Oftalmol. Derg 47, 285–291. https://doi.org/10.4274/tjo.72593.

Bassuk, A.G., Zheng, A., Li, Y., Tsang, S.H., Mahajan, V.B., 2016. Precision medicine: genetic repair of retinitis pigmentosa in patient-derived stem cells. Sci. Rep. 6, 19969. https://doi.org/10.1038/srep19969.

Belfort, M., Bonocora, R.P., 2014. Homing endonucleases: from genetic anomalies to programmable genomic clippers. Methods Mol. Biol. 1123, 1–26. https://doi.org/10.1007/978-1-62703-968-0_1.

Bennett, J., Ashtari, M., Wellman, J., Marshall, K.A., Cyckowski, L.L., Chung, D.C., McCague, S., Pierce, E.A., Chen, Y., Bennicelli, J.L., Zhu, X., Ying, G. -s., Sun, J., Wright, J.F., Auricchio, A., Simonelli, F., Shindler, K.S., Mingozzi, F., High, K.A., Maguire, A.M., 2012. AAV2 gene therapy readministration in three adults with congenital blindness. Sci. Transl. Med. 4 120ra15-120ra15. https://doi.org/10.1126/scitranslmed.3002865.

Bennett, J., Wellman, J., Marshall, K.A., McCague, S., Ashtari, M., DiStefano-Pappas, J., Elci, O.U., Chung, D.C., Sun, J., Wright, J.F., Cross, D.R., Aravand, P., Cyckowski, L.L., Bennicelli, J.L., Mingozzi, F., Auricchio, A., Pierce, E.A., Ruggiero, J., Leroy, B.P., Simonelli, F., High, K.A., Maguire, A.M., 2016. Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial. Lancet 388, 661–672. https://doi.org/10.1016/S0140-6736(16)30371-3.

Bidaut-Garnier, M., Monnet, E., Prongué, A., Montard, R., Gauthier, A.S., Desmarets, M., Mariet, A.S., Ratajczak, C., Binda, D., Saleh, M., Delbosc, B., 2016. Evolution of corneal graft survival over a 30-year period and comparison of surgical techniques: a cohort study. Am. J. Ophthalmol. 163, 59–69. https://doi.org/10.1016/j.ajo.2015.12.014.

Bogdanove, A.J., Voytas, D.F., 2011. TAL effectors: customizable proteins for DNA targeting. Science (80-.) 333, 1843–1846. https://doi.org/10.1126/science.1204094.

Boisguérin, P., Deshayes, S., Gait, M.J., O'Donovan, L., Godfrey, C., Betts, C.A., Wood, M.J.A., Lebleu, B., 2015. Delivery of therapeutic oligonucleotides with cell penetrating peptides. Adv. Drug Deliv. Rev. 87, 52–67. https://doi.org/10.1016/j.addr.2015.02.008.

Borrás, T., Brandt, C.R., Nickells, R., Ritch, R., 2002. Gene therapy for glaucoma: treating a multifaceted, chronic disease. Investig. Ophthalmol. Vis. Sci. 43 (8).

Borrás, T., Gabelt, B.T., Klintworth, G.K., Peterson, J.C., Kaufman, P.L., 2001. Non-invasive observation of repeated adenoviral GFP gene delivery to the anterior segment of the monkey eye in vivo. J. Gene Med. 3, 437–449. https://doi.org/10.1002/jgm.210.

Büning, H., Huber, A., Zhang, L., Meumann, N., Hacker, U., 2015. Engineering the AAV capsid to optimize vector-host-interactions. Curr. Opin. Pharmacol. 24, 94–104. https://doi.org/10.1016/j.coph.2015.08.002.

Cameron, P., Fuller, C.K., Donohoue, P.D., Jones, B.N., Thompson, M.S., Carter, M.M., Gradia, S., Vidal, B., Garner, E., Slorach, E.M., Lau, E., Banh, L.M., Lied, A.M., Edwards, L.S., Settle, A.H., Capurso, D., Llaca, V., Deschamps, S., Cigan, M., Young, J.K., May, A.P., 2017. Mapping the genomic landscape of CRISPR-Cas9 cleavage. Nat. Methods 14, 600–606. https://doi.org/10.1038/nmeth.4284.

Cencic, R., Miura, H., Malina, A., Robert, F., Ethier, S., Schmeing, T.M., Dostie, J., Pelletier, J., 2014. Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. PLoS One 9. https://doi.org/10.1371/journal.pone.0109213.

Charlesworth, C.T., Deshpande, P.S., Dever, D.P., Dejene, B., Gomez-Ospina, N., Mantri, S., Pavel-Dinu, M., Camarena, J., Weinberg, K.I., Porteus, M.H., 2018. Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. bioRxiv 243345. https://doi.org/10.1101/243345.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., Huang, B., 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155, 1479–1491. https://doi.org/10.1016/j.cell.2013.12.001.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., Doudna, J.A., 2017. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. Nature 550, 407–410. https://doi.org/10.1038/nature24268.

Chen, M., Xie, L., 2013. Features of recurrence after excimer laser phototherapeutic keratectomy for anterior corneal pathologies in North China. Ophthalmology 120, 1179–1185. https://doi.org/10.1016/j.ophtha.2012.12.001.

Chen, Y., Liu, X., Zhang, Y., Wang, H., Ying, H., Liu, M., Li, D., Lui, K.O., Ding, Q., 2016. A self-restricted CRISPR system to reduce off-target effects. Mol. Ther. 24, 1508–1510. https://doi.org/10.1038/mt.2016.172.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., Kim, J.S., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 24, 132–141. https://doi.org/10.1101/gr.162339.113.

Choi, V.W., McCarty, D.M., Samulski, R.J., 2005. AAV hybrid serotypes: improved vectors for gene delivery. Curr. Gene Ther. 5, 299–310. https://doi.org/10.2174/1566523054064968.

Christie, K.A., Courtney, D.G., Dedionisio, L.A., Shern, C.C., De Majumdar, S., Mairs, L.C., Nesbit, M.A., Moore, C.B.T., 2017. Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-16279-4.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., Kühn, R., 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. 33, 543–548. https://doi.org/10.1038/nbt.3198.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., Porteus, M.H., Baltimore, D., Miller, J.C., Sander, J.D., Wood, A.J., Christian, M., Zhang, F., Miller, J.C., Reyon, D., Boch, J., Moscou, M.J., Bogdanove, A.J., Stoddard, B.L., Jinek, M., Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V., Garneau, J.E., Deveau, H., Garneau, J.E., Moineau, S., Horvath, P., Barrangou, R., Makarova, K.S., Bhaya, D., Davison, M., Barrangou, R., Deltcheva, E., Sapranauskas, R., Magadán, A.H., Dupuis, M.E., Villion, M., Moineau, S., Deveau, H., Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., Almendros, C., Jinek, M., Doudna, J.A., Malone, C.D., Hannon, G.J., Meister, G., Tuschl, T., Certo, M.T., Mali, P., Carr, P.A., Church, G.M., Barrangou, R., Horvath, P., Brouns, S.J., Guschin, D.Y., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823. https://doi.org/10.1126/science.1231143.

Courtney, D.G., Atkinson, S.D., Allen, E.H.A., Moore, J.E., Walsh, C.P., Pedrioli, D.M.L., MacEwen, C.J., Pellegrini, G., Maurizi, E., Serafini, C., Fantacci, M., Liao, H., Irvine, A.D., McIrwin Lean, W.H., Tara Moore, C.B., 2014. SiRNA silencing of the mutant keratin 12 allele in corneal limbal epithelial cells grown from patients with

Meesmann's epithelial corneal dystrophy. Investig. Ophthalmol. Vis. Sci. 55, 3352–3360. https://doi.org/10.1167/iovs.13-12957.
Courtney, D.G., Moore, J.E., Atkinson, S.D., Maurizi, E., Allen, E.H.A., Pedrioli, D.M.L., McLean, W.H.I., Nesbit, M.A., Moore, C.B.T., 2016. CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting. Gene Ther. 23, 108–112. https://doi.org/10.1038/gt.2015.82.
Cox, D.B.T., Platt, R.J., Zhang, F., 2015. Therapeutic genome editing: prospects and challenges. Nat. Med. 21, 121–131. https://doi.org/10.1038/nm.3793.
Cradick, T.J., Fine, E.J., Antico, C.J., Bao, G., 2013. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. 41, 9584–9592. https://doi.org/10.1093/nar/gkt714.
Crosetto, N., Mitra, A., Silva, M.J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., Pasero, P., Rowicka, M., Dikic, I., 2013. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat. Methods 10, 361–365. https://doi.org/10.1038/nmeth.2408.
Daya, S., Berns, K.I., 2008. Gene therapy using adeno-associated virus vectors. Clin. Microbiol. Rev. https://doi.org/10.1128/CMR.00008-08.
de la Fuente, M., Raviña, M., Paolicelli, P., Sanchez, A., Seijo, B., Alonso, M.J., 2010. Chitosan-based nanostructures: a delivery platform for ocular therapeutics. Adv. Drug Deliv. Rev. https://doi.org/10.1016/j.addr.2009.11.026.
de la Fuente, M., Seijo, B., Alonso, M.J., 2008a. Bioadhesive hyaluronan–chitosan nanoparticles can transport genes across the ocular mucosa and transfect ocular tissue. Gene Ther. 15, 668–676. https://doi.org/10.1038/gt.2008.16.
De La Fuente, M., Seijo, B., Alonso, M.J., 2008b. Novel hyaluronic acid-chitosan nanoparticles for ocular gene therapy. Investig. Ophthalmol. Vis. Sci. 49, 2016–2024. https://doi.org/10.1167/iovs.07-1077.
DeWitt, M.A., Corn, J.E., Carroll, D., 2017. Genome editing via delivery of Cas9 ribonucleoprotein. Methods 121–122, 9–15. https://doi.org/10.1016/j.ymeth.2017.04.003.
DeWitt, M.A., Magis, W., Bray, N.L., Wang, T., Berman, J.R., Urbinati, F., Heo, S.-J., Mitros, T., Muñoz, D.P., Boffelli, D., Kohn, D.B., Walters, M.C., Carroll, D., Martin, D.I.K., Corn, J.E., 2016. Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. Sci. Transl. Med. 8 360ra134-360ra134. https://doi.org/10.1126/scitranslmed.aaf9336.
Deyle, D.R., Russell, D.W., 2009. Adeno-associated virus vector integration. Curr. Opin. Mol. Therapeut. 11, 442–447.
DiGiusto, D.L., Cannon, P.M., Holmes, M.C., Li, L., Rao, A., Wang, J., Lee, G., Gregory, P.D., Kim, K.A., Hayward, S.B., Meyer, K., Exline, C., Lopez, E., Henley, J., Gonzalez, N., Bedell, V., Stan, R., Zaia, J.A., 2016. Preclinical development and qualification of ZFN-mediated CCR5 disruption in human hematopoietic stem/progenitor cells. Mol. Ther. - Methods Clin. Dev. 3, 16067. https://doi.org/10.1038/mtm.2016.67.
Dinh, R., Rapuano, C.J., Cohen, E.J., Laibson, P.R., 1999. Recurrence of corneal dystrophy after excimer laser phototherapeutic keratectomy. Ophthalmology 106, 1490–1497. https://doi.org/10.1016/S0161-6420(99)90441-4.
Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H.W., Listgarten, J., Root, D.E., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. 34, 184–191. https://doi.org/10.1038/nbt.3437.
Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., Root, D.E., 2014. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat. Biotechnol. 32, 1262–1267. https://doi.org/10.1038/nbt.3026.
Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., Church, G.M., 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat. Methods 10, 1116–1121. https://doi.org/10.1038/nmeth.2681.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806–811. https://doi.org/10.1038/35888.
Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lécrivain, A.L., Bzdrenga, J., Koonin, E.V., Charpentier, E., 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. 42, 2577–2590. https://doi.org/10.1093/nar/gkt1074.
Friedland, A.E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., Marco, E., Gotta, G.M., Maeder, M.L., Kennedy, E.M., Kornepati, A.V.R., Sousa, A., Collins, M.A., Jayaram, H., Cullen, B.R., Bumcrot, D., 2015. Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. Genome Biol. 16, 257. https://doi.org/10.1186/s13059-015-0817-8.
Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.-J., Kii, E., Alt, F.W., 2014. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. 33, 179–186. https://doi.org/10.1038/nbt.3101.
Fu, A., Tang, R., Hardie, J., Farkas, M.E., Rotello, V.M., 2014. Promises and pitfalls of intracellular delivery of proteins. Bioconjug. Chem. https://doi.org/10.1021/bc500320j.
Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., Sander, J.D., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31, 822–826. https://doi.org/10.1038/nbt.2623.
Garba, A.O., Mousa, S.A., 2010. Bevasiranib for the treatment of wet, age-related macular degeneration. Ophthalmol. Eye Dis. 2, 75–83. https://doi.org/10.4137/OED.S4878.
Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., Liu, D.R., 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 1–27. https://doi.org/10.1038/nature24644.
Gilbert, L. a, Larson, M.H., Morsut, L., Liu, Z., Gloria, A., Torres, S.E., Stern-ginossar, N., Brandman, O., Whitehead, H., Doudna, J. a, Lim, W. a, Jonathan, S., 2013. CRISPR-mediated modular RNA-guided regualtion of transcription in eukaryotes. Cell 154, 442–451. https://doi.org/10.1016/j.cell.2013.06.044.CRISPR-Mediated.
Hafner, A., Langenbucher, A., Seitz, B., 2005. Long-term results of phototherapeutic keratectomy with 193-nm excimer laser for macular corneal dystrophy. Am. J. Ophthalmol. 140, 1–6. https://doi.org/10.1016/j.ajo.2005.03.052.
Hainzl, S., Peking, P., Kocher, T., Murauer, E.M., Larcher, F., Del Rio, M., Duarte, B., Steiner, M., Klausegger, A., Bauer, J.W., Reichelt, J., Koller, U., 2017. COL7A1 editing via CRISPR/cas9 in recessive dystrophic epidermolysis bullosa. Mol. Ther. https://doi.org/10.1016/j.ymthe.2017.07.005.
Hämäläinen, K.M., Kananen, K., Auriola, S., Kontturi, K., Urtti, A., 1997. Characterization of paracellular and aqueous penetration routes in cornea, conjunctiva, and sclera. Investig. Ophthalmol. Vis. Sci. 38, 627–634.
Han, K.E., Choi, S. Il, Kim, T.I., Maeng, Y.S., Stulting, R.D., Ji, Y.W., Kim, E.K., 2016. Pathogenesis and treatments of TGFBI corneal dystrophies. Prog. Retin. Eye Res. https://doi.org/10.1016/j.preteyeres.2015.11.002.
Hanna, C., Bicknell, D., O'Brien, J., 1961. Cell turnover in the adult human eye. Arch. Ophthalmol. 65, 695–698. https://doi.org/10.1001/archopht.1961.01840020697016.
Hauck, B., Chen, L., Xiao, W., 2003. Generation and characterization of chimeric recombinant AAV vectors. Mol. Ther. 7, 419–425. https://doi.org/10.1016/S1525-0016(03)00012-1.
Hauswirth, W.W., Aleman, T.S., Kaushal, S., Cideciyan, A.V., Schwartz, S.B., Wang, L., Conlon, T.J., Boye, S.L., Flotte, T.R., Byrne, B.J., Jacobson, S.G., 2008. Treatment of leber congenital amaurosis due to RPE65 mutations by ocular subretinal injection of adeno-associated virus gene vector: short-term results of a phase I trial. Hum. Gene Ther. 19, 979–990. https://doi.org/10.1089/hum.2008.107.
Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., Wilkens, A.B., Bacchetta, R., Tsalenko, A., Dellinger, D., Bruhn, L., Porteus, M.H., 2015. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nat. Biotechnol. 33, 985–989. https://doi.org/10.1038/nbt.3290.
Hentze, M.W., Kulozik, A.E., 1999. A perfect message: RNA surveillance and nonsense-mediated decay. Cell. https://doi.org/10.1016/S0092-8674(00)80542–80545.
Hill, J.C., 1994. Systemic cyclosporine in high-risk keratoplasty: short- versus long-term therapy. Ophthalmology 101, 128–133. https://doi.org/10.1016/S0161-6420(13)31253-6.
Hilton, I.B., Gersbach, C.A., 2015. Enabling functional genomics with genome engineering. Genome Res. https://doi.org/10.1101/gr.190124.115.
Hippert, C., Ibanes, S., Serratrice, N., Court, F., Malecaze, F., Kremer, E.J., Kalatzis, V., 2012. Corneal transduction by intra-stromal injection of AAV vectors in vivo in the mouse and Ex vivo in human explants. PLoS One 7. https://doi.org/10.1371/journal.pone.0035318.
Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L., Sontheimer, E.J., Thomson, J.A., 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. USA 110, 15644–15649. https://doi.org/10.1073/pnas.1313587110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1313587110.
Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., Zhang, F., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832. https://doi.org/10.1038/nbt.2647.
Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., Frock, R.L., 2016. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification–mediated high-throughput genome-wide translocation sequencing. Nat. Protoc. 11, 853–871. https://doi.org/10.1038/nprot.2016.043.
Hung, S.S.C., Chrysostomou, V., Li, F., Lim, J.K.H., Wang, J.H., Powell, J.E., Tu, L., Daniszewski, M., Lo, C., Wong, R.C., Crowston, J.G., Pébay, A., King, A.E., Bui, B.V., Liu, G.S., Hewitt, A.W., 2016. AAV-mediated CRISPR/cas gene editing of retinal cells in vivo. Investig. Ophthalmol. Vis. Sci. 57, 3470–3476. https://doi.org/10.1167/iovs.16-19316.
Irion, S., Luche, H., Gadue, P., Fehling, H.J., Kennedy, M., Keller, G., 2007. Identification and targeting of the ROSA26 locus in human embryonic stem cells. Nat. Biotechnol. 25, 1477–1482. https://doi.org/10.1038/nbt1362.
Iyama, T., Wilson, D.M., 2013. DNA repair mechanisms in dividing and non-dividing cells. DNA Repair (Amst) 12, 620–636. https://doi.org/10.1016/j.dnarep.2013.04.015.
Iyer, V., Shen, B., Zhang, W., Hodgkins, A., Keane, T., Huang, X., Skarnes, W.C., 2015. Off-target mutations are rare in Cas9-modified mice. Nat. Methods 12 479–479. https://doi.org/10.1038/nmeth.3408.
Jacobson, S.G., Cideciyan, A.V., Roman, A.J., Sumaroka, A., Schwartz, S.B., Heon, E., Hauswirth, W.W., 2015. Improvement and decline in vision with gene therapy in childhood blindness. N. Engl. J. Med. 372, 1920–1926. https://doi.org/10.1056/NEJMoa1412965.
Jasin, M., Haber, J.E., 2016. The democratization of gene editing: insights from site-specific cleavage and double-strand break repair. DNA Repair (Amst). https://doi.org/10.1016/j.dnarep.2016.05.001.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA – guided DNA endonuclease in adaptice bacterial immunity. Science 337, 816–822. https://doi.org/10.1126/science.1225829.
Jun, R.M., Tchah, H., Kim, T.I., Stulting, R.D., Jung, S.E., Seo, K.Y., Lee, D.H., Kim, E.K., 2004. Avellino corneal dystrophy after LASIK. Ophthalmology. https://doi.org/10.1016/j.ophtha.2003.06.026.
Karginov, F.V., Hannon, G.J., 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell. https://doi.org/10.1016/j.molcel.2009.12.033.
Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., Maehr, R., 2015. Functional annotation of native enhancers with a Cas9–histone demethylase fusion. Nat. Methods 12, 401–403. https://doi.org/10.1038/nmeth.3325.
Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.-I., Kim, J.-S., 2015.

Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat. Methods 12, 237–243. https://doi.org/10.1038/nmeth.3284.
Kim, K., Park, S.W., Kim, J.H., Lee, S.H., Kim, D., Koo, T., Kim, K.E., Kim, J.H., Kim, J.S., 2017. Genome surgery using Cas9 ribonucleoproteins for the treatment of age-related macular degeneration. Genome Res. 27, 419–426. https://doi.org/10.1101/gr.219089.116.
Kim, S., Kim, D., Cho, S.W., Kim, J., Kim, J.S., 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 24, 1012–1019. https://doi.org/10.1101/gr.171322.113.
Kim, T.-I., Roh, M.I., Grossniklaus, H.E., Kang, S.J., Hamilton, S.M., Schorderet, D.F., Lee, W.B., Kim, E.K., 2008. Deposits of transforming growth factor-beta-induced protein in granular corneal dystrophy type II after LASIK. Cornea 27, 28–32. https://doi.org/10.1097/ICO.0b013e318156d36d.
Klausner, E.A., Zhang, Z., Chapman, R.L., Multack, R.F., Volin, M.V., 2010. Ultrapure chitosan oligomers as carriers for corneal gene transfer. Biomaterials 31, 1814–1820. https://doi.org/10.1016/j.biomaterials.2009.10.031.
Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., Joung, J.K., 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529, 490–495. https://doi.org/10.1038/nature16526.
Klintworth, G.K., 2009. Corneal dystrophies. Orphanet J. rare Dis. Cornea Dystrophy Foundaion 4 (7). https://doi.org/10.1186/1750-1172-4-7.
Kocher, T., Peking, P., Klausegger, A., Murauer, E.M., Hofbauer, J.P., Wally, V., Lettner, T., Hainzl, S., Ablinger, M., Bauer, J.W., Reichelt, J., Koller, U., 2017. Cut and paste: efficient homology-directed repair of a dominant negative KRT14 mutation via CRISPR/cas9 nickases. Mol. Ther. https://doi.org/10.1016/j.ymthe.2017.08.015.
Kodama, Y., Nakamura, T., Kurosaki, T., Egashira, K., Mine, T., Nakagawa, H., Muro, T., Kitahara, T., Higuchi, N., Sasaki, H., 2014. Biodegradable nanoparticles composed of dendrigraft poly-l-lysine for gene delivery. Eur. J. Pharm. Biopharm. 87, 472–479. https://doi.org/10.1016/j.ejpb.2014.04.013.
Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R., 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424. https://doi.org/10.1038/nature17946.
Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., Nureki, O., Zhang, F., 2014. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583–588. https://doi.org/10.1038/nature14136.
Kotin, R.M., Linden, R.M., Berns, K.I., 1992. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J. 11, 5071–5078. https://doi.org/10.1182/blood-2008-10-181479.
Le Bourlais, C., Acar, L., Zia, H., Sado, P.A., Needham, T., Leverge, R., 1998. Ophthalmic drug delivery systems—recent advances. Prog. Retin. Eye Res. 17, 33–58. https://doi.org/10.1016/S1350-9462(97)00002-5.
Lee, K., Conboy, M., Park, H.M., Jiang, F., Kim, H.J., Dewitt, M.A., Mackley, V.A., Chang, K., Rao, A., Skinner, C., Shobha, T., Mehdipour, M., Liu, H., Huang, W., Lan, F., Bray, N.L., Li, S., Corn, J.E., Kataoka, K., Doudna, J.A., Conboy, I., Murthy, N., 2017. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. Nat. Biomed. Eng. https://doi.org/10.1038/s41551-017-0137-2.
Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., Chesnut, J.D., 2015. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J. Biotechnol. 208, 44–53. https://doi.org/10.1016/j.jbiotec.2015.04.024.
Lin, S., Staahl, B.T., Alla, R.K., Doudna, J.A., 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3, e04766. https://doi.org/10.7554/eLife.04766.
Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., Jaenisch, R., 2016. Editing DNA methylation in the mammalian genome. Cell 167, 233–247. e17. https://doi.org/10.1016/j.cell.2016.08.056.
Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D.F., Neri, M., Magnani, Z., Cantore, A., Lo Riso, P., Damo, M., Pello, O.M., Holmes, M.C., Gregory, P.D., Gritti, A., Broccoli, V., Bonini, C., Naldini, L., 2011. Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat. Methods 8, 861–869. https://doi.org/10.1038/nmeth.1674.
Lungwitz, U., Breunig, M., Blunk, T., Göpferich, A., 2005. Polyethylenimine-based non-viral gene delivery systems. Eur. J. Pharm. Biopharm. 247–266. https://doi.org/10.1016/j.ejpb.2004.11.011.
Ma, H., Tu, L.-C., Naseri, A., Huisman, M., Zhang, S., Grunwald, D., Pederson, T., 2016. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nat. Biotechnol. 34, 528–530. https://doi.org/10.1038/nbt.3526.
Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., Joung, J.K., 2013. CRISPR RNA-guided activation of endogenous human genes. Nat. Methods 10, 977–979. https://doi.org/10.1038/nmeth.2598.
Maguire, A.M., Simonelli, F., Pierce, E.A., Pugh, E.N., Mingozzi, F., Bennicelli, J., Banfi, S., Marshall, K.A., Testa, F., Surace, E.M., Rossi, S., Lyubarsky, A., Arruda, V.R., Konkle, B., Stone, E., Sun, J., Jacobs, J., Dell'Osso, L., Hertle, R., Ma, J., Redmond, T.M., Zhu, X., Hauck, B., Zelenaia, O., Shindler, K.S., Maguire, M.G., Wright, J.F., Volpe, N.J., McDonnell, J.W., Auricchio, A., High, K.A., Bennett, J., 2008. Safety and efficacy of gene transfer for Leber's congenital amaurosis. N. Engl. J. Med. 358, 2240–2248. https://doi.org/10.1056/NEJMoa0802315.
Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin, E.V., 2011. Evolution and classification of the CRISPR–Cas systems. Nat. Rev. Microbiol. 9, 467–477. https://doi.org/10.1038/nrmicro2577.
Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., Church, G.M., 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838. https://doi.org/10.1038/nbt.2675.
Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013b. RNA-guided human genome engineering via Cas9. Science (80-.) 339, 823–826. https://doi.org/10.1126/science.1232033.
Mao, S., Sun, W., Kissel, T., 2010. Chitosan-based formulations for delivery of DNA and siRNA. Adv. Drug Deliv. Rev. https://doi.org/10.1016/j.addr.2009.08.004.
Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., Ploegh, H.L., 2016. Corrigendum: increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. 34 210–210. https://doi.org/10.1038/nbt0216-210c.
Matthaei, M., Meng, H., Bhutto, I., Xu, Q., Boelke, E., Hanes, J., Jun, A.S., 2012. Systematic assessment of microneedle injection into the mouse cornea. Eur. J. Med. Res. 17, 19. https://doi.org/10.1186/2047-783X-17-19.
Mavilio, F., Pellegrini, G., Ferrari, S., Di Nunzio, F., Di Iorio, E., Recchia, A., Maruggi, G., Ferrari, G., Provasi, E., Bonini, C., Capurro, S., Conti, A., Magnoni, C., Giannetti, A., De Luca, M., 2006. Correction of junctional epidermolysis bullosa by transplantation of genetically modified epidermal stem cells. Nat. Med. 12, 1397–1402. https://doi.org/10.1038/nm1504.
Mohan, R.R., Sharma, A., Netto, M.V., Sinha, S., Wilson, S.E., 2005. Gene therapy in the cornea. Prog. Retin. Eye Res. https://doi.org/10.1016/j.preteyeres.2005.04.001.
Mohan, R.R., Tovey, J.C.K., Sharma, A., Tandon, A., 2012. Gene therapy in the cornea: 2005-present. Prog. Retin. Eye Res. https://doi.org/10.1016/j.preteyeres.2011.09.001.
Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740. https://doi.org/10.1099/mic.0.023960-0.
Monopoli, M.P., Åberg, C., Salvati, A., Dawson, K.A., 2012. Biomolecular coronas provide the biological identity of nanosized materials. Nat. Nanotechnol. 7, 779–786. https://doi.org/10.1038/nnano.2012.207.
Moore, R., Spinhirne, A., Lai, M.J., Preisser, S., Li, Y., Kang, T., Bleris, L., 2015. CRISPR-based self-cleaving mechanism for controllable gene delivery in human cells. Nucleic Acids Res. 43, 1297–1303. https://doi.org/10.1093/nar/gku1326.
Nettelbeck, D.M., Jérôme, V., Müller, R., 1998. A strategy for enhancing the transcriptional activity of weak cell type-specific promoters. Gene Ther. 5, 1656–1664. https://doi.org/10.1038/sj.gt.3300778.
Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156, 935–949. https://doi.org/10.1016/j.cell.2014.02.001.
NoRathi, V.M., Taneja, M., Murthy, S.I., Bagga, B., Vaddavalli, P.K., Sangwan, V.S., 2016. Phototherapeutic keratectomy for recurrent granular dystrophy in postpenetrating keratoplasty eyes. Indian J. Ophthalmol. 64, 140–144.
O'Callaghan, J., Crosbie, D.E., Cassidy, P.S., Sherwood, J.M., Flügel-Koch, C., Lütjen-Drecoll, E., Humphries, M.M., Reina-Torres, E., Wallace, D., Kiang, A.S., Campbell, M., Stamer, W.D., Overby, D.R., O'Brien, C., Tam, L.C.S., Humphries, P., 2017. Therapeutic potential of AAV-mediated MMP-3 secretion from corneal endothelium in treating glaucoma. Hum. Mol. Genet. 26, 1230–1246. https://doi.org/10.1093/hmg/ddx028.
Oliveira, A.V., Rosa da Costa, A.M., Silva, G.A., 2017. Non-viral strategies for ocular gene delivery. Mater. Sci. Eng. C. https://doi.org/10.1016/j.msec.2017.04.068.
Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., Liu, D.R., 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. 31, 839–843. https://doi.org/10.1038/nbt.2673.
Pellegrini, G., Golisano, O., Paterna, P., Lambiase, A., Bonini, S., Rama, P., De Luca, M., 1999. Location and clonal analysis of stem cells and their differentiated progeny in the human ocular surface. J. Cell Biol. 145, 769–782. https://doi.org/10.1083/jcb.145.4.769.
Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., Guilak, F., Crawford, G.E., Reddy, T.E., Gersbach, C.A., 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods 10, 973–976. https://doi.org/10.1038/nmeth.2600.
Petris, G., Casini, A., Montagna, C., Lorenzin, F., Prandi, D., Romanel, A., Zasso, J., Conti, L., Demichelis, F., Cereseto, A., 2017. Hit and go CAS9 delivered through a lentiviral based self-limiting circuit. Nat. Commun. 8. https://doi.org/10.1038/ncomms15334.
Pierce, E.A., Bennett, J., 2015. The status of RPE65 gene therapy trials: safety and efficacy. Cold Spring Harb. Perspect. Med 5. https://doi.org/10.1101/cshperspect.a017285.
Putnam, D., 2006. Polymers for gene delivery across length scales. Nat. Mater. 5, 439–451. https://doi.org/10.1038/nmat1645.
Qasim, W., Zhan, H., Samarasinghe, S., Adams, S., Amrolia, P., Stafford, S., Butler, K., Rivat, C., Wright, G., Somana, K., Ghorashian, S., Pinner, D., Ahsan, G., Gilmour, K., Lucchini, G., Inglott, S., Mifsud, W., Chiesa, R., Peggs, K.S., Chan, L., Farzaneh, F., Thrasher, A.J., Vora, A., Pule, M., Veys, P., 2017. Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. Sci. Transl. Med. 9 eaaj2013. https://doi.org/10.1126/scitranslmed.aaj2013.
Rajala, A., Wang, Y., Zhu, Y., Ranjo-Bishop, M., Ma, J.X., Mao, C., Rajala, R.V.S., 2014. Nanoparticle-assisted targeted delivery of eye-specific genes to eyes significantly improves the vision of blind mice in vivo. Nano Lett. 14, 5257–5263. https://doi.org/10.1021/nl502275s.
Rama, P., Bonini, S., Lambiase, A., Golisano, O., Paterna, P., De Luca, M., Pellegrini, G., 2001. Autologous fibrin-cultured limbal stem cells permanently restore the corneal surface of patients with total limbal stem cell deficiency. Transplantation 72, 1478–1485. https://doi.org/10.1097/00007890-200111150-00002.
Rama, P., Matuska, S., Paganoni, G., Spinelli, A., De Luca, M., Pellegrini, G., 2010. Limbal

stem-cell therapy and long-term corneal regeneration. N. Engl. J. Med. 363, 147–155. https://doi.org/10.1056/NEJMoa0905955.
Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., Koonin, E.V., Sharp, P.A., Zhang, F., 2015. In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186–191. https://doi.org/10.1038/nature14299.
Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., Zhang, F., 2013a. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell 154, 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021.
Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., Zhang, F., 2013b. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308. https://doi.org/10.1038/nprot.2013.143.
Richardson, C.D., Ray, G.J., Bray, N.L., Corn, J.E., 2016. Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. Nat. Commun. 7, 12463. https://doi.org/10.1038/ncomms12463.
Riley, M., Vermerris, W., 2017. Recent advances in nanomaterials for gene delivery—a review. Nanomaterials 7 (94). https://doi.org/10.3390/nano7050094.
Romanelli, L., Valeri, P., Morrone, L.A., Pimpinella, G., Graziani, G., Tita, B., 1994. Ocular absorption and distribution of bendazac after topical administration to rabbits with different vehicles. Life Sci. 54, 877–885. https://doi.org/10.1016/0024-3205(94)00624-5.
Santos, H.A., Mäkilä, E., Airaksinen, A.J., Bimbo, L.M., Hirvonen, J., 2014. Porous silicon nanoparticles for nanomedicine: preparation and biomedical applications. Nanomedicine 9, 535–554. https://doi.org/10.2217/nnm.13.223.
Schermer, A., Galvin, S., Sun, T.T., 1986. Differentiation-related expression of a major 64K corneal keratin in vivo and in culture suggests limbal location of corneal epithelial stem cells. J. Cell Biol. 103, 49–62. https://doi.org/10.1083/jcb.103.1.49.
Seitz, B., Lisch, W., 2011. Stage-related therapy of corneal dystrophies. Corneal Dystrophies 116–153. https://doi.org/10.1159/000324081.
Sharma, A., Tovey, J.C.K., Ghosh, A., Mohan, R.R., 2010. AAV serotype influences gene transfer in corneal stroma in vivo. Exp. Eye Res. 91, 440–448. https://doi.org/10.1016/j.exer.2010.06.020.
Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., Waterston, R.H., 2017. DNA sequencing at 40: past, present and future History of DNA sequencing technologies. Nat. Publ. Gr. https://doi.org/10.1038/nature24286.
Sheridan, C., 2017. CRISPR therapeutics push into human testing. Nat. Biotechnol. 35, 3–5. https://doi.org/10.1038/nbt0117-3.
Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., Zhang, F., 2016. Rationally engineered Cas9 nucleases with improved specificity. Science (80-.) 351, 84–88. https://doi.org/10.1126/science.aad5227.
Song, J., Yang, D., Xu, J., Zhu, T., Chen, Y.E., Zhang, J., 2016. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. Nat. Commun. 7, 10548. https://doi.org/10.1038/ncomms10548.
Spencer, B., Agarwala, S., Miskulin, M., Smith, M., Brandt, C.R., 2000. Herpes simplex virus-mediated gene delivery to the rodent visual system. Investig. Ophthalmol. Vis. Sci. 41, 1392–1401.
Srivastava, A., 2016. In vivo tissue-tropism of adeno-associated viral vectors. Curr. Opin. Virol. https://doi.org/10.1016/j.coviro.2016.08.003.
Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., Doudna, J.A., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62–67. https://doi.org/10.1038/nature13011.
Stoddard, B.L., 2011. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. Structure. https://doi.org/10.1016/j.str.2010.12.003.
Straub, C., Granger, A.J., Saulnier, J.L., Sabatini, B.L., 2014. CRISPR/Cas9-mediated gene knock-down in post-mitotic neurons. PLoS One 9. https://doi.org/10.1371/journal.pone.0105584.
Sullenger, B.A., Nair, S., 2016. From the RNA world to the clinic. Science (80-.) 352, 1417–1420. https://doi.org/10.1126/science.aad8709.
Swiech, L., Heidenreich, M., Banerjee, A., Habib, N., Li, Y., Trombetta, J., Sur, M., Zhang, F., 2015. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. Nat. Biotechnol. 33, 102–106. https://doi.org/10.1038/nbt.3055.
Tabebordbar, M., Zhu, K., Cheng, J.K.W., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A., Cong, L., Zhang, F., Vandenberghe, L.H., Church, G.M., Wagers, A.J., 2016. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science (80-.) 351, 407–411. https://doi.org/10.1126/science.aad5177.
Truong, D.J.J., Kühner, K., Kühn, R., Werfel, S., Engelhardt, S., Wurst, W., Ortiz, O., 2015. Development of an intein-mediated split-Cas9 system for gene therapy. Nucleic Acids Res. 43, 6450–6458. https://doi.org/10.1093/nar/gkv601.
Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., Joung, J.K., 2017. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. Nat. Methods 14, 607–614. https://doi.org/10.1038/nmeth.4278.
Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., Aryee, M.J., Joung, J.K., 2014. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187–197. https://doi.org/10.1038/nbt.3117.
Tsubota, K., Inoue, H., Ando, K., Ono, M., Yoshino, K., Saito, I., 1998. Adenovirus-mediated gene transfer to the ocular surface epithelium. Exp. Eye Res. 67, 531–538. https://doi.org/10.1006/exer.1998.0557.
Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., Gregory, P.D., 2010. Genome editing with engineered zinc finger nucleases. Nat. Rev. Genet. 11, 636–646. https://doi.org/10.1038/nrg2842.
van Overbeek, M., Capurso, D., Carter, M.M., Thompson, M.S., Frias, E., Russ, C., Reece-Hoyes, J.S., Nye, C., Gradia, S., Vidal, B., Zheng, J., Hoffman, G.R., Fuller, C.K., May, A.P., 2016. DNA repair profiling reveals nonrandom outcomes at cas9-mediated breaks. Mol. Cell 63, 633–646. https://doi.org/10.1016/j.molcel.2016.06.037.
van Putten, M., Hulsker, M., Nadarajah, V.D., van Heiningen, S.H., van Huizen, E., van Iterson, M., Admiraal, P., Messemaker, T., den Dunnen, J.T., t Hoen, P.A.C., Aartsma-Rus, A., 2012. The effects of low levels of dystrophin on mouse muscle function and pathology. PLoS One 7. https://doi.org/10.1371/journal.pone.0031937.
Van Putten, M., Hulsker, M., Young, C., Nadarajah, V.D., Heemskerk, H., Van Der Weerd, L., T Hoen, P.A.C., Van Ommen, G.J.B., Aartsma-Rus, A.M., 2013. Low dystrophin levels increase survival and improve muscle pathology and function in dystrophin/utrophin double-knockout mice. FASEB J 27, 2484–2495. https://doi.org/10.1096/fj.12-224170.
Van Putten, M., Van der Pijl, E.M., Hulsker, M., Verhaart, I.E.C., Nadarajah, V.D., van der Weerd, L., Aartsma-Rus, A., 2014. Low dystrophin levels in heart can delay heart failure in mdx mice. J. Mol. Cell. Cardiol. 69, 17–23. https://doi.org/10.1016/j.yjmcc.2014.01.009.
Van Tendeloo, V.F.I., Ponsaerts, P., Berneman, Z.N., 2007 Oct. mRNA-based gene transfer as a tool for gene and cell therapy. Curr. Opin. Mol. Ther. 9 (5), 423–431.
Vance, M., Llanga, T., Bennett, W., Woodard, K., Murlidharan, G., Chungfat, N., Asokan, A., Gilger, B., Kurtzberg, J., Samulski, R.J., Hirsch, M.L., 2016. AAV gene therapy for MPS1-associated corneal blindness. Sci. Rep. 6, 22131. https://doi.org/10.1038/srep22131.
Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Talkowski, M.E., Musunuru, K., 2014. Low incidence of Off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell 15, 27–30. https://doi.org/10.1016/j.stem.2014.04.020.
Vigneswara, V., Akpan, N., Berry, M., Logan, A., Troy, C.M., Ahmed, Z., 2014. Combined suppression of CASP2 and CASP6 protects retinal ganglion cells from apoptosis and promotes axon regeneration through CNTF-mediated JAK/STAT signalling. Brain 137, 1656–1675. https://doi.org/10.1093/brain/awu037.
Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., Jaenisch, R., 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. Cell 153, 910–918. https://doi.org/10.1016/j.cell.2013.04.025.
Wang, L., Xiao, R., Andres-Mateos, E., Vandenberghe, L.H., 2017. Single stranded adeno-associated virus achieves efficient gene transfer to anterior segment in the mouse eye. PLoS One 12. https://doi.org/10.1371/journal.pone.0182473.
Weiss, J.S., Møller, H.U., Aldave, A.J., Seitz, B., Bredrup, C., Kivelä, T., Munier, F.L., Rapuano, C.J., Nischal, K.K., Kim, E.K., Sutphin, J., Busin, M., Labbé, A., Kenyon, K.R., Kinoshita, S., Lisch, W., 2015. IC3D classification of corneal dystrophies—edition 2. Cornea 34, 117–159. https://doi.org/10.1097/ICO.0000000000000307.
Weleber, R.G., Pennesi, M.E., Wilson, D.J., Kaushal, S., Erker, L.R., Jensen, L., McBride, M.T., Flotte, T.R., Humphries, M., Calcedo, R., Hauswirth, W.W., Chulay, J.D., Stout, J.T., 2016. Results at 2 Years after gene therapy for RPE65-deficient leber congenital amaurosis and severe early-childhood-onset retinal dystrophy. Ophthalmology 123, 1606–1620. https://doi.org/10.1016/j.ophtha.2016.03.003.
Wiley, L.A., Burnight, E.R., Songstad, A.E., Drack, A.V., Mullins, R.F., Stone, E.M., Tucker, B.A., 2015. Patient-specific induced pluripotent stem cells (iPSCs) for the study and treatment of retinal degenerative diseases. Prog. Retin. Eye Res. https://doi.org/10.1016/j.preteyeres.2014.10.002.
Wu, H.M., Pan, S.R., Chen, M.W., Wu, Y., Wang, C., Wen, Y.T., Zeng, X., Wu, C.B., 2011. A serum-resistant polyamidoamine-based polypeptide dendrimer for gene transfection. Biomaterials 32, 1619–1634. https://doi.org/10.1016/j.biomaterials.2010.09.045.
Wu, X., Kriz, A.J., Sharp, P.A., 2014a. Target specificity of the CRISPR-Cas9 system. Quant. Biol. 2, 59–70. https://doi.org/10.1007/s40484-014-0030-x.
Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., Jaenisch, R., Zhang, F., Sharp, P.A., 2014b. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. 32, 670–676. https://doi.org/10.1038/nbt.2889.
Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X., Huang, P.Y., Daley, G., Church, G., 2013. Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 41, 9049–9061. https://doi.org/10.1093/nar/gkt555.
Yang, Y., Wang, L., Bell, P., McMenamin, D., He, Z., White, J., Yu, H., Xu, C., Morizono, H., Musunuru, K., Batshaw, M.L., Wilson, J.M., 2016. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. Nat. Biotechnol. 34, 334–338. https://doi.org/10.1038/nbt.3469.
Yin, H., Song, C.-Q., Dorkin, J.R., Zhu, L.J., Li, Y., Wu, Q., Park, A., Yang, J., Suresh, S., Bizhanova, A., Gupta, A., Bolukbasi, M.F., Walsh, S., Bogorad, R.L., Gao, G., Weng, Z., Dong, Y., Koteliansky, V., Wolfe, S.A., Langer, R., Xue, W., Anderson, D.G., 2016. Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. Nat. Biotechnol. 34, 328–333. https://doi.org/10.1038/nbt.3471.
Yin, H., Song, C.-Q., Suresh, S., Wu, Q., Walsh, S., Rhym, L.H., Mintzer, E., Bolukbasi, M.F., Zhu, L.J., Kauffman, K., Mou, H., Oberholzer, A., Ding, J., Kwan, S.-Y., Bogorad, R.L., Zatsepin, T., Koteliansky, V., Wolfe, S.A., Xue, W., Langer, R., Anderson, D.G., 2017. Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. Nat. Biotechnol. https://doi.org/10.1038/nbt.4005.
Yu, W., Mookherjee, S., Chaitankar, V., Hiriyanna, S., Kim, J.-W., Brooks, M., Ataeijannati, Y., Sun, X., Dong, L., Li, T., Swaroop, A., Wu, Z., 2017. Nrl knockdown by AAV-delivered CRISPR/Cas9 prevents retinal degeneration in mice. Nat. Commun. 8, 14716. https://doi.org/10.1038/ncomms14716.
Zangi, L., Lui, K.O., von Gise, A., Ma, Q., Ebina, W., Ptaszek, L.M., Später, D., Xu, H., Tabebordbar, M., Gorbatov, R., Sena, B., Nahrendorf, M., Briscoe, D.M., Li, R.A., Wagers, A.J., Rossi, D.J., Pu, W.T., Chien, K.R., 2013. Modified mRNA directs the fate of heart progenitor cells and induces vascular regeneration after myocardial

infarction. Nat. Biotechnol. 31, 898–907. https://doi.org/10.1038/nbt.2682.
Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E.V., Zhang, F., 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-cas system. Cell 163, 759–771. https://doi.org/10.1016/j.cell.2015.09.038.
Zinn, E., Vandenberghe, L.H., 2014. Adeno-associated virus: fit to serve. Curr. Opin. Virol 8, 90–97. https://doi.org/10.1016/j.coviro.2014.07.008.
Zischewski, J., Fischer, R., Bortesi, L., 2017. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. Biotechnol. Adv. https://doi.org/10.1016/j.biotechadv.2016.12.003.
Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.-Y., Liu, D.R., 2014. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat. Biotechnol. 33, 73–80. https://doi.org/10.1038/nbt.3081.

Paper II

3 Paper II - Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders

Kathleen A. Christie, David G. Courtney, Larry A. DeDionisio, Connie Chao Shern, Shyamsree De Majumdar, Laura C. Mairs, M. Andrew Nesbit, C.B.Tara Moore

The main aims of this paper were to:

1. Demonstrate gene disruption by NHEJ *in vivo*
2. Determine what proportion of the *TGFBI* missense mutations could be targeted by an allele-specific CRISPR/Cas9 targeting approach
3. Compare a PAM-specific approach and guide-specific approach and determine which confers more stringent allele-specificity when targeted to single base pair mutations
4. Assess whether the 5 most prevalent *TGFBI* mutations can be targeted by either approach

CONTRIBUTION

For this manuscript I performed the mutational analysis of the 62 *TGFBI* missense mutations. I cloned all plasmid constructs and generated all cleavage templates used throughout. I performed all dual-luciferase assays and *in vitro* digests. I formed a collaboration with Prof Feng Zhang for use of his mutant Cpf1 construct. I wrote the manuscript and prepared all figures.

SCIENTIFIC REPORTS

OPEN

Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders

Received: 29 June 2017
Accepted: 9 November 2017
Published online: 23 November 2017

Kathleen A. Christie[1], David G. Courtney[1], Larry A. DeDionisio[2], Connie Chao Shern[2], Shyamasree De Majumdar[1], Laura C. Mairs[1], M. Andrew Nesbit[1] & C. B. Tara Moore[1,2]

CRISPR/Cas9 holds immense potential to treat a range of genetic disorders. Allele-specific gene disruption induced by non-homologous end-joining (NHEJ) DNA repair offers a potential treatment option for autosomal dominant disease. Here, we successfully delivered a plasmid encoding *S. pyogenes* Cas9 and sgRNA to the corneal epithelium by intrastromal injection and acheived long-term knockdown of a corneal epithelial reporter gene, demonstrating gene disruption via NHEJ *in vivo*. In addition, we used *TGFBI* corneal dystrophies as a model of autosomal dominant disease to assess the use of CRISPR/Cas9 in two allele-specific systems, comparing cleavage using a SNP-derived PAM to a guide specific approach. *In vitro*, cleavage via a SNP-derived PAM was found to confer stringent allele-specific cleavage, while a guide-specific approach lacked the ability to distinguish between the wild-type and mutant alleles. The failings of the guide-specific approach highlights the necessity for meticulous guide design and assessment, as various degrees of allele-specificity are achieved depending on the guide sequence employed. A major concern for the use of CRISPR/Cas9 is its tendency to cleave DNA non-specifically at "off-target" sites. Confirmation that *S. pyogenes* Cas9 lacks the specificity to discriminate between alleles differing by a single base-pair regardless of the position in the guide is demonstrated.

The promise of personalised gene therapy has been brought nearer fruition with the recent advances in the field of genome engineering, particularly the development of Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/CRISPR associated protein (Cas) systems. CRISPR/Cas9 is an RNA guided endonuclease, that has been manipulated for use in mammalian cells to act as a two component system, requiring only a Cas9 nuclease and a single guide RNA (sgRNA)[1–3]. Cas9 can be directed to cut a desired sequence in the genome, provided it is directly upstream of a protospacer adjacent motif (PAM), by simply altering the guide RNA sequence (Fig. 1a). The site-specific sgRNA will direct the Cas9 nuclease to make a double strand break (DSB). The cell will then attempt to repair this damage, by either error-prone non-homologous end joining (NHEJ) or precise homology directed repair (HDR)[4], and it is by these different cellular responses that different forms of gene editing can be achieved.

NHEJ can be utilised to generate gene knockouts, due to the high frequency of frameshifting mutations generated[4,5]. Allele-specific gene disruption via NHEJ is a potential approach to treat dominant negative disorders, in which the causative gene is haplosufficient; this involves targeting the mutant allele alone for disruption, leaving the wild-type allele intact and restoring the phenotype[6–9]. This approach relies on the ability of the targeting system to unequivocally discriminate between wild-type and mutant sequence.

Although CRISPR/Cas9 holds immense promise, one caveat to the use of the system is that Cas9 nuclease has been shown to tolerate mismatches between the guide sequence and the target[10,11]. This can lead to off-targeting elsewhere in the genome or, indeed in this case, cleavage of the wild-type allele. Efforts have been made to increase the specificity of Cas9 and eliminate off-target cutting, including; use of truncated guides[12], Cas9 variants from other bacterial species to exploit more intricate PAMs[13], rationally engineering the Cas9 nuclease[14], and using the mutant sequence to induce specificity such as utilising a novel SNP-derived PAM[15,16].

In particular utilising a novel PAM is an attractive option for allele-specific editing. Previous work has demonstrated that when mutations result in a novel PAM, guide RNAs can be designed, utilising this new PAM, allowing only the mutant allele to be targeted, producing an allele-specific knockout[14].

[1]Biomedical Sciences Research Institute, Ulster University, Coleraine, Northern Ireland, BT52 1SA, UK. [2]Avellino Laboratories, Menlo Park, California, CA, 94025, USA. Correspondence and requests for materials should be addressed to C.B.T.M. (email: t.moore@ulster.ac.uk)

a.

Mutation	Associated Corneal Dystrophy	Codon Change
R124C	Classic Lattice Corneal Dystrophy Type I	CGC > TGC
R124H	Avellino Corneal Dystrophy	CGC > CAC
R124L	Reis-Bücklers Corneal Dystrophy	CGC > CTC
R555Q	Thiel-Behnke Corneal Dystrophy	CGG > CAG
R555W	Classic Granular Corneal Dystrophy Type I	CGG > TGG

b.

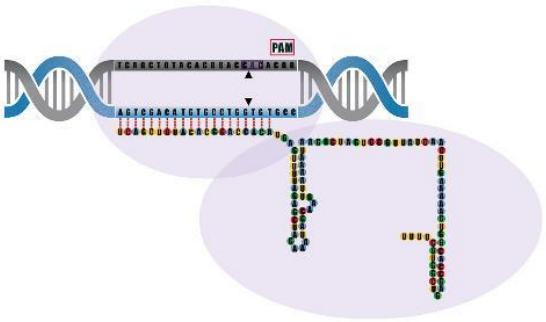


Figure 1. *S. pyogenes* Cas9 to treat dominant negative *TGFBI* corneal dystrophies. (**a**) Cas9 (purple outline) can be directed to cut any sequence in the genome (DNA target in grey), provided it is directly upstream of a protospacer adjacent motif known as PAM (pink box). This can be achieved by altering the 20 nucleotide guide sequence, which is associated with a 82 nucleotide scaffold. (**b**) 5 prevalent *TGFBI* mutations and their associated corneal dystrophy and codon change. (**c**) Schematic of the position of the 60 missense mutations across the *TGFBI* gene. The hotspots at exons 4, 11, 12 and 14 are evident, with exons 4 and 12 expanded to show the location of the 5 most prevalent *TGFBI* mutations; R124C, R124H, R125L, R555Q and R555W.

The cornea offers an ideal platform for testing personalised gene therapy, due to its immediate accessibility, small surface area and immune-privileged status. Collectively the corneal dystrophies represent a group of inheritable blinding diseases that alter the shape or transparency of the cornea. Currently mutations in 14 genes are associated with corneal dystrophies, 9 of them presenting with an autosomal dominant inheritance pattern[17]. Corneal dystrophies linked to these 9 genes predominantly result from missense mutations or small in-frame insertions or deletions that cause disease by a dominant negative effect of the mutant protein[17].

Transforming growth factor beta induced protein (*TGFBI*p) has been linked to a range of stromal or stromal-epithelial corneal dystrophies[18–20]. *TGFBI* is predominantly produced in the corneal epithelium and is transported to the stromal layer, where the mutant protein accumulates[21]. To-date a total of 60 missense mutations in *TGFBI* have been linked to various corneal dystrophies[22,23]. These mutations span the entire *TGFBI* gene but are clustered in hotspots found in exon 4, 11, 12 and 14.

Despite the wide spectrum of mutations, the vast majority of cases are due to 5 prevalent mutations found in either codon 124 (exon 4) or codon 555 (exon 12)[24,25] (Fig. 1b). These 5 mutations include; R124C, R124H, R124L, R555Q and R555W and as described account for the bulk of reported cases of *TGFBI* corneal dystrophies[26–28]. Remarkably, each of these mutations, differing by only a single amino acid, result in strikingly different protein aggregates with a very strong genotype-phenotype correlation.

To achieve complete allele-specificity for a particular *TGFBI* mutation, stringent fidelity is required as an almost perfect off-target site exists in the form of the wild-type allele, which, for the majority of *TGFBI* mutations, differs by only one base pair from the mutant. This report uses TGFBI corneal dystrophies as a model of autosomal dominant disease to assess the specificity of the CRISPR/Cas9 system for autosomal dominant disorders. An allele-specific approach to target the five most prevalent *TGFBI* corneal dystrophy mutations is investigated, which highlights the promiscuity of Cas9 and the need for a validated, highly specific approach, that will encompass all possible *TGFBI* mutations.

Results

***In vivo* corneal gene disruption induced by CRISPR/Cas9 gene editing and NHEJ-mediated DNA repair.** We utilised a previously reported reporter knock-in mouse (*Krt12+/luc2*), that exclusively expresses firefly luciferase (luc2) in the corneal epithelium under control of the keratin K12 promoter, to study corneal delivery and activity of CRISPR/Cas9-mediated gene editing in living animals. To target the luciferase gene, an sgRNA utilising a PAM site 61 nucleotides downstream of the luc2 start codon was designed (Fig. 2a) and validated using a dual-luciferase assay (Fig. 2b). Therapeutic efficacy of this CRISPR/Cas9 system in living cornea was assessed following a single intrastromal injection in Krt12+/luc2 mice. Luciferase activity was evaluated with daily measurements up to 1 week following injection and weekly thereafter, for an additional 5 weeks (Fig. 2c). Luciferase activity was significantly reduced from post-injection day 1, with maximal silencing of >99% achieved

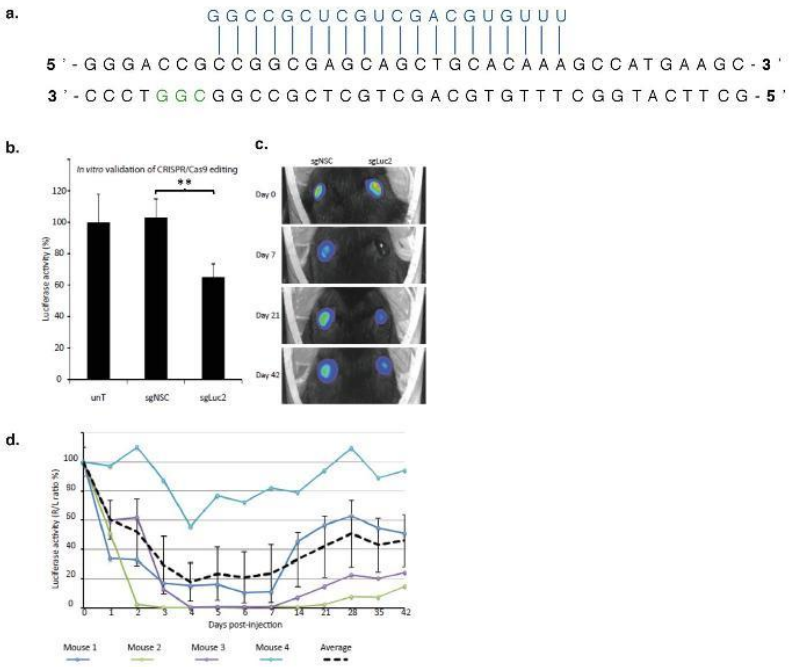


Figure 2. Sustained CRISPR/Cas9 mediated silencing of *luc2 in vivo*. (**a**) The short guide RNA (sgRNA) specific for *luc2* was designed to target the 5′ region of the gene, to increase the likelihood of inducing a frame-shifting deletion that would knock out luciferase activity by generating a premature termination codon. (**b**) An *in vitro* dual-luciferase assay demonstrated successful targeting of *luc2* by the sgLuc2 construct, as shown by a significant reduction in luciferase activity when normalized to untreated cells (data normalised against the untreated control = 100%). (**c**) Representative image of mice displaying a maximal reduction in *luc2* expression after injection with the sgLuc2 construct (right eye). This image was taken from the mouse represented by the green line in panel (d), below, at 7 days post treatment. (**d**) After treatment, the corneal luciferase activity of each mouse was quantified using a Xenogen IVIS live animal imager every day for 7 days, then every 7 days thereafter, for a total of 6 weeks. Luciferase activity for each treatment group expressed as a percentage of control (R/L ratio %).

at day 3 in one animal, and a maximal mean reduction of 82% ± 13% observed in 4 mice on day 4. Sustained silencing of luciferase expression was observed in 3 out of 4 mice over the entire monitoring period of the experiment (7 weeks), while in the remaining animal, luciferase inhibition persisted for 2 weeks (Fig. 2c).

Mutational analysis of *TGFBI* corneal dystrophy mutations. Currently the best characterised CRISPR/Cas9 system is that of *Streptococcus pyogenes* (*Sp*Cas9), which recognises a 5′-NGG-3′ PAM. An analysis of the 60 known *TGFBI* missense mutations (See Supplementary Table 1) was performed to determine if i) they generate a novel *S. pyogenes* PAM or ii) they have a *S. pyogenes* PAM nearby, placing the mutation within the seed region, defined here as the first 8 nucleotides immediately adjacent to the PAM[29–31].

19/60 mutations generate a novel *S. pyogenes* PAM, while 44/60 have a naturally occurring adjacent PAM site that places the mutation within an eight nucleotide seed region. When these figures are considered together, 20% of the *TGFBI* missense mutations are not targetable by *S. pyogenes* (Fig. 3a).

Analysis of the most prevalent *TGFBI* mutations in codon 124 and codon 555 revealed that none generated a novel *S. pyogenes* PAM, however all mutations had a *S. pyogenes* PAM within the first eight nucleotides of the target sequence (Fig. 3b).

Further to this, an analysis was conducted to determine if any of the prevalent *TGFBI* mutations generated a novel PAM with a CRISPR system from a different bacterial species. It was found that the R555W mutation generated a novel PAM with *Staphylococcus aureus* Cas9 (*Sa*Cas9), which recognises a 5′-NNGRRT-3′ PAM. In addition to this the R124L mutation generated a novel PAM with a mutant *Acidaminococcus* Cpf1 (*As*Cpf1) system[32], which is capable of recognising a 5′-VYCV-3′ PAM (Fig. 3c).

Validation of an *S. pyogenes* Cas9 PAM-specific approach. A PAM-specific approach has previously been shown to be an ideal way to achieve allele-specific editing[14]. To validate this approach a lattice corneal dystrophy-associated *TGFBI* mutation (L527R), was assessed[33]. The L527R mutation (c.1580 T > G) generates a novel PAM with *S. pyogenes*, (CTG > CGG) (Fig. 4a, top). A 20 nt sgRNA utilising the novel PAM was designed and an additional 20 nt sgRNA targeted to a naturally occurring PAM was designed as a positive control (Fig. 4a, bottom). Specificity was first assessed using a previously described *in vitro* dual-luciferase assay[7,9,34] in

a.

WT 124	5'-CACTCAGCTGTACACGGACCGCACGGAGAAGC-3'
R124C	5'-CACTCAGCTGTACACGGACTGCACGGAGAAGC-3'
R124H	5'-CACTCAGCTGTACACGGACCACACGGAGAAGC-3'
R124L	5'-CACTCAGCTGTACACGGACCTCACGGAGAAGC-3'
WT 555	5'-CACCAAGAGAACGGAGCAGACTCTTGGgtaaag-3'
R555Q	5'-CACCAAGAGAACAGAGCAGACTCTTGGgtaaag-3'
R555W	5'-CACCAAGAGAATGGAGCAGACTCTTGGgtaaag-3'

b.

WT 124	5'-GACCGCACGGAGAAGCTGAGGCCTGAGATGGA-3'
R124L	5'-GACCTCACGGAGAAGCTGAGGCCTGAGATGGA-3'
WT 555	5'-CCTTCCGAGCCCTGCCACCAAGAGAACGGAGCA-3'
R555W	5'-CCTTCCGAGCCCTGCCACCAAGAGAATGGAGCA-3'

c.

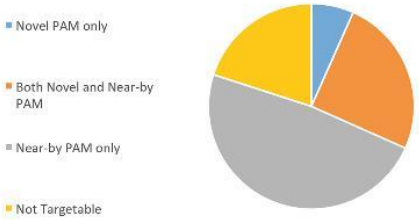


Figure 3. Analysis of *TGFBI* corneal dystrophy mutations in a CRISPR system Codons 124 or 555 shown in green, mutated base shown in red, nearest PAM to be utilised shown in blue and consequential guide sequence shown in orange. (**a**) Mutation analysis revealed that none of the prevalent *TGFBI* mutations generated a novel *S. pyogenes* PAM, however a naturally occurring PAM exists for all five mutations. For mutations in codon 124 the nearest downstream PAM places the mutated base at either position 3 or 4 of the guide sequence. For mutations in codon 555 the nearest downstream PAM places the mutated base at either position 7 or 8 of the guide sequence. (**b**) Mutational analysis revealed that R124L generates a novel PAM with a mutant *AsCpf1* that recognises a 5′-VYCV-3′ PAM. R124L generates a 5′-CTCA-3′ PAM. Further analysis revealed that R555W generates a novel PAM with *S.aureus* which is capable of recognising a 5′-NNGRRT-3′ PAM. R555W generates a 5′-GAGAAT-3′ PAM. (**c**) Venn diagram to illustrate the total number of *TGFBI* mutations that (i) generate a novel *S. pyogenes* PAM, (ii) have a near-by *S. pyogenes* PAM i.e. within the first 8 bp of the guide sequence, (iii) have both a novel and near-by *S. pyogenes* PAM or (iv) are not targetable by either approach.

which the two sgRNAs were co-expressed with either *S. pyogenes* Cas9, *S.aureus* Cas9 or *As*Cpf1 and a luciferase reporter containing a 50 bp region of either wild -type or mutant *TGFBI* sequence, which has been cloned into the multiple-cloning-site within the 3′UTR of *Luc2*. Cleavage of the *TGFBI* sequence within the reporter construct prevents transcription and processing of luciferase mRNA and results in an proportionate reduction of luciferase expression and therefore luciferase activity was measured as an indicator of sgRNA activity. The sgRNA utilising the novel PAM was shown to be highly specific, directing cutting of only the mutant *TGFBI* sequence, while both reporters were cleaved by the common sgRNA (Fig. 4b). In addition, an *in vitro* digestion using mutant 18 and 20 nt sgRNAs with a reporter containing either wild-type or mutant *TGFBI* sequence was carried out which confirmed the specificity observed in the dual-luciferase assay (Fig. 4c). Co-transfection with the mutant 18 and 20 nt sgRNAs only resulted in cleavage of the mutant reporter, the wild-type reporter template remained intact. Truncation of the guide did not appear to improve specificity.

Investigation of Cas9 orthologues *S.aureus* and *AsCpf1*. As none of the prevalent TGFBI mutations generated a novel PAM with *S. pyogenes* Cas9, alternative Cas9 orthologues were investigated. Although *S.aureus* Cas9 prefers a 5′-NNGRRT-3′ PAM, generated by the *TGFBI* R555W mutation (5′-GAGAAT-3′) (Fig. 3b). It has also been shown to recognise a 5′-NNGRRV-3′ PAM with comparable efficiencies[35], and this is present in the wild-type *TGFBI* sequence. Since *S.aureus* Cas9 prefers a guide length of either 21 nucleotides or 22 nucleotides[13]. both 21 nt and 22 nt guides utilising the novel *S.aureus* PAM were designed and targeted to both wild-type and mutant R555W *TGFBI* sequences. No significant knockdown was observed with either guide length and the mutant R555W guide was unable to distinguish between wild-type and mutant *TGFBI* sequence (Fig. 5a).

A mutant *As*Cpf1 was generated that has the capability of recognising a 5′-VYVC-3′ PAM, as generated by the *TGFBI* R124L mutation (5′-CTCA-3′) (Fig. 3b). A 20 nt guide was designed utilising the novel mutant *As*Cpf1 PAM and targeted to both the wild-type and mutant R124L TGFBI sequences. Although the mutant guide can distinguish between wild-type and mutant *TGFBI* sequence the knockdown efficiency is very low with a maximal knockdown of 20% (Fig. 5b).

Investigation of a guide-specific approach using *S. pyogenes* Cas9. As none of the most prevalent *TGFBI* mutations generated a novel PAM with *S. pyogenes* Cas9 and adequate specificity or efficiency could not

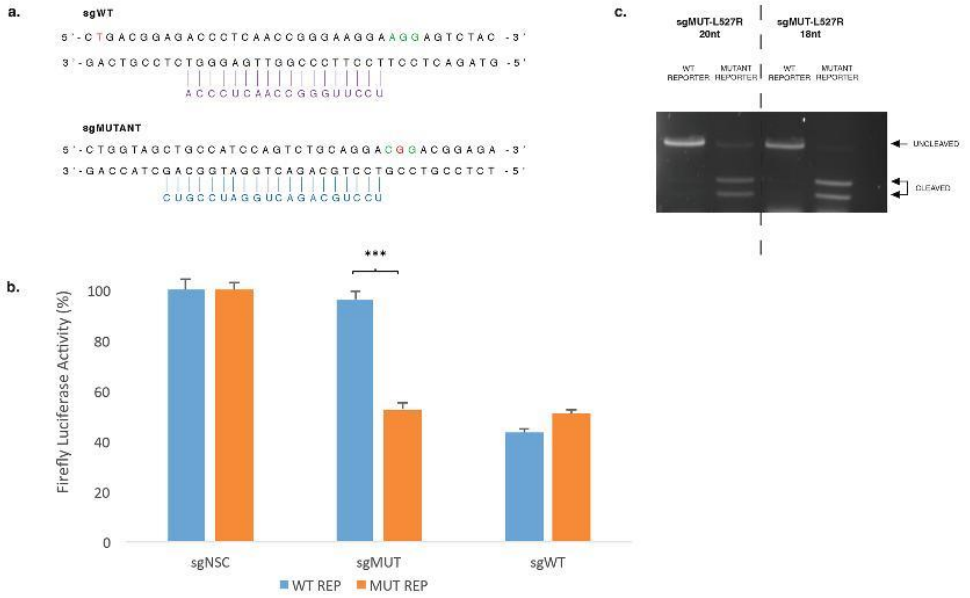


Figure 4. Allele-specific cleavage of L527R *TGFBI* mutation utilising a PAM-specific approach. (**a**) The L527R mutation (c.1580 T > G) is indicated in red and PAM utilised is shown in green. A 20 nt sgRNA targeted to a naturally occurring PAM was designed as a positive control (sgWT, purple –top of figure). A 20 nt sgRNA utilising the novel PAM, containing the L527R mutation, was designed (sgMUTANT, blue – bottom of figure). (**b**) Both sgWT and sgMUTANT were targeted to a luciferase reporter plasmid containing either a wild-type or mutant *TGFBI* sequence to determine potency and allele specificity. (**c**) An *in vitro* digestion with Cas9 protein complexed with a sgRNA utilising the novel L527R PAM was carried out to confirm the specificity observed. Mutant guides of both 20 and 18 nucleotides were tested. Uncropped gel images are available in Supplementary Figure 1.

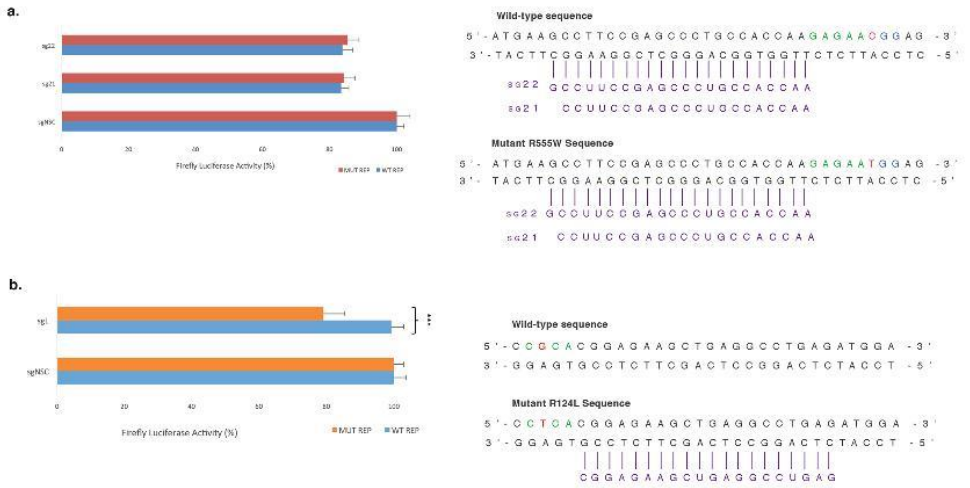


Figure 5. Evaluation of Cas9 orthologues in a PAM-specific system targeted to prevalent *TGFBI* mutations Guide RNA tested shown in purple, PAM utilised shown in green and mutation shown in red. (**a**) 22 and 21 nucleotide guides were designed to target the novel *S.aureus* Cas9 PAM generated by R555W. Both guide lengths were targeted to a luciferase reporter plasmid containing either a wild-type or mutant *TGFBI* sequence to determine potency and allele specificity. (**b**) A guide utilising the novel mutant *As*Cpf1 PAM generated by R124L was targeted to a luciferase reporter plasmid containing either a wild-type or mutant *TGFBI* sequence to determine potency and allele specificity.

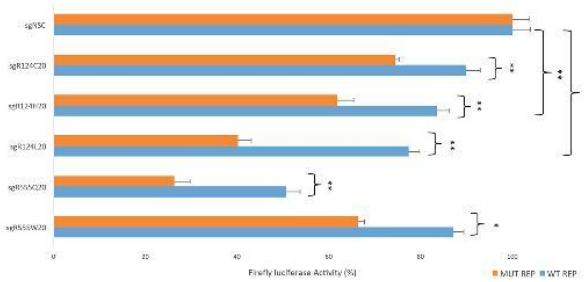


Figure 6. Investigation of a guide-specific approach to treat prevalent *TGFBI* mutations Using a guide-specific approach, 20 nucleotide guides for the 5 most prevalent TGFBI mutations (as shown in Fig. 3a) were targeted to wild-type and respective mutant sequence in a dual luciferase assay. The 5 guides cut with varying degrees of specificities and efficiencies. There was a significant difference between the wild-type and mutant sequence in all cases.

be achieved with Cas9 orthologues from other bacterial species, a guide-specific approach was explored; whereby the mutant guide differs from the wild-type sequence only by a single base pair. A dual-luciferase assay was employed to assess the specificity of a 20 nt guide for the 5 most prevalent TGFBI mutations; R124C, R124H, R124L, R555Q and R555W. Each guide was targeted to the wild-type and respective mutant sequence and the firefly luciferase activity was measured as an indicator of specificity (Fig. 6).

Cas9 directed by R124C sgRNA was able distinguish between wild-type and mutant sequence, although it cut with a low efficiency of 26%. R124H cut with an improved specificity and efficiency, although wild-type sequence was significantly cleaved (17%). R124L offered the most promising specificity profile, 60% cleavage of mutant sequence was observed in comparison to 23% of the wild-type sequence, however the wild-type sequence was still significantly cleaved when compared to the non-specific control. Although the R555Q guide directed efficient cleavage of the mutant reporter, the wild-type sequence was also substantially cut by 50%. Finally R555W preferentially cleaved mutant sequence, however the wild-type sequence was still cleaved by 10%.

Investigation of the effect of guide length on the specificity of *S. pyogenes* Cas9. Reports have indicated that truncating the length of the matching sequence within the guide to 18 nucleotides can reduce off-target cutting, while maintaining on-target efficiencies[12]. As none of the 20 nt guides provided adequate specificity an assessment of the effect of guide-length upon specificity using a dual-luciferase assay was conducted for the 5 most prevalent TGFBI mutations. Reports have shown that guide lengths <16 nt abolish cleavage activity[36,37]. For each mutation a range of guide lengths from 16–22 nucleotides were tested, each guide was targeted to the wild-type and respective mutant sequence and the firefly luciferase activity was measured as an indicator of specificity (Fig. 7).

For all mutations investigated the truncated guides did not provide a marked improvement of specificity, for most cases maximal discrimination occurred with guides 20 or 19 nucleotides in length. For R124C, a 20 nt guide seemed to confer allele-specificity, however no other guide length offered any adequate discrimination (Fig. 7a). In the case of the R555Q mutation guides in the 18–20 nt range did not offer sufficient discrimination, although, interestingly, the 21 nt guide provided convincing allele-specificity (Fig. 7d). R555W did not offer any considerable allele-specificity for any length tested (Fig. 7e). R124H and R124L displayed clear allele-specific cleavage, especially in the 18–20 nt sgRNA range, with minimal cutting of the wild-type sequence (Fig. 7b,c). Interestingly for the R124 mutations guide lengths of 21 nt seemed to impair cleavage activity in all cases.

Addition of 5′-GG to the 20 nt guide sequence. Standard design of sgRNA guides includes the addition of a guanine to the 5′ end of the guide sequence (5′-GX_{20}-3′) to help facilitate efficient transcription[4]. An alternative guide design of 5′-GGX_{20}-3′ has been reported to minimise off-target activity in certain cases, offering an improved specificity of Cas9[38]. This parameter was tested using the 5 most prevalent TGFBI mutations (Fig. 8). The additional guanine at the 5′ end of the guide sequence did not provide an improved specificity in any case. In some instances a reduction in on-target activity was observed, confirming that specificity is guide dependent.

***In vitro* digestion to confirm specificity of *S. pyogenes* Cas9.** *In vitro* digestion of either wild-type or mutant *TGFBI* sequence with Cas9 protein complexed with sgRNA was carried out to further assess the specificity profile of *S. pyogenes* Cas9 (Fig. 9). Guide lengths of 18 and 20 nucleotides were tested to evaluate the impact of truncating the guide sequence. For R124C the mutant 20 nt guide appeared to cut the mutant sequence more than the wild-type sequence. However, when truncated to 18 nt the mutant guide appeared to loose ability to distinguish between wild-type and mutant sequence, reflecting results from the dual luciferase assay. For R124H and R124L both mutant 20 nt and 18 nt guides appeared to clearly cut the mutant sequence preferentially over the wild-type sequence, again reflecting the dual-luciferase results. Interestingly, in both cases the wild-type guide appeared to result in more cleavage of the mutant sequence in comparison to the mutant sequence, although as the wild-type guide would not be implicated in a clinical setting it can be ignored. For R555Q and R555W the 20 nt or 18 nt guides did not confer allele-specificity under any conditions, cutting both wild-type and mutant

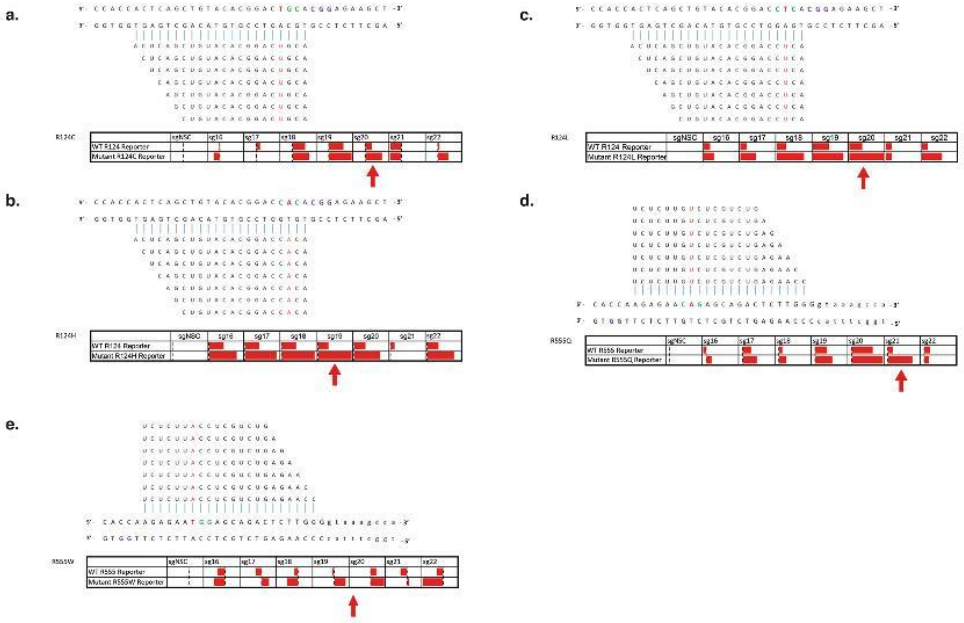


Figure 7. Guide-length screen to determine the effect on specificity of a guide-specific system Heatmaps showing varying degrees of knockdown observed via a dual luciferase assay when guides ranging in different lengths are targeted to the wild-type and respective mutant sequence. Specificity bars show knockdown when normalised to the non-specific control, with 100% being maximal knockdown observed. Maximal allele-specificity observed for each mutation indicated with a red arrow.

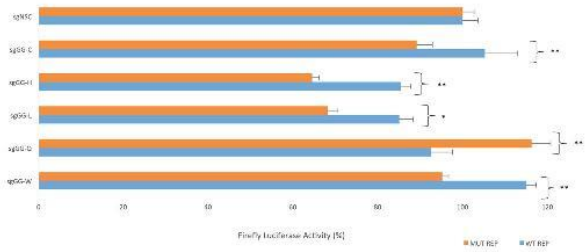


Figure 8. Effect of the addition of 5′-GG to the 20 nt guide sequence on specificity 5′-GG was prefixed to the 20 nucleotide guides for the 5 most prevalent TGFBI mutations (as shown in Fig. 3a) were targeted to wild-type and respective mutant sequence in a dual luciferase assay. The addition of 5′-GG did not improve specificity, in some cases it caused a reduction in on-target activity.

sequence equally, demonstrating mismatches in the distal region of the guide are less critical in determining specificity of Cas9.

Discussion

Dominant negative disorders that are the result of an accumulation of mutant protein can be targeted by allele-specific CRISPR mediated gene disruption via NHEJ. We have shown *in vivo* that gene disruption via NHEJ offers a viable approach to achieve gene silencing. Sustained knockdown of luciferase was observed in the corneal epithelium of reporter mice over several weeks in 3 out of 4 mice, following a single intrastromal injection of CRISPR/Cas9 components (Fig. 2c). Since the corneal epithelium is completely turned over every 1–2 weeks[39], our data suggests permanent editing took place within the corneal stem cell compartment following *in vivo* delivery of CRISPR/Cas9. By extension, CRISPR/Cas9 gene editing using an sgRNA specific to pathologic mutant alleles delivered by intrastromal injection has great potential for editing resident corneal stem cells as a permanent cure for dominant-negative corneal disorders. However, in order to translate this strategy to the clinic as a therapy the issue of specificity must be addressed.

The prevalent *TGFBI* mutations offer an interesting real-life scenario in which to test different approaches to allele-specific CRISPR/Cas9 gene therapy as the different causative mutations with different phenotypes associated with the same codon create different specificity profiles.

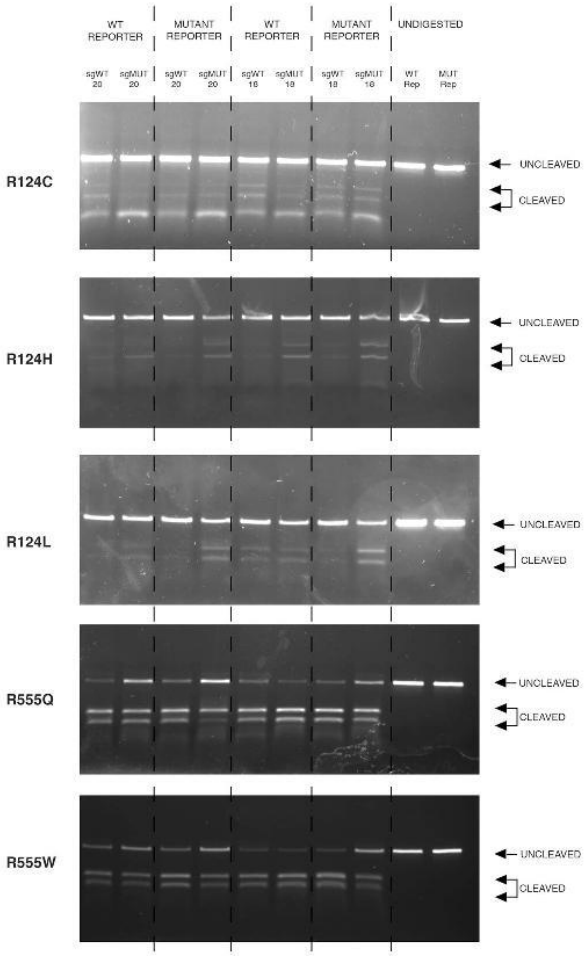


Figure 9. Confirmation of the specificity achieved using a guide-specific system targeted to prevalent *TGFBI* mutations *In vitro* digestion of either wild-type or respective mutant *TGFBI* sequence via Cas9 protein complexed with an sgRNA. Guides lengths of 20 and 18 nucleotides were assessed. Uncropped gel images are available in Supplementary Figure 2.

Published reports illustrate that the region immediately adjacent to the PAM is critical to specificity[25–27]. The documented importance of this region has led to it being coined as the 'seed' region. The Cas9:sgRNA complex will initially identify the correct PAM, and only once the PAM has been identified will the Cas9:sgRNA complex then test the complementarity between the guide and target DNA. The PAM proximal region, or seed region, is critical in this step and mismatches in this region will prevent the ternary complex forming and therefore cleavage will not occur[40]. The exact length of the seed region is unclear, with reports ranging from 5–12 nucleotides[29–31].

The *TGFBI* mutations investigated here gave the opportunity to investigate the extent of this seed region further. The mutations in codon 124 lie at guide positions 3 or 4, so are within the seed region of whichever definition, whereas, codon 555 mutations lie in guide positions 7 or 8, so can be considered either inside or outside the seed region. Accordingly, it was demonstrated that allele-specificity was achieved by guides targeting R124H and R124L mutations both found at position 3 of the guide. However, neither R124C, R555Q or R555W mutations found at positions 4, 7 and 8 respectively, were capable of adequate allele-specific cleavage. This confirms that the sequence immediately adjacent to the PAM is most critical in determining specificity of the guide, and mismatches are not well tolerated here. In contrast mismatches in positions 4, 7 or 8 of the guide are better tolerated and do not have as strong an influence on the fidelity of the guide[41].

In addition, it has been demonstrated that U-rich seeds are linked with a low knockdown efficiency, due to RNA polymerase III being terminated by U-rich sequences. Interestingly, the R555W mutation in which minimal knockdown was observed has a very U-rich seed with 4 U's within the first 6 bp: 3′-UCUCUU-5′[30]. Jiang *et al.* reported that mutations in positions ranging from position 1 in the guide to position 6 of the guide sequence abolish cleavage activity, except mutations at position 3[41]. These *TGFBI* results directly contradict this, as the R124H and R124L mutations exhibit clear allele-specific cleavage and both mutations are present at position 3

of the guide. Therefore, it is evident that restrictions most likely vary from one guide sequence to another and in each case should be individually assessed.

That single mismatches in guide sequences can be tolerated regardless of their position in the guide has been confirmed in other reports[11,42,38]. Contrary to initial reports using other genes, manipulation of the guide sequence, in the form of truncation or addition of extra guanine bases, did not provide improved specificity in any case. This is consistent with follow up reports that demonstrate truncated guides or additional guanines do not offer improved specificity in most cases[42,43]. An intriguing observation was that for all R124 mutations the guide length of 21 nt seemed to impair cleavage activity, it is unclear why this happens but we hypothesise it may alter structure of stability of the sgRNA.

To confirm whether the results observed *in vitro* could be directly translated to a real-life scenario, it would be compelling to test these guides *ex vivo* in patient derived primary cells or *in vivo* in a mutation-specific animal model. This would demonstrate the effectiveness of a combined *in vitro* dual-luciferase and cleavage assay as a preliminary screening stage to ensure guides with adequate specificity are utilised downstream in a clinical setting.

The use of CRISPR therefore has clear limitations in targeting specific disease-causing mutations. In circumstances when one is not tied to targeting a specific disease-causing mutation, the criteria for selecting an appropriate sgRNA can be outlined as; avoid selecting guides that have predicted off-targets directly followed by a PAM, high global sequence similarity, mismatches only in the PAM distal region and those that do not have maximal consecutive mismatches[10]. However, when designing an sgRNA to targeting a particular disease-causing mutation there is no flexibility (other than guide length) to meet these criteria. Consequently, a guide-specific treatment strategy is not suitable for targeting the mutant alleles which cause *TGFBI* corneal dystrophies, as an almost perfect off-target site exists in the form of the wild-type allele.

Although here, and in previous reports, a SNP-derived PAM approach has been shown to provide highly specific cleavage[15], this can only be applied to PAM-generating mutations. In the case of *TGFBI* corneal dystrophies, of the 60 causative mutations less than a third generate a novel *S. pyogenes* PAM. Therefore, even if the problems associated with Cas9/sgRNA delivery at present were overcome, the majority of patients with TGFBI would not have mutations that could be directly targeted.

Cas9 *S. pyogenes* orthologues are not as well characterised, therefore their off-target profiles are not as well understood as the that of *S. pyogenes*. In addition to this, they have much more intricate PAMs that occur much less frequently in the genome, reducing the fraction of *TGFBI* mutations that will result in a novel PAM. Furthermore, our results highlight another concern; even though the mutant SNP generated a novel PAM a non-canonical PAM existed in the wild-type sequence (Fig. 5b), meaning allele-specific cleavage could not be achieved. If non-canonical PAMs are considered within the analysis the number of targetable mutations would be even further reduced.

It is clear that individual guides perform with different cleavage efficiencies and specificity profiles. It is unrealistic to suggest a 60 allele-specific guide system as an effective treatment for *TGFBI* corneal dystrophies. A need for a highly-specific catch-all approach is apparent.

Materials and Methods

Oligonucleotides. All oligonucleotides used in this study were purchased from Integrated DNA Technologies. Sequences are listed in Supplementary Table 2.

Constructs. The *S. pyogenes* Cas9 vector plasmid used was pSpCas9(BB)-2A-Puro (PX459) V2.0, a gift from Feng Zhang (Addgene plasmid # 62988).The *S.aureus* Cas9 vector plasmid used was pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA, a gift from Feng Zhang (Addgene plasmid # 61591). The mutant *AsCpf1* used was kindly provided from Professor Feng Zhang, Broad Institute MIT. Wild-type *TGFBI* or mutant *TGFBI* guides were cloned into the various plasmids by standard molecular biology techniques. A detailed protocol is outlined by Ran *et al.*[4]. In brief, *S. pyogenes* Cas9 and mutant *As*Cpf1 were digested with *BbsI* (NEB Cat # R0539S) while *S.aureus* Cas9 was digested with *BsaI* (NEB Cat # R0535S). Guide sequences (shown in Supplementary Table 2) were annealed and cloned into the corresponding digested plasmid.

A firefly luciferase reporter plasmid was used to assess knockdown. The vector plasmid used was psiTEST-LUC-Target (York Bioscience Ltd, York, UK). 50 nucleotides of wild type *TGFBI* or mutant *TGFBI* sequence was cloned into the MCS by standard molecular biology techniques.

An expression construct for Renilla luciferase (pRL-CMV, Promega, Southampton, UK) was used for the dual-luciferase assay to normalize transfection efficiency. In brief, psiTEST-LUC-Target was digested with NheI and KpnI (NEB Cat # R0131S and # R0142S). Human wild-type or mutant *TGFBI* sequences (shown in Supplementary Table 2) were annealed and cloned into the digested plasmid.

Off-target analysis. Off-target and on-target scores were calculated using the 'Optimised CRISPR Design Tool', available online by the Zhang lab, MIT 2013 and 'Benchling's CRISPR Tool' available online by Benchling.

Dual-Luciferase Assay. A dual luciferase assay was used to determine the potency and allele specificity of the different guides previously described. HEK AD293 cells (Life Technologies) were co-transfected using Lipofectamine 2000 (Life Technologies) with a CRISPR plasmid, a firefly luciferase reporter plasmid and *Renilla* Luciferase expression plasmid. Cells were incubated for 72 hours, before being lysed and the activities of both *Firefly* luciferase and *Renilla* luciferase quantified.

Intrastromal Injection. Animals were used for the following experiments in accordance with the UK Animal Welfare Act; the experiments were approved by the Home Office (Scotland) and the DHSSPS (Northern Ireland). Prior to intrastromal injection of CRISPR components, mice were anaesthetised by intraperitoneal

injection with a mix of Hypnorm (25 mg/kg; VetaPharma Ltd, Leeds, UK) and Hypnovel (25 mg/kg; Roche, Hertfordshire, UK). In addition, topical anaesthetic (0.5%w/v Tetracaine Hydrochloride; Bausch & Lomb, Aubenas, France) was applied to the eye. Following injection, mice were allowed to recover in a heated cabinet and monitored for adverse effects until the anesthesia had worn off fully. Cas9/sgRNA constructs were delivered to the mouse cornea by intrastromal injection, as previously described (Courtney *et al.*[8]). Both a guide targeted to *Luc2* (sgLuc2 - right eye) and a non-specific control guide (sgNSC - left eye) were injected intrastromally in a total volume of 4 µl of PBS at a concentration of 500ng/µl.

Live animal imaging. All mice used for live imaging were aged between 12 and 25 weeks old. For imaging, mice were anaesthetised using 1.5–2% isoflurane (Abbott Laboratories Ltd., Berkshire, UK) in ~1.5 l/min flow of oxygen. A mix of luciferin substrate (30 mg/ml D-luciferin potassium salt; Gold Biotechnology, St. Louis, USA) mixed 1:1 w/v with Viscotears gel (Novartis, Camberley, UK) was dropped onto the eye of heterozygous Krt12 +/luc2 transgenic mice immediately prior to imaging. A Xenogen IVIS Lumina (Perkin Elmer, Cambridge, UK) was used to quantify luminescence. Live images of mice ($n = 4$) were taken every 24 hours for 7 days, then once every week thereafter for six weeks (42 days) in total. Quantification of luciferase inhibition was determined by calculating the right/left ratio, with values normalised to those at day 0 (as 100%).

***In vitro* digestion of circular plasmid and DNA template with purified *S. pyogenes* Cas9.** A double-stranded DNA template was prepared by amplifying a region of the luciferase reporter plasmid containing the desired sequence using the following primers: 5′-ACCCCAACATCTTCGACGCGGGC-3′ and 3′-TGCTGTCCTGCCCCACCCCA-5′. A cleavage reaction was set up by incubating 30 nM *S. pyogenes* Cas9 nuclease (NEB UK) with 30 nM synthetic sgRNA (Synthego) for 10 minutes at 25 °C. The Cas9:sgRNA complex was then incubated with 3 nM of DNA template at 37 °C for 1 hour. Fragment analysis was then carried out on a 1% agarose gel.

Statistical analysis. All error bars represent the S.E.M. unless stated otherwise. Significance was calculated using a Mann-Whitney test. Statistical significance wasset at $p < 0.05$. Variance was calculated among groups and deemed to be similar.

Data availability. No datasets were generated or analysed during the current study.

References

1. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**(6096), 816–21 (2012).
2. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science.* **339**(6121), 823–6 (2013)
3. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science.* **339**(6121), 819–23 (2013).
4. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* **8**(11), 2281–308 (2013).
5. Yin, H. *et al.* Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components *in vivo*. *Nat Biotechnol.* **34**(3), 328–33 (2016).
6. Liao, H. *et al.* Development of allele-specific therapeutic siRNA in Meesmann epithelial corneal dystrophy. *PLoS One.* **6**(12), e28582 (2011).
7. Courtney, D. G. *et al.* Development of allele-specific gene-silencing siRNAs for TGFBI Arg124Cys in lattice corneal dystrophy type I. *Invest Ophthalmol Vis Sci.* **55**(2), 977–85 (2014).
8. Courtney, D. G. *et al.* siRNA silencing of the mutant keratin 12 allele in corneal limbal epithelial cells grown from patients with Meesmann's epithelial corneal dystrophy. *Invest Ophthalmol Vis Sci.* **55**(5), 3352–60 (2014).
9. Allen, E. H. *et al.* Allele-specific siRNA silencing for the common keratin 12 founder mutation in Meesmann epithelial corneal dystrophy. *Invest Ophthalmol Vis Sci.* **54**(1), 494–502 (2013).
10. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* **31**(9), 827–32 (2013).
11. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol.* **31**(9), 822–6 (2013).
12. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol.* **32**(3), 279–84 (2014).
13. Ran, F. A. *et al. In vivo* genome editing using Staphylococcus aureus Cas9. *Nature.* **520**(7546), 186–91 (2015).
14. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science.* **351**(6268), 84–8 (2016).
15. Courtney, D. G. *et al.* CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both *in vitro* and *in vivo* KRT12 mutation-specific targeting. *Gene Ther.* **23**(1), 108–12 (2016).
16. Li, Y. *et al.* Exploiting the CRISPR/Cas9 PAM Constraint for Single-Nucleotide Resolution Interventions. *PLoS One.* **11**(1), e0144970 (2016).
17. Klintworth, G. K. Corneal dystrophies. *Orphanet J Rare Dis.* **4**(7), 1172-4-7 (2009).
18. Munier, F. L. *et al.* Kerato-epithelin mutations in four 5q31-linked corneal dystrophies. *Nat Genet.* **15**(3), 247–51 (1997).
19. Mashima, Y. *et al.* A novel mutation at codon 124 (R124L) in the BIGH3 gene is associated with a superficial variant of granular corneal dystrophy. *Arch Ophthalmol.* **117**(1), 90–3 (1999 Jan).
20. Yee, R. W. *et al.* Linkage mapping of Thiel-Behnke corneal dystrophy (CDB2) to chromosome 10q23-q24. *Genomics.* **46**(1), 152–4 (1997).
21. Han, K. E. *et al.* Pathogenesis and treatments of TGFBI corneal dystrophies. *Prog Retin Eye Res.* **50**, 67–88 (2016).
22. Weiss, J. S. *et al.* The IC3D classification of the corneal dystrophies. *Klin Monbl Augenheilkd.* **228**(Suppl 1), S1–39 (2011).
23. Stenson, P. D. *et al.* The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet.* **136**(6), 665–77 (2017).
24. Munier, F. L. *et al.* BIGH3 mutation spectrum in corneal dystrophies. *Invest Ophthalmol Vis Sci.* **43**(4), 949–54 (2002).
25. Poulaki, V. & Colby, K. Genetics of anterior and stromal corneal dystrophies. *Semin Ophthalmol.* **23**(1), 9–17 (2008).
26. Mashima, Y. *et al.* Association of autosomal dominantly inherited corneal dystrophies with BIGH3 gene mutations in Japan. *Am J Ophthalmol.* **130**(4), 516–7 (2000).
27. Han, K. E. *et al.* Clinical findings and treatments of granular corneal dystrophy type 2 (avellino corneal dystrophy): a review of the literature. *Eye Contact Lens.* **36**(5), 296–9 (2010 Sep).
28. Lee, J. H. *et al.* Prevalence of granular corneal dystrophy type 2 (Avellino corneal dystrophy) in the Korean population. *Ophthalmic Epidemiol.* **17**(3), 160–5 (2010).

29. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature.* **507**(7490), 62–7 (2014).
30. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol.* **32**(7), 670–6 (2014).
31. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol.* **31**(9), 839–43 (2013).
32. Gao, L. *et al.* Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol.* (2017).
33. Fujiki, K. *et al.* A new L527R mutation of the betaIGH3 gene in patients with lattice corneal dystrophy with deep stromal opacities. *Hum Genet.* **103**(3), 286–9 (1998).
34. Atkinson, S. D. *et al.* Development of allele-specific therapeutic siRNA for keratin 5 mutations in epidermolysis bullosa simplex. *J Invest Dermatol.* **131**(10), 2079–86 (2011).
35. Friedland, A. E. *et al.* Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol.* **16**(257), 015–0817-8 (2015).
36. Dahlman, J. E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol.* **33**(11), 1159–61 (2015).
37. Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nat Methods.* **12**(11), 1051–4 (2015).
38. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**(1), 132–41 (2014).
39. HANNA, C., BICKNELL, D. S. & O'BRIEN, J. E. Cell turnover in the adult human eye. *Arch Ophthalmol.* **65**, 695–8 (1961).
40. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell.* **156**(5), 935–49 (2014).
41. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol.* **31**(3), 233–9 (2013).
42. Josephs, E. A. *et al.* Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res.* **43**(18), 8924–41 (2015).
43. Zhang, J. P. *et al.* Different Effects of sgRNA Length on CRISPR-mediated Gene Knockout Efficiency. *Sci Rep.* **6**, 28566 (2016).

Acknowledgements

We are grateful to Prof Feng Zhang, Broad Institute MIT, USA for kindly providing a novel *As*Cpf1 mutant nuclease pre-publication. This work has been supported by Avellino Laboratories.

Author Contributions

Manuscript Preparation – K.C., M.A.N. and C.B.T.M. Experimental Procedures – K.C. D.C., L.D., C.C.S., S.D. and L.C.M. Data Analysis - K.C., D.C., M.A.N. and C.B.T.M.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-16279-4.

Competing Interests: C.B.T.M. is a consultant for Avellino Laboratories.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders

Kathleen A. Christie,[1] David G. Courtney,[1] Larry A. DeDionisio,[2] Connie Chao Shern,[1] Shyamsree De Majumdar,[1] Laura C. Mairs[1], M. Andrew Nesbit,[1] C.B.Tara Moore[1*]

Supplementary Table 1: Mutational analysis performed on the TGFBI corneal dystrophy mutations to determine which i) generated a novel *S.pyogenes* PAM or ii) had a near-by *S.pyogenes* PAM

	Lattice Type			Wild-type *TGFBI* sequence	Mutant *TGFBI* sequence	Novel *S.pyogenes* PAM	Novel *S.aureus* PAM	Near-by *S.pyogenes* PAM	Near-by *S.aureus* PAM
Missense	R496W	C>T	CGG>TGG	GTACGGGACCCTGTTCACGATGGACCGGGTGCTGACCCCCCAATGGGGAC	GTACGGGACCCTGTTCACGATGGACTGGGTGCTGACCCCCCAATGGGGAC			X (position 1)	X (position 13)
	R514P	G>C	CGC>CCC	ATGGATGTCCTGAAGGGAGACAATCGCTTTAGgtaattagttccatccccg	ATGGATGTCCTGAAGGGAGACAATCCCTTTAGgtaattagttccatccccg	X		X (position 5)	
	F515L	T>A	TTT>TTA	ATGTCCTGAAGGGAGACAATCGCTTTAGgtaattagttccatccccgggtg	ATGTCCTGAAGGGAGACAATCGCTTAAGgtaattagttccatccccgggtg			X (position 1)	X (position 19)
	T538P	A>C	ACA>CCA	GACCCTCAACCGGGGAAGGAGTCTACACAGTCTTTGCTCCCACAAATGAAGC	GACCCTCAACCGGGGAAGGAGTCTACCCAGTCTTTGCTCCCACAAATGAAGC	X		X (position 14)	X (position 19)
	V539D	T>A	GTC>GAC	CCTCAACCGGGGAAGGAGTCTACACAGTCTTTGCTCCCACAAATGAAGCCTT	CCTCAACCGGGGAAGGAGTCTACACAGACTTTGCTCCCACAAATGAAGCCTT			X (position 18)	
	F540S	T>C	TTT>TCT	CAACCGGGGAAGGAGTCTACACAGTCTTTGCTCCCACAAATGAAGCCTTCCGA	CAACCGGGGAAGGAGTCTACACAGTCTCTGCTCCCACAAATGAAGCCTTCCGA				
	P542R	C>G	CCC>CGC	GAAGGAGTCTACACAGTCTTTGCTCCCACAAATGAAGCCTTCCGAGCCC	GAAGGAGTCTACACAGTCTTTGCTCGCACAAATGAAGCCTTCCGAGCCC				
	A546D	C>A	GCC>GAC	CACAGTCTTTGCTCCCACAAATGAAGCCTTCCGAGCCCTGCCACCAAGAGA	CACAGTCTTTGCTCCCACAAATGAAGACTTCCGAGCCCTGCCACCAAGAGA			X (position 10)	X (position 20)
	F547S	T>C	TTC>TCC	AGTCTTTGCTCCCACAAATGAAGCCTTCCGAGCCCTGCCACCAAGAGAACG	AGTCTTTGCTCCCACAAATGAAGCCTCCCGAGCCCTGCCACCAAGAGAACG	X		X (position 1)	
	R548P	G>C	CGA>CCA	TCCCACAAATGAAGCCTTCCGAGCCCTGCCACCAAGAGAACGGAG	TCCCACAAATGAAGCCTTCCCAGCCCTGCCACCAAGAGAACGGAG	X		X (position 4)	
	P551Q	C>A	CCA>CAA	ACAAATGAAGCCTTCCGAGCCCTGCCACCAAGAGAACGGAGCAGACTCTTG	ACAAATGAAGCCTTCCGAGCCCTGCAACCAAGAGAACGGAGCAGACTCTTG			X (position 3)	
	L558R	T>G	CTC>CGC	GCCACCAAGAGAACGGAGCAGACTCTTGGgtaaagaccaacttaagtacacgtc	GCCACCAAGAGAACGGAGCAGACGCTTGGgtaaagaccaacttaagtacacgtc			X (position 3)	X (position 2)
	L558P	T>C	CTC>CCC	CTGCCACCAAGAGAACGGAGCAGACTCTTGGgtaaagaccaacttaagtac	CTGCCACCAAGAGAACGGAGCAGACCCTTGGgtaaagaccaacttaagtac	X		X (position 3)	X (position 2)
	L565P	T>C	CTT>CCT	GAGATGCCAAGGAACTTGCCAACATCCTGAAATACCACATT	GAGATGCCAAGGAACCTGCCAACATCCTGAAATACCACATT	X		X (position 7)	
	G594V	G>T	GGT>GTT	CTGGTGCGGCTAAAGTCTCTCCAAGGTGACAAGCTGGAAGTCAGCTTGgta	CTGGTGCGGCTAAAGTCTCTCCAAGTTGACAAGCTGGAAGTCAGCTTGgta			X (position 3)	
	V613G	G>T	GTT>GGT	TGTGGTGAGTGTCAACAAGGAGCCTGTTGCCGAGCCTGACATCATGGCCAC	TGTGGTGAGTGTCAACAAGGAGCCTGGTGCCGAGCCTGACATCATGGCCAC	X		X (position 2)	
	A620D	C>A	GCC>GAC	CCTGTTGCCGAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTCATC	CCTGTTGCCGAGCCTGACATCATGGACACAAATGGCGTGGTCCATGTCATC			X (position 7)	
	V624M	G>A	GTG>ATG	CTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAATGTTCTG	CTGACATCATGGCCACAAATGGCATGGTCCATGTCATCACCAATGTTCTG			X (position 1)	
	V625D	G>A	GTC>ATC	CTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAATGTTCTG	CTGACATCATGGCCACAAATGGCGTGATCCATGTCATCACCAATGTTCTG			X (position 12)	
Small Deletions	His572del	c.1714_1716delCAC		GGAACTTGCCAACATCCTGAAATACCACATTGGTGATGAAATCCTGGTTAG	GGAACTTGCCAACATCCTGAAATAC_ATTGGTGATGAAATCCTGGTTAG			X (position 3)	X (position 7)
	Val613del	c.1838_1849del12		TGTGGTGAGTGTCAACAAGGAGCCTGTTGCCGAGCCTGACATCATGGCCACA	TGTGGTGAGTGTCAACAAGGAGCCTG_ACATCATGGCCACAAATGGCGTGGTC			X (position 2)	
	Val624_Val625del	c.1870_1875delGTGGTC		CTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAATGTTCTGAGCCTC	CTGACATCATGGCCACAAATGGC_CATGTCATCACCAATGTTCTGCAGCCTC			X (position 9)	
	Lattice Intermediate type I/IIIA								
Missense	L518R	T>C	CTG>CCG	ttctgtgtgtgtatctacagCATGCTGGTAGCTGCCATCCAGTCTGCAGGA	ttctgtgtgtgtatctacagCATGCCGGTAGCTGCCATCCAGTCTGCAGGA	X			
	L518P	T>G	CTG>CGG	ttctgtgtgtgtatctacagCATGCTGGTAGCTGCCATCCAGTCTGCAGGA	ttctgtgtgtgtatctacagCATGCGGGTAGCTGCCATCCAGTCTGCAGGA	X	X		
	T538R	C>G	ACA>AGA	GACCCTCAACCGGGGAAGGAGTCTACACAGTCTTTGCTCCCACAAATGAAGCTT	GACCCTCAACCGGGGAAGGAGTCTACAGAGTCTTTGCTCCCACAAATGAAGCC		X	X (position 15)	X (position 20)
	N544S	A>G	AAT>AGT	GTCTACACAGTCTTTGCTCCCACAAATGAAGCCTTCCGAGCCCTGCCACCA	GTCTACACAGTCTTTGCTCCCACAAGTGAAGCCTTCCGAGCCCTGCCACCA			X (position 4)	
	N622H	A>C	AAT>CAT	GCCGAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAAT	GCCGAGCCTGACATCATGGCCACACATGGCGTGGTCCATGTCATCACCAAT			X (position 2)	
	G623D	G>A	GGC>GAC	GAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAATGTT	GAGCCTGACATCATGGCCACAAATGACGTGGTCCATGTCATCACCAATGTT			X (position 3)	
	H626R	A>G	CAT>CGT	CATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAATGTTCTG	CATCATGGCCACAAATGGCGTGGTCCGTGTCATCACCAATGTTCTG			X (position 16)	
	H626P	A>C	CAT>CCT	CATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAATGTTCTG	CATCATGGCCACAAATGGCGTGGTCCCTGTCATCACCAATGTTCTGCAGCCTC	X		X (position 16)	
Small Deletion	c.1618_1620delTTT		Phe540del	CAACCGGGGAAGGAGTCTACACAGTCTTTGCTCCCACAAATGAAGCCTTCCGA	CAACCGGGGAAGGAGTCTACACAGTC_GCTCCCACAAATGAAGCCTTCCGA			X (position 20)	
Small Insertion	c.1886_1894dupCCAATGTTC		Thr629_Asn630insAsnValPro	AATGGCGTGGTCCATGTCATCACCAATGTTCTGCAGCCTCCAGgtaagtgtcgcatcc	CATGTCATCACCAATGTTCCCAATGTTCTGCAGCCTCCAGgtaagtgtcgcatcc	X			
	Lattice Type I								
Missense	R124C	C>T	CGC>TGC	CACCACCACTCAGCTGTACACGGACCGCACGGAGAAGCTGAGGCCTGAGAT	CACCACCACTCAGCTGTACACGGACTGCACGGAGAAGCTGAGGCCTGAGAT			X (position 4)	X (position 13)
	V505D	T>A	GTC>GAC	GTGCTGACCCCCCCAATGGGGACTGTCATGGATGTCCTGAAGGGAGACAAT	GTGCTGACCCCCCCAATGGGGACTGACATGGATGTCCTGAAGGGAGACAAT			X (position 3)	X (position 2)
	I522N	T>A	ATC>AAC	atctacagCATGCTGGTAGCTGCCATCCAGTCTGCAGGACTGACGGAGACC	atctacagCATGCTGGTAGCTGCCAACCAGTCTGCAGGACTGACGGAGACC			X (position 1)	
	L569R	T>G	CTG>CGG	GATGCCAAGGAACTTGCCAACATCCTGAAATACCACATTGGTGATGAAATC	GATGCCAAGGAACTTGCCAACATCCGGAAATACCACATTGGTGATGAAATC	X		X (position 7)	
	L569Q	T>G	CTG>CGG	ATGCCAAGGAACTTGCCAACATCCTGAAATACCACATTGGTGATGAAATCC	ATGCCAAGGAACTTGCCAACATCCGGAAATACCACATTGGTGATGAAATCC	X		X (position 7)	
	H572R	A>G	CAC>CGC	GAACTTGCCAACATCCTGAAATACCACATTGGTGATGAAATCCTGGTTAGC	GAACTTGCCAACATCCTGAAATACCGCATTGGTGATGAAATCCTGGTTAGC			X (position 4)	X (position 8)
	T621P	A>C	ACA>CCA	GAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCA	GAGCCTGACATCATGGCCCCAAATGGCGTGGTCCATGTCATCA	X		X (position 5)	
	Lattice Type IIIA								
Missense	P501T	C>A	CCA>ACA	CACGATGGACCGGGTGCTGACCCCCCAATGGGGACTGTCATGGATGTCCT	CACGATGGACCGGGTGCTGACCCCCACAATGGGGACTGTCATGGATGTCCT			X (position 1)	X (position 4)
	A546T	G>A	GCC>ACC	CACAGTCTTTGCTCCCACAAATGAAGCCTTCCGAGCCCTGCCACCAAGAGA	CACAGTCTTTGCTCCCACAAATGAAACCTTCCGAGCCCTGCCACCAAGAGA			X (position 9)	
	A620P	G>C	GCC>CCC	GCCTGTTGCCGAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTCAT	GCCTGTTGCCGAGCCTGACATCATGCCCACAAATGGCGTGGTCCATGTCAT	X		X (position 8)	
	N622K	T>A	AAT>AAA	GCCGAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAAT	GCCGAGCCTGACATCATGGCCACAAAAGGCGTGGTCCATGTCATCACCAAT			X (position 5)	
	N622K	T>G	AAT>AAG	GCCGAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAAT	GCCGAGCCTGACATCATGGCCACAAAGGGCGTGGTCCATGTCATCACCAAT	X		X (position 5)	
Small Deletion	c.1879delG		Val627SerfsX44	CATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAATGTTCTGCAGCCTC	CATCATGGCCACAAATGGCGTGGTCCAT_TCATCACCAATGTTCTGCAGCCTC		X	X (position 2)	
	Granular								
Missense	V113I	G>A	GTT>ATT	AAACCTTTACGAGACCCTGGGAGTCGTTGGATCCACCACCACTCAGCTGTA	AAACCTTTACGAGACCCTGGGAGTCATTGGATCCACCACCACTCAGCTGTA			X (position 2)	X (position 1)
	D123H	G>C	GAC>CAC	ATCCACCACCACTCAGCTGTACACGGACCGCACGGAGAAGCTGAGGCCTGA	ATCCACCACCACTCAGCTGTACACGCACCGCACGGAGAAGCTGAGGCCTGA			X (position 7)	X (position 9)
	F547C	T>G	TTC>TGC	TTTGCTCCCACAAATGAAGCCTTCCGAGCCCTGCCACCAAGAGAACGGAG	TTTGCTCCCACAAATGAAGCCTGCCGAGCCCTGCCACCAAGAGAACGGAG			X (position 1)	
	L550P	T>C	CTG>CCG	CACAAATGAAGCCTTCCGAGCCCTGCCACCAAGAGAACGGAGCAGACTC	CACAAATGAAGCCTTCCGAGCCCCGCCACCAAGAGAACGGAGCAGACTC	X		X (position 1)	
	Granular Type I								
Missense	R124S	C>A	CGC>AGC	CCACCACCACTCAGCTGTACACGGACCGCACGGAGAAGCTGAGGCCTGAGAT	CCACCACCACTCAGCTGTACACGGACAGCACGGAGAAGCTGAGGCCTGAGAT			X (position 4)	
	R555W	C>T	CGG>TGG	TCCGAGCCCTGCCACCAAGAGAACGGAGCAGACTCTTGGGAGATGCCAA	TTCCGAGCCCTGCCACCAAGAGAATGGAGCAGACTCTTGGGAGATGCCAA		X	X (position 7)	
Small Deletion	c.310_311delCT		Leu103_Ser104del	cttctgctcctgcagCCCTACCACTCTCAAACCTTTACGAGACCCTGGGAG	cttctgctcctgcagCCCTACCACT_CAAACCTTTACGAGACCCTGGGAG		X	X (position 3)	
Small Indel	c.371_378delGCACGGAGinsTC		Asp123delins	CCACCACCACTCAGCTGTACACGGACCGCACGGAGAAGCTGAGGCCTGAGAT	CCACCACCACTCAGCTGTACACGGACCTCAAGCTGAGGCCTGAGATGGAGGG			X (position 7)	X (position 13)
	Corneal Dystrophy								
Missense	M502V	A>G	ATG>GTG	GACCGGGTGCTGACCCCCCCAATGGGGACTGTCATGGATGTCCTGAAG	GACCGGGTGCTGACCCCCCCAGTGGGGACTGTCATGGATGTCCTGAAG			X (position 1)	X (position 12)
	S516R	C>G	AGC>AGG	acatttctgtgtgtgtatctacagCATGCTGGTAGCTGCCATCCAGTCT	acatttctgtgtgtgtatctacagGATGCTGGTAGCTGCCATCCAGTCT	X	X	X (position 5)	
	L559V	T>G	TTG>GTG	CAAGAGAACGGAGCAGACTCTTGGgtaaagaccaacttaagtacacgtc	CAAGAGAACGGAGCAGACTCGTGGgtaaagaccaacttaagtacacgtc			X (position 1)	
	Epithelial Basement Membrane								
Missense	L509R	T>G	CTG>CGG	CCAATGGGGACTGTCATGGATGTCCTGAAGGGAGACAATCGCTTTAGgtaa	CCAATGGGGACTGTCATGGATGTCCGGAAGGGAGACAATCGCTTTAGgtaa	X		X (position 3)	
	R666S	G>C	AGG>AGC	cttcagGCTTCCCAGAGGTCTGTGCGACTAGgtgagtctggtctgggtt	cttcagGCTTCCCAGAGCTCTGTGCGACTAGgtgagtctggtctgggtt			X (position 4)	X (position 14)
	Reis-Buckler								
Missense	R124L	G>T	CGC>CTC	CCACCACCACTCAGCTGTACACGGACCGCACGGAGAAGCTGAGGCCTGAGAT	CCACCACCACTCAGCTGTACACGGACCTCACGGAGAAGCTGAGGCCTGAGAT			X (position 3)	
	L509P	T>C	CTG>CCG	CCAATGGGGACTGTCATGGATGTCCTGAAGGGAGACAATCGCTTTAG	CCAATGGGGACTGTCATGGATGTCCCGAAGGGAGACAATCGCTTTAG	X		X (position 3)	
	Keratoconus								
Missense	H403Q	T>G	CAT>CAG	GACAAGCCGGCCTCGGCAATCATCTCTCTGGAAGTGAGCGGTTGACCCT	GACAAGCCGGCCTCGGCAATCAGCTCTCTGGAAGTGAGCGGTTGACCCT			X (position 6)	
	G535Term	G>T	GGA>TGA	TGACGGAGACCCTCAACCGGGGAAGGAGTCTACACAGTCTTTGCTCCCACA	TGACGGAGACCCTCAACCGGGGAATGAGTCTACACAGTCTTTGCTCCCACA		X	X (position 5)	X (position 10)
	Avellino/GCD II								
Missense	R124H	G>A	CGC>CAC	CCACCACCACTCAGCTGTACACGGACCGCACGGAGAAGCTGAGGCCTGAGAT	CCACCACCACTCAGCTGTACACGGACCACACGGAGAAGCTGAGGCCTGAGAT			X (position 3)	X (position 13)
	Combined granular-lattice type								
Missense	M619K	T>A	ATG>AAG	GAGCCTGTTGCCGAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTC	GAGCCTGTTGCCGAGCCTGACATCAAGGCCACAAATGGCGTGGTCCATGTC			X (position 8)	
	Granular Type II								
Missense	R179Term	C>T	CGA>TGA	ATATGGTGGGCAGGCGAGTCCTGACTGATGAGCTGAAACACG	ATATGGTGGGCAGGTGA				
	Map-like								
Missense	G623R	G>C	GGC>CGC	GAGCCTGACATCATGGCCACAAATGGCGTGGTCCATGTCATCACCAATGTT	GAGCCTGACATCATGGCCACAAATCGCGTGGTCCATGTCATCACCAATGTT			X (position 4)	
	Schnyder								
Missense	E131D	G>T	GAG>GAT	GGACCGCACGGAGAAGCTGAGGCCTGAGATGGAGGGGCCCGGCAGCT	GGACCGCACGGAGAAGCTGATGCCTGAGATGGAGGGGCCCGGCAGCT			X (position 9)	
	Thiel-Behnke								
Missense	R555Q	G>A	CGG>CAG	TTCCGAGCCCTGCCACCAAGAGAACGGAGCAGACTCTTGGgtaaagaccaa	TTCCGAGCCCTGCCACCAAGAGAACAGAGCAGACTCTTGGgtaaagaccaa			X (position 8)	X (position 11)
					Total:	20/70	7 OUT OF 70	51/70	7 OUT OF 70

Supplementary Table 2: Oligo nucleotides used to clone the CRISPR/Cas9 constructs and luciferase reporter plasmids

Oligo sequence (5' to 3')	Oligo name
CACCGACTCAGCTGTACACGGACTGCA	sgR124C22ntTOP
AAACTGCAGTCCGTGTACAGCTGAGTC	sgR124C22ntBOTTOM
CACCGCTCAGCTGTACACGGACTGCA	sgR124C21ntTOP
AAACTGCAGTCCGTGTACAGCTGAGC	sgR124C21ntBOTTOM
CACCGTCAGCTGTACACGGACTGCA	sgR124C20ntTOP
AAACTGCAGTCCGTGTACAGCTGAC	sgR124C20ntBOTTOM
CACCGCAGCTGTACACGGACTGCA	sgR124C19ntTOP
AAACTGCAGTCCGTGTACAGCTGC	sgR124C19ntBOTTOM
CACCGAGCTGTACACGGACTGCA	sgR124C18ntTOP
AAACTGCAGTCCGTGTACAGCTC	sgR124C18ntBOTTOM
CACCGGCTGTACACGGACTGCA	sgR124C17ntTOP
AAACTGCAGTCCGTGTACAGCC	sgR124C17ntBOTTOM
CACCGCTGTACACGGACTGCA	sgR124C16ntTOP
AAACTGCAGTCCGTGTACAGC	sgR124C16ntBOTTOM
CACCG ACCACTCAGCTGTACACGGACCACA	sgR124H22ntTOP
AAACTGTGGTCCGTGTACAGCTGAGTGGTC	sgR124H22ntBOTTOM
CACCG CCACTCAGCTGTACACGGACCACA	sgR124H21ntTOP
AAACTGTGGTCCGTGTACAGCTGAGTGGC	sgR124H21ntBOTTOM
CACCG CACTCAGCTGTACACGGACCACA	sgR124H20ntTOP
AAACTGTGGTCCGTGTACAGCTGAGTGC	sgR124H20ntBOTTOM
CACCG ACTCAGCTGTACACGGACCACA	sgR124H19ntTOP
AAACTGTGGTCCGTGTACAGCTGAGTC	sgR124H19ntBOTTOM
CACCG CTCAGCTGTACACGGACCACA	sgR124H18ntTOP
AAACTGTGGTCCGTGTACAGCTGAGC	sgR124H18ntBOTTOM
CACCG TCAGCTGTACACGGACCACA	sgR124H17ntTOP
AAACTGTGGTCCGTGTACAGCTGAC	sgR124H17ntBOTTOM
CACCG CAGCTGTACACGGACCACA	sgR124H16ntTOP
AAACTGTGGTCCGTGTACAGCTGC	sgR124H16ntBOTTOM
CACCG ACTCAGCTGTACACGGACCTCA	sgR124L22ntTOP
AAACTGAGGTCCGTGTACAGCTGAGTC	sgR124L22ntBOTTOM
CACCG CTCAGCTGTACACGGACCTCA	sgR124L21ntTOP
AAACTGAGGTCCGTGTACAGCTGAGC	sgR124L21ntBOTTOM
CACCG TCAGCTGTACACGGACCTCA	sgR124L20ntTOP
AAACTGAGGTCCGTGTACAGCTGAC	sgR124L20ntBOTTOM
CACCG CAGCTGTACACGGACCTCA	sgR124L19ntTOP
AAACTGAGGTCCGTGTACAGCTGC	sgR124L19ntBOTTOM

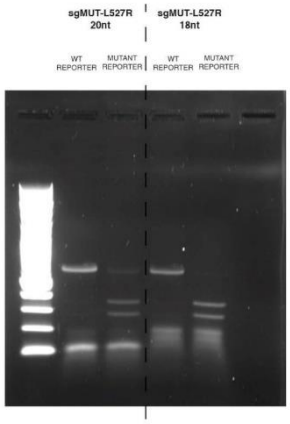
CACCG AGCTGTACACGGACCTCA	sgR124L18ntTOP
AAACTGAGGTCCGTGTACAGCTC	sgR124L18ntBOTTOM
CACCG GCTGTACACGGACCTCA	sgR124L17ntTOP
AAACTGAGGTCCGTGTACAGCC	sgR124L17ntBOTTOM
CACCG CTGTACACGGACCTCA	sgR124L16ntTOP
AAACTGAGGTCCGTGTACAGC	sgR124L16ntBOTTOM
CACCG CCAAGAGTCTGCTCCATTCTCT	sgR555W22ntTOP
AAACAGAGAATGGAGCAGACTCTTGGC	sgR555W22ntBOTTOM
CACCG CAAGAGTCTGCTCCATTCTCT	sgR555W21ntTOP
AAACAGAGAATGGAGCAGACTCTTGC	sgR555W21ntBOTTOM
CACCG AAGAGTCTGCTCCATTCTCT	sgR555W20ntTOP
AAACAGAGAATGGAGCAGACTCTTC	sgR555W20ntBOTTOM
CACCG AGAGTCTGCTCCATTCTCT	sgR555W19ntTOP
AAACAGAGAATGGAGCAGACTCTC	sgR555W19ntBOTTOM
CACCG GAGTCTGCTCCATTCTCT	sgR555W18ntTOP
AAACAGAGAATGGAGCAGACTCC	sgR555W18ntBOTTOM
CACCG AGTCTGCTCCATTCTCT	sgR555W17ntTOP
AAACAGAGAATGGAGCAGACTC	sgR555W17ntBOTTOM
CACCG GTCTGCTCCATTCTCT	sgR555W16ntTOP
AAACAGAGAATGGAGCAGACC	sgR555W16ntBOTTOM
CACCG CCAAGAGTCTGCTCTGTTCTCT	sgR555Q22ntTOP
AAACAGAGAACAGAGCAGACTCTTGGC	sgR555Q22ntBOTTOM
CACCG CAAGAGTCTGCTCTGTTCTCT	sgR555Q21ntTOP
AAACAGAGAACAGAGCAGACTCTTGC	sgR555Q20ntBOTTOM
CACCG AAGAGTCTGCTCTGTTCTCT	sgR555Q20ntTOP
AAACAGAGAACAGAGCAGACTCTTC	sgR555Q21ntBOTTOM
CACCG AGAGTCTGCTCTGTTCTCT	sgR555Q19ntTOP
AAACAGAGAACAGAGCAGACTCTC	sgR555Q19ntBOTTOM
CACCG GAGTCTGCTCTGTTCTCT	sgR555Q19ntTOP
AAACAGAGAACAGAGCAGACTCC	sgR555Q19ntBOTTOM
CACCG AGTCTGCTCTGTTCTCT	sgR555Q17ntTOP
AAACAGAGAACAGAGCAGACTC	sgR555Q17ntBOTTOM
CACCG GTCTGCTCTGTTCTCT	sgR555Q16ntTOP
AAACAGAGAACAGAGCAGACC	sgR555Q16ntBOTTOM
AGATGCGGAGAAGCTGAGGCCTGAG	sgR124L AsCpf1 TOP
AAAACTCAGGCCTCAGCTTCTCCGC	sgR124L AsCpf1 BOTTOM
CACCGCCTTCCGAGCCCTGCCACCAA	sgR555W sg21 S.aureus TOP
AAACTTGGTGGCAGGGCTCGGAAGGC	sgR555W sg21 S.aureus BOTTOM
CACCGGCCTTCCGAGCCCTGCCACCAA	sgR555W sg22 S.aureus TOP
AAACTTGGTGGCAGGGCTCGGAAGGCC	sgR555W sg22 S.aureus BOTTOM
ACCACCACTCAGCTGTACACGGACCGCACGGAGAAGCTGAGGCCTGAGATG	R124 Rep WT Top
CTAGCATCTCAGGCCTCAGCTTCTCCGTGCGGTCCGTGTACAGCTGAGTGGTGGTGTAC	R124 Rep WT Bottom
ACCACCACTCAGCTGTACACGGACTGCACGGAGAAGCTGAGGCCTGAGATG	R124C Rep Mutant Top
CTAGCATCTCAGGCCTCAGCTTCTCCGTGAGGTCCGTGTACAGCTGAGTGGTGGTGTAC	R124C Rep Mutant Bottom
ACCACCACTCAGCTGTACACGGACCACACGGAGAAGCTGAGGCCTGAGATG	R124H Rep Mutant Top

CTAGCATCTCAGGCCTCAGCTTCTCCGTGTGGTCCGTGTACAGCTGAGTGGTGGTGTAC	R124H Rep Mutant Bottom
ACCACCACTCAGCTGTACACGGACCTCACGGAGAAGCTGAGGCCTGAGATG	R124L Rep Mutant Top
CTAGCATCTCAGGCCTCAGCTTCTCCGTGAGGTCCGTGTACAGCTGAGTGGTGGTGTAC	R124L Rep Mutant Bottom
TTCCGAGCCCTGCCACCAAGAGAACGGAGCAGACTCTTGGGAGATGCCAAG	R555 Rep WT Top
CTAGCTTGGCATCTCCCAAGAGTCTGCTCCGTTCTCTTGGTGGCAGGGCTCGGAAGTAC	R555 Rep WT Bottom
TTCCGAGCCCTGCCACCAAGAGAACAGAGCAGACTCTTGGGAGATGCCAAG	R555Q Rep Mutant Top
CTAGCTTGGCATCTCCCAAGAGTCTGCTCTGTTCTCTTGGTGGCAGGGCTCGGAAGTAC	R555Q Rep Mutant Bottom
CTTCCGAGCCCTGCCACCAAGAGAATGGAGCAGACTCTTGGGAGATGCCAA	R555W MUT REP TOP
CTAGTTGGCATCTCCCAAGAGTCTGCTCCATTCTCTTGGTGGCAGGGCTCGGAAGGTAC	R555W MUT REP BOTTOM

Supplementary Table 3: Off-target scores for each guide RNA utilised

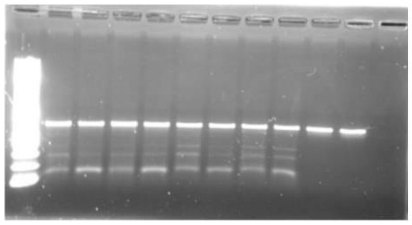
		Benchling		**MIT Optimised CRISPR Design Tool**
		Off-target score	On-target score	Off-target score
***S.pyogenes* Cas9 NGG PAM**	**R124WT**	95	83.7	94.4
	R124C	62	65.4	80.1
	R124H	70	81.4	79.5
	R124L	71	47.9	82
	R555WT	65.6	48.1	83
	R555Q	46.6	0.9	31
	R555W	56.7	3.7	50
***S.aureus* Cas9 NNGRRT PAM**	**R555W**	30.4	79.1	Program does not look for S.aureus Cas9 PAM
Mutant *As*Cpf1	**R124L**	Program does not look for mutant AsCpf1 PAM		

Supplementary Figure 1: Allele-specific cleavage of L527R TGFBI mutation utilising a PAM-specific approach

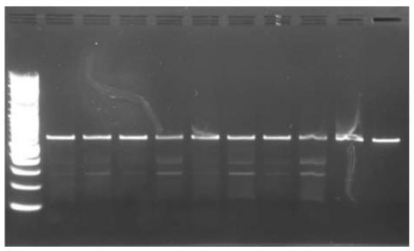


Supplementary Figure 2: Confirmation of the specificity achieved using a guide-specific system targeted to prevalent TGFBI mutations

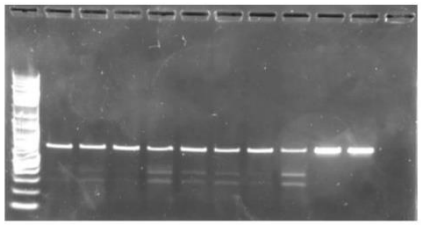
R124C:



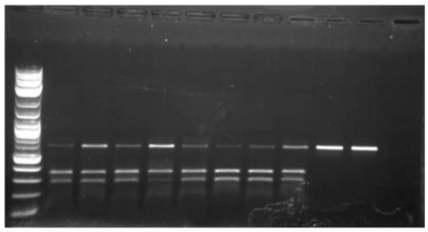
R124H:



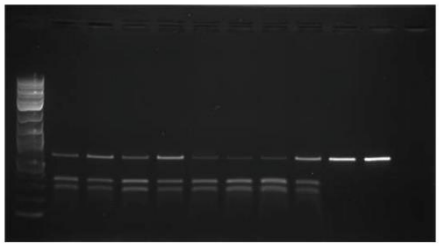
R124L:



R555Q:



R555W:



Paper III

Paper III

4 Paper III - ASNIP-CRISPR enables mutation independent allele-specific editing by Cas9

Kathleen A. Christie, Kevin Blighe, Marie Lukkassen, Zaheer Ali, Larry A. DeDionisio, Caroline Conway, Anton Lennikov, Connie Chao Shern, Rachelle E. Irwin, Doug Turnbull, John Marshall, Jan J. Enghild, Lasse D. Jensen, M. Andrew Nesbit, C.B.Tara Moore*

The main aims of this paper were to:

1. Determine if TGFBIp has a role in wound healing via a tail-fin regeneration assay
2. Identify SNPs with a MAF of >0.1 that have a PAM present on only one allele
3. Estimate the proportion of the East Asian population that could be targeted by ASNIP CRISPR utilising the SNPs identified in aim 2
4. Perform phased sequencing in a R124H *TGFBI* corneal dystrophy patient, determine if the patient has any of the SNPs identified in aim 2 and if they contain a novel PAM on the same allele as the mutation
5. Design guides utilising SNPs within the R124H patient genome that match the above criteria
6. Determine if these guides can achieve allele-specificity in an *in vitro* digest and in a patient derived lymphocyte cell line
7. Demonstrate the potential of generating an allele-specific dual-cut

Contribution

I formed a collaborator with Dr Lasse Jensen and travelled to Sweden to carry out the tail-fin regeneration assay. I performed mutational analysis on the *TGFBI* locus to identify SNPs with a MAF of >0.1. Using haplotype data from the 1000 Genomes I calculated the proportion of the East Asian population that could be targeted by this approach. I designed guides utilising the phased sequencing data. I generated a lymphocyte cell line by EBV transformation of PBMCs which I isolated from the whole blood of the R124H patient. I tested these guides in *in vitro* digests and performed all single and dual nucleofections. I extracted DNA and performed all PCR and qPCR reactions. I wrote the manuscript and prepared all figures and tables.

ASNIP-CRISPR enables mutation-independent allele-specific editing by Cas9 to treat autosomal dominant disease

Kathleen A. Christie[1], Kevin Blighe[2], Marie Lukkassen[3], Zaheer Ali[4], Larry A. DeDionisio[2], Caroline Conway[1], Anton Lennikov[4], Connie Chao Shern[1,2], Rachelle E. Irwin[1], Doug Turnbull[5], John Marshall[6], Jan J. Enghild[3], Lasse D. Jensen[4], M. Andrew Nesbit[1], C.B.Tara Moore[1,2*]

Affiliations

[1] Biomedical Sciences Research Institute, Ulster University, Coleraine, Northern Ireland, BT52 1SA, UK

[2] Avellino Laboratories, Menlo Park, California, CA 94025, USA

[3] Department of Molecular Biology and Genetics, Aarhus University, Aarhus, DK-8000, Denmark

[4] Unit of Cardiovascular Medicine, Department of Medical and Health Science, Linköping University, 851 85 Linköping, Sweden

[5] Genomics and Cell Characterization Core Facility, University of Oregon, OR 97403, USA

[6] Department of Genetics, UCL Institute of Ophthalmology, London, EC1V 9EL, UK

* To whom correspondence should be addressed

Biomedical Sciences Research Institute, University of Ulster, Coleraine, Northern Ireland BT52 1SA, United Kingdom. Tel no: +44(0)2870124577. Email: t.moore@ulster.ac.uk

Abstract

Corneal dystrophies are predominantly caused by autosomal dominant missense mutations. Silencing only the mutant allele offers a promising treatment strategy for these blinding diseases. Gene disruption by CRISPR/Cas9 provides a tool to permanently silence the mutant allele. In order to discriminate between wild-type and mutant alleles Cas9 must be able to detect a single base pair change. Allele-specific editing can be achieved by employing either a guide-specific approach, in which the missense mutation is found within the guide sequence; or a PAM-specific approach, in which the missense mutation generates a novel PAM. While both approaches have been shown to offer allele-specificity in certain contexts, in cases where a number of missense mutations are associated with a particular disease, such as *TGFBI* corneal dystrophies, it is neither possible nor realistic to target each mutation

individually. Here we demonstrate allele-specific, SNP-derived PAM, *in cis*, personalised (ASNIP) CRISPR is capable of achieving complete allele discrimination and propose it as a targeting approach for autosomal dominant disease. ASNIP CRISPR utilises natural variants in the target region that contain a PAM on one allele which lies *in cis* with the causative mutation. Consequently, the targeting approach is no longer constrained by the mutation, ensuring a highly specific guide can be selected. In addition, genetic variation has been shown to affect the target specificity of CRISPR. ASNIP CRISPR guide design will take into account the patient's individual genetic make-up allowing on and off target activity to be assessed in a personalised manner.

Introduction

Corneal dystrophies comprise a group of inherited, bilateral genetic eye diseases that affect the transparency or shape of the cornea, which can lead to progressive vision loss and eventually blindness[1]. Transforming growth factor β-induced (*TGFBI*) has been implicated as the causative gene in a number of epithelial and stromal corneal dystrophies. TGFBIp is an extracellular matrix (ECM) protein and through its interaction with integrins is involved in many key cellular processes, thus has been shown to have a role in wound healing, angiogenesis, cancer and inflammatory diseases[2,3]. Despite the fact that TGFBIp is ubiquitously expressed, mutations within *TGFBI* appear only to result in an adverse phenotype in the cornea, although the mechanism behind the accumulation of mutant TGFBIp in the cornea is incompletely understood. To date >60 different disease-causing missense mutations within *TGFBI* have been described*;* these mutations and the dystrophies associated with them are broadly known as *TGFBI* corneal dystrophies[4,5]. A very strong genotype-phenotype correlation exists between each missense mutation and the pattern of the mutant protein deposits that accumulate in the cornea. They are monogenic and are predominantly due to autosomal dominant missense mutations [4,5]. As such they are not

amenable to gene replacement therapy, as the production of mutant protein in the cornea will persist. Repair of the R124H missense mutation has been demonstrated in patient-derived primary corneal keratocytes[6]. However, template repair is considered relatively rare in most cell types[7]. Critically, patients that have an underlying *TGFBI* mutation but a seemingly quiet cornea who receive laser eye surgery will see a sudden emergence of corneal opacities [2,8–11], indicating a potential role for *TGFBI* in wound repair, suggesting complete knockout of *TGBFI* in the cornea would not be advisable. Due to the autosomal dominant inheritance pattern of the *TGFBI* corneal dystrophies selective ablation of only the mutant allele may be a viable treatment strategy. Heterozygous nonsense mutations in *TGFBI* associated with a normal phenotype have been reported[12,13], indicating that *TGFBI* is haplosufficient and disruption of only the mutant allele would not lead to a more severe phenotype. Allele-specific siRNAs targeted to a lattice corneal dystrophy (LCD1) (OMIM:122200) mutation R124C have been shown to achieve potent and specific knockdown of the mutant allele[14]. However, as knockdown of mutant protein expression by siRNA is only transient and would require continued application and, depending on the delivery route chosen, likely repeat injections into the eye, permanent disruption of the mutant allele would be an attractive alternative strategy. Non-homologous end-joining (NHEJ) mediated gene disruption by CRISPR/Cas9 provides a tool to permanently silence a gene [15]. *S.pyogenes* Cas9 searches the genome for a NGG protospacer adjacent motif (PAM), once a NGG PAM is encountered Cas9 will determine if the guide sequence supplied has complementarity with the flanking sequence. If there is global sequence similarity Cas9 will bind and generate a double-strand break (DSB) at this location. NHEJ, the DNA repair mechanism most often used in non-dividing cells, can introduce insertions and deletions (indels) at the repair site and can result in a frameshifting mutation leading to premature termination of translation and thus permanent disruption of the target gene. The majority of *TGFBI* missense mutations are

caused by single base pair changes, therefore to achieve allele-specific NHEJ mediated gene disruption of mutant *TGFBI,* Cas9 must be able to discriminate between wild-type and mutant alleles which differ by only a single base pair change. Allele-specific editing of missense mutations via CRISPR/Cas9 can be achieved by employing either a guide-specific approach, in which the missense mutation is found within the guide sequence; or a PAM specific approach, in which the missense mutation generates a novel PAM. Utilising a guide-specific approach has been shown to be promising, achieving good allele discrimination with certain mutations[16–18]. However, successful application of this approach requires the mutation of interest to have a usable PAM in close proximity, in addition to having a flanking guide sequence that has both a good on and off target cleavage profile. The position of the missense mutation within the guide sequence, and critically within the seed region, which is defined as the first 8-12nt in the guide sequence, has been shown to limit this approach, with a reduction in the allele discrimination observed the more distal the mutation is from the PAM[19]. Similarly, while exploitation of a novel PAM has been shown to confer stringent allele-specificity[15,20], only a fraction of missense mutations will generate a novel PAM[19]. While both approaches can be efficiently utilised in the context of certain mutations, they both highlight the limitations of a mutation dependent approach. In the case of *TGFBI* corneal dystrophies, >60 missense mutations are currently associated with disease, utilising either a guide-specific or PAM specific approach would require the design of >60 different guides that all have good on-target activity and low off-target activity, which is an insurmountable task as >1/3 of these missense mutations cannot be targeted by either approach and not all of the remaining mutations will offer guides with good on and off target profiles[19]. Here we demonstrate that allele-specific, SNP-derived PAM, *in cis*, personalised (ASNIP) CRISPR is capable of achieving stringent allele discrimination with wild-type *S.pyogenes* Cas9 and propose it as a targeting approach for autosomal dominant disease that can be implemented

independently of the mutation present. We present a work flow that allows allele-specificity to be achieved in cases where phase cannot be pre-determined[21]. ASNIP CRISPR utilises natural variants in the target region that are associated with a PAM on only allele that lie *in cis* with the causative mutation (supplementary figure 2). Consequently, the targeting approach is no longer constrained by the mutation. While employing common variants ensures that a pool of well tested guides can be used to treat the majority of individuals in a given population. Genetic variation has been shown to affect the target specificity of CRISPR[22,23], we present a workflow for genes associated with autosomal dominant disease that will allow guide design based on the patient's individual genetic make-up, therefore on and off target activity can be routinely assessed, in a personalised manner for every therapeutic application.

Materials and Methods

Caudal fin regeneration assay

Zebrafish were housed under standard conditions[24] at the Linköping University Zebrafish Core Facility. For caudal fin amputations, fish were anesthetized in 0.02 % tricaine and fins were cut using scalpel blades, care was taken to cut perpendicular to the anterior/posterior plane of the animal. 3 days post amputation fish were were anesthetized in 0.02 % tricaine and 3 nmole/ul morpholino 5'-GAGACGCATTGGGAACTCACAGTGG -3' (Gene Tools) was mixed with rhodamine red 1 ug/ul at ratio 9:1, was injected distal to each bone ray along the regenerating tissue on the dorsal side of the fin. Immediately following injection each side of the fin was electroporated using a 3 mm diameter tweezer electrode (BTX, Holliston, MA, USA) and electroporator (BTX, Holliston, MA, USA) the parameters used were ten consecutive 50 msec pulses, at 15 V with a 1 sec pause between pulses. 5 days post injection brightfield and fluorescent images were taken of the regenerated tails and the fins were

collected for LC-MS/MS. These studies were approved by the Linköping Research Animal Ethical Council.

LC-MS/MS of regenerated caudal fins

The regenerated caudal fins from seven fish were collected and separated into morpholino treated and untreated halves. Each sample was incubated in 8 M urea, 100 mM ammonium bicarbonate, pH 8 for 1h at room temperature. The samples were then reduced with 5 mM DTT for 30 min followed by alkylation with 25 mM iodoacetamide (IAA) for 30 min. Remaining IAA was quenched with 30 mM DTT and the sample diluted to 1 M Urea before addition of 2.5 μg trypsin. Digestion with trypsin was carried out O.N. at 37 °C. The digested samples were centrifuged (5 min, 17,000 ×g) before collection of 20 μl or 30 μl from untreated and treated samples respectively. These aliquots were desalted on homemade reversed-phase micro columns containing small plugs of Octadecyl C18 Solid Phase Extraction disks (Empore, 3M) and dissolved in 10 μl 0.1% formic acid (solvent A) before LC-MS/MS analysis. Nano-LC-MS/MS analyses were performed on an EASY-nLC II system (Thermo Scientific) connected to a Q Excative Plus mass spectrometer (ThermoFisher Scientific). The samples were injected and desalted on a trap column (2 cm × 75 μm inner diameter) and separated on an in-house pulled silica emitter (15 cm × 75 μm inner diameter). Both columns were packed with PreproSil-Pur C18-AQ 3 μm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The peptides were electrosprayed into the mass spectrometer by gradients from 5% solvent A to 35% solvent B (90% acetonitrile in 0.1% formic acid) over 50 min, followed by a quick increase to 100% solvent B over 10 min at a constant flow rate of 250 nl/min. Parallel reaction monitoring (PRM) analyses were performed in the targeted MS/MS mode, using Xcalibur software, with time-scheduled acquisition of the 12 peptides in +/- 5 min retention time windows (Supplementary Table 1) with 1.2 m/z isolation windows. The AGC target was 1e5, and maximum injection time was

150 ms. MS/MS scans were acquired at a resolution of 35,000 at m/z 200. A full mass spectrum followed after every 50 PRM scans using the following parameters: resolution of 70,000 at m/z 200, AGC target 3e6, m/z 250-2000, and maximum injection time of 200 ms.

Quantification of MS Data.

The raw files were converted to MGF files using RawConverter [25] and searched against zebrafish proteome (TrEMBL, and Swiss-Prot) using the Mascot search engine with the following parameters: MS tolerance of 10 ppm, MS/MS tolerance of 0.1 Da, trypsin with one missed cleavage, carbamidomethyl as fixed modification, and oxidized methionine as variable modification. One TGFBIp peptide was not identified and was excluded from the analysis. The raw files from the LC-MS/MS analysis were analysed in Skyline [26] and the extracted ion chromatograms (XIC) of three fragments of each peptide were extracted. The summed XIC of the eight TGFBIp peptides were normalized to the summed XIC of the three B-actin peptides. The significance of the results was tested by a two-tailed paired T-test.

Phased sequencing of R124H patient genome

Genomic DNA was extracted from 3 mls of whole blood with a MagAttract HMW DNA kit (QIAGEN, Hilden, Germany). DNA fragment lengths of approximately 45 kb were enriched for on a Blue Pippen pulsed field electrophoresis instrument (Sage Science, Beverly, MA, USA). Fragment sizes averaging 51,802 bps were confirmed with a Large Fragment kit on the Fragment Analyzer (Advanced Analytical, Ankeny, IA, USA). This high molecular weight (HMW) DNA (1 ng) was partitioned across approximately 1 million synthetic barcodes (GEMs) on a microfluidic Genome Chip with A Chromium™ System (10x Genomics, Pleasanton, CA, USA) according to the manufacturer's protocol. Upon dissolution of the Genome Gel Bead in the GEM, HMW DNA fragments with 16-bp 10x Barcodes along with attached sequencing primers were released. A standard library prep was performed

according to the manufacturer's instructions resulting in sample-indexed libraries using 10x Genomics adaptors. Prior to Illumina bridge amplification and sequencing, the libraries were analyzed on the Fragment Analyzer with the high sensitivity NGS kit. One lane of whole genome paired end short read (2 x 150 nt) sequencing was conducted on a HiSeq 4000 (Illumina, San Diego, CA, USA). The FASTQ files served as input into Long Ranger (10x Genomics) which was used to assemble, align and give haplotype phasing information.

TGFBI linkage disequilibrium analysis

Chromosome 5 1000 Genomes[27] Phase III data in gzipped variant call format (VCF)[28] for build GRCh37 / hg19 was downloaded from the Department of Biostatistics at the University of Washington in November 2014. Indels were left-aligned, multi-allelic calls split, and the data converted to binary call format (BCF) using BCFtools v1.3.1[29]. Variants spanning *TGFBI* (+/-1Kbp) were then extracted, also using BCFtools. The resulting dataset was then temporarily converted to plaint-text VCF to allow for the manual recoding of rs11348106 (a variant of interest) from an indel variant to a dummy single nucleotide variant to allow for later compatibility with downstream tools, before being converted back to BCF. From this dataset, sample groups were then extracted into separate BCF files for the following 1000 Genomes populations: CHB - Han Chinese in Beijing, China (n=103), EAS - East Asian super population (n=504), JPT - Japanese in Tokyo, Japan (n=104). Each file representing each population was then converted into a separate PLINK dataset using PLINK v1.90b3.38[30]. From PLINK, each dataset was then recoded into HaploView-compatible format using the options --chr 5 --from-bp 135364584 --to-bp 135399507 --snps-only no-DI –recodeHV. Recoded datasets (as PED files) were then read separately into HaploView v4.2[31] with default parameters: ignoring pairwise comparisons of markers > 500 Kbp apart; excluding individuals with > 50% missing genotypes. Within HaploView, from the 'Check Markers' tab, 24 variants of interest were selected. A LD heatmap plot was then output in

PNG format for each dataset from the 'LD Plot' tab. Colour scheme and numerical values for display were both set to 'R-squared'. The default method for identifying haploblocks, i.e., confidence intervals[32], was used. The different haplotypes for each identified haploblock were then output in PNG format from the 'Haploblocks' tab. Again, default parameters were used: only including haplotypes > 1%; connecting with thin lines if > 1%; connecting with thick lines if > 10%. All subsequent figure editing was performed using GNU Image Manipulation Program v2.8.16 and R Programming Language 3.5.1.

In vitro digestion to determine on-target specificity

A double-stranded DNA template was prepared by amplifying a region of the luciferase reporter plasmid containing the desired sequence using the primers listed in Supplementary Table 2. A cleavage reaction was set up by incubating 30nM *S.pyogenes* Cas9 nuclease (NEB UK) with 30nM synthetic sgRNA (Synthego) for 10 minutes at 25˚C. The Cas9:sgRNA complex was then incubated with 3nM of DNA template at 37˚C for 1 hour. Fragment analysis was then carried out on a 1% agarose gel.

Preparation of primary human PBMCs

A whole blood sample was collected from a patient with Avellino corneal dystrophy. PBMCs were isolated by centrifugation on a Ficoll density gradient. PBMCs were washed in RPMI 1640 media containing 20% FBS and incubated with EBV at 37˚C for 1 hour. After infection RPMI 1640 containing 20% FBS was added to a total volume of 3ml and 40µl of 1mg/ml phytohaemagglutinin was added. 1.5ml of the lymphocyte mixture was added to two wells of a 24-well plate and allowed to aggregate. Lymphoblastoids were cultured in RPMI 1640 media containing 20% FBS.

Nucleofection of lymphocyte cell line (LCL) with ribonucleoprotein (RNP) complexes

S.pyogenes Cas9 nuclease (NEB) and modified synthetic sgRNAs (Synthego) were complexed to form RNPs. RNPs were formed directly in the Lonza Nucleofector SF solution (SF Cell line 4D-Nucleofector X kit - Lonza), and incubated for 10 minutes at room temperature. Desired number of cells were spun down (300g x 5mins) and resuspended in Nucleofector solution. 5µl of each cell solution was added to 25µL of corresponding preformed RNPs, mixed and transferred to the nucleofector 16-well strip. The cells were electroporated using the 4D Nucleofector (Lonza) and program DN-100, cells were allowed to recover at room temperature for 5mins and 70µl of pre-warmed media was added to each well of Lonza strip to help recovery. The transfected cells were then transferred to 24-well plate with 200µl media. After 48hrs of incubation at 37°C, gDNA was extracted using the QIAmp DNA Mini Kit (Qiagen), the target region was PCR amplified using primer pairs listed in Supplementary Table 2 and targeted resequencing was performed.

Targeted resequencing across target locus

48 hours post nucleofection gDNA was extracted from cells and PCR amplified using primer pairs listed in Supplementary Table 2. PCR products were purified using the Wizard® PCR Preps DNA Purification System (Promega)and subjected to TruSeq PCR free library preparation. Samples then underwent paired end sequencing using an Illumina MiSeq instrument as per the manufacturer's instructions. For genomic DNA samples, paired FASTQ files first undewent read filtration and trimming with Trim Galore! V0.4.0 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (utilising Cutadapt v1.15 and FastQC v0.11.5), using default parameters and --qual 20 --length 70 –paired. Reads from human samples were then aligned to the reference genome GRCh38 / hg38 / GCA_000001405.15 (downloaded from the UCSC), using BWA v0.7.12 (mem algorithm with default settings) [34]. Aligned reads in SAM format were converted to BAM, sorted, and indexed with SAMtools v1.3.1[29]. PCR and optical duplicates were marked with Picard v1.119

(https://broadinstitute.github.io/picard/) and then expunged with SAMtools view function with parameter -F set to 0x400. Reads with MAPQ below 30 were also expunged using SAMtools view with parameter -q set to 30. Output BAMs were then sorted and indexed using SAMtools. Reads in each sample's BAM file were then split based on the SNP of interest. This was achieved using SAMtools view to first extract reads overlapping the target SNP region, and then dividing these into allele-specific reads by using the shell function grep -e to extract reads containing each SNP of interest flanked by 3 bases in both the 3' and 5' directions. Allele-specificity of the resulting reads was visually checked for each sample in IGV v2.3.97 (http://www.broadinstitute.org/igv). The number of properly-paired reads in each allele-specific BAM file with and without indels was then tabulated by using SAMtools view in combination with the shell function awk to filter on the CIGAR string. For example: Reads with indels: samtools view -f 0x02 Allele1.BAM | awk '$6 ~ "I|D"', reads with no indels: samtools view -f 0x02 Allele2.BAM | awk '$6 !~ "I|D"'. Separately, for each allele-specific BAM file, pindel v0.2.5b9 [35] was used to identify indels and substitutions using default settings. Output for each input file was then converted to VCF using pindel2vcf with default parameters plus --min_coverage 1 --het_cutoff 0.1 --hom_cutoff 0.9 to allow for low frequency variants to be retained. Output VCFs were bg-zipped and tab-indexed, and then BCFtools was used to filter out variants that did not have any genotype call by using BCFtools view function with --exclude-uncalled –min-ac=1.

Quantitative PCR

RT-qPCRs were performed using 1× LightCycler 480 SYBR Green I Master (Roche), 10 μM primers and 10ng gDNA. Reactions were run on the LightCycler 480 II (Roche), with an initial incubation step of 95°C, 10 minutes; followed by 45 cycles of 95°C for 10 seconds, 60°C for 10 seconds and 72°C for 10 seconds. Expression was normalised to β-actin, and relative expression was determined using the ΔΔCT method.

Results

Role of *Tgfbi* in zebrafish regenerating tail fin

To test the hypothesis that TGFBIp is involved in wound healing a well-established zebrafish tail fin regeneration assay, commonly used to assess wound healing, was implemented (Figure 1a)[36,37]. 3 days post amputation (dpa) the dorsal side of the fin was injected with a morpholino targeted to *tgfbi*, mixed with rhodamine red, and five days later (8 dpa) fin regeneration of the *tgfbi* morpholino injected dorsal side was compared to the uninjected ventral side. For all fins assessed, the dorsal Tgfbi morpholino-injected side of the fin regenerated slower than the uninjected ventral side. On average there was a 55.50% (± 5.52%) reduction in regenerating tissue of the Tgfbi morpholino-injected side when compared to the untreated, $p < 0.0001$. (Figure 1b). Knockdown of Tgfbi protein in injected tail fins was confirmed by liquid chromatography mass-spectrometry (LC-MS/MS), with a 25% decrease in Tgfbi protein between Tgfbi morpholino injected and control fins, $p = 0.0016$ (Figure 1c). The reduction of wound repair associated with loss of Tgfbi protein indicates an essential role of TGFBIp in wound repair and suggests maintenance of wild-type TGFBIp in the cornea is required to maintain normal repair responses. Therefore, complete knockout of *TGFBI* would be detrimental should the cornea be injured. Thus, allele-specific gene disruption is necessary, in which only the mutant allele is targeted for disruption, leaving the wild-type allele intact.

Identification of mutation-independent PAM-associated SNPS in the *TGFBI* Gene

The *TGFBI* gene, including untranslated regions (UTRs) and introns, covers ~35kb, and there are 17 coding exons. Mutations within *TGFBI* occur in exons 4-16 but are clustered in hotspots found in exons 4, 11, 12 and 14 (Supplementary figure 1). Previously all missense

mutations were analysed to determine if they were targetable for allele-specific NHEJ gene disruption by either a guide-specific or PAM specific approach utilising *S.pyogenes* Cas9; >1/3 were not targetable by either approach, in addition stringent allele-specificity could not be achieved for the 5 most prevalent mutations using a guide-specific approach, the specificity of Cas9 for the mutant allele varied for each mutation investigated and was dependent upon the position of the mutation in the guide sequence [19]. Thus the feasibility of an alternative mutation-independent strategy was explored; it was hypothesised that allele-specific targeting could be achieved by targeting non-disease causing SNPs that contain a PAM on only one allele, that lie *in cis* with the disease causing mutation. This approach was named allele-specific SNP-derived-PAM *in cis* personalised (ASNIP)CRISPR. To identify variants across the *TGFBI* locus suitable for the ASNIP-CRISPR approach, SNPs were filtered to keep those with a minor allele frequency (MAF) of > 0.1 across all of the individuals in the 1000 Genomes Project Phase 3 cohort, the remaining SNPs were then manually examined to determine which contain a PAM on only one allele. (Supplementary Figure 2a and Supplementary Table 3). If no suitable SNPs containing a PAM on only one allele, that lies *in cis* with the patient's mutation, are found across the coding region of *TGFBI* it may be necessary to utilise SNPs outside in the flanking regions of *TGFBI*. In the case of *TGFBI* there are no genes within a 50kb region both upstream and downstream, allowing SNPs to be utilised in these regions without disrupting flanking genes, if necessary. (Supplementary Figure 2b).

Haplotype Analysis of identified ASNIP SNPs across *TGFBI*

Granular corneal dystrophy type II (GCD2), commonly known as Avellino corneal dystrophy (OMIM: 607541), is extremely prevalent in East Asia, in the Korean population the prevalence is 1 in 870 people affected, while in China the reported prevalence is 1 in 400 people affected[38]; in order to understand if the SNPs identified using the average allele-

frequencies across all populations are suitable for the populations in which *TGFBI* corneal dystrophies are prevalent we performed haplotype analysis using the 1000 Genomes project phase 3 data for the East Asian population (EAS) (Figure 2a,d) in addition to sub-populations of Han Chinese in Beijing, China (CHB) (Figure 2b,e) and Japanese in Tokyo, Japan (JPT) (Figure 2c,f), all based on the 24 SNPs that contain a PAM on only one allele in the coding region of *TGFBI*, highlighted in red on Supplementary Table 3. 3 Linkage disequilibrium (LD) blocks were found in the 1000 Genomes EAS population; the first block (EAS-B1) spans 1kb within intron 1-2 (rs2237063- rs756462), the second block (EAS-B2) spans 21kb from intron 2-3 (rs11738979) to intron 13-14 (rs10064478) and finally the third block (EAS-B3) spans 1kb from intron 14-15 (rs6880837) to intron 15-16 (rs6865463) (Figure 2a). The recombination frequencies were 0.73 between EAS-B1 and EAS-B2, and 0.85 between EAS-B2 and EAS-B3, 3 haplotypes exist for EAS-B1, EAS-B2 and EAS-B3 which account for 100%, 98.3% and 99.9% of the haplotype frequencies for each block, respectively (Figure 2d). The same pattern of 3 LD blocks was found in the 1000 Genomes CHB population; the first block (CHB-B1) spans 1kb within intron 1-2 (rs2237063- rs756462), the second block (CHB-B2) spans 21kb from intron 2-3 (rs11738979) to intron 13-14 (rs10064478) and finally the third block (CHB-B3) spans 1kb from intron 14-15 (rs6880837) to intron 15-16 (rs6865463) (Figure 2b). The recombination frequencies were 0.80 between CHB-B1 and CHB-B2, and 0.81 between CHB-B2 and CHB-B3, 3 haplotypes exist for CHB-B1 and CHB-B3, while 4 haplotypes exist for CHB-B2, which account for 100%, 98.6% and 100% of the haplotype frequencies for each block, respectively (Figure 2e). 2 LD blocks were found in the 1000 Genomes JPT population; the first block spans 25kb from intron 1-2 (rs2237063) to intron 10-11 (rs6860369) and the second block spans 2kb from intron 13-14 (rs6880837) to intron 15-16 (rs6865463) (Figure 2c). The recombination frequency was 0.91 between JPT-B1 and JPT-B2, 6 haplotypes exist for JPT-B1 while 3 haplotypes exist for JPT-B2, which

account for 99.1% and 99.4% of the haplotype frequencies for each block, respectively (Figure 2f). Of the 62 missense *TGFBI* mutations known to date 61/62 occur between exons 4 and 14 (Figure 2), therefore for both EAS and CHB blocks 1 and 2 will span 61/62 known mutations, while for JPT blocks 1 and 2 will span <50% of all known mutations.

The haplotypes identified were then analysed using only the large haploblock (EAS-B2, CHB-B2, JPT-B1) to determine the approximate % of the population in which ASNIP CRISPR could be used to selectively disrupt only one allele. ASNIP CRISPR relies on variation across the target locus therefore the % of homozygotes in the population were calculated as these portions of the populations are instantly not targetable. Considering only the allele frequencies of block 2 from figure 2d,e,f the percentage of homozygous individuals was calculated for each population using the Hardy-Weinberg equation for multiple alleles; after exclusion of the homozygous individuals it was calculated that 66% of the EAS population are heterozygous for these alleles and therefore potentially targetable, while 67% of the CHB population and 71% of the JPT population are potentially targetable. Critically, this analysis was performed using one large haploblock in each population, individuals who are homozygous for this haploblock may not be homozygous for the remaining haploblocks, therefore the % of the population that is potentially targetable may be underestimated. The different combinations of haplotypes were then assessed for targeting capacity (Figure Supplementary Figure 3); the number of targetable SNPs in each allele was determined (Supplementary Figure 3a,b,c) and then all possible heterozygous combinations for block 2 were assessed to determine the targeting capacity of the heterozygous individuals in each population (Supplementary Figure 3 d,e,f). In the EAS population the fraction of the SNPs in block 2 that can be targeted across the different combination of heterozygotes ranges from 8 to 18/19 (Supplementary Figure 3d), 5 to 17/19 in CHB (Supplementary Figure 3e) and 2 to 16/17in JPT (Supplementary Figure 3f). This analysis reveals that even when only

considering haploblock 2 ASNIP CRISPR has the potential to target, for at least one position, all heterozygous combinations across all populations investigated, indicating that the 24 SNPs identified can be used to treat the majority of East Asian patients.

Guide design based on patient haplotype

In order to validate this approach in a real-life example we performed phased sequencing of a Japanese patient harbouring a R124H *TGFBI* mutation causative of granular corneal dystrophy type II (GCD2), also known as Avellino corneal dystrophy (OMIM: 607541), the phased sequencing revealed the R124H patient had JPT-B1H1 which co-segregated with JPT-B2H1 and JPT-B1H2 which co-segregated with JPT-B2H2, differing by one position in JPT-B1H1 as the patient was homozygous for the major allele (Figure 3a); this allowed identification of SNPs associated with a PAM on only one allele that lie *in cis* with the patient's R124H mutation. (Figure 3b, Supplementary Table 4). Based on the phased sequencing results a range of guides targeted to the mutant allele were then designed. (Figure 4a, Supplementary Table 5). The ability of wild-type *S.pyogenes* Cas9 to distinguish between 'PAM associated' and 'No PAM present' alleles was assessed firstly by *in vitro* digestion (Figure 4b); a PCR product containing either the allele associated with a PAM or the allele with no PAM present was incubated with ribonucleoprotein (RNP) complexes of Cas9 and sgRNA, digestion products were then electrophoresed on an agarose gel and intensity of the digested products revealed the *in vitro* specificity of each guide. Of the 12 ASNIP guides tested, 8 appeared to preferentially cleave the PAM associated allele while 4 appeared to have little activity at either the 'PAM associated' or 'No PAM' allele. We found that SNPs generating a non-canonical PAM, which is a PAM sequence other than NGG that can still act as a weak PAM for *S.pyogenes* Cas9 such as NAG or NGA [39,40], on the 'No PAM present' allele, only conferred partial discrimination at best. These *in vitro* results suggest that in order

to achieve stringent allele-specificity a SNP in which the allele not associated with a PAM is either NGC, NCG, NGT or NTG should be selected.

Allele-specificity of single ASNIP guides in R124H lymphocyte cell line

In order to generate a model to test the ASNIP CRISPR approach we generated a proliferating lymphocyte cell line (LCL) utilising peripheral blood mononuclear cells (PBMC's) from the patient harbouring the *TGFBI* R124H mutation on which phased sequencing had been performed. RNP complexes of Cas9 and 9 modified synthetic sgRNAs, of the ASNIP guides previously tested in the *in vitro* digestion, were individually nucleofected into the R124H LCLs. To determine the allele-specificity of each ASNIP guide targeted resequencing across the on-target region, where Cas9 is predicted to cleave, was performed. The target region for all 9 guides was amplified and subjected to deep sequencing, and computational analysis was performed to determine the indels that had occurred and with which allele they were associated. For all ASNIP guides screened, we found that they could all efficiently distinguish between 'PAM associated' and 'No PAM present' alleles (Figure 5a). On average only 3.7% of indels occurred on the allele not associated with a PAM, in comparison to 96.3% of indels on the allele that is associated with a PAM. Indicting that ASNIP CRISPR design can achieve stringent allele-specificity in a mutation-independent manner. However, in contrast to what was observed via the *in vitro* digest (Figure 4b), SNPs for which the 'No PAM' allele contained a known non-canonical PAM did not have reduced specificity compared to SNPs in which a known non-canonical PAM was not present on the 'No PAM' allele, indicating that the *in vitro* screen can merely be used as a predictive tool for specificity but no strong conclusions can be drawn. In addition, ASNIP guide rs6860369 appeared inactive in the *in vitro* screen but was active in a cellular context. For 8 out 9 ASNIP guides tested the predominant indels observed were 1 or 2bp insertions, which occurred 3 or 4bp upstream of the PAM (Figure 5b).

Increasing distance between dual-guides reduces the efficiency of a productive edit

ASNIP CRISPR provides an approach that removes the limitations of a mutation-dependent approach. All of the non-disease causing SNPs matching the ASNIP criteria were located in intronic regions, thus indels introduced by the single ASNIP guides do not have therapeutic potential. To overcome this an *in cis* dual-guide approach targeted to the mutant allele is required. Using phasing information, dual guides *in cis* with the mutation were designed to flank the coding region of *TGFBI*. The ASNIP approach allows the design of guides targeting any region of the gene, they are not constrained by the location of the mutation and, upon careful design, excision of the exons between the guides will result in a frameshift that will result in premature termination of translation and nonsense mediated decay (NMD) of the resultant mRNA and selective knockout of expression of the mutant allele. In order to reach a therapeutic threshold *in vivo* we must be able to excise the region between the dual-guides at a high frequency, however, the minimum reduction of TGFBIp in the cornea required to achieve a therapeutic effect is unknown. To our knowledge the maximum distance possible between the dual-guides with which an efficient deletion can still occur has not yet been reported, therefore the frequency of the deletion was estimated for dual combinations ranging from 419bp to 63,428bp. To estimate the frequency with which the deletion occurs we quantified the amplification across both target sites compared to an untreated sample, as primer binding sites should be removed if the region between both guides has been excised (Figure 6a). We hypothesised that the larger the distance between the dual-guides the less frequent the deletion would be, while we found that for the lengths tested, ranging from 419bp to 63,428bp, the greater the distance between the dual guides the less efficient the deletion was (Figure 6b), the decline in frequency was much more gradual than expected; the average of PCR 1 and PCR 2 for 419bp saw a 40% reduction in product, compared to 20% reduction for the average of PCR 1 and PCR 2 for 63,428bp, all dual-combinations used are

shown in supplementary table 6. While these results are encouraging for the use of a dual-cut to induce a therapeutically relevant deletion recent reports by Kosicki *et al* indicate that complex deletions and rearrangements may also be occurring, which may be contributing to the reduction in PCR product[41].

Allele-specific excision of coding region in *TGFBI* utilising an alternative dual cut approach

In some cases, the target SNPs described (Supplementary Table 5) lie substantial distances apart, up to >18kb (Supplementary table 7); as the efficiency of deletion drops with increasing intervening distance, additional guides were designed that lie closer to a particular ASNIP guide yet still allow excision of exons (Supplementary table 8 and Supplementary figure 4). In contrast to the PAM discriminatory guides these were not allele-specific, they were selected to target the intronic region of both alleles (Figure 6c). It was hypothesised that the ASNIP guide will only cut the mutant allele while the common-intronic guide will cut both alleles; on the mutant allele when both cuts are made on that chromosome the region between these cuts may be deleted, while on the wild-type allele a cut should only occur with the common-intronic guide which, provided meticulous design has been applied to avoid important regulatory elements, at most, will result in a small indel and should have no functional effect (Supplementary Figure 5). The efficiency of the dual-cut was assessed in cells transfected with pairs of RNP complexes; dual combinations with a maximum difference of <3.5 kb, ranging in size from 602bp to 4008bp were tested (Supplementary Table 9), in line with previous results we found that small increments in distance had no significant effect on the efficiency of the deletion. On average the reduction of PCR 1 and PCR 2, and hence deletion, when compared to an untreated samples, was 38.87% ± 6.34% for PCR 1 (Figure 6d, shown in blue) and 33.64% ± 2.76% for PCR 2 (Figure 6e, shown in blue); the variation between reduction efficiencies was not significant and can be attributed to the

fact that not all guide sequences perform at equal efficiencies, therefore one dual combination may cut more efficiently than the other.

Improving the efficiency of deletion by the addition of a 50:50 ssODN

Richardson *et al* demonstrated that the addition of a non-homologous single-stranded DNA template can increase gene disruption by altering repair outcomes [42]. Therefore, we hypothesised that the introduction of a single-stranded oligonucleotide containing 50bp of the sequence flanking each cut site (50:50 ssODN) would encourage the DNA repair mechanisms to excise the region in between. When the ssODNs were co-transfected with the RNP complexes the average reduction of PCR 1 was 52.23% ± 3.77%, compared to transfection of only RNP complexes with which the reduction was 39.44% ± 5.21%, $p<0.01$ (Figure 6d, shown in orange); for PCR 2 the average reduction achieved with ssODN co-transfected with RNPs was 48.79% ± 3.43%, compared to transfection of only RNPs with which the reduction was 33.73% ± 3.43%, $p<0.01$ (Figure 6e, shown in orange). The significant increase in efficiency observed indicates it is possible to generate an efficient dual-cut within a cell, therefore a dual-cut approach could be used to achieve a therapeutic effect.

Discussion

Corneal dystrophies are a group of blinding disorders that affect the shape or transparency of the cornea, they are very attractive candidates for gene therapy due to the cornea's small surface area, accessibility and immune privilege status[43]. The autosomal dominant inheritance observed with the majority of corneal dystrophies make them an unsuitable target for traditional gene replacement[4,5], due to continued production of the mutant protein in the cornea; thus, a strategy that ablates the mutant protein is required. The critical role of *TGFBI* in wound repair is evident from the sudden emergence of corneal opacities in patients harbouring an underlying *TGFBI* mutation following laser eye surgery [2,8,10], in addition Rawe

et al reported the upregulation of TGFBIp following injury to the rabbit cornea[11] The key role of TGFBIp in wound repair was underlined by the effect that Tgfbi knockdown had upon tail-fin regeneration in zebrafish, an established assay for monitoring wound repair; this is consistent with previous reports that Tgfbi is upregulated in a regenerating zebrafish tail fin[44]; transcriptome profiling of the tail-fin regeneration in zebrafish again revealed that *tgfbi* was differentially expressed during regeneration[45], in addition Smad 1/5/8 proteins, involved in TGF-β signalling, were reported to be phosphorylated during fin regeneration therefore capable of initiating downstream signalling, further indicative of a critical role of the TGF-β signalling pathway and thus *TGFBI* in the wound healing process. These reports demonstrate that the role of *TGFBI* in wound repair is analogous in zebrafish, rabbit and human and thus suggests that a *TGFBI* corneal dystrophy gene therapy strategy that completely knocks out *TGFBI* in the cornea would not be a viable option as the resultant cornea would be incapable of repair following injury. As haplosufficiency of *TGFBI* has been demonstrated by the report of heterozygous nonsense mutations in *TGFBI* associated with a normal phenotype[12,13], the development of a stringent allele-specific targeting strategy for the corneal dystrophies, that mitigates the effect of the mutant allele and leaves the wild-type allele intact, is a potential treatment strategy for these blinding dystrophies. Currently >60 *TGFBI* corneal dystrophy causing missense mutations have been described, we have previously reported more than one third of these mutations are not targetable by either a guide- or PAM specific approach[19] therefore a strategy based on individual targeting of each mutation would provide an incomplete approach to treat these dystrophies; it is well documented that wild-type *S.pyogenes* Cas9 can tolerate single base-pair mismatches within the seed region of the spacer [19,46–48], conversely mutations within the PAM are much less tolerated and have been shown to impair the cleavage efficiency of Cas9[49–51], indicating that, in the case of wild-type *S.pyogenes* Cas9, a novel PAM approach will be the ideal choice. While the application of

Cas12a or variant nucleases with altered PAM specificities will broaden the potential of a guide-specific approach[17,50,52–55], they will still require individual design for each mutation; due to this we have chosen to by-pass mutation-specific optimisation and develop a streamlined mutation independent approach utilising the well characterised wild-type *S.pyogenes* Cas9, however, the nucleases with altered PAM specificities may broaden the targeting capacity of ASNIP CRISPR and allow a dual-guide combination, highly specific for the mutant allele but close in proximity, to be designed.

ASNIP CRISPR provides a promising alternative to mutation dependent approaches that can be used to treat any patient affected with an autosomal dominant monogenic disease irrespective of their causative mutation. Here we suggest a work-flow that will allow a completely personalised design for each patient to ensure both safe and effective guides are selected; for the gene of interest the SNPs matching the ASNIP criteria must first be determined, SNPs with a MAF of >0.1 that contain a PAM on one allele should be selected and subjected to haplotype analysis to ensure that the haplotypes present differ at enough positions in heterozygous combinations that a reasonable proportion of the population will be treatable by ASNIP CRISPR, this will provide a pool of guides based on commonly occurring SNPs that can be used to treat the majority of a given population. For each patient, phased sequencing of their genome should be undertaken, which will then indicate which guides in the pool of pre-designed guides are suitable for that patient. These guides can be tested in a patient-derived cell line, such as a LCL, or primary fibroblasts to assess the genomic consequences of the selected guide in a personalised manner before treating the patient. In this new era of personalised medicine where progress will be made with great caution whole genome sequencing (WGS) will undoubtedly be a requisite for any patient undergoing gene editing therapies, in order to fully comprehend the success or failure of such therapies; whole-genome phased sequencing allows the design of guides *in cis* with the

mutation for autosomal disease but will also aid in the understanding of outcomes should unwanted off-targets effects or chromosomal translocations occur. Importantly, ASNIP CRISPR can be adopted for other autosomal dominant diseases for cases were phase cannot be pre-determined. This study acts as a proof of concept for ASNIP CRISPR to treat autosomal dominant disease. However, in order to induce a functional effect, further considerations for guide design must be applied. It is widely accepted that provided a premature stop codon resides ≥50-55 nucleotides upstream of the 3' most exon-exon junction then the exon-junction complex will not be removed and thus nonsense mediated decay (NMD) will be induced[56,57]; thus, in order to induce NMD and selectively degrade the mutant allele guides that target early in the transcript will be most desirable, however depending on the phasing data this may not always be possible. When this concept is applied to *TGFBI* a premature termination codon no later than 50-55 nucleotides in exon 15 will result in NMD; this is evident from phenotype associated mutations in *TGFBI*. All mutations associated with *TGFBI* corneal dystrophies are missense mutations or in frame indels with the exception of a frameshifting single base deletion at codon 626 reported by Munier *et al* [58], the result of this frameshift mutation is the addition of 43 missense amino acids and premature termination at codon 669, which is less than the required 50-55nt distance from the 3' most exon-exon junction, therefore NMD is predicted not to occur and the nonsense transcript is translated. In addition, if common intronic guides are required, such as those described here to increase deletion efficiency, care must be taken to ensure any indels they may induce do not disrupt any regulatory elements. In a similar approach used to target the Huntington gene (*HTT*), common intronic guides were found to affect expression of the normal allele due to the targeting of intronic transcription factor binding sites[59], however in contrast to our approach these guides were designed to target intron 1 where they are more likely to affect regulation of transcription. The use of common-intronic guides assumes that any small indels induced in

an intronic region will have no functional effect. However, Kosicki *et al* recently reported single guides targeted to intronic regions produced deletions of up to 2kb at significant frequencies; they demonstrate that transfection of 10 different guides singly, located 263–520 bp from the nearest exon, caused a 8–20% reduction in their gene of interest, while 2 guides > 2 kb away caused a 5–7% loss of their gene of interest[41]. While this would indicate that, provided they are highly allele-specific, single ASNIP guides could have a functional effect by inducing larger deletions, it raises concerns about the types of alterations that Cas9 generates and whether or not current detection methods are capturing a complete picture of the changes induced.

This study focusses on the issue of on-target allele-specificity in relation to the *TGFBI* corneal dystrophies; however, for translation to the clinic a number of key hurdles will have to be overcome, including genome wide specificity and efficiency of delivery to the correct cells in the cornea. Furthermore, potent targeting of the correct cell population must be achieved; the majority of TGFBIp is produced in the corneal epithelium, the epithelium is continually turned over and repopulated via the limbal epithelial stem cells (LESCs), thus, in order to permanently correct the *TGFBI* corneal dystrophies efficient delivery to the LESCs must be achieved. The continued development of engineered nucleases with improved specificity will further enhance the ASNIP CRISPR approach [55,60–62]. In order to minimise off-target cleavage, non-viral delivery systems that allow delivery of Cas9 mRNA or Cas9 protein, that would not persist in the cell for a long time, would be most desirable. Current efforts to develop a non-viral system to deliver these components have been promising and it is likely they will bring about the next frontier in genome editing[63–66]. The era of personalised genome editing has progressed with unprecedented pace, other hurdles such as efficiency and delivery remain to be addressed, however ASNIP CRISPR offers a promising strategy to achieve the necessary degree of on-target allele-specificity in a mutation independent manner.

Acknowledgements

We would like to thank Synthego, specifically Kevin Holden and Travis Maures, for their helpful discussions and contributions. This work has been supported by Avellino Labs.

Author disclosure statement

C.B.Tara Moore is a consultant for Avellino Labs

References

1. Klintworth GK. Corneal dystrophies. Orphanet J rare Dis Cornea Dystrophy Foundaion 2009;4:7.

2. Han KE, Choi S Il, Kim TI, et al. Pathogenesis and treatments of TGFBI corneal dystrophies. *Progress in Retinal and Eye Research* 2016;50:67–88.

3. Lakshminarayanan R, Vithana EN, Chai SM, et al. A novel mutation in transforming growth factor-beta induced protein (TGFβIp) reveals secondary structure perturbation in lattice corneal dystrophy. Br J Ophthalmol 2011. DOI: 10.1136/bjophthalmol-2011-300651.

4. Weiss JS, Møller HU, Aldave AJ, et al. IC3D Classification of Corneal Dystrophies—Edition 2. Cornea 2015;34:117–159.

5. Weiss JS, Moller HU, Lisch W, et al. The IC3D Classification of the Corneal Dystrophies. Cornea 2008;27:S1–S42.

6. Taketani Y, Kitamoto K, Sakisaka T, et al. Repair of the TGFBI gene in human corneal keratocytes derived from a granular corneal dystrophy patient via CRISPR/Cas9-induced homology-directed repair. Sci Rep;7 . Epub ahead of print 2017. DOI: 10.1038/s41598-017-16308-2.

7. Richardson CD, Kazane KR, Feng SJ, et al. CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. Nat Genet 2018;50:1132–1139.

8. Aldave AJ, Sonmez B, Forstot SL, et al. A Clinical and Histopathologic Examination of Accelerated TGFBIp Deposition After LASIK in Combined Granular-Lattice Corneal Dystrophy. Am J Ophthalmol 2007;143:416–419.

9. Jun RM, Tchah H, Kim TI, et al. Avellino corneal dystrophy after LASIK. *Ophthalmology* 2004;111:463–468.

10. Kim T-I, Roh MI, Grossniklaus HE, et al. Deposits of transforming growth factor-beta-induced protein in granular corneal dystrophy type II after LASIK. Cornea 2008;27:28–32.

11. Rawe IM, Zhan Q, Burrows R, et al. Beta-Ig: Molecular cloning in situ hybridization in corneal tissues. Investig Ophthalmol Vis Sci 1997;38:893–900.

12. Sakimoto T, Kanno H, Shoji J, et al. A novel nonsense mutation with a compound heterozygous mutation in TGFBI gene in lattice corneal dystrophy type I. Jpn J Ophthalmol 2003;47:13–17.

13. Song JS, Lim DH, Chung ES, et al. Mutation analysis of the TGFBI gene in consecutive korean patients with corneal dystrophies. Ann Lab Med 2015;35:336–340.

14. Courtney DG, Atkinson SD, Moore JE, et al. Development of allele-specific gene-silencing siRNAs for TGFBI Arg124Cys in lattice corneal dystrophy type i. Investig Ophthalmol Vis Sci 2014;55:977–985.

15. Courtney DG, Moore JE, Atkinson SD, et al. CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting. Gene Ther 2016;23:108–112.

16. Gao X, Tao Y, Lamas V, et al. Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. Nature 2018;553:217–221.

17. Li P, Kleinstiver BP, Leon MY, et al. Allele-Specific CRISPR-Cas9 Genome Editing of the Single-Base P23H Mutation for Rhodopsin-Associated Dominant Retinitis Pigmentosa. Cris J 2018;1:55–64.

18. György B, Lööv C, Zaborowski MP, et al. CRISPR/Cas9 mediated disruption of the Swedish APP allele as a therapeutic approach for early-onset Alzheimer's disease. Mol Ther - Nucleic Acids 2018;11:429–440.

19. Christie KA, Courtney DG, Dedionisio LA, et al. Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. Sci Rep;7 . Epub ahead of print 2017. DOI: 10.1038/s41598-017-16279-4.

20. Bakondi B, Lv W, Lu B, et al. In vivo CRISPR/Cas9 gene editing corrects retinal dystrophy in the S334ter-3 rat model of autosomal dominant retinitis pigmentosa. Mol Ther 2016;24:556–563.

21. Shin JW, Kim K-H, Chao MJ, et al. Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. Hum Mol Genet 2016;ddw286.

22. Lessard S, Francioli L, Alfoldi J, et al. Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. Proc Natl Acad Sci 2017;114:E11257–E11266.

23. Scott DA, Zhang F. Implications of human genetic variation in CRISPR-based therapeutic genome editing. Nat Med 2017;23:1095–1101.

24. Folkesson M, Sadowska N, Vikingsson S, et al. Differences in cardiovascular toxicities associated with cigarette smoking and snuff use revealed using novel zebrafish models. Biol Open 2016;5:970–978.

25. He L, Diedrich J, Chu YY, et al. Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. Anal Chem 2015;87:11361–11367.

26. MacLean B, Tomazela DM, Shulman N, et al. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 2010;26:966–968.

27. 1000 genome project consortium. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.

28. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156–2158.

29. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987–2993.

30. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 2007;81:559–575.

31. Barrett JC, Fry B, Maller J, et al. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–265.

32. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science (80-) 2002;296:2225–2229.

33. Hsiau T, Maures T, Waite K, et al. Inference of CRISPR Edits from Sanger Trace Data. bioRxiv 2018;251082.

34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–1760.

35. Ye K, Schulz MH, Long Q, et al. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2009;25:2865–2871.

36. Hyde DR, Godwin AR, Thummel R. In vivo Electroporation of Morpholinos into the Regenerating Adult Zebrafish Tail Fin. J Vis Exp . Epub ahead of print 2012. DOI: 10.3791/3632.

37. Thummel R, Bai S, Sarras MP, et al. Inhibition of zebrafish fin regeneration using in vivo electroporation of morpholinos against fgfr1 and msxb. Dev Dyn 2006;235:336–346.

38. Chao-Shern C, Me R, Dedionisio LA, et al. Post-LASIK exacerbation of granular corneal dystrophy type 2 in members of a Chinese family. Eye 2018;32:39–43.

39. Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 2013;31:827–832.

40. Zhang Y, Ge X, Yang F, et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Sci Rep;4 . Epub ahead of print 2014. DOI: 10.1038/srep05405.

41. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol . Epub ahead of print 2018. DOI: 10.1038/nbt.4192.

42. Richardson CD, Ray GJ, Bray NL, et al. Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. Nat Commun 2016;7:12463.

43. Moore CBT, Christie KA, Marshall J, et al. Personalised genome editing - The future for corneal dystrophies. *Progress in Retinal and Eye Research*, 2018 . Epub ahead of print 2018. DOI: 10.1016/j.preteyeres.2018.01.004.

44. Page L, Polok B, Bustamante M, et al. Bigh3 is upregulated in regenerating zebrafish fin. Zebrafish 2013;10:36–42.

45. Schebesta M, Lien CL, Engel FB, et al. Transcriptional profiling of caudal fin regeneration in zebrafish. ScientificWorldJournal 2006;6:38–54.

46. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 2013;31:822–826.

47. Cho SW, Kim S, Kim Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res 2014;24:132–141.

48. Josephs EA, Kocak DD, Fitzgibbon CJ, et al. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. Nucleic Acids Res 2015;43:8924–8941.

49. Anders C, Niewoehner O, Duerst A, et al. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 2014;513:569–573.

50. Kleinstiver BP, Prew MS, Tsai SQ, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 2015;523:481–485.

51. Tsai SQ, Zheng Z, Nguyen NT, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol 2014;33:187–197.

52. Zetsche B, Gootenberg JS, Abudayyeh OO, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell 2015;163:759–771.

53. Gao L, Cox DBT, Yan WX, et al. Engineered Cpf1 variants with altered PAM specificities. Nat Biotechnol 2017;35:789–792.

54. Kleinstiver BP, Prew MS, Tsai SQ, et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol 2015;33:1293–1298.

55. Kleinstiver BP, Pattanayak V, Prew MS, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature 2016;529:490–495.

56. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends in Biochemical Sciences* 1998;23:198–199.

57. Popp MW, Maquat LE. Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. *Cell* 2016;165:1319–1332.

58. Munier FL, Frueh BE, Othenin-Girard P, et al. BIGH3 mutation spectrum in corneal dystrophies. Investig Ophthalmol Vis Sci 2002;43:949–954.

59. Monteys AM, Ebanks SA, Keiser MS, et al. CRISPR/Cas9 Editing of the Mutant Huntingtin Allele In Vitro and In Vivo. Mol Ther 2017;25:12–23.

60. Slaymaker IM, Gao L, Zetsche B, et al. Rationally engineered Cas9 nucleases with improved specificity. Science (80-) 2016;351:84–88.

61. Chen JS, Dagdas YS, Kleinstiver BP, et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. Nature 2017;550:407–410.

62. Anderson KR, Haeussler M, Watanabe C, et al. CRISPR off-target analysis in genetically engineered rats and mice. *Nature Methods*, 2018, pp. 1–3.

63. Zuris JA, Thompson DB, Shu Y, et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat Biotechnol 2014;33:73–80.

64. Lee K, Conboy M, Park HM, et al. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. Nat Biomed Eng . Epub ahead of print 2017. DOI: 10.1038/s41551-017-0137-2.

65. Yin H, Song C-Q, Dorkin JR, et al. Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. Nat Biotechnol 2016;34:328–333.

66. D'Astolfo DS, Pagliero RJ, Pras A, et al. Efficient intracellular delivery of native proteins. Cell 2015;161:674–690.

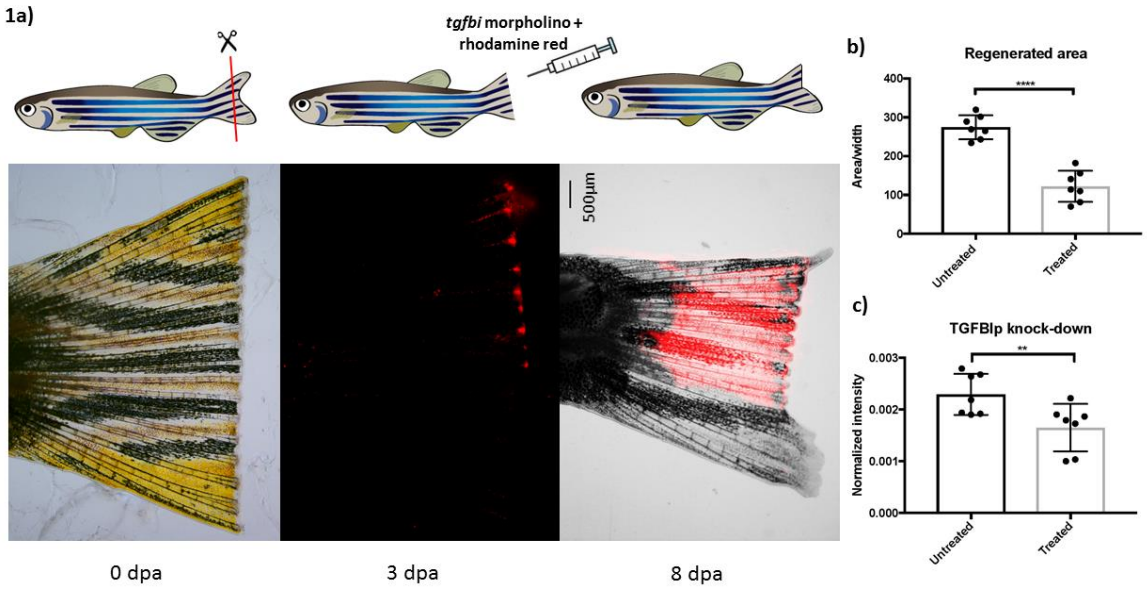


Figure 1: Effect of Tgfbi knockdown on zebrafish tail-fin regeneration a) At day 0 the tail fin was amputated perpendicular to the anterior/posterior plane of the fish, 3 dpa *tgfbi* morpholino mixed with rhodmaine red was injected distal to each bone array on the dorsal side of the fin. 5 dpa regenerated fins were imaged and regeneration of fins deficient in Tgfbi were compared to untreated fins **b)** The area of the treated and untreated fins were calculated and normalised to the width of the fin, regeneration in the Tgfbi deficient fin was significantly impaired in comparison to the untreated fin (p=****), indicating TGFBI has a critical role in wound repair **c)** LC-MS/MS was used to confirm a reduction of Tgfbi in the regenerating fins, Tgfbi was significantly reduced in the treated fins compared to the untreated fins (p=**).

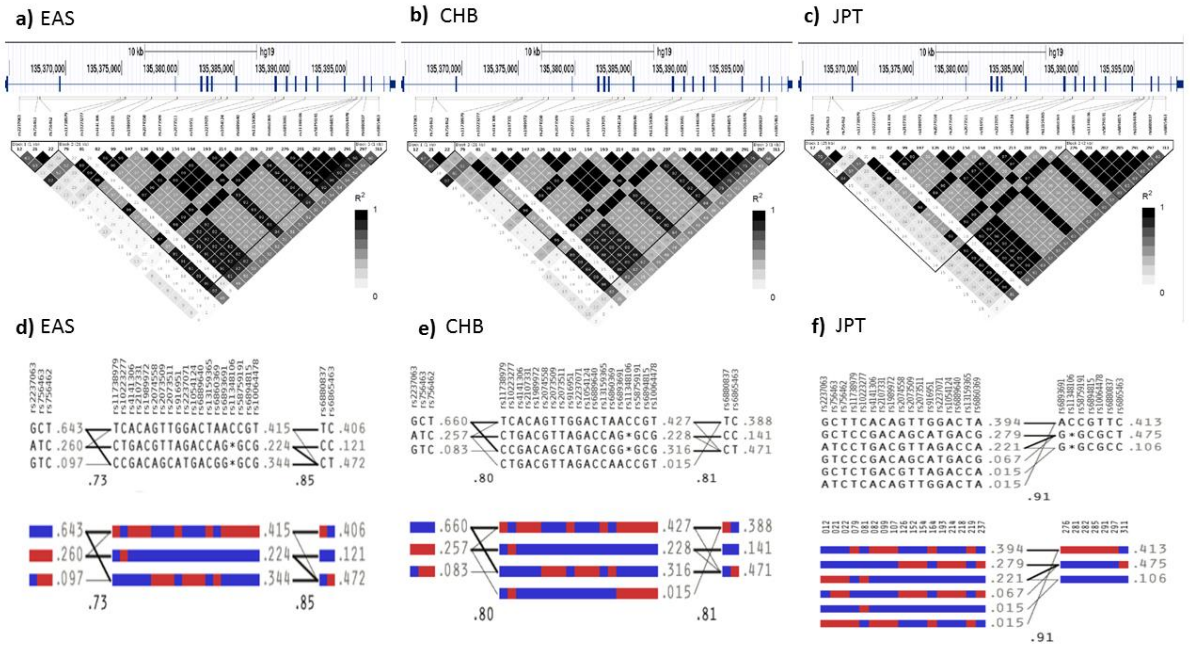


Figure 2: Linkage disequilibrium (LD) of the identified ASNIP SNPs in the coding region of the *TGFBI* gene – plots were generated using the 1000 Genomes Project phase 3 data for East Asian (EAS), Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT) populations, black indicates very strong LD, while a white indicates no LD a) LD plot showing the association between the 24 SNPs containing a PAM on only allele identified across the *TGFBI* locus for the EAS population; 3 LD blocks were found, the first block (EAS-B1) spans 1kb within intron 1-2 (rs2237063- rs756462), the second block (EAS-B2) spans 21kb from intron 2-3 (rs11738979) to intron 13-14 (rs10064478) and finally the third block (EAS-B3) spans 1kb from intron 14-15 (rs6880837) to intron 15-16 (rs6865463) b) LD plot showing the association between the 24 SNPs containing a PAM on only allele identified across the *TGFBI* locus for the CHB population; 3 LD blocks were identified, ; the first block (CHB-B1) spans 1kb within intron 1-2 (rs2237063- rs756462), the second block (CHB-B2) spans 21kb from intron 2-3 (rs11738979) to intron 13-14 (rs10064478) and finally the third block (CHB-B3) spans 1kb from intron 14-15 (rs6880837) to intron 15-16 (rs6865463) c) LD plot showing the association

between the 24 SNPs containing a PAM on only allele identified across the *TGFBI* locus for the JPT population, 2 LD blocks were found in the 1000 Genomes JPT population; the first block spans 25kb from intron 1-2 (rs2237063) to intron 10-11 (rs6860369) and the second block spans 2kb from intron 13-14 (rs6880837) to intron 15-16 (rs6865463) d,e,f) Haplotype frequencies of the identified ASNIP SNPs in the *TGFBI* gene in the d) EAS e) CHB and f) JPT populations. The blue indicates the major allele and red indicates the minor allele, numbers next to each haplotype bar are haplotype frequencies, in the crossing areas a value of multiallelic D' is shown to represent the level of recombination between the two blocks.

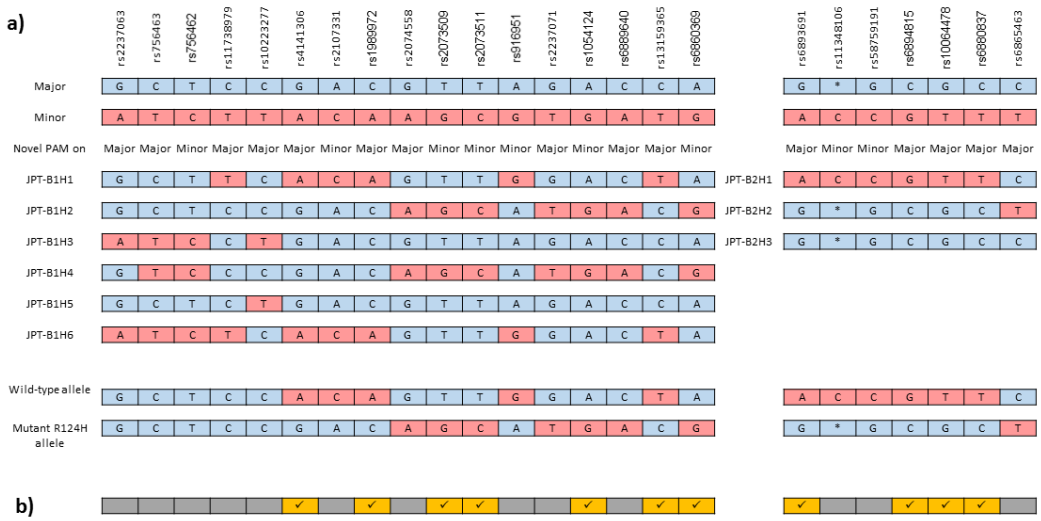


Figure 3: Haplotype analysis of R124H Japanese patient a) Phased sequencing, of a Japanese corneal dystrophy patient harbouring a R124H mutation, revealed the patients haplotype blocks; comparison to our haplotype analysis of the *TGFBI* locus in the Japanese population revealed the patient had JPT-B1H1 which co-segregated with JPT-B2H1 and JPT-B1H2 which co-segregated with JPT-B2H2, the patient differed at one position (rs11738979) in JPT-B1H1 as the patient was homozygous for the major allele. Blue indicates the major allele and red indicates the minor allele. b) The determination of the R124H patients haploblocks by phased sequencing allowed the identification of SNPs that contain a PAM on only the allele associated with the R124H mutation. Yellow shading and ticks indicate a combination of haplotypes that generate a heterozygote at this position, offering only one PAM-generating allele, therefore providing a potential SNP for ASNIP CRISPR, grey indicates that either there is no PAM-associated allele present on either haplotype or each haplotype has the same PAM-associated allele at this position.

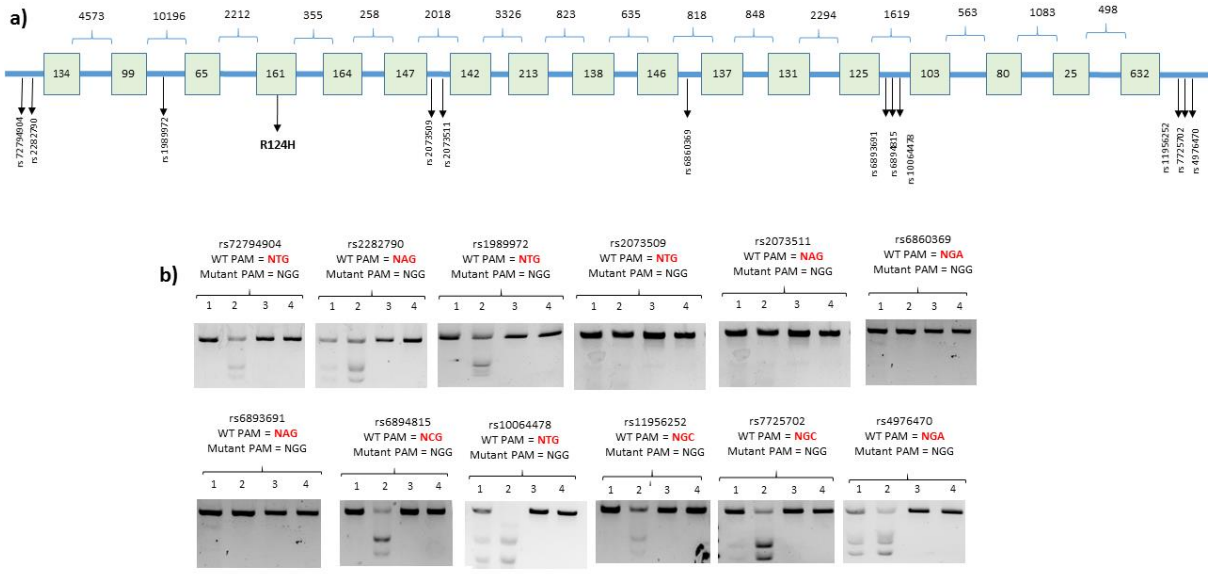
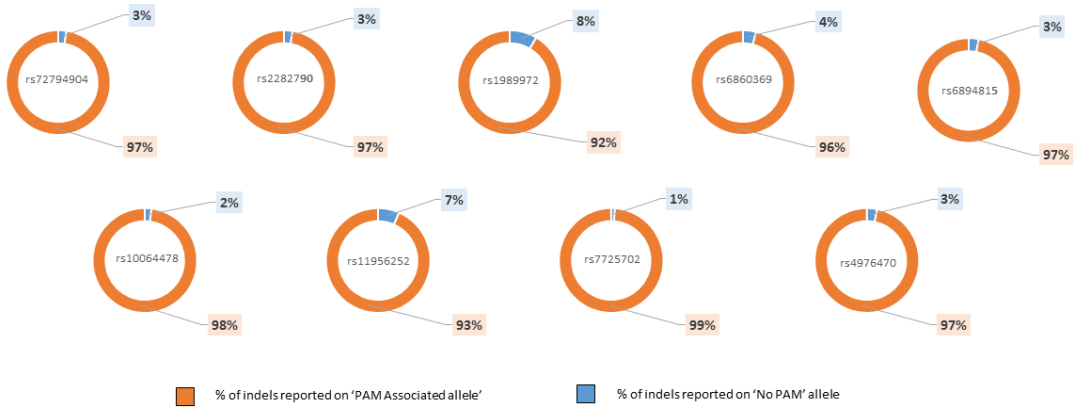


Figure 4: *In vitro* assessment of ASNIP guides allele-specificity a) Based on the phased sequencing data 12 ASNIP guides were designed that i) are associated with a PAM only on one allele ii) lie in cis with the R124H mutation and iii) have good scores using the *in silico* tools. Coding exons in *TGFBI* are shown by the green boxes, length in base pairs of each exon and intronic regions are indicated across the region. Location of the 12 ASNIP guides and the R124H mutation are depicted by black drop-down arrows b) Initially *in vitro* digests were used to determine the allele-specificity of the 12 ASNIP guides. RNP complexes were incubated with templates containing 'No-PAM allele' or 'PAM-associated allele' sequences for the respective SNPs, for each digest lane 1 = 'No PAM' digested, lane 2 = 'PAM-associated' digested, lane 3 = 'No PAM' undigested, lane 4 = 'PAM associated' undigested. Undigested bands in the treated samples were normalised to undigested samples and were quantified to determine cleavage efficiency. SNPs that have a non-canonical PAM (NAG/NGA) on the wild-type allele appear to be less specific than SNPs that read NGC/NCG/NGT/NTG on the wild-type allele.

a)



b)

SNP	Alleles	Most frequent indels detected (SNP in bold, PAM underlined, indels in red)	Type of indel	Guide Sequence
rs72794904	G,T,	CACAGGTCAGGATCTATACCATGT-GGCTGGGCCTGTTTCCTCCAAGAGTTTCCAGATCGT	1bp deletion	GGATCTATACCATGTGGGCT
rs2282790	A,G,	CCCCAGGGCTAGCAGTGCCAAGTAACCTGACGGGTGATTAATAGATGCTTGGGTAAGTATCA	1bp insertion	TAGCAGTGCCAAGTAACTGA
rs1989972	A,C,	GTTTGGGGGGACTTGACCTAATCCCACATCCCAGCCCCAGTAATACAGCCCTGGAATTTAT	2bp insertion	AGGGCTGTATTACTGGGGCT
rs6860369	A,G,	GCTCTGAACAAATCAGGAGGCCCCTCGTGGGAGTATAACCAGTCCTTTCTTTCTCTGTCCC	1bp insertion	CAAATCAGGAGGCCCCTCGT
rs6894815	C,G,	TAGCTTAAAAGAGGCACTTCCCCTGTCTTCCATTTCCAGCTGTTCCACCTACTTAGCTTAAAAGAGGCACTTCCCCTGTCTTCAGTCTCAGT	53bp insertion	GAGACTGAGACTGAAGACAG
rs10064478	G,T,	GGTGGCACATGCCTGTAATCACAGCTTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACC	1bp insertion	TGCCTGTAATCACAGCTACT
rs11956252	C,G,	GCCCTAACAGTCCCTGCAGTGCTGGAGCTCCCCATCCAACTTGGGGAGGCGATGGAGGTCACT	2bp insertion	CATCGCCTCCCCAAGTGATG
rs7725702	C,G,	GACTCCCTAACTGAGAAAGGTCACCCCCCTGGGTCAGCAGGCCCCAACTCAGAGAGAACATCT	2bp insertion	AACTGAGAAAGGTCACCCCT
rs4976470	A,G,	GGCTTCCTCCCGTGACATGTGGGGATTTATGGTAACTACAATTCAAGATGAGATTTGGGTGG	1bp insertion	CCCGTGACATGTGGGGATTA

Figure 5: Allele-specificity of ASNIP guides tested in a patient-derived cell line a) LCLs were transfected with RNPs for each of the 12 ASNIP guides. Targeted resequencing across the on-target cut site was used to determine the allele-specificity of each guide. Orange indicates % of indels that occurred on the 'PAM-associated' allele and blue indicated % of indels that occurred on the 'No PAM' allele. On average 96.3% of indels were observed on the 'PAM associated' allele with only 3.7% of indels occurring on the 'No PAM' allele. b) The most frequent indels detected for each ASNIP guide, the SNP is shown in bold, the PAM is underlined and the indel is depicted in red.

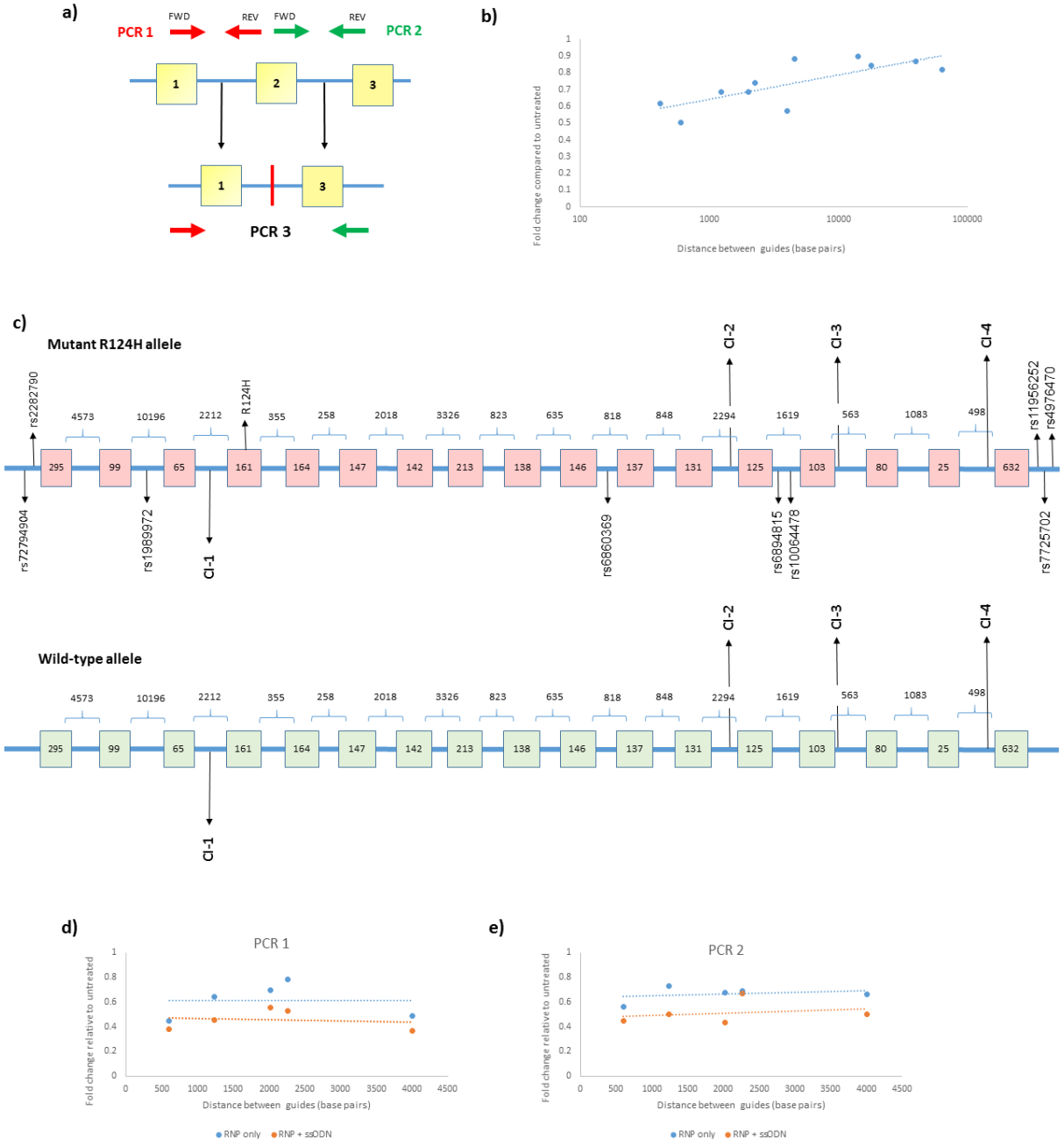
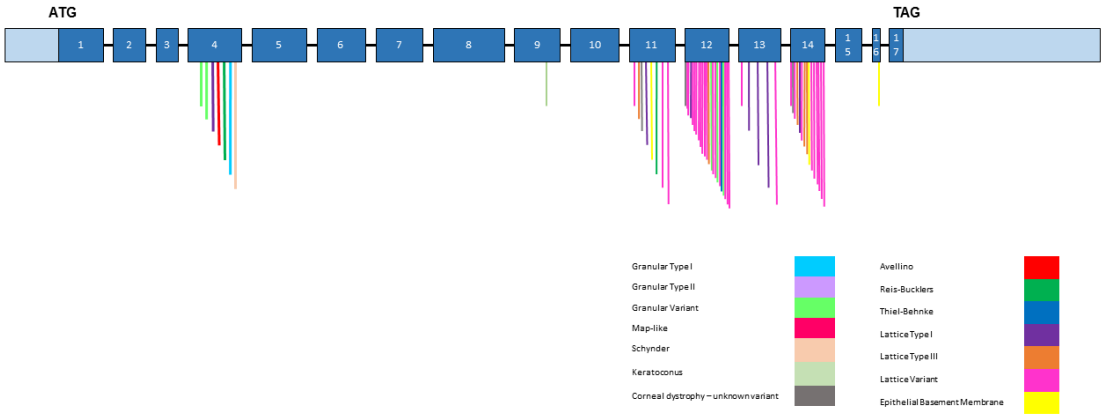


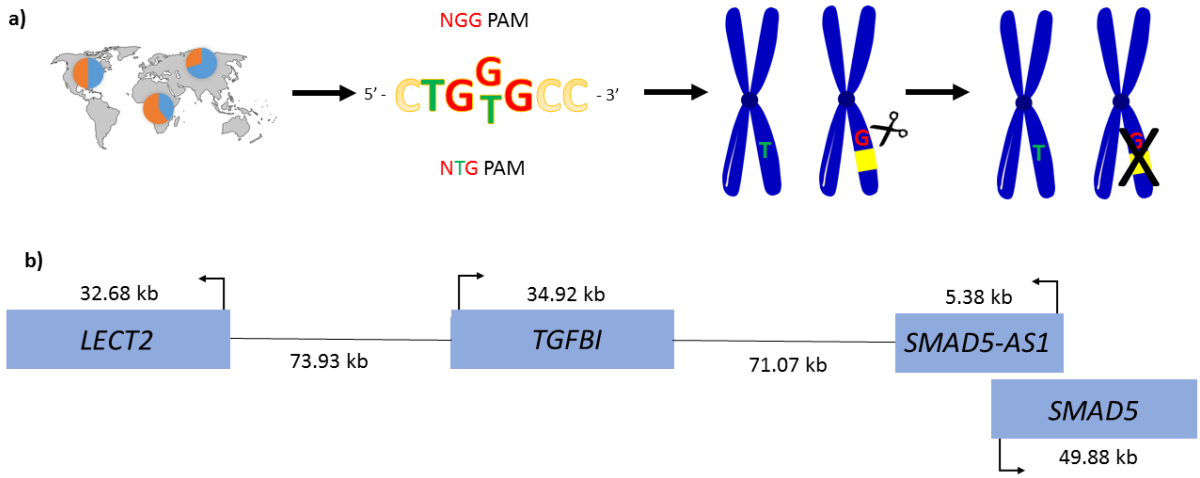
Figure 6: Allele-specific dual-guide approach targeted to *TGFBI* a) Schematic to show quantification of dual-cut – dual-guides are shown by the black arrows, PCR amplification across target site 1 is denoted as PCR 1 and shown by the red arrows, PCR amplification across target site 2 is denoted as PCR 2 and shown by the green arrows. If both guides cut in the same cell and intervening region is excised then the binding sites for PCR 1 REV and PCR 2 FWD is removed, thus the abundance of PCR 1 and PCR 2 compared to untreated cells should be

reduced if a frequent productive edit has occurred b) Fold change for the average of PCR 1 and PCR 2 compared to untreated is shown for dual combinations ranging in size from 419bp to 63,428bp. The efficiency of the deletion reduces as the distance between the guides increases, indicating the closer the dual-guides are the more frequent the productive edit is c) Schematic showing the predicted cleavage of each guide utilised. As the ASNIP guides are substantial distances apart common-intronic guides are designed to overcome this limitation, the common-intronic guides are shown in bold and are predicted to cleave both alleles, while the ASNIP guides should only cleave the mutant allele d,e) The addition of a ssODN containing 50bp of each flanking region in the dual-guide combination was shown to increase the frequency of a productive edit. The fold change of PCR 1 (d) and PCR 2 (e) in cells transfected with RNP only is shown in blue and of fold change of PCR 1 and PCR 2 in cells transfected with RNP + ssODN is shown in orange. The addition of a 50:50 ssODN appears to increase the frequency of dual-cut events ($p<0.01$).

4.3 Supplementary data

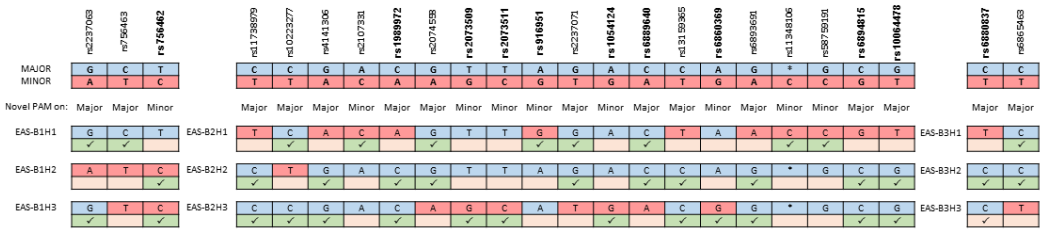


Supplementary Figure 1: Mutation hotspots across *TGFBI* – The *TGFBI* gene, including untranslated regions (UTRs) (shown by light blue shading) and introns (shown by black interlinking lines), covers ~35kb, and there are 17 coding exons (shown by dark blue shading). To-date 62 missense mutations within *TGFBI* have been associated with corneal dystrophies, each mutation is depicted by a single drop-down line and the colours correspond to the dystrophy the mutation is associated with, described in the colour coded key. These missense mutations are found in exons 4 to 16 of the gene; however, the majority of mutations are clustered in hot-spots in exons 4, 11, 12 and 14 of the gene.

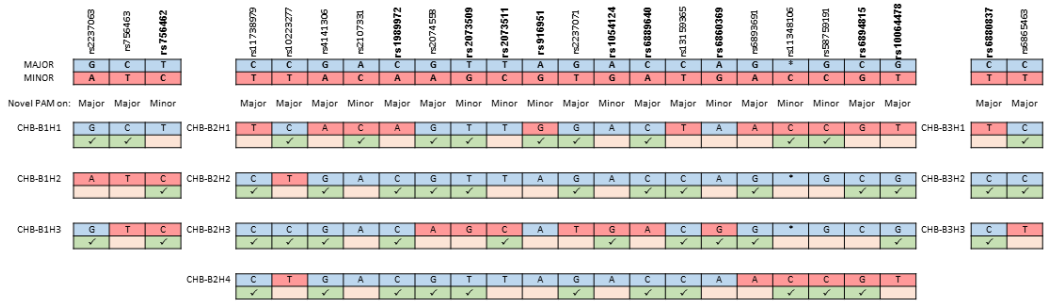


Supplementary Figure 2: ASNIP design based on the *TGFBI* locus a) Mutational workflow for ASNIP CRISPR. Non-disease causing mutations within the *TGFBI* locus with a minor allele frequency (MAF) of >0.1 are identified, these SNPs are then analysed to determine if they generate a novel *S.pyogenes* PAM (NGG), guides are designed and prospective guides are ran through *in silico* design programs to determine the most promising pool of guides. **b)** Genomic location of the *TGFBI* locus. There are no flanking genes within a >70kb region either up or downstream of *TGFBI*, indicating gene editing events are not likely to disrupt neighbouring genes when considering a linear genomic DNA context.

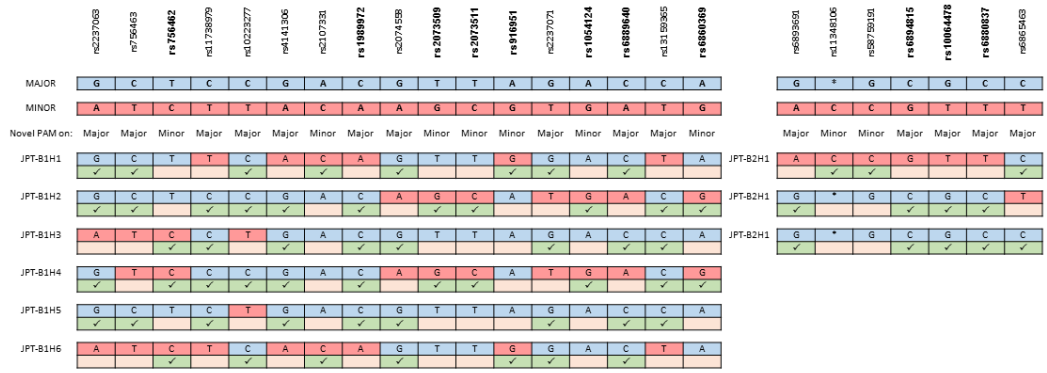
a) EAS



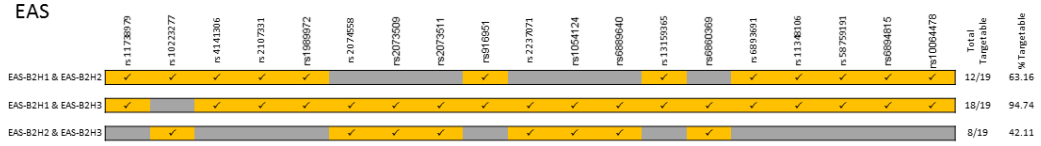
b) CHB



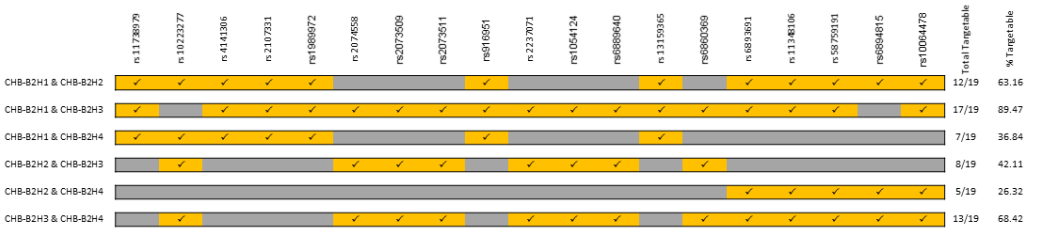
c) JPT

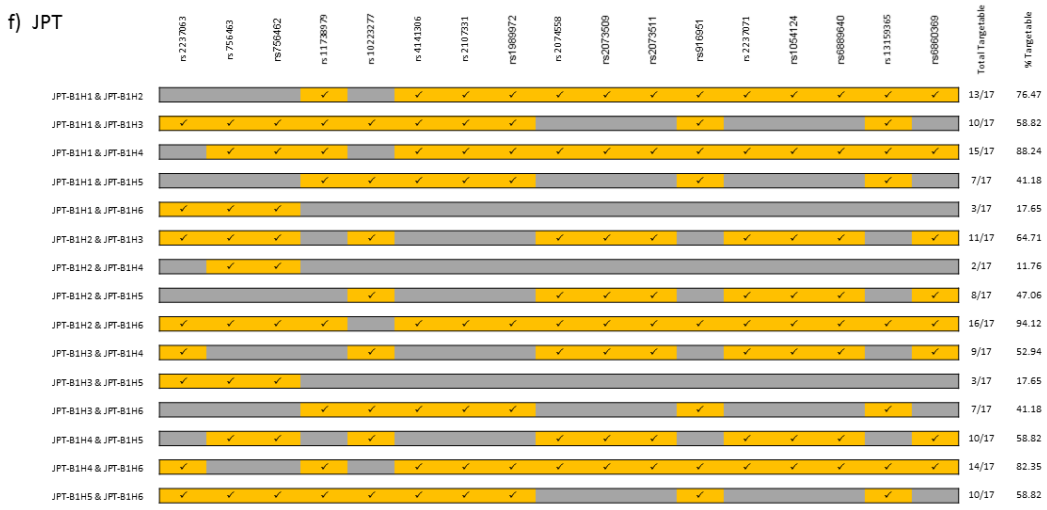


d) EAS

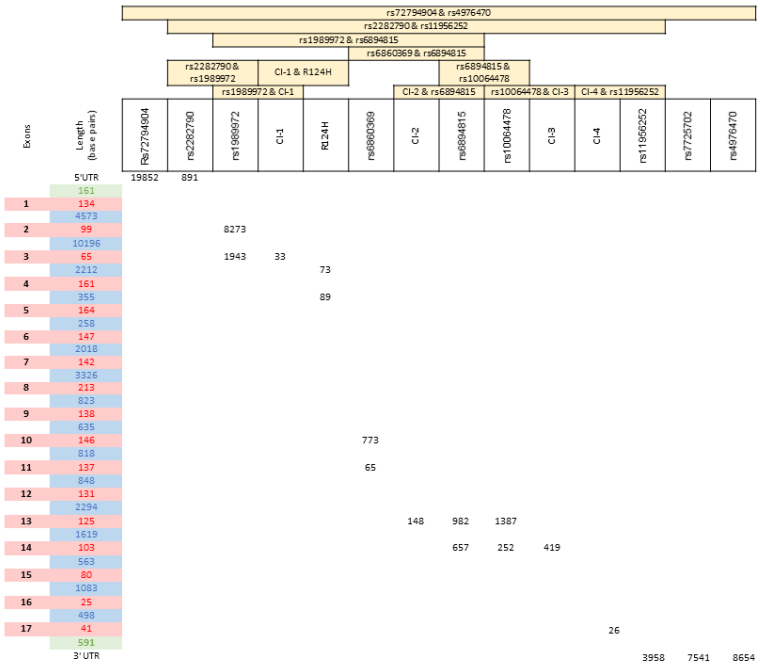


e) CHB

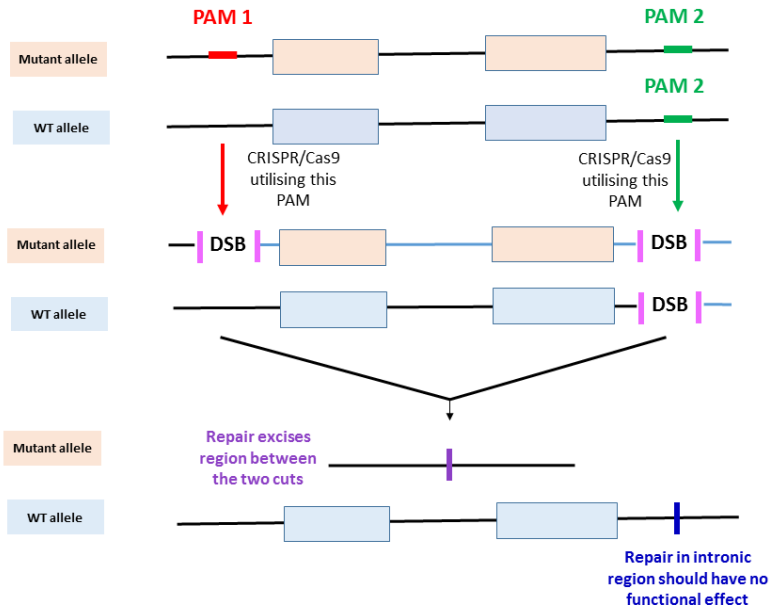




Supplementary Figure 3: Haplotype analysis to determine proportion of population targetable by ASNIP CRISPR a,b,c): SNPs across each haplotype in the **a)** EAS population **b)** CHB population and **c)** JPT population were assessed to determine which have the PAM generating allele and which are not targetable. Blue indicates the major allele and red indicates the minor allele, green indicates a PAM-generating allele is present while orange indicates it is not targetable at this position **d,e,f)** All possible heterozygous haplotype combinations for the **d)** EAS population **e)** CHB population and **f)** JPT population were assessed using only the large haploblock (EAS-B2, CHB-B2 and JPT-B1) to determine the total % of each haplotype that has targetable SNPs. Yellow shading and ticks indicate a combination of haplotypes that generate a heterozygote at this position, offering only one PAM-generating allele, therefore providing a potential SNP for ASNIP CRISPR, grey indicates that either there is no PAM-associated allele present on either haplotype or each haplotype has the same PAM-associated allele at this position.



Supplementary Figure 4: Schematic to shown the different dual combinations used and their location relative to the exon/intron regions in *TGFBI*. Exons are shown in red, introns are shown in blue, each guide is shown vertically along the top of the figure, each dual combination is shown by a yellow box spanning the region between the two guides, the drop-down numbers from each guide indicate the distance of the guide relative to the closet exon or intron.



Supplementary Figure 5: The greater the distance between the dual-guides the less frequent the productive edit will be, the distance between the original ASNIP guides is substantial therefore to overcome this limitation common-intronic guides were designed. The 'No PAM' allele is shown in blue and the 'PAM associated' allele is shown in orange. The ASNIP guide (shown in red) is predicted to only cut the 'PAM associated' allele while the common-intronic guide (shown in green) should cut both alleles; on the 'PAM associated' allele when both cuts are made in a cell the region between these cuts may be excised resulting in a productive edit (shown in purple), while on the 'No PAM' allele a cut should only occur with the common-intronic guide which should have no functional effect (shown in dark blue).

Supplementary table 1: Scheduled inclusion list for PRM method

Protein	Peptide sequence	Mass (m/z)	CS* (z)	Start (min)	End (min)
TGFBIp	TQGPNV**C**AMQK	617,2894	2	10,4	20,4
TGFBIp	I**C**GKPTVISYE**CC**PGYEK	720,9969	3	17,5	27,5
TGFBIp	G**C**PAALPLVNIYK	708,3894	2	31,4	41,4
TGFBIp	VITAITNDVNSIIDTDDDLDTLR	844,7607	3	40,4	50,4
TGFBIp	TLLELAEGSVVTTAAK	534,9698	3	33,6	43,6
TGFBIp	DAGLNDHLVGSESVTLLAPLNEAFKDK	714,1212	4	38,1	48,1
TGFBIp	SLYHGQELETLGGLK [a]	548,9579	3	23,8	33,8
TGFBIp	YANMFLVDSILTPPQGTVMDVLK	851,4428	3	48,2	58,2
TGFBIp	FSTLVGAIQK	532,3109	2	22,3	32,3
B-actin	VAPEEHPVLLTEAPLNPK	652,0263	3	23,8	33,8
B-actin	L**C**YVALDFEQEMGTAASSSSLEK	846,0576	3	38	48
B-actin	DLYANTVLSGGTTMYPGIADR	739,0282	3	34,5	44,5

*CS: Charge state, [a] not included in quantitative analysis

Supplementary Table 2: List of oligo nucleotides used

Oligo Name	Oligo Sequence (5' - 3')
Cleavage template FWD	ACCCCAACATCTTCGACGCGGGC
Cleavage template REV	TGCTGTCCTGCCCCACCCCA
rs72794904 956bp FWD	GGCAGTGTATTTCTTTCAGAGGA
rs72794904 956bp REV	GAGCCGAGATCATGCCACT
rs72794904 238bp FWD	CCAAGTGCCAGTCAATCCTG
rs72794904 238bp REV	TGCAAGAGAGGACATCAATTTGA
rs2282790 748bp FWD	GGCCTCAGAGCAGGTATCAC
rs2282790 748bp REV	TAGGTCCCTTAGGCCTCCTG
rs2282790 240bp FWD	TGGGCTACGGATCTTCCCAA
rs2282790 240bp REV	CATCTCTGCAACAGTACCTGC
rs1989972 708bp FWD	GTTCAGCTCCCTTGCGGTAT
rs1989972 708bp REV	CAGGCTATTGTCTTGGGACTCA
rs1989972 249bp FWD	GCCCTGACATGAGGACTTTGA
rs1989972 249bp REV	CCAGCTAAATCCAGGGAGAGC
rs6860369 762bp FWD	GGGGCCTCTCTAACCGTTCT
rs6860369 762bp REV	GCCGGGCAAGAAAACAAACT
rs6860369 215bp FWD	TCCCAGCCTTAATAACCCATCC
rs6860369 215bp REV	GGTCCATCGTGAACAGGGTC
rs6894815 797bp FWD	ATAGATTTGCCCTGGGTGGG
rs6894815 797bp REV	AAGAAAAACAGAGTAGTGGTTGAAA
rs6894815 241bp FWD	GGCCTGAGATAGATTTGCCC
rs6894815 241bp REV	CTCAGTCCTCACAGCAGTGTAT
rs10064478 961bp FWD	TCCCCAGTCTAACACAGGAC
rs10064478 961bp REV	GAGGCAGGACTGAGGTTCAA
rs10064478 150bp FWD	AAAATTAGCTGGGCGTGGTG
rs10064478 150bp REV	TGGAGTTTCAATCTTGTCGCC
rs11956252 741bp FWD	AGCCAGGAGAGAAAGTCATGG
rs11956252 741bp REV	TCCCCCAACTAAAACCCTCC
rs11956252 210bp FWD	CACCCACTTGTGGTTGGGGA
rs11956252 210bp REV	CCCCACCCTCTTCATTCTTCAG
rs7725702 702bp FWD	GGCTCCTTCAGTCAACAAGGT
rs7725702 702bp REV	TCCCTCACCCTCCGATTCTG
rs7725702 247bp FWD	TCTTCTCAGGAAAGCAGGGTG
rs7725702 247bp REV	CTCCCCAGAAGGGTTAGAGG
rs4976470 829bp FWD	ATGTAGCCTCAAATCCCAGCC
rs4976470 829bp REV	GCACACCTGACTATGGCTCT
rs4976470 168bp FWD	GCAACAGATCAAGTGACACCT
rs4976470 168bp REV	GGGGCTTGATATGGTTTGGC
R124H 988bp FWD	TGAGTTCACGTAGACAGGCA
R124H 988bp REV	ACAGCTTAAACCCCAGAAACCA
R124H 187bp FWD	CCTTTACGAGACCCTGGGAG

R124H 187bp REV	GTTCCCCATAAGAGTCCCCC
CI-1 703bp FWD	CCAGTTGGTTGGCTGTAGGT
CI-1 703bp REV	ATCCCATCGGCTCTCTAGCA
CI-1 73bp FWD	TCCAGCAGGTGAATGAATCC
CI-1 73bp REV	TACTCCTCTCTCCCACCATTCC
CI-2 925bp FWD	CTGGAAAGGTCCCTGGCTTT
CI-2 925bp REV	GGCTCACAGAGCAAGTGTCA
CI-2 117bp FWD	TGCTTTGTGTCCTCTGACCAT
CI-2 117bp REV	AGTGGTCACCCCTGAAATGAA
CI-3 736bp FWD	GTTGCCGAGCCTGACATCAT
CI-3 736bp REV	CGCAAACCTAGCAGGCATCT
CI-3 173bp FWD	GACACATTGCTCTTTGCGGA
CI-3 173bp REV	GAGAGGCAGGACTGAGGTTC
CI-4 818bp FWD	TCAGAACAGCAGGGTGACTTG
CI-4 818bp REV	CCAGCTGTGCAAGGGCTTTA
CI-4 253bp FWD	AGAAAACCAGAACATCGGGC
CI-4 253bp REV	TGGTGCATTCCTCCTGTAGTG

chromStart	chromEnd	SNP	Alleles	Allele Frequencies	Strand
135314845	135314846	rs72793185	C,T,	0.855431,0.144569,	+
135314876	135314880	rs373839451	-,TAAG,	0.143970,0.856030,	+
135314879	135314880	rs2346012	A,G,	0.706270,0.293730,	+
135315167	135315168	rs17169582	A,G,	0.855631,0.144369,	+
135316486	135316487	rs2158351	G,T,	0.160942,0.839058,	+
135318158	135318158	rs559931571	-,T,	0.765974,0.234026,	+
135319760	135319761	rs257480	A,T,	0.661542,0.338458,	+
135322719	135322720	rs6868908	A,G,	0.214657,0.785343,	+
135323867	135323868	rs6874348	A,G,	0.665335,0.334665,	+
135327490	135327491	rs1859295	C,T,	0.502596,0.497404,	+
135327678	135327679	rs17688533	A,G,	0.728634,0.271366,	+
135328361	135328362	rs10076250	A,G,	0.647364,0.352636,	+
135329914	135329915	rs12520800	G,T,	0.387181,0.612819,	+
135330053	135330054	rs17739831	C,N,T,	0.367472,0.000767,0.631761,	+
135331391	135331391	rs113921691	-,T,	0.171126,0.828874,	+
135333390	135333391	rs6881712	A,T,	0.326278,0.673722,	+
135333505	135333506	rs12332587	A,G,	0.118610,0.881390,	+
135334846	135334847	rs10074474	A,G,	0.143770,0.856230,	+
135334927	135334928	rs10074539	A,G,	0.143770,0.856230,	+
135335228	135335229	rs10079806	A,C,	0.852436,0.147564,	+
135335396	135335396	rs10522532	-,GTGT,	0.622204,0.377796,	+
135335578	135335579	rs6882087	A,G,	0.898363,0.101637,	+
135335676	135335677	rs11747904	A,T,	0.408946,0.591054,	+
135336560	135336561	rs13157444	A,G,	0.459864,0.540136,	+
135337231	135337232	rs57104529	C,G,	0.288738,0.711262,	+
135338598	135338599	rs916950	C,T,	0.260383,0.739617,	+
135339463	135339464	rs17740150	C,G,	0.167931,0.832069,	+
135342086	135342087	rs2525490	A,G,	0.577276,0.422724,	+
135343546	135343547	rs9327738	C,T,	0.160144,0.839856,	+
135344162	135344163	rs6892697	A,G,	0.501398,0.498602,	+
135345183	135345184	rs72794904	G,T,	0.186302,0.813698,	+
135345816	135345817	rs9327739	C,T,	0.152157,0.847843,	+
135351182	135351183	rs72794907	A,G,	0.184704,0.815296,	+
135352328	135352329	rs7728408	A,T,	0.229233,0.770767,	+
135354323	135354324	rs4976360	A,T,	0.336262,0.663738,	+
135355037	135355038	rs6894906	A,G,	0.319289,0.680711,	+
135355436	135355437	rs6895177	A,G,	0.148789,0.851211,	+
135357723	135357724	rs146020713	-,T,	0.323482,0.676518,	+
135360737	135360738	rs35901765	C,T,	0.699081,0.300919,	+
135361140	135361141	rs35636600	A,C,	0.246206,0.753794,	+
135362549	135362550	rs34098140	C,T,	0.636781,0.363219,	+
135362572	135362573	rs4976459	C,G,	0.658946,0.341054,	+
135362681	135362682	rs10463536	C,T,	0.601038,0.398962,	+
135362716	135362716	rs111308112	-,CATT,	0.403954,0.596046,	+
135362719	135362720	rs55821461	C,T,	0.409145,0.590855,	+
135363874	135363875	rs2282790	A,G,	0.365415,0.634585,	+
135364189	135364190	rs17169707	C,T,	0.189297,0.810703,	+
135366135	135366136	rs2237063	A,G,	0.225240,0.774760,	+

135367219	135367220	rs2237065	A,G,	0.303115,0.696885,	+
135367602	135367603	rs2237066	C,T,	0.659145,0.340855,	+
135367756	135367757	rs756463	C,T,	0.666134,0.333866,	+
135367944	135367945	rs756462	C,T,	0.309505,0.690495,	+
135374314	135374315	rs10053962	C,T,	0.771166,0.228834,	+
135375041	135375042	rs11738979	C,T,	0.610024,0.389976,	+
135375330	135375331	rs10223277	C,T,	0.794728,0.205272,	+
135375472	135375473	rs4141306	A,G,	0.425519,0.574481,	+
135375596	135375597	rs739866	A,G,	0.225839,0.774161,	+
135375604	135375605	rs739867	A,G,	0.281550,0.718450,	+
135377348	135377349	rs2107331	A,C,	0.549121,0.450879,	+
135377565	135377566	rs7719624	C,T,	0.449880,0.550120,	+
135377729	135377730	rs2282791	G,T,	0.516973,0.483027,	+
135377801	135377802	rs1989972	A,C,	0.425919,0.574081,	+
135378238	135378239	rs1989973	C,G,	0.889577,0.110423,	+
135378363	135378363	rs540142018	-,A,	0.693291,0.306709,	+
135380058	135380059	rs2074558	A,G,	0.373203,0.626797,	+
135380763	135380764	rs6897320	C,T,	0.609625,0.390375,	+
135383356	135383357	rs2073508	A,G,	0.204673,0.795327,	+
135383376	135383377	rs2073509	G,T,	0.404952,0.595048,	+
135383429	135383430	rs2073510	A,G,	0.630591,0.369409,	+
135383892	135383893	rs2073511	C,T,	0.404952,0.595048,	+
135384080	135384081	rs45554435	A,G,	0.335064,0.664936,	+
135384442	135384443	rs916951	A,G,	0.634784,0.365216,	+
135384844	135384845	rs6596281	A,T,	0.406550,0.593450,	+
135385315	135385316	rs17169753	C,T,	0.406550,0.593450,	+
135385699	135385700	rs1060433	C,T,	0.406550,0.593450,	+
135385777	135385778	rs1137550	C,T,	0.368810,0.631190,	+
135386023	135386024	rs10706409	-,A,	0.679113,0.320887,	+
135386729	135386730	rs2237070	A,G,	0.647564,0.352436,	+
135386752	135386753	rs2237071	G,T,	0.664537,0.335463,	+
135386799	135386800	rs2237072	C,T,	0.664736,0.335264,	+
135387802	135387803	rs17169768	A,G,	0.613019,0.386981,	+
135388662	135388663	rs1054124	A,G,	0.612819,0.387181,	+
135389424	135389425	rs6889640	A,C,	0.369808,0.630192,	+
135389432	135389433	rs13159365	C,T,	0.637181,0.362819,	+
135391325	135391326	rs6860369	A,G,	0.595847,0.404153,	+
135391373	135391374	rs1133170	C,T,	0.719848,0.280152,	+
135392425	135392426	rs4669	C,T,	0.420527,0.579473,	+
135392734	135392735	rs7727725	A,T,	0.420327,0.579673,	+
135393137	135393138	rs17689879	C,T,	0.637580,0.362420,	+
135393196	135393197	rs6871571	A,G,	0.579473,0.420527,	+
135395432	135395433	rs6893691	A,G,	0.385982,0.614018,	+
135395625	135395626	rs10036667	C,T,	0.794529,0.205471,	+
135395825	135395826	rs11348106	-,C,	0.577276,0.422724,	+
135395826	135395827	rs58759191	C,G,	0.422524,0.577476,	+
135395863	135395864	rs6894815	C,G,	0.577276,0.422724,	+
135396083	135396084	rs10042825	A,T,	0.422324,0.577676,	+
135396291	135396292	rs10064478	G,T,	0.577875,0.422125,	+
135396451	135396452	rs13168506	A,G,	0.404353,0.595647,	+

135396467	135396468	rs13188659	A,T,	0.593650,0.406350,	+
135396668	135396669	rs6880837	C,T,	0.591054,0.408946,	+
135397701	135397702	rs6886556	C,T,	0.438698,0.561302,	+
135397784	135397785	rs6865463	C,T,	0.597843,0.402157,	+
135400034	135400035	rs13189180	A,G,	0.628395,0.371605,	+
135400056	135400057	rs10043360	A,T,	0.156949,0.843051,	+
135400380	135400381	rs45543842	A,G,	0.369808,0.630192,	+
135401118	135401119	rs59239478	C,T,	0.826078,0.173922,	+
135401677	135401678	rs17169786	A,G,	0.612620,0.387380,	+
135402851	135402852	rs11956252	C,G,	0.471046,0.528954,	+
135403528	135403529	rs6899012	A,G,	0.471446,0.528554,	+
135403764	135403765	rs6880582	G,T,	0.154153,0.845847,	+
135403850	135403851	rs34319360	A,G,	0.842851,0.157149,	+
135404172	135404173	rs9986124	G,T,	0.528754,0.471246,	+
135404612	135404613	rs9986287	C,T,	0.476438,0.523562,	+
135404659	135404660	rs10051650	G,T,	0.419928,0.580072,	+
135405333	135405333	rs372125340	-,A,	0.374401,0.625599,	+
135406458	135406459	rs7725702	C,G,	0.556310,0.443690,	+
135406533	135406534	rs7725447	A,G,	0.434505,0.565495,	+
135406780	135406781	rs2881285	C,T,	0.635583,0.364417,	+
135407571	135407572	rs4976470	A,G,	0.543331,0.456669,	+
135407746	135407747	rs6892173	C,G,	0.508016,0.491984,	+
135408324	135408325	rs4976471	A,T,	0.449081,0.550919,	+
135409013	135409014	rs6861956	C,T,	0.449081,0.550919,	+
135409123	135409124	rs12521108	A,G,	0.353235,0.646765,	+
135410862	135410863	rs11742191	A,G,	0.587460,0.412540,	+
135411280	135411281	rs11749522	C,T,	0.588858,0.411142,	+
135412194	135412195	rs10079215	A,G,	0.555711,0.444289,	+
135412674	135412675	rs35137944	A,G,	0.555711,0.444289,	+
135413025	135413026	rs7724672	A,G,	0.555711,0.444289,	+
135414454	135414455	rs4246798	A,G,	0.524760,0.475240,	+
135414509	135414510	rs4246799	A,G,	0.471046,0.528954,	+
135414865	135414866	rs17169806	C,T,	0.624201,0.375799,	+
135414892	135414893	rs34134607	C,T,	0.224840,0.775160,	+
135415063	135415064	rs62365993	A,G,	0.623802,0.376198,	+
135415299	135415300	rs2346018	A,C,	0.373802,0.626198,	+
135415725	135415726	rs2346019	A,G,	0.538339,0.461661,	+
135416546	135416547	rs9327740	A,G,	0.896965,0.103035,	+
135417202	135417203	rs4976364	A,C,	0.584465,0.415535,	+
135417897	135417898	rs12653557	G,T,	0.457069,0.542931,	+
135418716	135418717	rs4976472	C,G,	0.543730,0.456270,	+
135419158	135419159	rs4976473	A,C,	0.534744,0.465256,	+
135419340	135419341	rs13159052	A,C,	0.453874,0.546126,	+
135420142	135420143	rs5871594	-,A,	0.485623,0.514377,	+
135420944	135420944	rs56382516	-,A,	0.722444,0.277556,	+
135422382	135422383	rs72794938	C,T,	0.786542,0.213458,	+
135422442	135422443	rs11242311	C,T,	0.573882,0.426118,	+
135422507	135422507	rs34835264	-,A,	0.575679,0.424321,	+
135422597	135422598	rs11242312	A,G,	0.544329,0.455671,	+
135422697	135422698	rs10900843	A,G,	0.543730,0.456270,	+

135422737	135422738	rs10900844	A,G,	0.455671,0.544329,	+
135422747	135422748	rs72794940	A,G,	0.781949,0.218051,	+
135422863	135422864	rs11242313	A,G,	0.543930,0.456070,	+
135423028	135423029	rs11242314	C,T,	0.534744,0.465256,	+
135424755	135424756	rs13186426	A,C,	0.478834,0.521166,	+
135424847	135424849	rs4035982	-,AT,	0.468650,0.531350,	+
135427081	135427082	rs1008345	G,T,	0.648163,0.351837,	+
135427460	135427460	rs142812848	-,TA,	0.743610,0.256390,	+
135429019	135429020	rs7715300	A,G,	0.844050,0.155950,	+
135434182	135434182	rs72338288	-,A,	0.585863,0.414137,	+
135435800	135435801	rs7720483	C,T,	0.466254,0.533746,	+
135436979	135436980	rs12519122	C,G,	0.476238,0.523762,	+
135439728	135439729	rs6863438	A,G,	0.474441,0.525559,	+
135439739	135439740	rs17691375	A,G,	0.364018,0.635982,	+
135440363	135440364	rs12521857	A,G,	0.327276,0.672724,	+
135441173	135441174	rs13182074	C,T,	0.533546,0.466454,	+
135441312	135441313	rs17748071	A,G,	0.179113,0.820887,	+
135441558	135441559	rs12515040	C,T,	0.344649,0.655351,	+
135443166	135443166	rs369404371	-,GT,	0.686302,0.313698,	+
135443622	135443623	rs740371	C,G,	0.468051,0.531949,	+
135444985	135444986	rs7726617	G,T,	0.660949,0.339051,	+
135446553	135446554	rs17169841	C,T,	0.338458,0.661542,	+
135447745	135447746	rs34082824	C,T,	0.356430,0.643570,	+
135448004	135448005	rs35809977	-,T,	0.166334,0.833666,	+
135448504	135448505	rs2346361	G,T,	0.556110,0.443890,	+

30bp Flanking Sequence	Novel PAM
CACCACTGTACTCCAGCCTGGACGATAGAGCAAGACTCCATCTCAGCAATAAATAAATAAA	No
AAGACTCCATCTCAGCAATAAATAAATAAATAAGTAAATAAATAAATAAATAAAGTGTTATGAT	No
ACTCCATCTCAGCAATAAATAAATAAATAAGTAAATAAATAAATAAATAAAGTGTTATGAT	No
TGAATATGAATGAAGGTGCTTCTCAGAGCCATGCATTAGAACTCTATCACAATATATCTAC	No
TGTCAGTTGTTGAAATAGTTGTGTAACAATTTCCCACTATGATTGCACATTTCTGTTTCTG	No
ATTAGAGAGAAAATAGATTGGTAGCAATACTTTTTTTTTTTTCAATTTACTCTCCTTTT	No
AATTCTGTGAAGAAAGTCATTGATAGCTTGATGGGGATGGCATTGAATCTGTAAATTACCT	No
GGCCCTGGTTATTTTTAATCTTGTGGGGACGTTGTATTTACAAAATGTCTGTAGAAATAGA	No
ACAATGCATAATAATCACCTCAGGGCAAATAAGGCATCTATCCCCTCAAGTACCTGTCCCT	No
TGGTTGACTGCAAGGCAAAAATAATACTCATGCACTTTGTGCCTTGAACTTTATGGCAATA	No
TGCCCCTTGGTAATATCTAGTTCATCAAATATTTCATTTTTATTGTTGTTTTTCTTGCAAG	No
TCCATTTTGATGTACATAAAATATATCTAAACAAATTCAGCTTTGACATAGGGGTTATCGG	No
AGCATGTGGGAAGTCTGCTCTGTGTCATTTTTCCATTAAAACGCAGTCCTAACAGGAGAAT	No
AATAGGTTCTGCGAGGGATCTTGTCCTATATACAAGTAAAAGTAGAAAGATTGGTATTTAA	No
TATATTAGCACAAATAAAGTAAAACAAGTGTTTTTTTTTAGAAAGTAGTTCTATAAGGA	No
ACGATTTTGAGATTCACCCATGTTATAGCATGTTATGGATAATTTGTTCTTTTTTTAGCT	No
GATAGACATTTGGGCTGTTTCCAGTTTGGTGTATTACAAGTACAAATGCCTTTTCTGATTC	No
TCACTGGAGTGACAGCCAGTCTTATTTACGGGTCCCACTTACACTCAAAGGGAGGAGATGA	Yes
GGAATCTTAGGACCCATCTTAGAATTCTGCGCATTGCACATGGTTTTTCCTTTGACTTCCT	No
GAGGCCTGCTCCTGGAATGGATGTTTGGTGAGTCATTAGGGCTGTTCACATACATTCTGTT	No
TATCATTCTGAGCTGAAGTTTTAAGAGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGT	No
ATGGACTTGTGGCACAGATGACAGAAATAAATCTTTGTTGTTTAAAGCCCCTAAGATTTAG	No
GCTGGTACAAAAAGGACAGTATCACAGTTATTCTGGCTGCCACCTTCTCTTTTTCTCCGGA	No
GGGGGACCGATAGAGGGTACACAGGCATAGATCACCCTCAAGCAGAGACTACAACTGGTTA	Yes
GTTTACAGCCCTGGCTGCAGGCATCCTGCCGGGTTGGACTGCCTTGCTTTTCAGAGTAGGG	Yes
ATTCCTTTTTATACCTTTCATTTTTTTTAGTGATCCAAATAGGAAACTTTTTCATTCATTG	No
TTCAGAAGCCAAGGACAGCTTAGGGAAATAGCAAAGGGCTAAGCATAAAATCTCAGAATTC	Yes
CCTCCAGCTCCAACATTCATCATCTTTCTCGGCTGCTGTGATTCAGTGGGAACACTGCAGA	Yes
AAGGATGATATTCCTGAGAAGACAAGTCAACAAACGTAGCCTGTGTGCATTCAGTTTACAT	No
ACCTAATACATCCACTCTTGTCGTTTTCCAGGAGGCAGAGGAACCTCGCCTGCTTCACCAT	Yes
CACAGGTCAGGATCTATACCATGTGGGCTGTGCCTGTTTCCTCCAAGAGTTTCCAGATCGT	Yes
TTGAATCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCATGCCACTGCACTTCAGCCTGGG	Yes
GCAGAAATTTGGGCAAGAAATGGAGAAGAAGACACCAGACTGTGGTGGGCATGTGGGAAAC	No
TAAAATTATCTCATATAATCCTCCTAATAGTCCTGTAAGTAGGTATTATTCTTATTCTCAA	No
TAATCTTTTAGAGCATTCCTAGCTGGAAACTTTTCCCGGAAATCCATAAGTTAAGTGCTCC	No
TGATGAGATGTTTTTACCCCAAATCCAGCCGTGTCTTTCCCTGTTGTTTAGGATTATTAGA	No
ACTGATTTAGAGTGATCAACCTATGTCCCCAAGAGTAAGTTACAGAAACTAAATTTTTAAAA	Yes
CATGCATGCACTGTTTTTTTTTTTCCTACTGTTTTTTTTTTCCTTCTGCTTTCTTCCCTGC	No
ATGCTGTTTTGCAGATCACATTTTGAGTGGCAAGACTGTGGAAAATCCTTGAGAAATCAAT	No
GAAAAGGCAGGCCTGGTGTCAGCTGGGCTGCAGATGCCAGCTCTCCCACCAACAGGCCCAG	No
CCCAGCACAGGCATCCCTTCCTGCCAGCTATGAGCCTCGAGGTTAGCTCTACTCCCCCTCC	No
CCAGCTATGAGCCTCGAGGTTAGCTCTACTCCCCCTCCCTAACCCTGCATGCCCAAGGGGT	Yes
CCTCCTCTCCACTAGCTTGATCACTCCCCATGCAGGCCCTCAGTTGCTTTATGCTCTCAGT	No
GCCCTCAGTTGCTTTATGCTCTCAGTAGGCCATTCCTCCTCCAGTGCCCACACTCTCTCCCTTC	Yes
CTCAGTTGCTTTATGCTCTCAGTAGGCCCTCCTCCAGTGCCCACACTCTCTCCCTTCTCCT	Yes
CCCCAGGGCTAGCAGTGCCAAGTAACTGACAGGTGATTAATAGATGCTTGGGTAAGTATCA	Yes
CAGGTGGGGTGGGGGTGGAGGGATTAGAGATTGAGGAGCTGGGGAGGGTGGTCAGCTCCTG	No
CTAGGCCTTTCAGGAGTTTGGGGCTCTGGCGGAGAGGGCCTGCTGGGAGCACATCTGGCCA	Yes

CCATTTTACAGGAGGTGAAACTGCAGCTTAGTGAGGTAGAGAGTGACTTAGTTCAGACACA	No
CTGCTGCCACAAGGACAGCAGCAGTGGAAACATTCAGCAAAGGAATGTTGGAGCCACATCC	No
CTTCTGGGATTCTGTAACAATAAATAGGACCGGGGGCTGGAGTATGGCCAGCAAGGACTCT	Yes
TGGTCCCTTCTCCAGCCTTCACTTCTCTTGTCCCTAGATCCTTACATGGATTCATTAATGC	Yes
CTGTAATCCCAGCTACTCAGGAGGCTGAGGCGGGAGAATTGCTTGAACCCAGGAGGCAGAG	No
CTGGTAAGCTGAGGAGGTCTGTCCACTTCCTTTTGCTGCCCCCAGGGGGTATCAAGCCTGG	Yes
CTTGGGAACTGAGGAGACACAGTCAGCCTCCAGGAGTGCCCAAAATGCCCTCACATGCTGC	Yes
CTGTAGTGGTTGAGGCCTTTGTTGGTAGACAGTAGGTTAAAGCAAGCCATGATTTTCTATT	Yes
CAGGCAAGCAGTCCATGGGCCATGTCAGATGTCTAGACGTTATGGGTCTGTGTTTGCTCTG	No
CAGTCCATGGGCCATGTCAGATGTCTAGACGTTATGGGTCTGTGTTTGCTCTGCCATTCCT	No
AACCCTTACAGCTTTCTTTCTGATTCTATCCTGAGTTACTCTACTCCAAGCTGAGACTTTT	Yes
TGAGGCCACTTGAGCTGTTCAGCTCCCTTGCGGTATTTTGGGGATGGAACTCAGAAGCCAA	No
TAACACTATTCCTCCAACTCTGCTTCAGCATCTCCATGGATTTTCACACAGACACTTTAGG	No
GTTTGGGGGGACTTGACCTAATCCCACATCCCAGCCCCAGTAATACAGCCCTGGAATTTAT	Yes
CTGAAGATAAAATTGTGAGTCAATCAAGATGAGTCCCAAGACAATAGCCTGTTTAGCCCTT	No
AGAGGAAGGAAGGAGGAGGAAGCAAACAGGAAAAAAAAAGAATGTGCATAGCTTGTCACT	No
TACCACACTGGTGGAGTAGACTCCAACTGTGGCCTGTCCATGCCCTTCCCAGCAGGCACAG	Yes
TTGAATCTGTGGGTGAAGAACCCACAGATACGAAGGGCCAACTGTATTGGCTATTTTTTTA	No
TTCCTTGCCTCCCCTGGAAAGGTCAGTGGTGTGTGGCTGCAGCAGCACAGTGTCCTCTGAG	No
GGTCAGTGGTGTGTGGCTGCAGCAGCACAGTGTCCTCTGAGCCCTGGACCTGCACTGTGGC	Yes
ACTGTGGCTTCCAGAGGTGGCAGTTCCCACATGGGGTACTAGAATAAATGGCCTATCAGGC	No
CGCAGGTGTGGATGGCTGTTAGCTGGGAGCCTCGCTGTCTAAGCTCCTCTCCCATGCTTTT	Yes
TCCTTCCCTCTTCTGACCCTCCATTTTGCCGATCTTTCCTTCTTATAACACATACTTACTA	No
GAGGGGAGAGTAATAGCAAAGGCTCAGGGCAGGAAGGGCAAGGGAGAGGCCAGTGGGTGAG	Yes
GGAAGAATGAATAGAATCAGAGAAGCAAAGTAAGAGGGAAGAGCAGAGAGGACAGCAGGGA	No
TCCTGCTGCTGCCTCATTTGTGCAGCTAGATTGAGCCCAAGACCTGCTCTGGTCCAAGATG	No
GCTTCACTATTCTTCTCTGTGGCTAGGGGATTTATGGATAAACCAAAATTACAGTTAAAAA	No
ACAGTGACTCACGCCTTTAATATCAGCACTTTGGGAGGACAAGGTGGGCGGATCACCTGAG	No
CTGGGTGACACAGCGACACTCCGTCTCAAGAAAAAAAAAAAAAAAAACAGTTATAGTAGTC	No
CACAGCACAAAATGGGGAATGAGGGCGGGCATTGGGACACACATAGCCTTAAGGGGCCCAA	No
GGCGGGCATTGGGACACACATAGCCTTAAGGGGCCCAAAGGCTTTTAGAACTGTATTCCCT	Yes
GAACTGTATTCCCTATTAAAACATGATTTGCACAGAGCACATTCTTTGCTTTGGAGACCTC	No
CCTCCTTGACTGGTCCTTGCATTTGCCTCCATCCAGCCTGTCTGGGCTCTCCGAGGCAATG	No
TGCTGAAGCCATCGTTGCGGGGCTGTCTGTAGAGACCCTGGAGGGCACGACACTGGAGGTG	Yes
CAGAGGCTGATCTCTGCCTAACTGAGCTCACCTCTCCTCCCTCTCCTCCTGACTGGTTAGA	Yes
GATCTCTGCCTAACTGAGCTCACCTCTCCTCCCTCTCCTCCTGACTGGTTAGATTTTCTAG	Yes
GCTCTGAACAAATCAGGAGGCCCCTCGTGGAAGTATAACCAGTCCTTTCTTTCTCTGTCCC	Yes
CTTTCTCTGTCCCTCTTCTGTGCAGAGCCTCTGCATTGAGAACAGCTGCATCGCGGCCCAC	No
CCTCAACCGGGAAGGAGTCTACACAGTCTTTGCTCCCACAAATGAAGCCTTCCGAGCCCTG	No
GGAGGATGAGAGCAGGAACCAGGGAGGTCATGAGCCTTGGACAAGGGCACAGAACAGCAGC	No
GAGGATGTTTGGCAGGGGATCTAGTGGTTACGGGTGGCTAAGAAAAATGAGGAAGGTAAGA	No
GAGTATCTTGCAGCCTGTGTTGGGAGGATTAAATAGGATGCCACACACAGGGCCAGGCAGA	No
AGCTCCAGAAATCTCCCTGGCTGCACCTGCAGAGGCCACTGACCCCTCTGTGGAGGGACCG	Yes
TGACTTTAGCAGGTCATTCAAGAATCTCCTCGCACCTGGTTTCAGATGCTGGGGTCCTGTC	No
TGCCCAGAGCAGGAAGCCTGTCTTCCATTTCCAGCTGTTCCACCTACTTAGCTTAAAAGAG	Yes
GCCCAGAGCAGGAAGCCTGTCTTCCATTTCCAGCTGTTCCACCTACTTAGCTTAAAAGAGG	Yes
TCCACCTACTTAGCTTAAAAGAGGCACTTCGCCTGTCTTCAGTCTCAGTCTCAGTCTCCTC	Yes
CTGTTTCATTCACCAGAGTATCCCCAGTCTAACACAGGACTTGGCATATGAAAAGTGTTCA	No
GGTGGCACATGCCTGTAATCACAGCTACTCTGGAGGCTGAGGCAGGAGAATCACTTGAACC	Yes
AACGAAAACTGTTCAGTAAACACTTGCTGAATGAATAAAATAAATATATAAATGTATAAAT	No

TAAACACTTGCTGAATGAATAAAATAAATATATAAATGTATAAATAAATGCTCTACTTTCA	No
CCCCACTGACTCTGCAGCCAGTCCTTTTCTTCATGTGGCAGTTGGTGGAGAGAAGAAAAAC	Yes
GGTGAGGCTGGGGCTCTCCTGGGCACTGTATGTATTCTGGATACAGGGATACTGGGCTCGC	No
CCTTGCCCCAGCCCCACCTCCCTCTCAAACCCTCTCTGGCTCTTTCTGAGCTTCCTTTCCT	Yes
CCCATCCCCTCTGTGCCAACCCACATTCAGATTCCTTCCCCGGCTCCCGTAATCTCTGGCA	Yes
ACATTCAGATTCCTTCCCCGGCTCCCGTAATCTCTGGCATCTAGAATATCCTCAGGACTCT	No
GGCTGGCCCACTTTCCTAGAGAATGGGACAGACCTCCTTCCCACCCACACCCATCTCTGCC	No
TACACACACTGCACTTTACCAAGATGACCTCGGAAACCAAAGAGGTGATCAGCATAAGTTT	No
GGAAGAAGGGGAATGGGCTCTTCTTAGTCCACTTCTGTCTTTGCTGTCACTCTGGGAATAC	No
GCCCTAACAGTCCCTGCAGTGCTGGAGCTCCCCATCACTTGGGGAGGCGATGGAGGTCACT	Yes
TTTGAGGCTGAATAATATTCCATTGCATATTATATACCACTCACTTTGTTTATTCACTCATC	No
TTAATTTGGGAGGGGGGGGAATTGCTATAGTGTTTCCATTGTGGCTGCACTATGTCACATT	Yes
AGTTGCTCCACATCCTCTCTAATACTTGTTATTTTCTGGTTTTTGGTAATAACCATTCTAT	No
TTTACTCTGTTGACAGTGTCCTTTGATGCATAAACATCTTTAATTTAGATGAAGTTCACAA	No
ACCACACTGTTTTGATTATTGTAGCCTTGGCGTAAGTTTTAAAATTAGTAAGTATGAGTCC	No
GTAAGTATGAGTCCTTCAACTTTGTTCTTCTTTTTCAAAATTGTTGTGGCTATTGGAAATT	No
AAAATTCAGAGCAGGGAAGAAGCTACTGGTAAAAAAAAAAAAAAAAGACTAAGAGTCTCCT	No
GACTCCCTAACTGAGAAAGGTCACCCCTGGGTCAGCAGGCCCCAACTCAGAGAGAACATCT	Yes
GGTAAATGGCTGATGAGTCACACTGCAGAAAGGCTGCTTCATCTTTAGCAGGTGACTCCAC	Yes
TCTCCTGTCAGTGGGGCGGGTGAGTCCCTGCGGCTTCTTGGCAACAGCTTATGAATAATTC	No
GGCTTCCTCCCGTGACATGTGGGGATTATGGTAACTACAATTCAAGATGAGATTTGGGTGG	Yes
CTTCAAAGCATCAAGCTTAGACATTGCACACAGTAGGTCATGCATAAGCATAGAACAAGTA	No
AAATTCAACTTGAATTTCATTTCCTCTAAGATGGAAATTCAGGGTCTTAATGAATTAATTT	No
TACAAAGTTGCTGGTTTGAAGAAAGAGGGGCATGAGAAAAAAGAATGCAGGCAGCCTCAGG	No
CTGGATTCTCCACTAACCTAAATTAGTTCGGAGAGGGAGTCTCCTCTCGCTCCTCCAGAGA	Yes
CTCCCAGCTCAGTGTTTCTCAAACTTGTTCAGATGAAGATTCTGACTCGAGAGGTCTGGGA	Yes
TTCTTTGCTCAAAGGCAAAACTCTGTTTCCCCAGGGGGCCCAGTCAGGTTGAGCAGTAAGA	Yes
ATGTCTATACCATCCCCAGGAGAGCTAGAGGAAAGAAGAGCCTTTTGACTCTGAATGTCCT	Yes
AAGGAACAAGAGAGCTGGACCTCGGGATTGGGGGCAGAAGTGGTCCCTTCTGCTTTCTCGG	Yes
AGAGTTCCCTTTGAGCAGGAAGAGGAGAGTGCTTTCTCCATTTCATCTTCTGCCAAATACA	No
AATAGTCACACTAGTGTTCTTTAAAAACGCGGGTTCCTGCGCACCACCCGGTGATTCTGAT	Yes
TCTGATTCAGCTGTCTGGGGTGGGTCGGGGAAGGAAGCTGAAGTTTTTAAAAGACCCTCAG	Yes
GGAGGTGCTCGCTCCTGATGTTCGCTGTTCCCAAACCCCTATAAGGATCCTTCAGTCCAAC	Yes
TTCCCAAACCCCTATAAGGATCCTTCAGTCCAACCTCTTGGGTGGGGAACAGAGGGGGTCC	Yes
CGGGCATAGGAGGGGAGGTGCACACTGGAGAGGAGGCGCGCGGGACCGGAGGCGCACACCG	Yes
CAAGAAGACGATTCTGGATTGTAGGCCGCCCGAGAGCGCCCCCAGGAGCCCATCACTGCGA	Yes
TGCGAACAGGTGAGAATAGCGTAGGTACAGGCCGCCAGGGAGGAAGAAACTTGGAACTTTC	Yes
GGTGCGGGGGGGCGTGTGGGCCGTCTACCTAGGTCCAGCAGCCAGGCTGCTGAGGAGTACC	Yes
TTAACAAAAAGTGACATATTCAGGAACAGAACCAAAGTTTGAACTCTCAAGAAAGAGTTGC	Yes
CCCTGTTAGGGCTAGACATTGAAGGTTTTTTTGTTTGTTTGTTTTTTTCCATGAGTTTACC	No
CCTCTCTCCTCTCCACACACTCTCCTTACTCTTTGTTGCATCACCCTACGTCTTTCCACCA	No
CATAATCTTCACCAAACTCTGCAAGGTAGAACCTTTTATTTGCTGGATGAGTAAACTGAGG	Yes
GTTCAATCTCATGCTTGTGCAATGGTGGTACTTCTCAGTGGAGAAATGTATACTGGCTAAA	No
GTCTAAACATTGTTGAGAAGAAGCAGTGGCAAAAAAAAAAAAAGTTGAAAATACATTAGAG	No
ATTTTAACATCAGGAATGCCTGATTTTTTTTAAAAAAAAACCTCCCAGCAAACTAGGAAT	No
AAAATCTTTGTAAATGGTATACCTGGTAAACGGCTTGTATACAGAATAAAAAATAACTCAA	No
ATAGTGAGAAAACAAACAGCTCATTTTTTTCCATGGGCCAAAGATGTGAATAGACATTTCA	Yes
AGAAGATATATGATAGCAAATAAGCACATGAAAAAAAAAAATCTCAGCATTGTTAGTCTT	No
ATACTACTCCACATTTATTACAATGTTTGGAACAAAAGGCCATTTTAAGTGTTGGCAAGGA	Yes
TCATCACTTTGGAAGAGAAAGAGTTTGGCAATTTCTTAAAAAGCTAAACATACACCTATCA	No

AAGCTAAACATACACCTATCACATGGTCCAGCCATTCTATAAGATATTTACTCAAGGGAAA	No
TACACCTATCACATGGTCCAGCCATTCTATAAGATATTTACTCAAGGGAAAAGAAAGCATA	No
CCAAACTGGAAACGACTCAAATGTTCCTAAACAGATCAACGGATAAACAATCTGTGTTATA	No
ATGAAAACAGTACATACTGTATTGCTCCATCTATATAAAACTCTAAAAAATGCAAGCTAAT	No
ACCTTGTCCTCCCCCTTCATCCCATGGGATACTCCTCTTCACTTCCCACAAAATCCAACCA	Yes
TGGCCAGTGCCAGGGTTGCTAAATACTTCAATATATCACTTTCGCTGGACCCCAAAGAGTAA	No
AATTTTATTGTTCTATTATATATATATATAGAGAGAGAGAGTTCTATATATATATTGTTCT	No
TGTGTGTGTGTGTGTTTGTGTGTGTGTGTGTATATATATATATATGTATATATATGGAGA	No
TCCTGGTACAGCGAGAAACATGGTTTTTACATAAGAATGATAACTTGGCTGGACATACAAT	No
TCTTGAGGTTCTAAGTACGGTATTTTTCTGAAAAAAAAAAAAAAATGGTTTTACTGAAAG	No
TCTGACCACCCAGGCTAGATTCTGTCCTCTCATTGAGATTTTTCACTGTACTTTATACTAC	No
AGGTAAGAAAAGTGCTTTGCCTAGAGCCTGGTATAAGTTTTCAATAGGCATTAAGCATATT	Yes
AGAATTAGTTCTATACCATGTGGCTCAAAAGCAGCACTTTGGACTGAATGCTACAGGGGCC	No
TATACCATGTGGCTCAAAAGCAGCACTTTGGACTGAATGCTACAGGGGCCACACACAAAGC	Yes
TTGTCTTTGCCTGTCACTTCCCCCTTCCCCGGCTGCAGGATTCCTGTCTCACAACTACCCT	Yes
CATTTCTGGAGAAGGATGGTGGTCACGGACCTGGGCTTGATCTGCCGACTCTCTGTGCCTC	Yes
GCAGCAGCAGTGATGGGGCTTTTTTCCTGCGTCCACCAGCCATCTGATACACACCCCAAAG	No
CAAAGTCCTGTATTACCTTATCCCCTTCCTTGCCCGGGCTGACTTTCCCTGCTTACCTCTG	No
ACGAGATTGATGAGGGAATTTGATAAAATAGTGTGTGTGTGTGTGTGTGTGTGTGTGT	No
AGTCTTGTCCACCTGCTGGGCCATGGAGCAGTGTCATGGAATCTCCAGGGAGCCTTTTATT	No
AGCAGAGGGGTGGTTGTCTGGCCACCTGAATGCAGGGGCAGCCCCCTTCTTTTTCTTCCCA	Yes
AGAATAGGGACTGCATCTGCTTGTCCATGGTTGTACACCAGTTACATGAGCAATGCTTGTG	No
ATGGGAGTGAATTTCTTTCTTTCTTTCTTTTTTTTTTTGAGACAGAGTCTCATTCTGTCA	No
TGATCTGCCTGCCTCGGCCTCCCAAAGTGCTTAGGATTACAGGTGTGAGCCACCACACCCG	No
TTCACATGGGCCCTTGCAAAAGGAGCATGTGTTTCTATGAGAGGTGGGGCCAAAGAGATTG	No

PAM 5' - 3'	WT SEQ PAM 5'-3'	Non-canonical PAM In WT seq?
GGG	GAG	YES - NAG
AGG	AGA	YES - NGA
CGG(-) OR CGG (+)	CCG(-) OR CCG	YES - NCG OR NCG
TGG	TGT	YES - NGT
CGG	CAG	YES - NAG
AGG	AAG	YES - NAG
GGG OR TGG	GTG	NO - NTG OR NGT
AGG	CGA	YES - NGA
CGG	CAG	YES - NAG
GGG	GGC	YES - NGC
TGG	GGG (shifted guide)	YES - NGA
AGG	AGA	YES - NGA
CGG	CAG	YES - NAG
CGG	CAG	YES - NAG

CGG	CAG	YES - NAG
AGG	GGA	YES - NGA
AGG	AAG	YES - NAG
TGG	TAG	YES - NAG
CGG	CAG	YES - NAG
AGG	ATG	NO - NTG
GGG OR TGG	GTG OR TTG	NO - NTG X2
TGG	TAG	YES - NAG
AGG OR GGG	AGT OR GTG	NO - NGT OR NTG
AGG	AAG	YES - NAG
CGG	CAG	YES - NAG
AGG OR GGG	AGT OR GTG	NO - NGT OR NTG
TGG	TAG	YES - NAG
AGG	AGT	YES - NGT
GGG	GGA	YES - NGA
GGG	GGA	YES - NGA
CGG	CAG	YES - NAG
TGG	TGA	YES - NGA
TGG	TCG	YES - NCG
GGG OR GGG	GCG OR GGC	YES - NCG OR NCG
CGG	CTG	NO - NTG

TGG	TGA	YES - NGA
GGG OR AGG	GAG OR AAG	YES - NAG OR NGA
AGG	AGA	YES - NGA
GGG OR GGG	CGG OR GCG	YES - CGG OR NCG
GGG	GTG	NO - NTG
GGG	GGC	YES - NGC
AGG OR GGG	AAG OR AGG	YES - NAG OR NGG
TGG	TGA	YES - NGA
CGG	CGA	YES - NGA
CGG	CAG	YES - NAG
GGG OR TGG	GAG OR TGA	YES - NAG OR NGA
AGG	AGA	YES - NGA
TGG OR GGG	TGA OR GAG	YES - NAG OR NGA
CGG	CAG	YES - NAG
GGG	GGA	YES - NGA
GGG OR TGG	GAG OR TGA	YES - NAG OR NGA
TGG	TAG	YES - NAG
AGG OR GGG	AGA OR GAG	YES - NAG OR NGA
CGG	CTG	NO - NTG
AGG	AGA	YES - NGA
TGG	TAG	YES - NAG
GGG	GGT	YES - NGT
GGG	GGT	YES - NGT
TGG	TGA	YES - NGA
GGG	GGA	YES - NGA

AGG	AGT	YES - NGT
TGG	TGC	YES - NGC
TGG	TGA	YES - NGA
CGG	CGA	YES - NGA
AGG	AAG	YES - NAG
AGG	ATG	NO - NTG

Supplementary Table 3: Table showing the ASNIP CRISPR mutational analysis for the *TGFBI* locus. Initially SNPs in the 50kb flanking regions and across the *TGFBI* coding region were filtered to leave only those with a MAF of >0.1. Each SNP and flanking sequence was then individually assessed to determine if it generates a novel *S.pyogenes* Cas9 PAM. Those that did generate a PAM were then further investigated to determine if a non-canonical PAM exists on the alternative allele. These SNPs (both with and without non-canonical PAMs on the alternative allele) were then cross-checked to the phased sequencing data from the R124H Avellino corneal dystrophy patient to determine if the PAM generating SNP lies *in cis* with the R124H mutation. Guides were then designed for those that are associated with a PAM on the same chromosome as the mutation. Guide sequences were then inputed into the in silico MIT CRISPR and Benchling design tools and sgRNAs were synthesised for those that generated the best on and off target scores.

chromStart	chromEnd	SNP	Alleles	Novel PAM	PAM 5' - 3'
135314845	135314846	rs72793185	C,T,	No	
135314876	135314880	rs373839451	-,TAAG,	No	
135314879	135314880	rs2346012	A,G,	No	
135315167	135315168	rs17169582	A,G,	No	
135316486	135316487	rs2158351	G,T,	No	
135318158	135318158	rs559931571	-,T,	No	
135319760	135319761	rs257480	A,T,	No	
135322719	135322720	rs6868908	A,G,	No	
135323867	135323868	rs6874348	A,G,	No	
135327490	135327491	rs1859295	C,T,	No	
135327678	135327679	rs17688533	A,G,	No	
135328361	135328362	rs10076250	A,G,	No	
135329914	135329915	rs12520800	G,T,	No	
135330053	135330054	rs17739831	C,N,T,	No	
135331391	135331391	rs113921691	-,T,	No	
135333390	135333391	rs6881712	A,T,	No	
135333505	135333506	rs12332587	A,G,	No	
135334846	135334847	rs10074474	A,G,	Yes	GGG
135334927	135334928	rs10074539	A,G,	No	
135335228	135335229	rs10079806	A,C,	No	
135335396	135335396	rs10522532	-,GTGT,	No	
135335578	135335579	rs6882087	A,G,	No	
135335676	135335677	rs11747904	A,T,	No	
135336560	135336561	rs13157444	A,G,	Yes	AGG
135337231	135337232	rs57104529	C,G,	Yes	CGG(-) OR CGG (+)
135338598	135338599	rs916950	C,T,	No	
135339463	135339464	rs17740150	C,G,	Yes	TGG
135342086	135342087	rs2525490	A,G,	Yes	CGG
135343546	135343547	rs9327738	C,T,	No	
135344162	135344163	rs6892697	A,G,	Yes	AGG
135345183	135345184	rs72794904	G,T,	Yes	GGG OR TGG
135345816	135345817	rs9327739	C,T,	Yes	AGG
135351182	135351183	rs72794907	A,G,	No	
135352328	135352329	rs7728408	A,T,	No	
135354323	135354324	rs4976360	A,T,	No	
135355037	135355038	rs6894906	A,G,	No	
135355436	135355437	rs6895177	A,G,	Yes	CGG
135357723	135357724	rs146020713	-,T,	No	
135360737	135360738	rs35901765	C,T,	No	
135361140	135361141	rs35636600	A,C,	No	
135362549	135362550	rs34098140	C,T,	No	
135362572	135362573	rs4976459	C,G,	Yes	GGG
135362681	135362682	rs10463536	C,T,	No	
135362716	135362716	rs111308112	-,CATT,	Yes	TGG
135362719	135362720	rs55821461	C,T,	Yes	AGG
135363874	135363875	rs2282790	A,G,	Yes	CGG
135364189	135364190	rs17169707	C,T,	No	
135366135	135366136	rs2237063	A,G,	Yes	CGG

135367219	135367220	rs2237065	A,G,	No	
135367602	135367603	rs2237066	C,T,	No	
135367756	135367757	rs756463	C,T,	Yes	CGG
135367944	135367945	rs756462	C,T,	Yes	AGG
135374314	135374315	rs10053962	C,T,	No	
135375041	135375042	rs11738979	C,T,	Yes	AGG
135375330	135375331	rs10223277	C,T,	Yes	TGG
135375472	135375473	rs4141306	A,G,	Yes	CGG
135375596	135375597	rs739866	A,G,	No	
135375604	135375605	rs739867	A,G,	No	
135377348	135377349	rs2107331	A,C,	Yes	AGG
135377565	135377566	rs7719624	C,T,	No	
135377729	135377730	rs2282791	G,T,	No	
135377801	135377802	rs1989972	A,C,	Yes	GGG OR TGG
135378238	135378239	rs1989973	C,G,	No	
135378363	135378363	rs540142018	-,A,	No	
135380058	135380059	rs2074558	A,G,	Yes	TGG
135380763	135380764	rs6897320	C,T,	No	
135383356	135383357	rs2073508	A,G,	No	
135383376	135383377	rs2073509	G,T,	Yes	AGG OR GGG
135383429	135383430	rs2073510	A,G,	No	
135383892	135383893	rs2073511	C,T,	Yes	AGG
135384080	135384081	rs45554435	A,G,	No	
135384442	135384443	rs916951	A,G,	Yes	CGG
135384844	135384845	rs6596281	A,T,	No	
135385315	135385316	rs17169753	C,T,	No	
135385699	135385700	rs1060433	C,T,	No	
135385777	135385778	rs1137550	C,T,	No	
135386023	135386024	rs10706409	-,A,	No	
135386729	135386730	rs2237070	A,G,	No	
135386752	135386753	rs2237071	G,T,	Yes	AGG OR GGG
135386799	135386800	rs2237072	C,T,	No	
135387802	135387803	rs17169768	A,G,	No	
135388662	135388663	rs1054124	A,G,	Yes	TGG
135389424	135389425	rs6889640	A,C,	Yes	AGG
135389432	135389433	rs13159365	C,T,	Yes	GGG
135391325	135391326	rs6860369	A,G,	Yes	GGG
135391373	135391374	rs1133170	C,T,	No	
135392425	135392426	rs4669	C,T,	No	
135392734	135392735	rs7727725	A,T,	No	
135393137	135393138	rs17689879	C,T,	No	
135393196	135393197	rs6871571	A,G,	No	
135395432	135395433	rs6893691	A,G,	Yes	CGG
135395625	135395626	rs10036667	C,T,	No	
135395825	135395826	rs11348106	-,C,	Yes	TGG
135395826	135395827	rs58759191	C,G,	Yes	TGG
135395863	135395864	rs6894815	C,G,	Yes	GGG OR GGG
135396083	135396084	rs10042825	A,T,	No	
135396291	135396292	rs10064478	G,T,	Yes	CGG
135396451	135396452	rs13168506	A,G,	No	
135396467	135396468	rs13188659	A,T,	No	
135396668	135396669	rs6880837	C,T,	Yes	TGG
135397701	135397702	rs6886556	C,T,	No	
135397784	135397785	rs6865463	C,T,	Yes	GGG OR AGG
135400034	135400035	rs13189180	A,G,	Yes	AGG
135400056	135400057	rs10043360	A,T,	No	

135400380	135400381	rs45543842	A,G,	No	
135401118	135401119	rs59239478	C,T,	No	
135401677	135401678	rs17169786	A,G,	No	
135402851	135402852	rs11956252	C,G,	Yes	GGG OR GGG
135403528	135403529	rs6899012	A,G,	No	
135403764	135403765	rs6880582	G,T,	Yes	GGG
135403850	135403851	rs34319360	A,G,	No	
135404172	135404173	rs9986124	G,T,	No	
135404612	135404613	rs9986287	C,T,	No	
135404659	135404660	rs10051650	G,T,	No	
135405333	135405333	rs372125340	-,A,	No	
135406458	135406459	rs7725702	C,G,	Yes	GGG
135406533	135406534	rs7725447	A,G,	Yes	AGG OR GGG
135406780	135406781	rs2881285	C,T,	No	
135407571	135407572	rs4976470	A,G,	Yes	TGG
135407746	135407747	rs6892173	C,G,	No	
135408324	135408325	rs4976471	A,T,	No	
135409013	135409014	rs6861956	C,T,	No	
135409123	135409124	rs12521108	A,G,	Yes	CGG
135410862	135410863	rs11742191	A,G,	Yes	CGG
135411280	135411281	rs11749522	C,T,	Yes	GGG OR TGG
135412194	135412195	rs10079215	A,G,	Yes	AGG
135412674	135412675	rs35137944	A,G,	Yes	TGG OR GGG
135413025	135413026	rs7724672	A,G,	No	
135414454	135414455	rs4246798	A,G,	Yes	CGG
135414509	135414510	rs4246799	A,G,	Yes	GGG
135414865	135414866	rs17169806	C,T,	Yes	GGG OR TGG
135414892	135414893	rs34134607	C,T,	Yes	TGG
135415063	135415064	rs62365993	A,G,	Yes	AGG OR GGG
135415299	135415300	rs2346018	A,C,	Yes	CGG
135415725	135415726	rs2346019	A,G,	Yes	AGG
135416546	135416547	rs9327740	A,G,	Yes	TGG
135417202	135417203	rs4976364	A,C,	Yes	GGG
135417897	135417898	rs12653557	G,T,	No	
135418716	135418717	rs4976472	C,G,	No	
135419158	135419159	rs4976473	A,C,	Yes	GGG
135419340	135419341	rs13159052	A,C,	No	
135420142	135420143	rs5871594	-,A,	No	
135420944	135420944	rs56382516	-,A,	No	
135422382	135422383	rs72794938	C,T,	No	
135422442	135422443	rs11242311	C,T,	Yes	TGG
135422507	135422507	rs34835264	-,A,	No	
135422597	135422598	rs11242312	A,G,	Yes	GGG
135422697	135422698	rs10900843	A,G,	No	
135422737	135422738	rs10900844	A,G,	No	
135422747	135422748	rs72794940	A,G,	No	
135422863	135422864	rs11242313	A,G,	No	
135423028	135423029	rs11242314	C,T,	No	
135424755	135424756	rs13186426	A,C,	Yes	AGG
135424847	135424849	rs4035982	-,AT,	No	
135427081	135427082	rs1008345	G,T,	No	
135427460	135427460	rs142812848	-,TA,	No	
135429019	135429020	rs7715300	A,G,	No	
135434182	135434182	rs72338288	-,A,	No	
135435800	135435801	rs7720483	C,T,	No	
135436979	135436980	rs12519122	C,G,	Yes	TGG

135439728	135439729	rs6863438	A,G,	No	
135439739	135439740	rs17691375	A,G,	Yes	TGG
135440363	135440364	rs12521857	A,G,	Yes	CGG
135441173	135441174	rs13182074	C,T,	Yes	AGG
135441312	135441313	rs17748071	A,G,	No	
135441558	135441559	rs12515040	C,T,	No	
135443166	135443166	rs369404371	-,GT,	No	
135443622	135443623	rs740371	C,G,	No	
135444985	135444986	rs7726617	G,T,	Yes	AGG
135446553	135446554	rs17169841	C,T,	No	
135447745	135447746	rs34082824	C,T,	No	
135448004	135448005	rs35809977	-,T,	No	
135448504	135448505	rs2346361	G,T,	No	

Reference SNP	Alternative SNP	Haplotype: 0|1 = reference SNP on haplotype 1, 1|0 = reference SNP on haplotype 2	In phase with mutant SNP? (0|1, require alternative SNP i.e it's on haplotype 2)	Guide Sequence
SNP required in red				
G	A	0|1	No	
A	G	0|1	Yes	GATAGAGGGTACACAGGCAT
G	C	0|1	Could use either. C *in cis*	AAAAGCAAGGCAGTCCAACC
G	A	1|0	Yes	CCAACATTCATCATCTTTCT
G	A	1|0	Yes	ATCCACTCTTGTCGTTTTCC
T	G	0|1	Yes	GGATCTATACCATGTGGGCT
C	T	1|1	Homozygous Alternative SNP	AGGATCTATACCATGTGGGC
A	G	1|1	Homozygous Alternative SNP	
C	G	1|0	Yes	GGCATGCAGGGTTAGGGAGG
-	CCATT	0|1	Yes	TGTGGGCACTGGAGGAGGAA
C	T	0|1	No	
A	G	0|1	Yes	TAGCAGTGCCAAGTAACTGA

T	C	1|1	Homozygous Alternative SNP	
A	G	0|1	Yes	TTGAGGCCTTTGTTGGTAGA
C	A	0|1	No	
A	C	0|1	Yes	AGGGCTGTATTACTGGGGCT
				CAGGGCTGTATTACTGGGGC
G	A	0|1	No	
T	G	0|1	Yes	TGTGTGGCTGCAGCAGCACA
				GTGTGTGGCTGCAGCAGCAC
T	C	0|1	Yes	GGAGAGGAGCTTAGACAGCG
A	G	1|0	No	
G	T	0|1	No	
A	G	0|1	Yes	CATCGTTGCGGGGCTGTCTG
C	A	0|1	No	
C	T	1|0	No	
A	G	0|1	Yes	CAAATCAGGAGGCCCCTCGT
A	G	0|1	Yes	AATCTCCCTGGCTGCACCTG
C	-	0|1	No	
C	G	0|1	No	
G	C	0|1	Yes	GAGACTGAGACTGAAGACAG
				TGAGACTGAGACTGAAGACA
T	G	0|1	Yes	TGCCTGTAATCACAGCTACT
T	C	0|1	Yes	TCTCTCCACCAACTGCCACA
C	T	0|1	No	
A	G	1|0	No	

C	G	1|0	Yes	CATCGCCTCCCCAAGTGATG
				CCATCGCCTCCCCAAGTGAT
G	C	1|0	Yes	AACTGAGAAAGGTCACCCCT
A	G	1|0	No	
G	A	1|0	Yes	CCCGTGACATGTGGGGATTA
G	A	1|0	Yes	TCCACTAACCTAAATTAGTT
A	G	0|1	Yes	CAGTGTTTCTCAAACTTGTT
C	T	0|1	No	
G	A	1|0	No	
G	A	1|0	Yes	GAGAGCTGGACCTCGGGATT
				AGAGAGCTGGACCTCGGGAT
G	A	1|0	Yes	ACTAGTGTTCTTTAAAAACG
A	G	1|0	No	
C	T	0|1	No	
C	T	1|1	Homozygous Alternative SNP	
A	G	0|1	Yes	GAGGGGAGGTGCACACTGGA
C	A	0|1	No	GGAGGGGAGGTGCACACTGG
G	A	1|0	Yes	GGTGAGAATAGCGTAGGTAC
A	C	1|0	No	
A	C	1|0	No	
C	T	1|0	Yes	CTATTCACATCTTTGGCCCA
A	G	1|0	No	
A	C	1|0	No	

G	A	0|1	No	
T	G	1|0	No	

Supplementary Table 4: Haplotype analysis of the *TGFBI* locus following phased sequencing of the R124H patient, allowing identification of SNPs that contain a PAM on only the allele associated with the mutation.

Supplementary table 5: Guide sequences of 12 ASNIP guides - designed based on the phased sequencing results of the R124H Japanese Avellino corneal dystrophy patient

SNP	Guide Sequence (5'-3')	MIT CRISPR	Benchling On-target	Benchling Off-target
rs72794904	GGATCTATACCATGTGGGCT	76	44.5	74.4
rs2282790	TAGCAGTGCCAAGTAACTGA	74	62.9	71.7
rs1989972	AGGGCTGTATTACTGGGGCT	68	41.2	66.1
rs2073509	GTGTGTGGCTGCAGCAGCAC	41	39.5	42.6
rs2073511	GGAGAGGAGCTTAGACAGCG	74	64.8	72.1
rs6860369	CAAATCAGGAGGCCCCTCGT	82	63.6	80.8
rs6893691	AATCTCCCTGGCTGCACCTG	52	58.7	51
rs6894815	GAGACTGAGACTGAAGACAG	41	69.3	45
rs10064478	TGCCTGTAATCACAGCTACT	55	51	60.1
rs11956252	CATCGCCTCCCCAAGTGATG	77	59.9	75.5
rs7725702	AACTGAGAAAGGTCACCCCT	73	62.5	70.4
rs4976470	CCCGTGACATGTGGGGATTA	78	37.9	75.7

Supplementary table 6: Table depicting all dual-guide combinations used

Guide combinations	Guide 1 (5' - 3')	Guide 2 (5' - 3')	Distance apart (base pairs)	Coding region excised (base pairs)	Frameshift?	
rs6894815 & rs10064478	GAGACTGAGACTGAAGACAG	TGCCTGTAATCACAGCTACT	419	Only removes intronic region	N/A	✓
rs10064478 & CI-3	TGCCTGTAATCACAGCTACT	CCCAGTTTTCTGTATTCGCG	602	103	34.33333333	?
CI-2 & rs6894815	TCACAACGTTGAGTATACAG	GAGACTGAGACTGAAGACAG	1238	125	41.66666667	✓
rs1989972 & CI-1	AGGGCTGTATTACTGGGGCT	CACCAACAGGCAAGGCCCGG	2021	65	21.66666667	×
CI-1 & R124H	CACCAACAGGCAAGGCCCGG	TCAGCTGTACACGGACCACA	2268	Cut site in exon, difficult to predict	Unknown	✓
CI-4 & rs11956252	AGAAGTTGGTAACGTCAAAT	CATCGCCTCCCCAAGTGATG	4008	632	210.6666667	✓
rs6860369 & rs6894815	CAAATCAGGAGGCCCCTCGT	GAGACTGAGACTGAAGACAG	4582	393	131	×
rs2282790 & rs1989972	TAGCAGTGCCAAGTAACTGA	AGGGCTGTATTACTGGGGCT	14131	394	131.3333333	✓
rs1989972 & rs6894815	AGGGCTGTATTACTGGGGCT	GAGACTGAGACTGAAGACAG	18081	1444	481.3333333	✓
rs2282790 & rs11956252	TAGCAGTGCCAAGTAACTGA	CATCGCCTCCCCAAGTGATG	39771	0	0	✓
rs72794904 & rs4976470	GGATCTATACCATGTGGGCT	CCCGTGACATGTGGGGATTA	63428	2051	683.6666667	✓

Supplementary table 7: Distance in base pairs between dual combinations consisting of only ASNIP guides

Guide Combo	Distance	
rs72794904 & rs2282790	18691	Both in 5' UTR
rs2282790 & rs1989972	14,131	
rs1989972 & rs6860369	13524	
s6860369 & rs6894815	4582	
rs6894815 & rs10064478	419	
rs10064478 & rs11956252	6560	
rs11956252 & rs7725702	3607	Both in 3' UTR
rs7725702 & rs4976470	1113	Both in 3' UTR

Supplementary table 8: Common-intron guide sequences

Common intronic guides	Guide sequences (5' - 3')
Common Intron 1 (CI-1)	CACCAACAGGCAAGGCCCGG
Common Intron 2 (CI-2)	TCACAACGTTGAGTATACAG
Common Intron 3 (CI-3)	CCCAGTTTTCTGTATTCGCG
Common Intron 4 (CI-4)	AGAAGTTGGTAACGTCAAAT

Supplementary table 9: Guide sequences for the common-intronic guides

Guide combinations	Guide 1 (5' - 3')	Guide 2 (5' - 3')	Distance apart (base pairs)
rs10064478 & CI-3	TGCCTGTAATCACAGCTACT	CCCAGTTTTCTGTATTCGCG	602
CI-2 & rs6894815	TCACAACGTTGAGTATACAG	GAGACTGAGACTGAAGACAG	1238
rs1989972 & CI-1	AGGGCTGTATTACTGGGGCT	CACCAACAGGCAAGGCCCGG	2021
CI-1 & R124H	CACCAACAGGCAAGGCCCGG	TCAGCTGTACACGGACCACA	2268
CI-4 & rs11956252	AGAAGTTGGTAACGTCAAAT	CATCGCCTCCCCAAGTGATG	4008

Paper IV

5 Paper IV- Delivery of CRISPR/Cas9 by a dual-AAV system to treat FECD

Kathleen A. Christie, Eleonora Maurizi, Caroline Conway, Marie Lukkassen, Paul S. Cassidy, Shyamasree De Majumdar, Kevin Blighe, Davide Schiroli, Laura C. Mairs, Hildegard Büning, Pete Humphries, Colin Willoughby, M. Andrew Nesbit, C.B.Tara Moore

The main aims of this paper were to:

1. Identify an AAV vector that could efficiently transduce the corneal endothelium *in vivo* via an intracameral injection
2. Design an allele-specific sgRNA using the *Col8a2* mutation, causative of Fuchs' endothelial corneal dystrophy
3. Determine the allele-specificity of this guide *in vitro*
4. Package this guide into a dual-AAV vector system and deliver packaged vectors via an intracameral injection
5. Determine if indels occurred in the cornea *in vivo*

Contributions

I carried out all HCE-S and B4G12 transductions and prepared all samples for FACS analysis. I performed all intracameral injections in mice for both AAV-GFP and AAV-Cas9 & AAV-sgRNA. I performed all imaging using the *in vivo* IVIS imager. I wrote the manuscript and made all of the figures.

Delivery of CRISPR/Cas9 by a dual-AAV system to treat FECD

Kathleen A. Christie[1], Eleonora Maurizi[1], Caroline Conway[1], Marie Lukkassen[3], Paul S. Cassidy[4], Shyamasree De Majumdar[1], Kevin Blighe[2], Davide Schiroli[1], Laura C. Mairs[1], Hildegard Büning, Pete Humphries[4], Colin Willoughby[1], M. Andrew Nesbit[1], C.B.Tara Moore[1,2*]

Affiliations

[1] Biomedical Sciences Research Institute, Ulster University, Coleraine, Northern Ireland, BT52 1SA, UK

[2] Avellino Laboratories, Menlo Park, California, CA 94025, USA 1

[3] Department of Molecular Biology and Genetics, Aarhus University, Denmark

[4] Ocular Genetics Unit, Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin, Ireland

[5] Institute for Experimental Hematology, Hannover Medical School, Hannover, Germany

* To whom correspondence should be addressed

Biomedical Sciences Research Institute, University of Ulster, Coleraine, Northern Ireland BT52 1SA, United Kingdom. Tel no: +44(0)2870124577. Email: t.moore@ulster.ac.uk

Abstract

Gene therapy offers a promising treatment strategy for the corneal dystrophies, however the inability to achieve efficient and potent delivery to the corneal layers poses a substantial hurdle for the translation of these therapies to the clinic. We compared 3 AAV serotypes for their ability to transduce human corneal epithelial and endothelial cells *in vitro*. Furthermore, we tested the ability of these 3 serotypes to transduce the corneal layers *in vivo* following an intracameral injection. We demonstrate AAV-2/9 has the ability to successfully transduce all corneal layers following an intracameral injection, identifying a vector capable of delivering gene therapy reagents to all corneal layers. Finally, we package *S.pyogenes* CRISPR/Cas9 into AAV-2/9 in a dual-vector system and demonstrate it is possible to achieve 25.7% editing efficiency in the whole cornea. These findings indicate both efficient delivery and editing by CRISPR/Cas9 is possible for all of the corneal layers providing hope for these blinding dystrophies.

Introduction

The cornea is an avascular, transparent tissue found in the anterior segment of the eye. There are a number of corneal dystrophies that all act to alter the shape or transparency of the cornea[1]. Fuchs' endothelial corneal dystrophy (FECD) is a common, age-related, inherited degenerative disease of the corneal endothelium which in advanced disease affects all layers of the cornea. It is identified by the presence of corneal guttae, which are excrescences of Descemet's membrane. These corneal guttae are associated with a progressive loss of corneal endothelial cells, the loss in endothelial cells below a critical threshold results in the inability of the corneal endothelium to successfully dehydrate the stroma; causing fluid accumulation in the stroma and the development of painful epithelial bullae leading to corneal clouding and a reduction in visual acuity.[2] In the US, approximately 5% of Caucasians over 40 years of age exhibit corneal guttae which may develop to corneal decompensation.[3] There are two categories of FECD; early-onset which presents with symptoms in the first decade and late-onset which presents with symptoms in the sixth decade.[4] Within these two categories there are 8 subtypes, in all subtypes of FECD the genetic locus has been identified, however the causative gene has not been elucidated in all cases.[5] (Table 1)

Due to the monogenic, penetrant genetics of these dystrophies they are an ideal target for gene editing solutions. A requisite for efficient gene editing is the selection of a robust delivery vehicle. Adeno associated virus (AAV) has been the most widely used vector in ocular gene replacement therapy thus far; the success of AAV in gene augmentation is largely due to its lack of immune response and persistence in the nucleus, AAV-2 has been extensively used in clinical trials for the treatment of Leber's congenital amaurosis[6]. A multitude of AAV vectors of various serotypes have been shown to successfully transduce all layers of the cornea. Of a study testing AAV-2/1, AAV-2/2, AAV-2/5 and AAV-2/8 on both human corneas *ex vivo* and mouse *in vivo*, AAV-2/8 was found to be most efficient in both cases.[7] The different serotypes

were delivered via an intrastromal injection and AAV-2/8 was found to achieve long-term transgene expression in the stroma keratocytes. AAV-2/9 and AAV-2/8 have shown successful transduction of the superficial cells of the stromal layer after epithelial debridement, with limited transduction of the same cells by AAV-2/6. [8] Following an intrastromal injection in human corneas *ex vivo* an AAV-8 and AAV-9 chimeric capsid vector (AAV8G9) was shown to successfully transduce the stromal layers in addition to some endothelial cells.[9] AAV-2/9 has been shown to successfully transduce the cornea endothelium in mouse *in vivo*, via an intracameral injection.[10] A novel synthetic AAV, Anc80L65, was shown to efficiently transduce the corneal stroma and endothelial cells. [11]

CRISPR/Cas9 has proven to be a robust tool for mammalian gene editing and indeed ocular gene editing in the retina[12]. CRISPR/Cas9 gene editing requires a i) Cas9 nuclease ii) single guide RNA (sgRNA) and iii) protospacer adjacent motif (PAM), which for *S.pyogenes* Cas9 is 5' – NGG – 3'. Cas9 will search the genome for the NGG PAM, it will then determine if the sgRNA binds specifically to the adjacent sequence. If the sequences are sufficiently complementary the Cas9 will bind and generate a double strand break (DSB), different gene editing outcomes can be achieved due to the repair processes of the DSB. Early onset FECD is caused by dominant negative missense mutations in *Col8a2*. Due to the dominant negative nature of these mutations ablation of only the mutant allele would potentially be a viable treatment strategy.

Here we demonstrate successful transduction of all corneal layers by an intracameral injection of AAV-2/9, in addition we demonstrate gene editing of a *Col8a2* mutation causative of early-onset FECD. This is the first study to our knowledge of efficient gene editing of the corneal endothelium and acts as a proof of concept study for gene editing in the corneal endothelium.

Results

Assessment of transduction efficiency of AAV-GFP serotypes *in vitro*

There have been a plethora of studies using AAV to target the anterior segment of the eye but there is little uniformity in both the vectors used, the mode of delivery and whether the study was carried out *in vitro*, *ex vivo* or *in vivo*. Thus we have tested 3 different single-stranded AAV serotypes in our delivery systems to determine the most efficient vector for the corneal layers. AAV-2 is the most characterised AAV serotype so we decided to test AAV-2 in addition to 2 AAV hybrids containing the AAV-2 genome and AAV-5 and AAV-9 caspids. The 3 vectors tested in this manuscript will be denoted as AAV-2/2, AAV-2/5 and AAV-2/9. Initially the 3 AAV-GFP vectors were transduced with two different multiplicities of infection (MOI 1 000 and MOI 10 000) into human corneal epithelial cells (HCE-S) and human corneal endothelial cells (B4G12), 48 hours after transduction the % of GFP+ cells were determined as a measure of transduction efficiency. For HCE-S AAV-2/2 achieved 41.65% transduction at a MOI of 10,000 compared to 9.25% and 0.1% with AAV-2/5 and AAV-2/9 respectively (Figure 1a). While for B4G12s AAV-2/5 achieved 63.28% transduction at a MOI of 10,000 compared to 32.73% and 21.8% with AAV-2/2 and AAV-2/9 respectively (Figure 1b).

Assessment of transduction efficiency of AAV-GFP serotypes *in vivo*

However, several studies have reported that *in vitro* and *in vivo* tropisms can vary substantially[13–15]; thus we decided to proceed to *in vivo* with all aforementioned vectors. AAV serotypes expressing GFP were injected by intracameral injection; *in vivo* imaging confirmed successful transduction of the anterior segment by all 3 vectors at day 4 (Figure 2 a-c), at day 7 fluorescence was not detectable by AAV-2/5(Figure 2d), suggesting that the cornea itself was not transduced and that the original signal came from the aqueous drainage through the trabecular meshwork into Schlemm's canal. The IVIS is able to detect a fluorescence signal from the anterior surface but it is necessary to confirm from which corneal layer the signal originates, AAV-2/9 appeared to have sustained expression (Figure 2c and f) and fluorescent

microscopy confirmed AAV-2/9 to successfully transduced all corneal layers (Figure 2 g) and this vector was chosen for the packaging of gene-editing reagents.

Mutational analysis of L450W mutation and evaluation of allele-discrimination *in vitro*

Mutational analysis of L450W, a dominant negative missense mutation in *Col8a2,* revealed that the TTG>TGG change resulted in a novel PAM for *S.pyogenes* Cas9 (Figure 3a), it has previously been reported allele-specific editing can be achieved by designing a sgRNA utilising a novel PAM[16–18]. Thus a sgRNA utilising the PAM created by L450W was designed (sgL450W) and the ability of sgL450W to distinguish between wild-type and mutant alleles of *Col8a2* was determined via an *in vitro* digestion. sgL450W appeared to selectively cleave the mutant *Col8a2* allele *in vitro* and was chosen to be packaged into AAV-2/9 (Figure 3b).

Allele specific editing of L450W mutation *in vivo*

All of the necessary components for CRISPR/Cas9 gene editing exceed the packaging capacity of AAV (<5 kb), therefore a dual-AAV CRISPR/Cas9 system was employed in which *S.pyogenes* Cas9 and sgL450W were packaged separately (Figure 4a). The dual-AAV system was co-injected into L450W knock-in mice by intracameral injection and indels were detected via TIDE analysis[19], which revealed 25.7% of indels in the whole cornea (Figure 4b).

Discussion

The corneal epithelium is known to be a rapidly dividing tissue, with complete turnover of the murine corneal epithelium taking around 3 weeks[20] while in the human it is reported to take 1-2 weeks[21], however the corneal endothelium consists of non-dividing cells[22]. In the case of the corneal epithelium successful gene disruption by NHEJ would require gene editing to occur in the limbal epithelial stem cells (LESC) as it will be these cells that repopulate the tissue, LESCs are located deep in the palisades of Vogt therefore efficient targeting of these cells may prove challenging[23]; in contrast, as the corneal endothelium is non-dividing once a productive edit

occurs within a cell the mutant protein will no longer be produced for the lifetime of that cell as such the corneal endothelium is a promising tissue to target for gene disruption via NHEJ.

Here we target *Col8a2* as a proof of concept for gene editing in the corneal endothelium; mutations in *COL8A2* cause early-onset Fuchs corneal dystrophy and exhibit mendelian inheritance and complete penetrance, therefore would appear to be a good candidate for gene therapy. However, as *COL8A2* only has 2 exons and the disease causing mutations occur in the last exon, restrictions of nonsense mediated decay (NMD) quality checks would limit gene disruption via NHEJ as a targeting strategy for this gene[24,25]. Such that, NMD will only occur if the premature stop codon resides ≥50-55 nucleotides upstream of the 3' most exon-exon junction, as the *COL8A2* mutation is located in the last exon NMD would not occur thus a truncated protein will be produced that could potentially exacerbate the phenotype. Thus, this study acts as a proof-of-concept for viral gene editing in the corneal endothelium that can be applied to more suitable causative mutations. A trinucleotide repeat expansion (TNR) in transcription factor 4 (TCF4) is the most prevalent cause of FECD, 70% of FECD cases are due to this TCF4 TNR. Delivery of a dual-guide approach to remove this expansion would offer a promising avenue for treatment.

A pressing concern of viral delivery of gene-editing reagents is the long-term expression of the transgene. The persistent expression of Cas9 may increase the risk of off-target cleavage at genomic sites other than the intended target. One possible strategy to overcome this limitation would be to drive expression of the transgene by an inducible promoter. O'Callaghan *et al* demonstrated it was possible to transduce the corneal endothelium with AAV-2/9 expressing MMP-3 under control of a doxycycline-inducible promoter and induce expression by topical application of doxycycline[10]. In addition, in a dual-AAV system it is possible to package Cas9 in one vector and a multiplexed CRISPR array in the other vector, containing guide sequences

for both the intended target and the Cas9 protein itself[26,27]. Both Cas9-sgRNA complexes will be produced, allowing on-target cleavage to be achieved in addition to the inactivation of Cas9 thus limiting its expression.

Due to the cornea's natural function as a structural barrier to the entry of foreign bodies into the eye, delivery to the cornea poses a substantial challenge. Delivery of gene editing reagents via AAV offers a promising approach to overcome this hurdle and facilitate gene editing strategies for the corneal dystrophies.

Acknowledgements

This work has been supported by Avellino Labs.

Methods

Cell culture and transduction of AAV serotypes

B4G12 cells represent a model of differentiated human corneal endothelial cells[28], culture flasks were coated with 10 µg/ml laminin (Sigma) and 10 mg/ml chondroitin sulfate (Sigma), cells were cultured in Human Endothelial-SFM supplemented with 10 ng/ml human recombinant bFGF (Thermo). HCE-S, a spontaneously immortalised human corneal epithelial cell line (a gift from J.T. Daniels, Institute of Ophthalmology, University College London, UK)[29] were grown in DMEM medium (GlutaMAX; Invitrogen, UK) supplemented with 10% Fetal Bovine Serum (FBS)(Thermo Fisher, UK). Cells were incubated at 37° C with 5% CO2 and passaged following standard laboratory procedures. For viral vector transduction, cells were infected with AAV serotypes at a multiplicity of injection of 1,000 or 10,000 viral genomes/cell. Cells were harvested 48 hours after transduction and transduction efficiency was determined by measuring % of GFP+ cells via flow cytometry (Beckman Coulter).

Intracameral injection of AAV serotypes

To assess delivery of AAV to the cornea injections were performed on wild-type C57BL/6 mice, to assess allele-specific gene editing injections were performed on L450W *Col8a2* knock-in mice previously reported[30]. Animals were anaesthetised by intra-peritoneal injection of ketamine and xylazine. Pupils were dilated using one drop of tropicamide and phenylephrine on each eye. 1.5 µl of virus at a stock titre of $1x10^{13}$ vector genomes per ml was initially back-filled into a glass needle (ID1.0 mm, WPI) attached to a Hamilton 10uL Syringe (ESSLAB). An additional 1 µl of air was then withdrawn into the needle. Animals were injected intracamerally just above the limbus. The air bubble prevented the reflux of virus/aqueous back through the injection site when the needle was removed. Fucidic gel was applied topically following injection as an antibiotic agent.

IVIS *in vivo* imaging of fluorescence

To assess delivery of AAV to the cornea, experiments measuring fluorescence were performed on wild-type C57BL/6 mice, all mice used for live imaging were aged between 12 and 25 weeks old. For imaging, mice were anaesthetised using 1.5–2% isoflurane (Abbott Laboratories Ltd., Berkshire, UK) in ~1.5 l/min flow of oxygen. A Xenogen IVIS Lumina (Perkin Elmer, Cambridge, UK) was used to quantify fluorescence.

Fluorescent microscopy

Eyes were enucleated 7 days post-injection of AAV, was washed in PBS and fixed in 4% paraformaldehyde overnight at 4 °C. Fixed eyes were frozen in PolyFreeze (Sigma) in an isopropanol bath immersed in liquid nitrogen and cryosectioned (CM 1900, Leica Microsystems) at 12 µm thick sections. Sections were gathered onto Superfrost plus slides (VWR).

In vitro digestion to determine on-target specificity

A double-stranded DNA template was prepared by amplifying a region of the luciferase reporter plasmid containing either wild-type or mutant *Col8a2* sequence using the primers listed in Supplementary Table 1. A cleavage reaction was set up by incubating 30nM *S.pyogenes* Cas9 nuclease (NEB UK) with 30nM synthetic sgRNA (Synthego) for 10 minutes at 25˚C. The Cas9:sgRNA complex was then incubated with 3nM of DNA template at 37˚C for 1 hour. Fragment analysis was then carried out on a 1% agarose gel.

Sanger sequencing across target locus

7 days post injection gDNA was extracted from the whole cornea (QIAamp DNA Mini Kit, Qiagen) and PCR amplified using primer pairs listed in Supplementary Table 1. PCR products were purified using the Wizard® PCR Preps DNA Purification System (Promega). PCR products then underwent Sanger sequencing.

References

1. Klintworth, G. K. Corneal dystrophies. Orphanet J. rare Dis. Cornea Dystrophy Foundaion 4, 7 (2009).

2. Borboli, S. & Colby, K. Mechanisms of disease: Fuchs' endothelial dystrophy. Ophthalmology Clinics of North America 15, 17–25 (2002).

3. Lorenzetti, D. W. C., Uotila, M. H., Parikh, N. & Kaufman, H. E. Central Cornea Guttata. Am. J. Ophthalmol. 64, 1155–1158 (1967).

4. Vedana, G., Villarreal, G. & Jun, A. S. Fuchs endothelial corneal dystrophy: Current perspectives. Clinical Ophthalmology 10, 321–330 (2016).

5. Weiss, J. S. et al. IC3D Classification of Corneal Dystrophies—Edition 2. Cornea 34, 117–159 (2015).

6. Pierce, E. A. & Bennett, J. The status of RPE65 gene therapy trials: Safety and efficacy. Cold Spring Harb. Perspect. Med. 5, (2015).

7. Hippert, C. et al. Corneal transduction by intra-stromal injection of AAV vectors in vivo in the mouse and Ex vivo in human explants. PLoS One 7, (2012).

8. Sharma, A., Tovey, J. C. K., Ghosh, A. & Mohan, R. R. AAV serotype influences gene transfer in corneal stroma in vivo. Exp. Eye Res. 91, 440–448 (2010).

9. Vance, M. et al. AAV Gene Therapy for MPS1-associated Corneal Blindness. Sci. Rep. 6, 22131 (2016).

10. O'Callaghan, J. et al. Therapeutic potential of AAV-mediated MMP-3 secretion from corneal endothelium in treating glaucoma. Hum. Mol. Genet. 26, 1230–1246 (2017).

11. Wang, L., Xiao, R., Andres-Mateos, E. & Vandenberghe, L. H. Single stranded adeno-associated virus achieves efficient gene transfer to anterior segment in the mouse eye. PLoS One 12, (2017).

12. DiCarlo, J. E., Mahajan, V. B. & Tsang, S. H. Gene therapy and genome surgery in the retina. J Clin Invest 128, 2177–2188 (2018).

13. Lipkowitz, M. S. et al. Transduction of Renal Cells in Vitro and in Vivo by Adeno-Associated Virus Gene Therapy Vectors. J. Am. Soc. Nephrol. 10, 1908–1915 (1999).

14. Richter, M. et al. Adeno-associated virus vector transduction of vascular smooth muscle cells in vivo instrumented dogs Adeno-associated virus vector transduction of vascular smooth muscle cells in vivo. Physiol. Genomics 2, 117–127 (2000).

15. Mohan, R. R., Tovey, J. C. K., Sharma, A. & Tandon, A. Gene therapy in the Cornea: 2005-present. Progress in Retinal and Eye Research 31, 43–64 (2012).

16. Courtney, D. G. et al. CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting. Gene Ther. 23, 108–112 (2016).

17. Li, Y. et al. Exploiting the CRISPR/Cas9 PAM constraint for single-nucleotide resolution interventions. PLoS One 11, (2016).

18. Christie, K. A. et al. Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. Sci. Rep. 7, (2017).

19. Brinkman, E. K., Chen, T., Amendola, M. & Van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 42, (2014).

20. Collinson, J. M. et al. Clonal analysis of patterns of growth, stem cell activity, and cell movement during the development and maintenance of the murine corneal epithelium. Dev. Dyn. 224, 432–40 (2002).

21. Hanna, C. et al. Cell Turnover in the Adult Human Eye. Arch. Ophthalmol. 65, 695–698 (1961).

22. Bourne, W. M., Nelson, L. I. L. & Hodge, D. O. Central corneal endothelial cell changes over a ten-year period. Investig. Ophthalmol. Vis. Sci. 38, (1997).

23. Tseng, S. C. Concept and application of limbal stem cells. Eye (Lond). 3 (Pt 2), 141–157 (1989).

24. Popp, M. W. & Maquat, L. E. Leveraging rules of nonsense-mediated mRNA decay for genome engineering and personalized medicine. Cell 165, 1319–1332 (2016).

25. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. Trends in Biochemical Sciences 23, 198–199 (1998).

26. Chen, Y. et al. A Self-restricted CRISPR System to Reduce Off-target Effects. Mol. Ther. 24, 1508–1510 (2016).

27. Moore, R. et al. CRISPR-based self-cleaving mechanism for controllable gene delivery in human cells. Nucleic Acids Res. 43, 1297–1303 (2015).

28. Valtink, M., Gruschwitz, R., Funk, R. H. W. & Engelmann, K. Two clonal cell lines of immortalized human corneal endothelial cells show either differentiated or precursor cell characteristics. Cells. Tissues. Organs 187, 286–94 (2008).

29. Notara, M. & Daniels, J. T. Characterisation and functional features of a spontaneously immortalised human corneal epithelial cell line with progenitor-like characteristics. Brain Res. Bull. 81, 279–286 (2010).

30. Meng, H. et al. L450W and Q455K Col8a2 knock-in mouse models of fuchs endothelial corneal dystrophy show distinct phenotypes and evidence for altered autophagy. Investig. Ophthalmol. Vis. Sci. 54, 1887–1897 (2013).

Table 1: Subtypes of Fuchs' endothelial corneal dystrophy and causative genes

Corneal Dystrophy		Genetic Locus	Genetic Known	Causative Gene
FECD	Early Onset	1q34.3-p32 (FECD 1)	Yes	*COL8A2*
	Late Onset	13pter-q12.3 (FECD 2)	Some cases	Unknown
		18q21.2-q21.3 (FECD3)		*TCF4*
		20p13-q12 (FECD 4)		*SLC4A11*
		5q33.1-q35.2 (FECD 5)		Unknown
		10p11.2 (FECD 6)		*ZEB1*
		9p24.1-p22.1 (FECD 7)		Unknown
		15q25 (FECD 8)		*AGBL1*

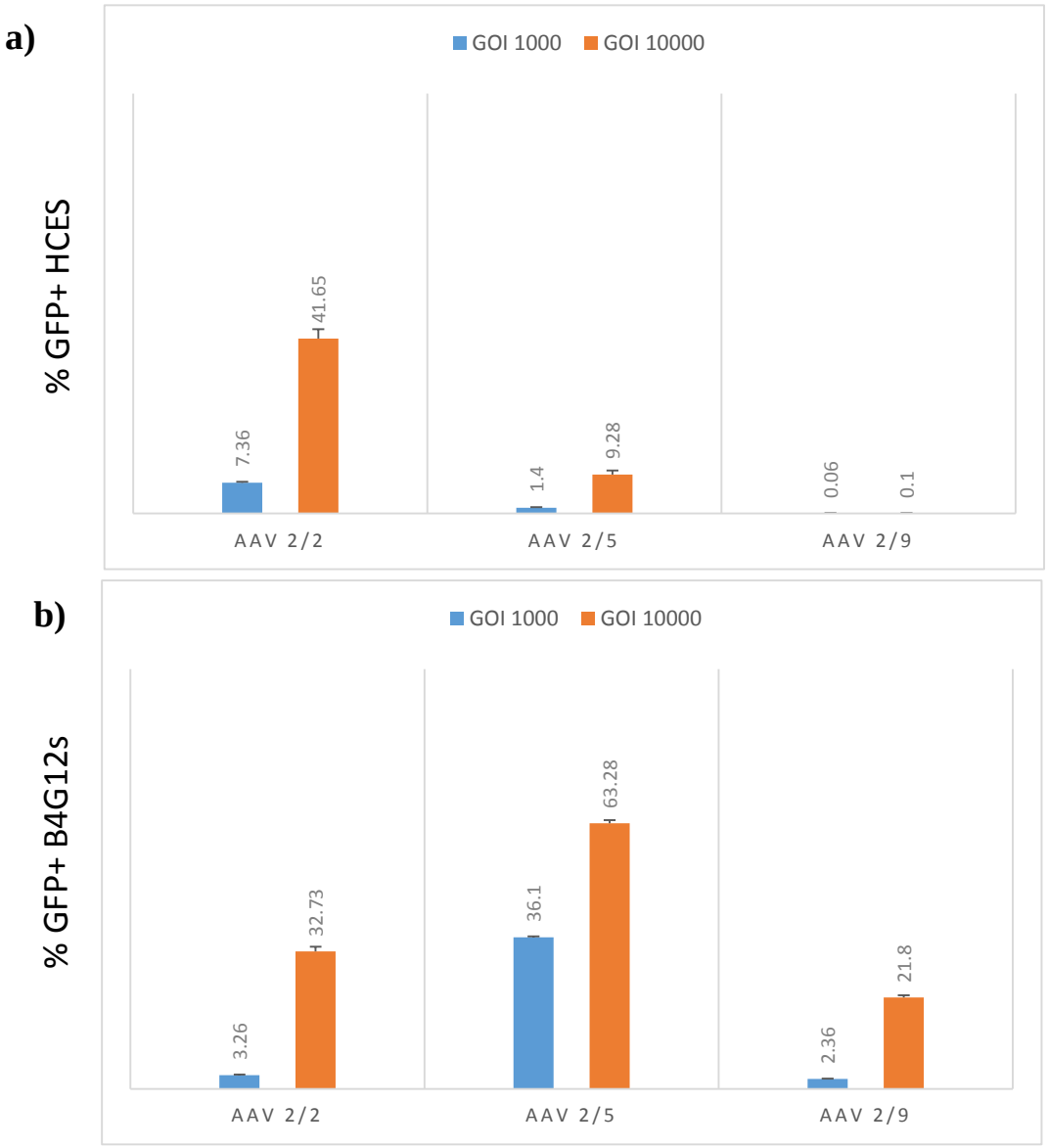


Figure 1: a) Assessment of transduction efficiency of AAV-GFP serotypes in Human Corneal Epithelial cells (HCES) – HCES were transduced with AAV serotypes (AAV 2/2, AAV 2/5, AAV 2/9) carrying GFP as the transgene at GOI 1000 and GOI 10000. 48 hours after transduction %GFP+ cells were measured using flow cytometry, n=3. AAV 2/5 appeared to transduce HCES with the highest efficiency. **b) Assessment of transduction efficiency of AAV-GFP serotypes in Human Corneal Endothelial cells (B4G12s)** – B4G12s were transduced with AAV serotypes (AAV 2/2, AAV 2/5, AAV 2/9) carrying GFP as the transgene at GOI 1000 and GOI 10000. 48 hours after transduction %GFP+ cells were measured using flow cytometry, n=3. AAV 2/5 appeared to transduce B4G12s with the highest efficiency.

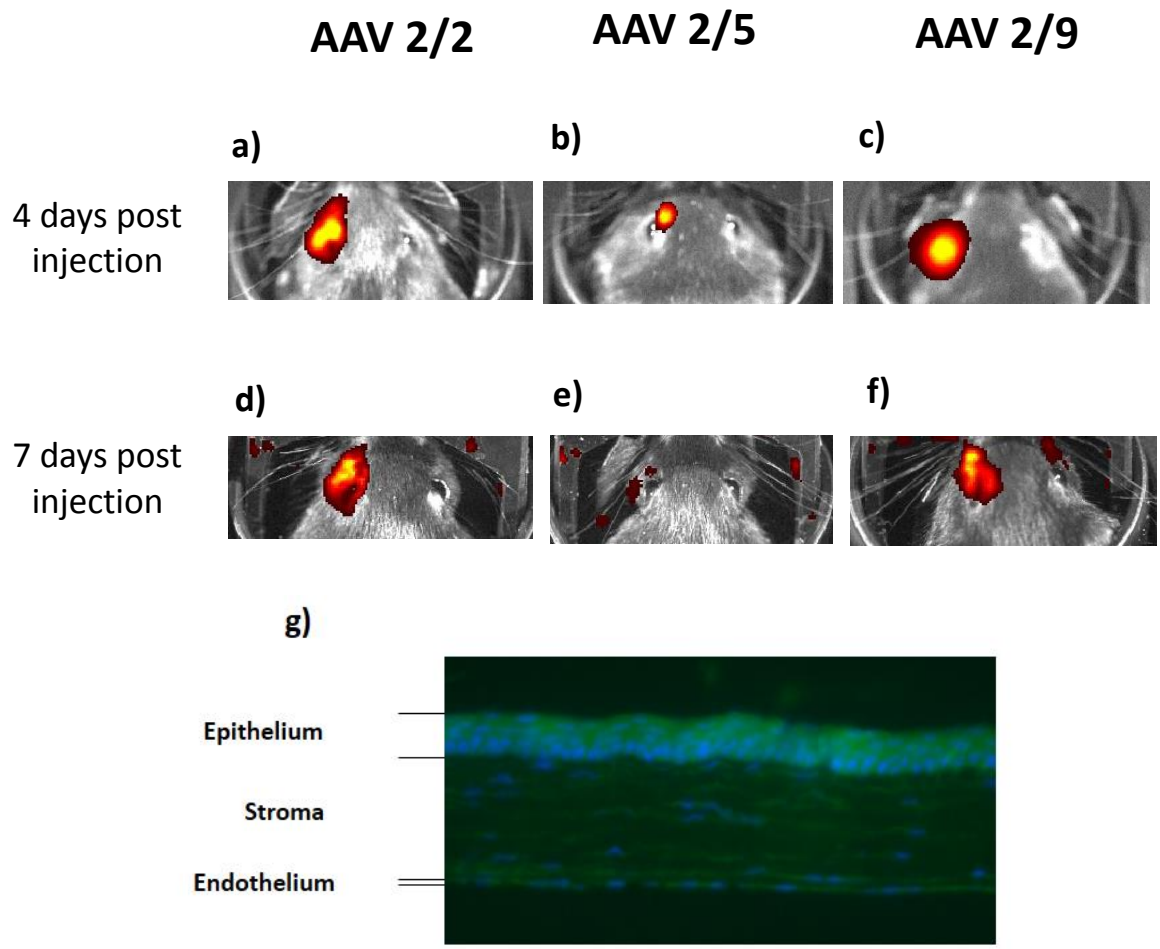
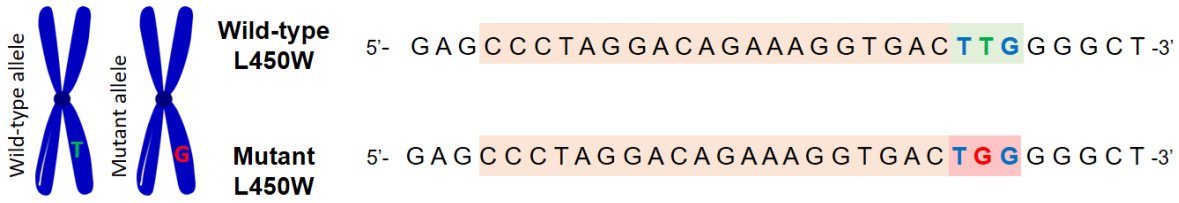


Figure 2: a-f) *In vivo* imaging of 3 AAV-GFP serotypes delivered via intracameral injection into the mouse eye – AAV serotypes (AAV 2/2, AAV 2/5, AAV 2/9) containing GFP as the transgene were all delivered via an intracameral injection into the left eye and the right remained untreated as a control. Using an IVIS *in vivo* imager the presence of GFP was detected in real-time 4 and 7 days post injection. The red and yellow colour represents the intensity of the fluorescence and not the colour of the fluorescent signal. At day 7 AAV 2/2 and AAV 2/9 both showed ability to successfully transduce the ocular surface. **g) Fluorescent microscopy revealed AAV 2/9 successfully transduced all corneal layers following an intracameral injection** – 7 days post injection the eyes were harvested for cryosectioning and fluorescent microscopy. AAV 2/9 demonstrated potent transduction of all corneal layers, providing a single vector for all corneal dystrophies.

a)



b)

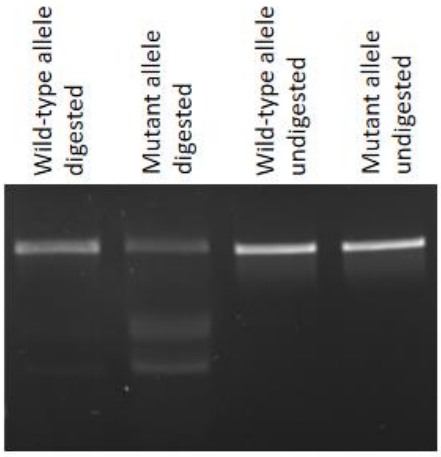


Figure 3: a) Schematic to show novel PAM generated with the L450W FECD mutation – TTG>TGG results in the creation of a novel *S.pyogenes* PAM (5' – NGG – 3') on the mutant allele indicated by the red box, at this position on the wild-type allele there is a non-canonical PAM NTG shown by the green box. A guide was designed utilising this novel PAM, shown by the orange box, Cas9 should only recognise NGG as a PAM therefore allele-specific cleavage should be achieved. **b) Assessment of allele discrimination of sgLW via an *in vitro* digest -** An *in vitro* digest was used to initially test the allele discrimination of the designed sgRNA, the mutant allele appears to be preferentially cleavage when compared to the wild-type allele.

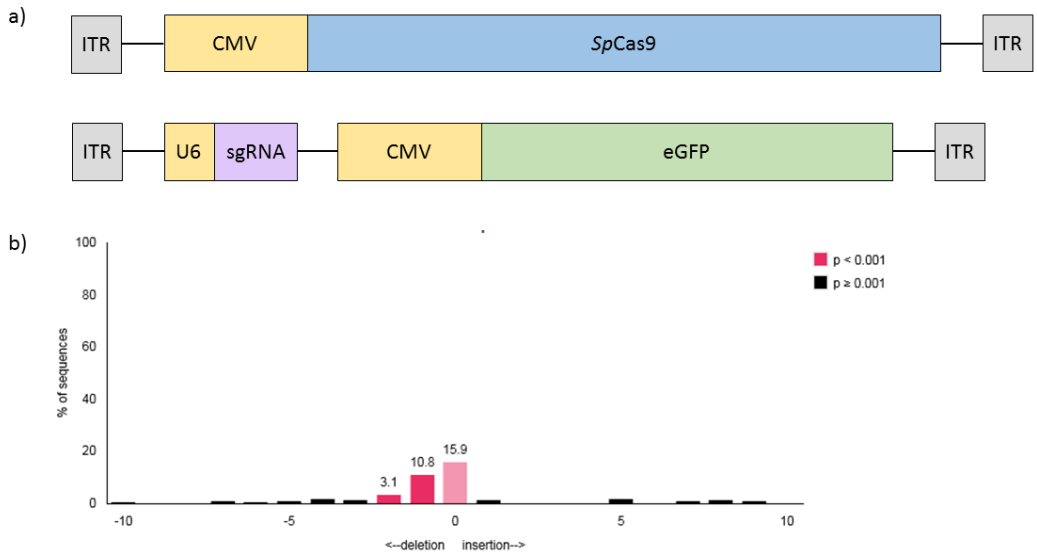


Figure 4: a) Schematic of the dual-AAV-2/9 CRISPR/Cas9 vectors utilised b) Output of TIDE analysis performed on a PCR product generated on whole cornea DNA extracted following co-injection of the dual-AAV CRISPR/Cas9 vectors.

Paper V

6 Paper V - Gene editing in the context of an increasingly complex genome

Blighe K*, DeDionisio L*, Christie KA*, Shareef S, Kakouli-Duarte T, Chao-Shern C, Harding V, Kelly RS, Stebbing J, Castellano L, Chawes B, Shaw JA, Lasky-Su JA, Nesbit MA, Moore CBT

* These authors are co-first authors

The main aims of this paper were to:

1. Highlight the complexities of the genome in relation to regulation of expression and role in disease mechanism
2. Encourage readers to progress from a 'linear' interpretation of the genome to that which encompasses a greater appreciation of the folding of the DNA molecule, DNA-RNA and -protein interactions, etc., and how these regulate expression and contribute to disease mechanism.
3. Focus on gene editing via CRISPR/Cas9 as a key technique that is likely to bring about the next frontier.

Contribution

I authored the 'Complex genetics, complex disease: room for gene editing?' section and the 'Ocular diseases' sub-section in the 'Complex genetics: a focus on 3 disease areas' section.

Blighe et al. BMC Genomics (2018) 19:595
https://doi.org/10.1186/s12864-018-4963-8

BMC Genomics

REVIEW Open Access

Gene editing in the context of an increasingly complex genome

K. Blighe[1,8,9*†], L. DeDionisio[3†], K. A. Christie[2†], B. Chawes[5], S. Shareef[7], T. Kakouli-Duarte[6], C. Chao-Shern[2,3], V. Harding[4], R. S. Kelly[1], L. Castellano[4,10], J. Stebbing[4], J. A. Lasky-Su[1], M. A. Nesbit[2] and C. B. T. Moore[2,3*]

Abstract

The reporting of the first draft of the human genome in 2000 brought with it much hope for the future in what was felt as a paradigm shift toward improved health outcomes. Indeed, we have now mapped the majority of variation across human populations with landmark projects such as 1000 Genomes; in cancer, we have catalogued mutations across the primary carcinomas; whilst, for other diseases, we have identified the genetic variants with strongest association. Despite this, we are still awaiting the genetic revolution in healthcare to materialise and translate itself into the health benefits for which we had hoped. A major problem we face relates to our underestimation of the complexity of the genome, and that of biological mechanisms, generally. Fixation on DNA sequence alone and a 'rigid' mode of thinking about the genome has meant that the folding and structure of the DNA molecule —and how these relate to regulation— have been underappreciated. Projects like ENCODE have additionally taught us that regulation at the level of RNA is just as important as that at the spatiotemporal level of chromatin.

In this review, we chart the course of the major advances in the biomedical sciences in the era pre- and post the release of the first draft sequence of the human genome, taking a focus on technology and how its development has influenced these. We additionally focus on gene editing via CRISPR/Cas9 as a key technique, in particular its use in the context of complex biological mechanisms. Our aim is to shift the mode of thinking about the genome to that which encompasses a greater appreciation of the folding of the DNA molecule, DNA- RNA/protein interactions, and how these regulate expression and elaborate disease mechanisms.

Through the composition of our work, we recognise that technological improvement is conducive to a greater understanding of biological processes and life within the cell. We believe we now have the technology at our disposal that permits a better understanding of disease mechanisms, achievable through integrative data analyses. Finally, only with greater understanding of disease mechanisms can techniques such as gene editing be faithfully conducted.

Keywords: Gene editing, Genomic complexity, Genome, Transcriptome, Epigenome, Sequencing technology development, Complex genetics, CRISPR, Integrated omics

Background

Life is more complex than we had previously thought. We have mapped the entire healthy human genome [1, 2] but many unanswered questions and challenges remain in terms of the genome's relationship with disease [3–5]. Indeed, when former President Clinton exited the White House to announce the first draft of the human genome, his words were met with the belief that we had made a paradigm shift toward a better understanding of human disease, with DNA being likened by Clinton to "*the language in which God created life*" [6]. Fast approaching 20 years since that announcement from the White House in June, 2000, and it may feel as if the fanfare that accompanied the occasion was premature. Perspective is a luxury, though, and although it can feel like research in the biological and medical sciences ('biomedical sciences') since that time has been slower than expected, we have nevertheless made huge progress, even looking far beyond the genome.

* Correspondence: kevinblighe@outlook.com; t.moore@ulster.ac.uk
†K. Blighe, L. DeDionisio and K. A. Christie contributed equally to this work.
[1]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA, USA
[2]Biomedical Sciences Research Institute, University of Ulster, Coleraine, Northern Ireland BT52 1SA, UK
Full list of author information is available at the end of the article

Indeed, international landmark projects such as the encyclopaedia of DNA elements in the human genome (ENCODE) [7] and functional annotation of the mammalian genome (FANTOM) [8] have shone much light on life's complexity through their studies on the transcriptome and epigenome, confirming the earliest conclusions by Lander and colleagues in their summary of the first human genome sequence [2]: "*The potential numbers of different proteins and protein–protein interactions are vast, and their actual numbers cannot readily be discerned from the genome sequence. Elucidating such system-level properties presents one of the great challenges for modern biology*". The challenge to which Lander alludes is still very much felt today, and these words are being confirmed as we delve even further into disease mechanisms and pathobiology.

Table 1 breast cancer *CCND1* locus. Status: unsolved

In breast cancer, germline SNPs at 11q13 in the vicinity of *CCND1* have puzzled researchers for decades. Cyclin D1 (CCND1) is key to cancer development: over-expression of CCND1 has been found in numerous cancers, whilst repression of CCND1 impairs homologous recombination-mediated DNA repair, making cells more sensitive to damaging agents. From GWAS, rs614367 is one of the SNPs most associated with ER+ (oestrogen-positive) breast cancer ($p = 10^{-39}$) [187]. The only problem with rs614367 is that it is located in a large intergenic region, upstream of *CCND1* - its function and how it alters CCND1 expression remains unknown. A separate study then found more intergenic SNPs at 11q13 in linkage disequilibrium with the original SNP, rs614367. These newly-identified SNPs are located within known enhancers and silencers of *CCND1*: PRE1 and PRE2 (putative regulatory elements 1 and 2) [188]. Their role is thought to be in modulating the binding of the ELK4 and GATA3 transcription factors, most likely modifying transcription of *CCND1*. Conclusion: The exact mechanism is still yet to be understood.

The genome

Projects like ENCODE [7] and FANTOM [8] provide evidence that it's no longer sufficient to think of DNA as the Holy Grail. Despite this, much focus and attention is still given to the genome and its usage in tackling disease through 'genomic medicine' and 'personalized medicine' [9–12]. However, there is doubt [13–15], and it has become apparent that simply knowing the sequence of DNA is not enough to fully understand disease and to drive us forward.

To take the focus completely away from the genome is to diminish its importance in disease, and we are not implying that we should ever ignore what the genome may be telling us; yet, it is clear that reading *just* the genomic sequence is not enough. Further evidence of this comes from projects such as The Cancer Genome Atlas (TCGA) [16] and International Cancer Genome Consortium (ICGC) [17], who, combined, now have the whole genome sequence of thousands of tumour-normal pairs across multiple cancers. Such information allows us to catalogue the main genes implicated in each cancer [18–21] but leaves us far from completely understanding the underlying mechanisms that are at play. For example, genome-wide association studies (GWAS) have for many years done very well at finding strong associations between SNPs and diseases of all types [22]. However, it is important to realise that the majority (roughly 95%) of statistically significant GWAS SNPs are not found in coding regions and instead lie in regions of regulatory DNA [23], a truth that leaves us to merely hypothesise on what the underlying mechanisms may be (see Table 1 for an example in breast cancer). Regretfully, GWAS have also been difficult to replicate [24–26], with Colhoun and colleagues specifically alluding to the complexity of disease traits as an issue [27]. Other issues include poor study design in both the initial and replication study as the chief causes, including small sample sizes and insufficient power, lack of comparability between cases and controls, and ignoring underlying population structure [28]. As of writing (March, 2017), the The National Human Genome Research Institute (NHGRI) [29] lists 35,329 GWAS hits reaching genome-wide significance, spanning > 1700 diseases or phenotypes, ranging from severe acne to World class endurance athleticism, variant Creutzfeldt-Jakob Disease (vCJD) to Sjögren's syndrome, etc. Despite these large efforts, our knowledge of the genetic basis of many traits is still incomplete [5]. Indeed, complete reliance on studies looking at a set of finely mapped SNPs, as in GWAS, ought to be reconsidered for future studies [30, 31].

In genomics, currently, many studies have shifted focus to rare variants in the belief that these will help us to better understand disease. The Department of Health in England has also launched a company, Genomics England, who are in the process of sequencing the genomes of patients recruited from within the National Health Service (NHS). The emphasis of Genomics England is on the study of rare diseases and the contribution of genomic variants to these (Genomics England, available from: http://www.genomicsengland.co.uk [Accessed March 4, 2017]). With the aim of sequencing 100,000 genomes, this project will undoubtedly add much to our knowledge of rare variants and rare disease but, as per other landmark sequencing projects, it will equally leave us with many questions and not bring us much closer to fully understanding disease mechanisms. The hypothesis that rare variants even contribute greatly to disease must be brought into question, and it has been [32–36]. Results from recent studies infer that complex phenotypes and diseases are in fact brought about by a mixture of both common and rare variants, each with different effect sizes [37–41]. Additionally, as monogenic diseases appear to be in the minority, with most phenotypic traits and diseases appearing to be dictated by complex genetics, sequencing projects will never advance our knowledge of these to a great extent

without thinking beyond the genome. Unfortunately, we can neither abandon these genome sequencing efforts because the information they provide is complementary to everything observed elsewhere in the cell.

The transcriptome

Including knowledge of the transcriptome with that of the genome can help to hone down the list of genomic regions that are likely to be implicated in disease and, as we'll see, the transcriptome and genome are inextricably connected. Again, in cancer, studies looking at gene expression in the past have been very successful in both segregating cancer into subtypes and also identifying the key oncogenic drivers of each [42–44]; yet, despite this, these still fail to complete our understanding of the underlying biological mechanisms for most findings. In fact, the results from ENCODE [7] prove to us that regulation at the level of the transcriptome is just as complex as that at the level of the genome, a finding echoed elsewhere in an earlier study by Mercer et al. [45]. Indeed, the original estimate on the number of protein coding genes upon the completion of the Human Genome Project (HGP) was 30,000–40,000 [2], which is a reasonable estimate, but it fails to take into account the now almost 200,000 identified transcripts and their splice isoforms that code for a messenger RNA (mRNA) that are either protein coding or have regulatory potential [7]. In fact, we now realise that only a small fraction —up to 2%— of the genome is actually transcribed into mRNA and then translated into protein [5]. Surprisingly, a much larger fraction —up to 70%— is transcribed into mRNA but not translated into protein - these are the non-coding RNAs (ncRNAs). Although for most of these ncRNAs the function (if any) remains unknown, some have been known for a long time, such as X-inactive specific transcript (XIST), which acts as an effector in female chromosome X inactivation [46]. Others, such as HOX transcript antisense RNA (HOTAIR), are strongly implicated in cancer [47]. In addition, regulation at the level of the transcriptome is intertwined with that of both itself and the genome through ncRNA interactions [48] —including micro-RNA (miRNA) [49], antisense RNA [50], long intergenic non-coding RNA (lincRNA) [51–53], etc.— and also further afield at the level of chromatin [54] and the proteome.

One could make the argument that the complexity of the transcriptome, in fact, far supersedes that of the genome due to the almost innumerable number of potential RNA interactions that can occur between DNA, proteins, and other RNA species, echoing Lander's earlier words. Transcription at a given locus is also quantifiable, with different levels of a transcript having potentially key roles in determining pathway and cell-type lineages (e.g. Sox2, Oct4, and Nanog) [55], and also functioning as buffers and dictating the transcription of other RNA species, as is seen with antisense RNA [50]. Antisense RNA transcripts are of particular interest because they stump the long held belief that transcription only occurs on a particular DNA strand. As transcription factors and enhancers do not know the rules that we believe they follow and merely bind to wherever there is an accessible matching motif, be it on the coding or non-coding strands, transcription on both strands can be expected. At certain genomic regions, transcription may even be physically 'blocked' when the same gene is being transcribed concurrently on both the sense and antisense strands as both RNA polymerases collide [50].

Many techniques are available to begin the undoubtedly difficult task of unravelling this transcriptomic complexity. For example, chromatin isolation by RNA purification sequencing (ChIRP-seq) can be used to determine regions of DNA that are bound by a RNA of interest [54], whilst crosslinking, ligation, and sequencing of hybrids (CLASH) [56] is capable of determining RNA-RNA binding. RNA-protein interactions can also be determined through multiple other techniques including RNA immunoprecipitation sequencing (RIP-seq) [57–59] (further techniques can be found in Table 2). The transcriptome is neither static within an organism and differs across different tissues and cells [8] – one could make the argument that each cell has, in fact, a unique profile, with a 'gradient' of transcription across the entire human organism's 1 trillion cells. The differences between each cell are brought about by a combination of the genetic code and both epigenetic and intrinsic and extrinsic environmental interactions, which slightly modify the transcriptional programme from one cell to the next in a gradient-like fashion.

Chromatin structure and folding

The transcriptome and its innumerable potential interactions operate within the spatiotemporal confines of densely-packed chromatin, i.e., DNA tightly wound around histones, which is itself ever changing in relation to cell cycle processes [60] and in preparation and response to transcription [61, 62]. Although research at the level of chromatin is still not a primary interest for many research groups, we are nevertheless now beginning to better appreciate the 3-dimensional structure and folding of the DNA molecule and the role that this plays in regulation and disease mechanisms. DNA 'accessibility' is also key, as much of the genome remains inaccessible to the cytosol, thus, shielding these regions —including any binding motifs within them— from transcription factors and other proteins.

Mercer and Mattick provide an outstanding review of genomic complexity, highlighting the importance of DNA-protein interactions and ncRNAs in, literally, shaping the genome and regulating gene expression in diverse ways [63]. The ability to capture the 3-dimensional

Table 2 A gambit of technological methods to interrogate the genome's complexity in every possible way

Broad area	Technique	Investigates	Description	Citation
RNA transcription, translation, and binding	ChIRP-seq	RNA-DNA binding	Chromatin Isolation by RNA purification sequencing (ChIRP-seq) is used to determine regions of the genome that are bound by a specific RNA species.	[54]
	CLASH	RNA-RNA binding	Crosslinking, Ligation, And Sequencing of Hybrids (CLASH) is capable of determining RNA-RNA binding interactions.	[56]
	GRO-seq	Active RNA transcription	Global Run-On sequencing (GRO-seq) determines the sites in the genome at which active transcription is occurring by targeting transcriptionally-engaged RNA polymerases.	[189]
	NET-seq	Active RNA transcription	Native elongating transcript sequencing (NET-seq) determines, at nucleotide resolution, the sites in the genome at which active transcription is occurring by targeting the 3'ends of nascent transcripts associated with RNA polymerases.	[190]
	Ribo-seq	Active RNA translation	Ribosome sequencing (Ribo-seq) is capable of identifying ribosome-bound messenger RNAs (mRNAs), i.e., mRNAs that are under active translation.	[191]
	TRAP-seq	Active RNA translation	Translating Ribosome Affinity Purification sequencing (TRAP-seq) quantifies all mRNAs that are associated with 80s ribosome.	[192]
	RIP-seq	RNA–protein binding	RNA Immunoprecipitation sequencing (RIP-seq) is used to determine RNA species that are bound to a RNA binding protein (RBP) of interest.	[57–59]
	HITS-CLIP	RNA-protein binding	High Throughput Sequencing Crosslinking and Immunoprecipitation (HITS-CLIP) is used to determine RNA species that are bound to a RBP of interest. HITS-CLIP is similar to RIP-seq with an added in vivo UV crosslinking step that improves specificity at the RNA-protein boundary.	[193]
	PAR-CLIP	RNA-protein binding	Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) determines RNA species that are bound to a RBP of interest. PAR-CLIP improves on HITS-CLIP and RIP-seq through the inclusion photoreactive ribonucleoside analogs, which further improves specificity at the RNA-protein boundary during crosslinking.	[194]
	iCLIP	RNA-protein binding	Individual-nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP) determines RNA species that are bound to a RBP of interest, and provides base-level specificity at the RNA-protein boundary.	[195]
	PARE-seq	miRNA target RNA	Parallel Analysis of RNA Ends sequencing (PARE-seq) looks at the 5′ ends of polyadenylated products of miRNA-mediated mRNA decay to identify miRNA-target RNA pairs.	[196, 197]
	TIF-seq PEAT	RNA transcript isoforms	Transcript Isoform Sequencing (TIF-seq) allows for the identification of transcript isoforms by mapping their exact 5′ start and 3'end boundaries.	[198, 199]
RNA form and structure	SHAPE-seq	RNA secondary and tertiary conformation	Selective 2′-Hydroxyl Acylation analyzed by Primer Extension sequencing (SHAPE-seq) utilizes SHAPE chemistry followed by multiplexed paired-end deep sequencing of primer extension products and bioinformatic analysis using a maximum likelihood model to infer secondary and tertiary RNA structure.	[200]
	PARS	RNA secondary structure	Parallel analysis of RNA structure (PARS) determines RNA secondary structure simultaneously for thousands of RNA molecules via enzymatic footprinting with different RNAses.	[201]
	Frag-seq	RNA secondary structure	Fragmentation sequencing (Frag-seq) determines RNA secondary structure transcriptome-wide via P1 endonuclease, which cleaves single-stranded nucleic acids.	[202]
	ICE	RNA inosines	Inosine Chemical Erasing (ICE) identifies inosines on RNA species in the context of adenosine-to-inosine (A-to-I) conversion, a post-transcriptional modification that diversifies the transcriptome in various pathways.	[203]
	MeRIP-seq	RNA methylation of the N^6 position of adenosine (m^6A)	Methylated RNA Immunoprecipitation sequencing (MeRIP-Seq) identifies RNA species with methylation of the N^6 position of adenosine (m^6A), a post-transcriptional RNA modification.	[204]
	Cap-seq / CIP-TAP	RNA 5′ capping	Cap sequencing (Cap-seq) and Calf Intestinal alkaline Phosphatase Tobacco Acid Pyrophosphatase (CIP-TAP) both enrich for the 5′ ends of Pol II RNA species and differ based on the following: Cap-seq is selective for long-capped RNAs; CIP-TAP is selective for capped small RNAs (csRNAs). Both therefore define Pol II transcription start sites (TSSs).	[205, 206]

Table 2 A gambit of technological methods to interrogate the genome's complexity in every possible way *(Continued)*

Broad area	Technique	Investigates	Description	Citation
DNA-protein interactions	DNase-seq	Global mapping of active regulatory chromatin, i.e., nucleosome-depleted	DNase-seq identifies regulatory regions by targeting DNase I hypersensitive (HS) sites.	[207]
	FAIRE-seq	Global mapping of active regulatory chromatin, i.e., nucleosome-depleted	Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE-seq) identifies regions of active chromatin that coincide with DNase I HS sites and others.	[208, 209]
	MNase-seq (MAINE-seq)	Global mapping of histone-bound DNA, i.e., nucleosome positioning	MNase-Assisted Isolation of Nucleosomes Sequencing (MAINE-seq) identifies histone-bound DNA via digestion by micrococcal nuclease (MN).	[210]
	ATAC-seq	Global mapping of both active regulatory chromatin and histone-bound DNA	Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) identifies regions of DNA via hyperactive Tn5 transposase, which inserts adapters into accessible regions of chromatin.	[211]
	ChIA-PET	Detects global chromatin interactions and infers 3-D structure	Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) isolates chromatin interactions by formaldehyde cross-linking, sonication, and then chromatin immunoprecipitation (ChIP). Paired chromatin DNA fragments are then connected with linkers.	[212]
	3-C, 4-C, 5-C, Hi-C	Captures interactions within and between chromosomes and infers 3-D structure	Chromosome conformation capture (3C), chromosome conformation capture on chip (4C), 3C-carbon copy (5C), and high-throughput chromosome conformation capture are methods used to identify chromatin interactions at short ranges between 2 loci (3C) or long ranges via multiple loci (Hi-C).	[213–216]
Sequence rearrangements	RC-seq	Retrotransposon insertions	Retrotransposon Capture sequencing (RC-seq) enriches for mobile the 5′ and 3′ termini of mobile genetic elements.	[217, 218]
	TN-seq / INseq	Mariner transposon insertions	Transposon sequencing (TN-seq) and Insertion sequencing (INseq) study the Himar I Mariner transposon.	[219, 220]
	TC-seq	DNA double strand break-mediated rearrangements	Translocation Capture sequencing (TC-seq) identifies AID-dependent chromosomal rearrangements.	[221, 222]

structure of a portion of chromatin can be achieved through chromosome conformation capture (3C) technology [64] - other, more complex, ways of interrogating chromatin and its interactions, including chromosome conformation capture on chip (4C), chromosome conformation capture carbon copy (5C), and high-throughput chromosome conformation capture (Hi-C), are mentioned in Table 2. Achieving this genome-wide to produce a 'structural reference chromatin', akin to the feats achieved by the HGP and ENCODE for the genome and transcriptome, respectively, is currently over-ambitious and poses a major challenge [63]. Moreover, based on what we now understand, DNA in its chromatin state is a 'fluid' molecule —not 'fixed' and static— that is constantly altering its structure inside the nucleus in relation to protein, ncRNA, and environmental interactions.

The inherent genetic makeup of each individual's genome —mainly in terms of copy number variation, SNPs, short tandem repeats, retrotransposons, etc. — would additionally translate to subtle variation in chromatin structure. Trying to delineate this level of subtlety could only be accurately predicted by entering the realm of quantum chemistry and by shifting the view of DNA from being a sequence of letters to that of a large, complex, deoxyribonucleic molecule, as it was when it was first discovered [65], which interacts with proteins and other nucleic acids in the cytosol via diverse electrochemical and electromagnetic interactions. Such work is currently being done in the quantum chemical and mechanical sciences [66–68], but is currently not a primary focus of this review. In addition, although trying to model an entire human DNA molecule in this way would be useful, it is computationally unfeasible.

With a greater appreciation of the importance and complexity of the genome, transcriptome, and epigenome, one can thus begin to imagine a very dynamic environment within the cytosol —a cellular 'microcosm' of activity—, whereby transcription is a pervasive process with transcription factors binding at numerous loci in the genome and initiating transcription where the electromagnetic potential, i.e. 'binding strength', mediated via certain DNA motifs or interactions with other proteins, is sufficiently strong such that transcription of downstream targets can ultimately occur - where the binding is not sufficiently strong, transcription of targets may be weak or not occur at all; an environment where the 'pillars' that give chromatin its shape and form, i.e., histones, are responding to environmental stressors [69] in a cell type-specific manner and, in this way, increasing or decreasing the accessibility —or 'opening up' or 'closing' loops— of certain DNA regions to factors in the cytosol, thus modifying expression profiles; finally, an environment where chemical modification of DNA bases, e.g., the addition of methyl groups (or 'methylation') is again brought about via environmental interactions and which actively hampers the expression of genes by, in part, reducing the binding of transcription factors [70, 71].

The technology that has driven research

A historical perspective: C.1980s onwards

Much of the challenge for understanding the mechanisms that drive the structure and function of nucleic acid, i.e., DNA and RNA, are limited by available technology. Although we now have numerous ways of interrogating the secrets of the genome (Table 2), automated sequencers utilising the dideoxy-sequencing method of Sanger [72] have been relied upon for DNA sequence information since 1977. The first successful automated sequencing runs utilised the Applied Biosystems (ABI) 370A and sequenced two cDNA clones encoding the muscarinic cholinergic receptor and the ß-adrenergic receptor within a rat heart cDNA library [73] - at the time, it was claimed that one sequencer could obtain > 30,000 bases with five overnight sequencing runs. Given the fact that the haploid human genome is approximately 3.5 billion bases-pairs, in 1987 sequencing one human genome on 100 of these instruments would have taken 5000 days or 13.7 years, with a cost of undoubtedly astronomic proportions.

Thus, whilst sequencing the cellular genome was first discussed as early as 1984 [74] and was a chief goal of the HGP [75], clearly no one intended to sequence an entire human genome with the ABI 370A on a routine basis. However, innovations ensued, detection methods were enhanced with the advent of capillary electrophoresis [76] and, in 2001, with multiple high throughput DNA sequencers (ABI 3700) running in tandem, the human genome was sequenced in two efforts [1, 2] with roughly 90–95% genomic coverage, and in a relatively short amount of time: 15 months [2] and 9 months [1].

These efforts provided for a momentous event in our quest to understand DNA, colloquially referred to as 'the code of life', and they provided impetus to sequence and understand DNA at an even quicker pace in the future. Whilst saying this, the first attempt to then move beyond ABI's automated sequencer was not driven by efforts to sequence the human genome; rather, "*to discover and understand the function and variation of genes*" [77]. The term massively parallel signature sequencing (MPSS) was used to describe a sequencing platform that would become the prototype for what was to follow as we entered the twenty-first century [77]. This platform was able to sequence millions of DNA strands at one time in conjunction with in vitro cloning of cDNA on microbeads. The instrument employed an innovative system that utilised a charge-coupled device (CCD) detector followed by image processing of fluorescent signals corresponding to each of the 4 deoxynucleotides.

The method harnessed biochemical and enzymatic reactions to deliver short tags that were 16 to 20 bases long, referred to as 'signature sequences'. This approach, developed as an alternative to the highly variable probe hybridising methods of microarray chips [78] was known, previous to MPSS, as serial analysis of gene expression (SAGE), which originally relied on short tags of 9 nucleotide bases [79]. Each of these methods —MPSS, SAGE, and the hybridisation method of arrayed cDNA libraries (microarrays)— relied upon previous knowledge of the mRNA sequences that code for the genes of interest. These platforms in a strict sense were not and are not DNA sequencers in the same way that a sequencer is defined today. Thus, it was impractical to expect MPSS to be able to carry out de novo sequencing on the genome of biological organisms that had not yet been deciphered.

In 2005 and 2006, after years of academic research into improved biochemical processes, two sequencing platforms emerged: the 454 sequencer [80] and the Illumina/Solexa Genome Analyzer, which both utilised sequencing by synthesis (SBS). This method, outlined in Hyman [81], involves the detection of the base-by-base addition of each of the 4 nucleotide bases facilitated by a biochemically engineered DNA polymerase. The detection method utilised in the 454 sequencer [80] takes advantage of the release of pyrophosphate (PPi), which occurs after the addition of each base, and then becomes the substrate for a coupled enzymatic reaction with *luciferase* that results in the release of light [82]. Another group at the University of Cambridge developed a platform that involved a novel single molecule approach with a laser detection system [83] that utilised nucleotides adapted with florescent and reversible 3′ terminator moieties, which in effect preserved the viability of the growing DNA molecule as it was replicated from the double-stranded template. This sequencing method became the driving force behind the technology spawned by engineers at Solexa, later acquired by Illumina [84]. A similar detection method involving fluorescently-labelled nucleotide bases was developed by a group at Columbia University [85, 86]. At the time, several competing technologies were attempting to replace the dideoxy Sanger sequencing method, then considered the gold standard for DNA sequencing [87].

What was driving this profusion of technological innovation? The goal for all of the competing technologies was to introduce a massively parallel sequencing platform that could sequence a genome in a matter of days instead of months. Thus, one could argue that we have had such an intense interest in the relationship of DNA sequence to disease due in part to the fact that the first technological successes that came out were specifically designed to read DNA sequence quickly, reminiscent of the series of technological advances that came from Apollo Program. Indeed, the concept of the 'personal genome', which envisions a world where everyone can have their genome sequenced for as little as $1000 [88], has propelled much of the change and innovation that has occurred during the past 15 years. While the technologies introduced by 454 Life Sciences in 2005 and Illumina/Solexa in 2006 demonstrated a remarkable ability to sequence DNA at a rate that was orders of magnitude faster than the ABI sequencers, they did not deliver the $1000 genome.

Then, in 2008, Baylor College of Medicine reported the sequencing of Dr. James Watson's complete genome with the 454 sequencing platform to a depth of 7.4-fold [89] - it took 2 months and cost less than US$1 million. Comparative bioinformatics revealed 3.3 million SNPs and structural variation in Dr. Watson's genome. Also in 2008, in a report outlining the SBS method first developed by Balasubramanian and Klenerman [83] at Cambridge, the genome of a male Yoruba from Nigeria was sequenced to >30× with the Genome Analyzer (Illumina/Solexa) [84], taking 8 weeks to complete at a cost of US$250,000.

Modern technological advances: C.2010 onward

The utilitarian needs that serve to advance technology often result in unanticipated discoveries that carry research in new directions. Pacific Biosciences (PacBio) developed a platform based on single-molecule real-time (SMRT) sequencing that was able to successfully sequence very long fragments of DNA [90]. In 2010, it was recognised that the SMRT technology would be able to secure read lengths greater than 1 Kbp, which far surpassed the capability of the SBS method at that time, i.e., 100-150 bp (Genome Analyzer) and 330 bp (Roche 454) [87]. Soon thereafter, the SMRT technology was utilised in a de novo sequencing method to demonstrate its ability to sequence the entire genome of a bacteria using only a single, long insert shotgun DNA library [91]. The mean length of the reads for this work was 5777 bp with a mean accuracy of 99.9%. Prior to this research conducted by Chin et al. [91], the SMRT platform was already deemed valuable as a tool for microbial phylogenetic profiling. The platform has inherent advantages over Sanger and Roche 454 for sequencing the 16S ribosomal RNA (rRNA) genes within microbial populations, which require longer reads to give finer resolution [92]. Due to the fact that the SMRT platform gives reads that are four times longer than the 454 platform and does not require a library amplification step, the cost was at that time significantly less than other sequencing technologies.

In addition to the recent proliferation of research conducted in the field of microbial profiling, longer read sequencing technologies have been utilised in attempts to produce haplotype-resolved genome sequences, i.e.

haplotype phasing. The need for this type of sequence information becomes apparent when considering hereditary disorders, which are invariably linked to the haplotype and mode of inheritance [93]. In addition to SMRT, Oxford Nanopore Technologies (ONT) also developed a platform that provides haplotype phasing; however, high error rates seen in both of these platforms proved to be a difficult hurdle to move past when it was discovered that PCR-chimera formation was not detected by software assembly programs [94]. An alternative approach to increasing the read length to gain long contiguous reads is to manipulate the upfront library preparation with a method that assigns a molecular barcode to very long (> 50 Kbp) DNA fragments, which are then sequenced with a short read NGS platform. This approach ensures that excessive chimera formation will not take place. After sequencing, bioinformatic algorithms assemble the fragments into a haplotype-resolved genomic sequence, e.g., 10× sequencing (10× Genomics, Pleasanton, USA). This method (from c.2015), along with single cell DNA and RNA sequencing, represents the current state of the art in terms of technological advances in sequencing since the HGP in 2000, and involves the attachment of several million synthetic barcodes —each to one DNA fragment within the genome of interest—, which can then furnish a de novo assembly of any genome and incidentally provide the haplotype phasing of that genome [95].

Regarding the role of PCR and NGS, it is important to grasp that, for most if not all sequencing methods, DNA amplification is a necessary preliminary step in order to increase the detection signal, whether that signal will originate from the excitation of a fluorescently labelled molecule (e.g. SBS), emitted light resulting from an enzymatic reaction (e.g. via PPi release), or the disruption of an electrical current (e.g. ONT). However, PCR-driven amplification will result in artefacts such as chimera formation, mentioned above, as well as random base modification errors [96]. To overcome base errors, NGS methods are designed to sequence at great depths of coverage to ensure that these errors —and indeed base-calling errors due to the sequencing process itself— can be bioinfomatically removed from the final data, or at best reduce their influence. For example, thresholds can be set for a minimum sequencing read depth over each base position during variant calling to ensure that errors retain less influence. On the other hand, PCR-chimera formation cannot be entirely eliminated from any NGS method without specific algorithms designed to target each region of interest within the sequencing data in order to computationally identify the chimeric events. Of importance, however, the length of the PCR amplicon affects the prevalence of chimera formation, with shorter PCR amplicons resulting in lower numbers of chimeric sequences. In saying this, when NGS is utilised to gain insight into the presence of SNPs without regard to how these variants relate to one another, in terms of haplotypes, then chimeric artefacts do not pose the same problem as when a definitive haplotype phasing determination is the goal.

Cutting edge gene editing technology

As technological advances progressed for probing the genome and far beyond this, and as knowledge contributed by academic settings about disease association variants and disease biomarkers accumulated at enormous rates, the desire to actually introduce modifications to the '*language in which God created life*' became a goal of some research groups, with controversy [97, 98]. Presently, the leading gene editing system involves CRISPR (clustered regularly interspaced short palindromic repeats)/Cas, which has been demonstrated to cleave the genome at endogenous loci in human and mouse cells [99], and to facilitate chromosomal rearrangements through sequence-specific DNA double-strand breaks (DSBs) [100] (Fig. 1). This type of gene editing often requires that the target sites be located on the same allele (cis) and it is crucial to examine the entire genome for unintended off target effects in particular when gene editing is applied for clinical applications [101]. While there have been well designed assays to determine off target effects [102], such methods do not directly sequence the entire genome of cells that have undergone CRISPR gene editing. Thus, modern technology that can produce a haplotype-resolved whole genome has much utility in the realm of gene editing, both pre- and post-experimentation.

Main text

Complex genetics, complex disease: Room for gene editing?

The CRISPR/Cas system has provided an unprecedented ability to delve further into the complexity of the genome and is a technique that is being widely discussed

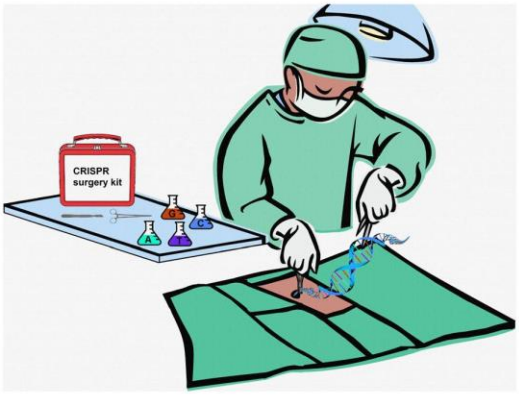


Fig. 1 'Surgery' by CRISPR

across different areas, including disease control in agriculture (see Table 3 for oversight on CRISPR and bees), drug manufacturing, 'de-extinction', vector control, food production, and others [103]. The ability to direct the Cas nuclease in a sequence-specific manner by simply altering a 20 nt guide sequence has permitted a cost-effective, high-throughput way to perform genome-wide analysis. Indeed, numerous large scale CRISPR/Cas9 knockout screens have been employed to generate loss-of-function mutations which allow functional characterisation of all annotated genetic elements [102, 104–108]. These screens have been implemented across a wide range of disciplines and have identified many promising hits, including: essential genes for cell viability, genes that confer resistance to current drug therapies, miRNAs involved in cell growth, potential cancer, and anti-viral drug targets etc. [104, 105, 107].

However, these screens have also highlighted a major issue, with researchers finding little correlation between the results from CRISPR/Cas9-driven screens and those previously carried out using techniques such as RNA interference (RNAi) [109]. A recent CRISPR/Cas9 screen for essential genes involved in tumour growth revealed that the MELK protein known to be essential in tumour growth does not drive cell proliferation in cancer cells as previously thought [110]. As CRISPR/Cas9 and RNAi mediate their effects by different mechanisms, it does not seem irrational that they can yield different results, although, drawing conclusions from contradictory results is problematic. RNAi has a well-documented tendency for off-target effects [111–115]. This underlines the need to validate results by complementary shRNA and CRISPR/Cas9 screening approaches to produce a more comprehensive analysis [105].

Table 3 Crisis 'bee'. Status: imminent problem

In recent years, domesticated honeybees (*Apis mellifera*) and commercially-reared bumblebees (*Bombus terrestris*) have become increasingly important in global crop production by enhancing pollination [223], as global agriculture faces the major challenge to maintain food security to feed an ever-increasing human population. The challenge grows bigger by the severe declines suffered by these pollinators due to land use change, causing habitat loss, fragmentation, degradation and resource diversity [224], pesticides [225], introduction of alien species for crop pollination and honey production, causing decline on native pollinators [226], and with these, introduction of bee pests and pathogens [227]. Despite extensive research efforts, no single factor has been identified as the definitive cause of bee colony decline [228, 229], and it is likely that the interaction amongst all these factors constitutes the driver for the bee losses. At global level, however, most managed *A. mellifera* colonies are infected with the ecto-parasite mite Varroa destructor, while other important bee pathogens (e.g. *Nosema spp.* and several viruses) display global distributions [227]. This points to the significance of these parasites and pathogens in interacting anywhere in the world with other bee colony decline factors, thus intensifying the problem.

The arrival of the powerful gene editing tool, CRISPR [230], could aid towards the alleviation of the situation, particularly now that we have access to honeybee [231] and bumblebee [232] genomes. Certain bee populations practice 'hive hygiene' by removing sick and infested bee larvae, and such populations are less likely to succumb to parasite pathogens [103]. Conclusion: Identification of genes associated with the hygiene behaviour and editing them in less hygienic populations would help enhance health of hives globally.

The generation of a catalytically inactive —or 'dead'— Cas9 (dCas9) introduced the possibility of fusing functional proteins to dCas9, allowing targeting in a sequence-specific manner without initiating a double strand break [116]. This has led to the generation of innovative adaptations of the CRISPR system that have greatly expanded the molecular biology toolkit and advanced both the scope and effectiveness of genome editing. Further, an inventive strategy termed 'CRISPR-X' has created a novel and rapid approach to investigate protein function [117]. It involves fusion of dCas9 to activation-induced cytidine deaminase (AID), which mediates somatic cellular hypermutation (SHM). This can be used to rapidly generate a diverse library of mutants with improved or novel functions, which can then be investigated. Another approach utilises the same enzyme to achieve 'base-editing' [118]. This provides a novel programmable way to directly change a mutated base at a greater efficiency than point mutations by homology-directed repair. However, as previously described, to get a full appreciation of complex disease, we need to look beyond the genome level. To facilitate this investigation, researchers have now generated adaptations to the CRISPR system that allow interrogation of both the transcriptome and epigenome.

CRISPR and the transcriptome

Transcriptional regulation provides a powerful approach to further the understanding of gene function and regulatory networks. However, the mechanism of transcriptional regulation in eukaryotic cells is complex and involves the interaction of many different transcription factors at DNA regulatory elements that can span large regions of DNA [119]. Previous techniques such as RNAi have been employed to investigate transcriptional repression but, as mentioned, they are prone to off-target effects that can complicate the interpretation. In addition, RNAi is limited to targeting protein coding transcripts only, whereas CRISPR interference (CRISPRi) involves the fusion to a repressive KRAB effector domain [120], thus allowing transcriptional repression beyond the coding sequence to include miRNAs, lincRNAs, ncRNAs, etc. Alternatively, fusion of dCas9 to transcriptional activation domains such as VP64 can be used to upregulate gene expression, known as CRISPR activation (CRISPRa) [120, 121].

Building on this initial approach, transcriptional activation in a real-life scenario was considered, whereby transcriptional factors act in synergy with multiple co-factors. This hypothesis resulted in a CRISPR complex termed 'Synergistic Activation Mediator' (SAM) [122]. SAM combines VP64 with additional activation

domains to further achieve higher levels of activation. The capacity to upregulate selected genes offers vast possibilities for reprogramming cellular identity in addition to understanding gene function. Furthermore, whilst wild-type Cas9 can be utilised to implement loss-of-function genome-wide screens, no technology was available previously that allows large-scale gain-of-function (GOF) screens to be conducted in a reliable and cost-effective way. Indeed, SAM was previously utilised for genome-scale transcriptional activation and resulted in the identification of genes that, upon GOF, may have resulted in resistance to a BRAF inhibitor [122].

CRISPR and the epigenome

The epigenome is a complex regulatory layer that acts in concert with the underlying DNA sequence to result in the immense array of variation that exists between cells. The epigenome has well documented strong links to disease status, for example, in its role in imprinting disorders and neurological disease [123, 124]. For many diseases, the problems may lie within this additional regulatory layer rather than the genomic sequence itself. Until now, progress in the field of epigenetics has been limited by the availability of appropriate molecular biology techniques to investigate the functional impact of deposition or removal of chromatin modifications [125]. Recent developments utilise dCas9 nuclease as a targeting domain fused to chromatin-modifying enzymes such as *Dnmt3a*, *Tet1*, *Lsd1*, or *Hat* catalytic domain of p300 [126–128]. This introduces an innovative capability to add or remove chromatin modifications in a site-specific manner, providing new insight into the downstream effects on chromatin state and gene expression of specific sequences, offering a better understanding of the role that epigenetics plays in disease. In addition, dCas9 has now been fused to EGFP or a combination of fluorescent proteins which has been called CRISPRainbow [129, 130]. This provides an insightful approach to visualise the native chromatin. The spatiotemporal organisation and dynamics of chromatin have a direct role in the functional output of genome function, and the ability to track real-time in a site-specific manner will provide another dimension of our understanding of the chromatin structure. Although these advancements introduce a new realm of possibilities for the field of epigenetics, such as advanced cellular reprogramming and functional studies, epigenome editing is still in very early stages. The effect of a stably bound Cas9 nuclease may itself affect the chromatin state and chromatin modifications, thus complicating interpretation [125]. Indeed, although much remains to be elucidated about the chromatin modification network, these advances offer promising steps in unravelling the complexity of the genome.

CRISPR in a therapeutic setting

Thus, whilst it is clear that the genome engineering revolution is fast living up to its potential, and that the wild-type CRISPR/Cas system, along with the ever-growing list of adaptations, has massively expanded our ability to investigate the genome to a new depth, two central issues persist: specificity and delivery. For CRISPR/Cas9 to be used in a therapeutic setting, these two issues need to be thoroughly addressed. Off-target cleavage is a known caveat of the CRISPR/Cas system, with many groups reporting indels at off-target sites [131, 132]. However, it is clear that initial guide-design is absolutely critical in achieving both good on-target cleavage in addition to low levels of off-target cleavage [133–135]. An attempt to rationally engineer Cas9 in order to improve the specificity has led to the development of high-fidelity Cas9 (HF-Cas9), enhanced Cas9 (eCas9), and hyper-active Cas9 variant (HypaCas9) - in all cases off-target cleavage was greatly reduced [136–138].

Furthermore, orthologues of *S. pyogenes* Cas9 from different species can be considered, which recognise more intricate PAMs (protospacer adjacent motifs) and thus have a reduced number of off-target sites within the genome [139]. Following the emergence of Cas9 for use in mammalian cells, an additional Class II nuclease, Cas12a, formerly known as Cpf1, was discovered [140]. Cas12a offers several distinct differences compared to Cas9, such as its use of T-rich PAMs and its generation of staggered-end double strand breaks with 5′ overhangs. Interestingly, Cas12a has been shown to be more specific than *S. pyogenes* Cas9, offering a promising alternative [141, 142].

Another hurdle to overcome is the delivery of the CRISPR/Cas system. For productive gene editing, an optimal delivery vehicle should be highly specific and efficient for a particular cell type, not produce an immune response, exhibit minimal genotoxicity and, in order to minimise off-target effects, the expression of the cargo should not persist for an extended period of time. Currently, no vehicle exists that meets all of these requirements; however, the field of gene-editing is nascent and the potential delivery options are continually evolving; therefore it is likely the current limitations of delivery vehicles will be overcome. Current strategies for delivery of CRISPR/Cas9 components have been extensively reviewed by Glass et al. [143].

Genome editing can additionally be only implemented in a setting where there exists a high level of understanding of the underlying disease mechanism. We now focus on 3 major disease areas in which genome editing could be applicable.

Complex genetics: A focus on 3 disease areas

Asthma

Asthma is a heterogeneous syndrome characterised by chronic airway inflammation, airway hyperresponsiveness

and intermittent airway obstruction that result in recurrent episodes of breathlessness, wheeze and cough. Asthma is emblematic of a truly complex genetic disease thought to develop through the interaction of multiple genetic loci and environmental factors and is estimated to affect approximately 300 million worldwide [144]. Asthma most often debuts during early childhood and it is currently the most common chronic disease in childhood [145] - its heritability is estimated to be up to 70% [146, 147].

The earliest childhood asthma disease-gene mapping approaches, including linkage and candidate gene based studies, had mixed results, resulting in identification of only a handful of reproducible loci. However, the advent of technical and statistical methods for comprehensive GWAS has identified numerous reproducible asthma-susceptibility loci including *ORMDL3, IL1RL1, WDR36, PDE4D, DENND1B, RAD50, IL13, IL18R1, SMAD3, HLA-DQB1, GSDMB, IL33, IL2RB, RORA, HLA-DPA1, IL6R, LRRC32, C11orf30, TNIP1* [146, 148–150]. More recently, two consortia, one European (GABRIEL) [151] and one North-American (EVE) [152], conducted independent large-scale meta-analyses of nearly all available asthma GWAS data, reporting striking overlap in the abovementioned loci, which predominantly reside in regulatory regions of the genome and are involved in immune regulation, which is an integral part of asthma pathogenesis. However, as has been observed in virtually all complex diseases, the asthma loci identified to date explain only a small proportion of the total observed heritability of the disease, suggesting that novel approaches are required to identify the additional risk variants underlying this 'missing heritability'.

The first childhood asthma GWAS identified common regulatory variants at and near the *ORMDL3*/*GSDMB*/*ZPBP2* loci on chromosome 17q21 in three populations of European ancestry, a finding that has now been confirmed in various ethnic groups. The 17q21 locus has been shown to increase the risk for an early onset, non-atopic phenotype through alterations of the sphingolipid metabolisms, resulting in bronchial hyperresponsiveness [153]. The understanding of the underlying biology of how this asthma locus operates will provide an avenue for development of new asthma drugs in the near future (see Table 4).

More recently, a genome-wide association study identified *CDHR3* as a novel susceptibility locus for early childhood asthma with severe exacerbations [154]. The *CDHR3* gene is highly expressed in airway epithelium and was, in a subsequent study, shown to be a rhinovirus C receptor of importance for both binding and replication of the virus [155]. Thus, novel therapeutics targeting this specific gene product may alleviate the burden of acute virus-induced exacerbations in children with the risk variant.

Table 4 Childhood asthma and the 17q21 locus. Status: partially solved

Childhood asthma is the most common chronic childhood disorder with up to 50% of all children experiencing asthma-like symptoms before the age of 6 years, and 15% being diagnosed with persistent asthma during school-age [233]. Asthma is considered a heterogeneous syndrome consisting of several endophenotypes with distinct clinical features, divergent underlying molecular causes, and different prevention and treatment options [234]. There is a substantial genetic contribution to asthma susceptibility and studies have revealed more than 100 implicated genes.
Importantly, one of the first GWAS studies focusing on childhood onset asthma discovered a risk locus at 17q21, increasing the risk of asthma by 20% [235], which has since then been robustly replicated across different ethnicities in large meta-GWAS consortia [151, 152]. Thereafter, it was shown that genetic risk variants in the 17q21 locus up-regulate transcription of the *ORMDL3* gene in EBV-transformed lymphoblastoid cell lines [235] and that rs12936231 is the functional SNP, which, via allele-specific changes in chromatin binding of the insulator protein CTCF, is responsible for *ORMDL3* expression [236]. However, the mechanistic link between the *ORMDL3* gene and asthma susceptibility was unknown. Further studies showed that the ORMDL3 protein is expressed in airway epithelium cells [237] and that ORMDL3 and other related orm proteins in the endoplasmic reticulum have a major role in sphingolipid homeostasis via inhibition of serine palmitoyltransferase (SPT), which is the rate-limiting enzyme in de novo sphingolipid biosynthesis [238, 239]. This finding triggered the hypothesis that the *ORMDL3* gene increases the risk of asthma through the sphingolipid metabolism [153], which has been confirmed in mouse studies showing that decreased sphingolipid biosynthesis in lung epithelial tissue [240] and SPT knockout [241] associate with airway hyper-reactivity via altered levels of ceramides, sphingosine-1-phosphate and sphingomyelins, subsequently affecting lung magnesium homeostasis.
Conclusion: Our understanding of the underlying biology of the initial GWAS discovery of 17q21 as a strong childhood asthma susceptibility locus has led to the recognition that the ORMDL3 protein, the SPT enzyme, and the sphingolipid metabolism are important players in airway reactivity and asthma pathogenesis, which may lead to novel therapeutics targeting this pathway. However, it is still unknown exactly how the sphingolipid homeostasis is regulated by expression of *ORMDL3* and external environmental perturbants, but this presumably involves a network of multiple interconnected mechanisms that can be disentangled by metabolomics studies.

Another important field in asthma genetics is pharmacogenomics, which is the study of the role of genetic determinants in the variable, inter-individual response to medications. Pharmacogenomic studies are of particular interest as up to one-half of children with asthma do not respond to treatment with inhaled β2-agonists, leukotriene modifiers, or inhaled corticosteroids. There has been numerous studies and findings, including *ADRB2* [156] and *CRISPLD2*, which has been shown to regulate the anti-inflammatory effects of corticosteroids in airway smooth muscle cells [157].

All of the above findings highlight how genetic studies in asthma have provided important and clinically-applicable knowledge that may be utilised by CRISPR in the future.

Ocular disorders

Ocular genetic disease offers distinct benefits as a test bed in the field of genome engineering. A high proportion of the causative genes in ocular diseases have been elucidated and are due to a single mutation in a single

gene [158, 159]. In addition, the eye offers unique anatomical and physiological qualities that make it amenable to treatment; it is easily accessible, has a small surface area and holds an immune-privileged status making ocular diseases an ideal system in which to develop CRISPR/Cas9 gene therapy [160].

Gene-therapy for recessive retinal diseases caused, largely, by loss-of-function mutations is more advanced than for therapies for dominant, gain-of-function diseases. There are several on-going clinical trials for retinal diseases including choroideremia, Leber congenital amaurosis (LCA), Retinitis pigmentosa, Usher syndrome, and Stargardt disease [161–165]. These therapies all employ a gene-replacement strategy in which a functional copy of the gene is introduced to target cells by either adeno-associated virus (AAV) or lentiviral vectors.

Gene-replacement is not always a viable approach as vector carrying capacity restricts the spectrum of disorders that can be treated and, while lentivirus has a larger carrying capacity, the potential for it to integrate into the genome raises safety concerns. A much more attractive treatment strategy would be to correct the defect itself, utilising the novel CRISPR technology. Editas Medicine have a clinical trial planned for LCA in which CRISPR will be targeted to delete a cryptic splice site and restore normal splicing. They have subsequently announced future plans for a similar trial targeted to Usher Syndrome.

An innovative allele-specific approach emerged when Courtney el al. [166] identified the potential to utilise a mutation that generates a novel PAM to achieve allele-specificity. Although this work focused on corneal dystrophy, the technique has also been exploited for use in retinal disease by Bakondi et al. [167]. This approach provided a highly specific treatment strategy for certain autosomal dominant disorders. As the CRISPR technology develops at a rapid pace it is conceivable that soon an array of therapeutics will materialise that will allow safe and efficient correction of a range of genetic defects.

The future for ocular disorders looks bright and, as we begin to understand the integral players and interactions of complex disease, treatment strategies via genome editing technologies will become apparent. The previous optimisation groundwork using well characterised disease as models will allow for a smooth translation to treatment.

Cancer

In the field of cancer, the primary issue in the future will surround tumour heterogeneity and how this will complicate treatment strategies [168]. The revelation that a single tumour biopsy represents, in fact, multiple distinct tumour cell populations [169] was a pivotal moment in the field of cancer research. Since the discovery, a variety of studies have additionally confirmed that metastases from the primary tumour are invariably representative of only one or more sub-populations [170]. The concept of clonal evolution in cancer has been around since 1976 [171] and has been adopted in the field in order to explain these recent findings [172, 173]. This comes as a startling realisation when one considers the implications for personalised medicine: whilst we may be capable of identifying a metastatic clone with a key driver mutation and eradicating this with a specific drug or therapy (if available), in the situation where the primary tumour is highly heterogeneous, by eradicating the initial metastatic clone we may be merely paving the way for a different clone to rise up, which may necessitate an entirely different treatment strategy [168, 172]. Thus, tumour heterogeneity and the driver of this, genomic instability, have been other key focuses of research and will continue to be.

Identification and functional validation of such driver mutations amongst the large number of passenger mutations is thus an ongoing challenge. Genome editing technology such as CRISPR/Cas9 is going some way to address these challenges. It is now possible to reproduce the complex genome states observed in human tumours, such as translocations and inversions, as well as point mutations and deletions, in both cell lines and mouse models. Until recently, cancer mouse models were both laboriously slow and costly to generate, requiring the injection of genetically modified embryonic stem cells into blastocytes. CRISPR has enabled the generation of knockout and knock-in mouse models in as little as four weeks, developing both germline and somatic mutation mouse models.

Taking breast cancer as just one example, CRISPR has facilitated the discovery of point mutations conferring endocrine therapy resistance and, in doing so, has enabled researchers to understand the mechanism by which this happens [174]. Further, CRISPR-engineered mouse models have been used to identify the secondary mutations that confer resistance to PARP inhibitors in *BRCA1* and *BRCA2* mutant cancers, which are initially responsive [175]. Others have shown that in a *HER2* positive model, a CRIPSR-induced mutation within an amplified *HER2* region instead confers a dominant negative effect, resulting in cell growth inhibition via the MAPK/ERK axis, with no effect on HER2 protein levels [176]. That this response is potentiated by PARP inhibition, and is a distinct pathway from current HER2 therapies like Trastuzumab, gives some idea of the potential of CRISPR-mediated engineering in identifying new targets for therapy. However, whilst cancer research has been catapulted by the discovery of CRISPR, the reality remains that delivery of Cas9 continues to be a significant obstacle in both the generation of cancer mouse models and the delivery of therapeutic Cas9 guide RNA systems to treat cancer.

Another potential application of CRISPR in cancer could be as a companion technology to 'blood biopsy' based methods. The release of circulating free DNA (cfDNA) from tumour cells, i.e., circulating tumour DNA (ctDNA), can be a consequence of different physiological and pathological process such as apoptosis, necrosis, or active secretion (Fig. 2). In cancer patients, the released DNA may carry specific alterations within the fragment such as genetic and/or epigenetic modifications, which include methylation, loss of heterozygosity (LOH), and tumour-specific mutations in oncogenes and tumour suppressor genes [177]. In this regard, cfDNA from the blood of cancer patients —and also circulating tumour cells (CTCs)— could be exploited for not just diagnosis and prognosis [178, 179] but also help to identify targets for CRISPR-mediated treatment of the primary tumour. After CRISPR therapeutic intervention, cfDNA analysis could equally be used to monitor the effectiveness of the therapy, as it has been documented that, post-surgery, cfDNA and miRNA levels decrease to those found in healthy individuals [180, 181]; however, when the levels of cfDNA do not change, it might show that residual tumour cells exist [182].

Conclusions

Our desire to achieve a greater understanding of the genome in the past 3 decades has been the main driver of technological development in this area. Now that we have achieved a greater understanding, we are realising that the genome is not the end of the line, in terms of understanding disease. In fact, one could argue that simply understanding DNA has opened a Pandora's Box and that the real work has only just begun. Thankfully, the technological advances that have allowed us to understand the genome have indirectly given us opportunities to study beyond the genome, specifically at the transcriptome and epigenome (see Table 2 for a list of these), and further beyond these.

One striking revelation from the deluge of data that has already been produced in the biomedical sciences is that it points out just how much we don't yet understand about disease and how much work there is still to

Fig. 2 Is there utility for CRISPR via circulating tumour DNA detection?

Table 5 Cardiovascular disease and gene editing. Status: gene editing's clinical utility in the cardiovascular realm

Cardiovascular disease (CVD) consists of acute coronary syndrome (ACS), acute myocardial infarction (AMI), angina, arrhythmia, atherosclerosis, congestive heart failure (CHF), coronary artery disease (CAD), myocardial ischemia, etc. In the USA, per year, approximately 700,000 people suffer their first AMI and 500,000 experience a second or recurrent AMI, with 1.7 million being hospitalised annually due to ACS [242]. Clinical laboratories play a vital role in detecting and characterising risk of cardiovascular diseases and there is already a gambit of tests available for this purpose. For example, cardiac troponin is an important test for detecting myocardial injury, whilst B-type natriuretic peptide (BNP) and N-terminal portion of proBNP are used to detect CHF and risk for an acute event. Numerous other biomarkers are used to monitor various cardiovascular conditions.
However, not all biochemical tests are accurate. For example, it is known that half of AMIs occur in individuals with normal lipid panels [242]. The lipid panel (total, LDL, and HDL cholesterol, as well as triglycerides) —in addition to apolipoproteins (ApoA1 and ApoB), Lp(a), hsCRP, homocysteine, and Lp-pla2— are used to manage and monitor CHD. These tests can all be run using commercially-available reagents on various biochemical analysers, some of which may provide inaccurate results, possibly due to the complexity and stability of lipid molecules [243]. To improve the quality of results, alternative and more accurate methods have been developed to measure subclasses of HDL and LDL, such as: 1, β-quantification method [244], i.e., the reference method according to The U.S. National Cholesterol Education Program (NCEP); 2, gradient gel electrophoresis (GGE) [245, 246]; 3, vertical auto profile (VAP) [247]; 4, nuclear magnetic resonance spectroscopy (NMR) [245]; 5, ion mobility (IM) [248]; 6, high performance liquid chromatography (HPLC) [245].
Advances in the management of patients with cardiovascular disease through improved pharmacologic therapy have lessened impact; however, various limitations including patient compliance, side effects, and the need for repeat procedures keep patients in symptomatic status [249]. Gene and stem cell therapies in conjunction have shown promise in animal models of myocardial ischemia [249]. CRISPR/Cas9 gene editing of the loss-of-function proprotein convertase subtilisin/kexin type 9 (*PCSK9*) has also proven to reduce LDL cholesterol levels and protect against cardiovascular disease [250]. The major advantage of gene therapy is that, in a single administration, permanent benefits can be obtained, and with the advent of molecular research, further genes associated with lipoproteins and CVD risk have been discovered, e.g. *APOA1*, *APOA5*, *APOE*, *CETP*, *GALNT2*, *LIPC*, LPL, and *MLXIPL* [251], which may prove future targets of gene therapies.
Current gene therapy clinical trials have proven short-term safety; however, long term surveillance over a period of decades is still under investigation. Also, the cost-effectiveness of gene therapy has to be considered due to the laborious nature of the procedures. Current pharmacological approaches may still be more favourable in terms of cost benefit ratio [249], albeit in terms of cardiovascular disease treatment.

be done. Indeed, biological data is complex, having diverse internal structures that scientists have struggled to interpret using traditional methods and approaches [183], and whereas we are attempting to define how life within the cell functions in a relatively short space of time in order to better understand disease, life itself has had millions of years for various processes to diversify and become 'fixed', which has given us the wide diversity of life that we now see. The main players in this diversity are the genome, transcriptome, epigenome, and environment, with the amount of possible configurations between these being limitless.

Many diseases are therefore complex because life itself is complex, and we are still waiting to see major improvements in healthcare in the era of 'big data' that modern technology has allowed us to produce [184–186]. We don't claim that a complete understanding of life within the cell will help us to eradicate disease - we may understand disease much better but people will still age and develop illness. In cardiovascular disease, for example, a vast array of methods already exist and we are already knowledgeable on how to prevent these diseases from occurring (see Table 5) - would adding knowledge from the genome significantly reduce cardiovascular deaths?

In order to see significant improvement in healthcare utilising genomic, transcriptomic, and epigenomics data, there must be greater interdisciplinary cross talk between scientists. This includes, but is not limited to, physicians, clinical geneticists, computational biologists, and policy makers. New and recent technology can help to improve treatment, but only in the context of an understanding of disease mechanisms. We must minimise scenarios in which uncertainty enters the healthcare market, particularly in relation to critical techniques such as gene editing. Would it be feasible to excise a 'disease allele' if the exact mechanism of functioning of the allele in question was misunderstood? There is hope in terms of data science: integrating omics data can assist in fully defining disease mechanisms (see Table 6), which opens up the door to 'safe' gene editing.

Table 6 T-cell acute lymphoblastic leukaemia. Status: solved

In T-cell acute lymphoblastic leukaemia (T-ALL), 25% of cases exhibit high expression of the *TAL1* oncogene, which is due to a large deletion occurring at 1q33 that brings the coding sequences of *TAL1* (a transcription factor) in proximity to the promoter of *STIL*, a ubiquitously-expressed gene. This results in the ubiquitous/over- expression of *TAL1* and drives cancer. In many cases of T-ALL, however, overexpression of *TAL1* is observed without the large deletion – in these cases, H3K27ac binding (a marker of an enhancer region) is also found upstream of *TAL1*. Despite this information, the exact mechanism of disease had remained elusive for many years in these cases. Mansour and colleagues [252] observed these cases and found small heterozygous insertion variants of varying lengths in the same region as the previously found H3K27ac marks. The insertion variants, they found, were introducing new binding sites for the MYB transcription factor family, resulting in the over-expression of *TAL1* and the driving of cancer.

Conclusion: The Mansour study shows how data from DNA, RNA, and DNA-binding interactions can be used in combination to clearly define a disease mechanism. In this example, observing the intergenic upstream insertion variants (DNA), the heightened expression of *TAL1* (RNA), or the acetylation marks (DNA-binding interactions) alone would not explain the mechanism of disease. The Mansour study, however, although difficult and summing up years of work and studies, was made relatively easier by the fact that only a single gene was involved: *TAL1*. Thus, technically, no expert analytics or bioinformatics input was required. However, for complex diseases like most other cancers, cardiovascular diseases, etc., describing disease mechanisms is made extremely difficult by the fact that there can be any number of variants —be they SNPs, insertions, deletions, translocations, or copy number variants— involved in augmenting risk of the disease, with none on their own contributing a large amount to the disease phenotype. Thus, for complex diseases, there is much room for computational methods to be introduced in order to assist in clearly defining diseases mechanisms, but it involves a greater appreciation away from solely the genome.

Abbreviations

3C: Chromosome conformation capture; 4C: Chromosome conformation capture on chip; 5C: Chromosome conformation capture carbon copy; AAV: Adeno-associated virus; ABI: Applied Biosystems; ACS: Acute coronary syndrome; AID: Activation-induced cytidine deaminase; AMI: Acute myocardial infarction; ATAC-seq: Assay for Transposase Accessible Chromatin sequencing; A-to-I: Adenosine-to-inosine; BNP: B-type natriuretic peptide; CAD: Coronary artery disease; Cap-seq: Cap sequencing; CCD: Charge-coupled device; CCND1: Cyclin D1; cfDNA: Circulating free DNA; CHF: Congestive heart failure; ChIA-PET: Chromatin Interaction Analysis by Paired-End Tag sequencing; ChIP: Chromatin immunoprecipitation; ChIRP-seq: Chromatin isolation by RNA purification sequencing; CIP-TAP: Calf Intestinal alkaline Phosphatase Tobacco Acid Pyrophosphatase; CLASH: Crosslinking, ligation, and sequencing of hybrids; CRISPR: Clustered regularly interspaced short palindromic repeats; CRISPRa: CRISPR activation; CRISPRi: CRISPR interference; csRNAs: Capped small RNAs; CTCs: Circulating tumour cells (CTCs); ctDNA: Circulating tumour DNA; CVD: Cardiovascular disease; dCas9: Dead Cas9; DNase I HS site: DNase I hypersensitive site; DNase-seq: DNase I HS site sequencing; DSB: Double-strand break; eCas9: Enhanced Cas9; ENCODE: ENCyclopedia Of DNA Elements in the human genome; FAIRE-seq: Formaldehyde-Assisted Isolation of Regulatory Elements sequencing; FANTOM: Functional ANnoTation Of the Mammalian genome; Frag-seq: Fragmentation sequencing; GGE: Gradient gel electrophoresis; GOF: Gain-of-function; GRO-seq: Global Run-On sequencing; GWAS: Genome-Wide Association Studies / Study; HF-Cas9: High-fidelity Cas9; HGP: Human Genome Project (HGP); Hi-C: High-throughput chromosome conformation capture; HITS-CLIP: High Throughput Sequencing Crosslinking and Immunoprecipitation; HOTAIR: HOX transcript antisense RNA; HPLC: High performance liquid chromatography; HypaCas9: Hyper-active Cas9; ICE: Inosine Chemical Erasing; ICGC: International Cancer Genome Consortium; iCLIP: Individual-nucleotide resolution UV cross-linking and immunoprecipitation; INseq: Insertion sequencing; LCA: Leber congenital amaurosis; lincRNA: Long intergenic non-coding RNA; LOH: Loss of heterozygosity; M6A: Methylation of the N6 position of adenosine; MAINE-seq: MNase-Assisted Isolation of Nucleosomes Sequencing; MeRIP-seq: Methylated RNA Immunoprecipitation sequencing; miRNA: micro-RNA; MN: Micrococcal nuclease; MPSS: Massively parallel signature sequencing; mRNA: Messenger RNA; ncRNA: non-coding RNA; NET-seq: Native elongating transcript sequencing; NHGRI: The National Human Genome Research Institute; NHS: National Health Service; NMR: Nuclear magnetic resonance; ONT: Oxford Nanopore Technologies; PacBio: Pacific Biosciences; PAM: Protospacer adjacent motifs; PAR-CLIP: Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation; PARE-seq: Parallel Analysis of RNA Ends sequencing; PARS: Parallel analysis of RNA structure; PCSK9: Proprotein convertase subtilisin/kexin type 9; PPi: Pyrophosphate; PRE1 / PRE2: putative regulatory element 1 / 2; RBP: RNA binding protein; RC-seq: Retrotransposon Capture sequencing; Ribo-seq: Ribosome sequencing; RIP-seq: RNA immunoprecipitation sequencing; RNAi: RNA interference; rRNA: Ribosomal RNA; SAGE: Serial analysis of gene expression; SAM: Synergistic Activation Mediator; SBS: Sequencing by synthesis; SHAPE-seq: Selective 2'-Hydroxyl Acylation analyzed by Primer Extension sequencing; SHM: Somatic cellular hypermutation; SMRT: Single-molecule real-time; SPT: Serine palmitoyltransferase; T-ALL: T-cell acute lymphoblastic leukaemia; TCGA: The Cancer Genome Atlas; TC-seq: Translocation Capture sequencing; TIF-seq: Transcript Isoform Sequencing; TN-seq: Transposon

sequencing; TRAP-seq: Translating Ribosome Affinity Purification sequencing; TSS: Transcription start site; US NCEP: US National Cholesterol Education Program; VAP: Vertical auto profile; vCJD: Variant Creutzfeldt-Jakob Disease; XIST: X-Inactive Specific Transcript

Acknowledgements
Many thanks to John Mattick (Genomics England & Garvan Institute of Medical Research) and David Guttery (University of Leicester) for their advice on shaping the structure of the review.

Authors' contributions
KB conceived the original idea to compose the review, formed and managed the collaboration, wrote the background, conclusions, and Table 6, provided additional text to link all contributors' sections together, produced the artworks, and provided final editing across all sections. LDD wrote the section on technology, and Table 2 together with KB. KAC, MAN, and CBTM wrote the section on gene editing and CRISPR, and ocular genetics. SS, VH, LC, and JS wrote the section on cancer and Table 1 together with KB. TK-D wrote Table 3 on CRISPR's utility in bees. CCS wrote Table 5 on cardiovascular disease. BC, JAL-S, and RSK jointly wrote the section on asthma and Table 4. All authors have reviewed and approved the final version of the review.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA, USA. [2]Biomedical Sciences Research Institute, University of Ulster, Coleraine, Northern Ireland BT52 1SA, UK. [3]Avellino Laboratories, Menlo Park, CA 94025, USA. [4]Imperial College London, Division of Cancer, Department of Surgery and Cancer, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK. [5]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark. [6]Institute of Technology Carlow, Department of Science and Health, Kilkenny Road, Carlow, Ireland. [7]University of Raparin, Ranya, Kurdistan Region, Iraq. [8]Department of Cancer Studies and Molecular Medicine, Robert Kilpatrick Clinical Sciences Building, Leicester Royal Infirmary, Leicester LE2 7LX, UK. [9]Bill Lyons Informatics Centre, UCL Cancer Institute, University College London, WC1E 6DD, London, UK. [10]JMS Building, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK.

Received: 26 June 2017 Accepted: 26 July 2018
Published online: 08 August 2018

References

1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. Science. 2001;291(5507):1304–51.
2. International Human Genome Sequencing C. Initial sequencing and analysis of the human genome. Nature. 2001;409:860.
3. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annu Rev Med. 2012;63(1):35–61.
4. Schatz MC. Biological data sciences in genome research. Genome Res. 2015;25(10):1417–22.
5. Venter JC, Smith HO, Adams MD. The sequence of the human genome. Clin Chem. 2015;61(9):1207–8.
6. Clinton WJ. In 'June 2000 White House Event'. The White House Office of the Press Secretary. 2000. https://www.genome.gov/10001356/june-2000-white-house-event/.
7. The EPC. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57.
8. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, et al. The transcriptional landscape of the mammalian genome. Science. 2005;309(5740):1559–63.
9. Guttmacher AE, Collins FS. Genomic Medicine — A Primer. N Engl J Med. 2002;347(19):1512–20.
10. Varmus H. Getting ready for gene-based medicine. N Engl J Med. 2002; 347(19):1526–7.
11. Chan IS, Ginsburg GS. Personalized medicine: progress and promise. Annu Rev Genomics Hum Genet. 2011;12(1):217–44.
12. Green ED, Guyer MS, National Human Genome Research I. charting a course for genomic medicine from base pairs to bedside. Nature. 2011;470:204.
13. Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle — will we get our wish? N Engl J Med. 2008;358(2):105–7.
14. McGuire AL, Burke W. Raiding the medical commons: an unwelcome side effect of direct-to-consumer personal genome testing. JAMA : the journal of the American Medical Association. 2008;300(22):2669–71.
15. Feero WG, Guttmacher AE, Collins FS. Genomic medicine — an updated primer. N Engl J Med. 2010;362(21):2001–11.
16. The Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45:1113.
17. The International Cancer Genome C. International network of cancer genome projects. Nature. 2010;464:993.
18. Stratton M. Exploring the genomes of cancer cells: progress and promise. Science. 2011;331(6024):1553–8.
19. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012;486:400.
20. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45:1127.
21. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502:333.
22. Witte JS. Genome-wide association studies and beyond. Annu Rev Public Health. 2010;31(1):9–20.
23. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 337(6099):1190–5.
24. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genetics In Medicine. 2002;4:45.
25. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet. 2003;33:177.
26. Manolio TA, Collins FS. The HapMap and genome-wide association studies in diagnosis and therapy. Annu Rev Med. 2009;60(1):443–56.
27. Colhoun HM, McKeigue PM, Smith GD. Problems of reporting genetic associations with complex outcomes. Lancet. 2003;361(9360):865–72.
28. Studies N-NWGoRiA. Replicating genotype–phenotype associations. Nature. 2007;447:655.
29. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). Nucleic Acids Res. 2017;45(Database issue):D896–901.
30. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003;33:228.
31. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. Cell. 2017;169(7):1177–86.
32. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease–common variant… or not? Hum Mol Genet. 2002;11(20): 2417–23.
33. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40:695.
34. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev. 2009;19(3):212–9.
35. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet. 2010;11:415.

36. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet. 2012;13:135.
37. Alves MM, Sribudiani Y, Brouwer RWW, Amiel J, Antiñolo G, Borrego S, Ceccherini I, Chakravarti A, Fernández RM, Garcia-Barcelo M-M, et al. Contribution of rare and common variants determine complex diseases—Hirschsprung disease as a model. Dev Biol. 2013;382(1):320–9.
38. Diogo D, Kurreeman F, Stahl Eli A, Liao Katherine P, Gupta N, Greenberg Jeffrey D, Rivas Manuel A, Hickey B, Flannick J, Thomson B, et al. Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. Am J Hum Genet. 2013;92(1):15–27.
39. Yang J, Wang S, Yang Z, Hodgkinson CA, Iarikova P, Ma JZ, Payne TJ, Goldman D, Li MD. The contribution of rare and common variants in 30 genes to risk nicotine dependence. Mol Psychiatry. 2014;20:1467.
40. Fritsche LG, Igl W, Bailey JNC, Grassmann F, Sengupta S, Bragg-Gresham JL, Burdon KP, Hebbring SJ, Wen C, Gorski M, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat Genet. 2015;48:134.
41. Gorski MM, Blighe K, Lotta LA, Pappalardo E, Garagiola I, Mancini I, Mancuso ME, Fasulo MR, Santagostino E, Peyvandi F. Whole-exome sequencing to identify genetic risk variants underlying inhibitor development in severe hemophilia a patients. Blood. 2016;127(23):2924–33.
42. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747.
43. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci. 2001;98(19):10869–74.
44. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486:346.
45. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, Mattick JS, Rinn JL. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol. 2011;30:99.
46. Sundeep K. Recent advances in X-chromosome inactivation. J Cell Physiol. 2011;226(7):1714–8.
47. Gutschner T, Diederichs S. The hallmarks of cancer. RNA Biol. 2012;9(6):703–19.
48. Mattick JS. RNA regulation: a new genetics? Nat Rev Genet. 2004;5:316.
49. Lai EC. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet. 2002;30:363.
50. Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nat Rev Genet. 2013;14:880.
51. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81(1):145–66.
52. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009;10:155.
53. Wang Kevin C, Chang Howard Y. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011;43(6):904–14.
54. Chu C, Qu K, Zhong Franklin L, Artandi Steven E, Chang Howard Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011;44(4):667–78.
55. Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J, Martinez Arias A. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biol. 2009;7(7): e1000149.
56. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D. Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. Proc Natl Acad Sci U S A. 2011;108(24):10010–5.
57. Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT. Genome-wide identification of Polycomb-associated RNAs by RIP-seq. Mol Cell. 2010;40(6):939–53.
58. Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods. 2008;5:613.
59. Penalva LOF, Tenenbaum SA, Keene JD. Gene Expression Analysis of Messenger RNP Complexes. In: Schoenberg DR, editor. mRNA Processing and Metabolism: Methods and Protocols. Totowa, NJ: Humana Press; 2004. p. 125–34.
60. O'Sullivan RJ, Kubicek S, Schreiber SL, Karlseder J. Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. Nature Structural &Amp; Mol Bio. 2010;17:1218.
61. Shebzukhov YV, Horn K, Brazhnik KI, Drutskaya MS, Kuchmiy AA, Kuprash DV, Nedospasov SA. Dynamic changes in chromatin conformation at the TNF transcription start site in T helper lymphocyte subsets. Eur J Immunol. 2014;44(1):251–64.
62. Eberharter A, Becker PB. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. 2002;3(3):224–9.
63. Mercer TR, Mattick JS. Understanding the regulatory and transcriptional complexity of the genome through structure. Genome Res. 2013;23(7):1081–8.
64. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes Dev. 2012;26(1):11–24.
65. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature. 1953;171(4356):737–8.
66. Šponer J, Šponer JE, Petrov AI, Leontis NB. Quantum chemical studies of nucleic acids: can we construct a bridge to the RNA structural biology and bioinformatics communities? J Phys Chem B. 2010;114(48):15723–41.
67. Harrison JG, Zheng YB, Beal PA, Tantillo DJ. Computational approaches to predicting the impact of novel bases on RNA structure and stability. ACS chemical biology. 2013;8(11) https://doi.org/10.1021/cb4006062.
68. Koch T, Shim I, Lindow M, Ørum H, Bohr HG. Quantum mechanical studies of DNA and LNA. Nucleic Acid Therapeutics. 2014;24(2):139–48.
69. Fang L, Wuptra K, Chen D, Li H, Huang S-K, Jin C, Yokoyama KK. Environmental-stress-induced chromatin regulation and its heritability. Journal of carcinogenesis & mutagenesis. 2014;5(1):22058.
70. Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MSI, Kawaji H, Lassmann T, Harbers M, Forrest ARR, Bajic VB. Effects of cytosine methylation on transcription factor binding sites. BMC Genomics. 2014;15:119.
71. Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, Shin J, Cox E, Rho HS, Woodard C, et al. DNA methylation presents distinct binding sites for human transcription factors. eLife. 2013;2:e00726.
72. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463–7.
73. Gocayne J, Robinson DA, FitzGerald MG, Chung FZ, Kerlavage AR, Lentes KU, Lai J, Wang CD, Fraser CM, Venter JC. Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic receptors obtained by automated DNA sequence analysis: further evidence for a multigene family. Proc Natl Acad Sci U S A. 1987;84(23):8296–300.
74. Dulbecco R. A turning point in cancer research: sequencing the human genome. Science. 1986;231(4742):1055–6.
75. Hood L, Rowen L. The human genome project: big science transforms biology and medicine. Genome Medicine. 2013;5(9):79.
76. Luckey JA, Drossman H, Kostichka AJ, Mead DA, D'Cunha J, Norris TB, Smith LM. High speed DNA sequencing by capillary electrophoresis. Nucleic Acids Res. 1990;18(15):4417–21.
77. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol. 2000;18:630.
78. Audic S, Claverie J-M. The significance of digital gene expression profiles. Genome Res. 1997;7(10):986–95.
79. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science. 1995;270(5235):484–7.
80. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376.
81. Hyman ED. A new method of sequencing DNA. Anal Biochem. 1988;174(2): 423–36.
82. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 1996;242(1):84–9.
83. Li H, Ren X, Ying L, Balasubramanian S, Klenerman D. Measuring single-molecule nucleic acid dynamics in solution by two-color filtered ratiometric fluorescence correlation spectroscopy. Proc Natl Acad Sci U S A. 2004;101(40):14425–30.
84. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008; 456(7218):53–9.

85. Ju J, Kim DH, Bi L, Meng Q, Bai X, Li Z, Li X, Marma MS, Shi S, Wu J, et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proc Natl Acad Sci U S A. 2006;103(52):19635–40.
86. Guo J, Xu N, Li Z, Zhang S, Wu J, Kim DH, Sano Marma M, Meng Q, Cao H, Li X, et al. Four-color DNA sequencing with 3′-O–modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. Proc Natl Acad Sci. 2008;105(27):9145–50.
87. Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet. 2009;11:31.
88. Shendure J, Mitra RD, Varma C, Church GM. Advanced sequencing technologies: methods and goals. Nat Rev Genet. 2004;5:335.
89. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y-J, Makhijani V, Roth GT, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature. 2008;452:872.
90. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323(5910):133–8.
91. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013;10:563.
92. Fichot EB, Norman RS. Microbial phylogenetic profiling with the Pacific biosciences sequencing platform. Microbiome. 2013;1(1):10.
93. Mostovoy Y, Levy-Sakin M, Lam J, Lam ET, Hastie AR, Marks P, Lee J, Chu C, Lin C, Džakula Ž, et al. A hybrid approach for de novo human genome sequence assembly and phasing. Nat Methods. 2016;13:587.
94. Laver TW, Caswell RC, Moore KA, Poschmann J, Johnson MB, Owens MM, Ellard S, Paszkiewicz KH, Weedon MN. Pitfalls of haplotype phasing from amplicon-based long-read sequencing. Sci Rep. 2016;6:21746.
95. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. Genome Res. 2017;27(5):757–67.
96. Potapov V, Ong JL. Examining sources of error in PCR by single-molecule sequencing. PLoS One. 2017;12(1):e0169774.
97. Hildt E. Human Germline interventions–think first. Front Genet. 2016;7:81.
98. Cribbs AP, Perera SMW. Science and bioethics of CRISPR-Cas9 gene editing: an analysis towards separating facts and fiction. The Yale Journal of Biology and Medicine. 2017;90(4):625–34.
99. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819–23.
100. Blasco Rafael B, Karaca E, Ambrogio C, Cheong T-C, Karayol E, Minero Valerio G, Voena C, Chiarle R. Simple and Rapid In Vivo Generation of Chromosomal Rearrangements using CRISPR/Cas9 Technology. Cell Rep. 2014;9(4):1219–27.
101. Wiles MV, Qin W, Cheng AW, Wang H. CRISPR–Cas9-mediated genome editing and guide RNA design. Mamm Genome. 2015;26(9):501–10.
102. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013;153(4):910–8.
103. Reardon S. The CRISPR zoo. Nature. 2016;531(7593):160–3.
104. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;343(6166):84–7.
105. Deans RM, Morgens DW, Ökesli A, Pillay S, Horlbeck MA, Kampmann M, Gilbert LA, Li A, Mateo R, Smith M, et al. Parallel shRNA and CRISPR-Cas9 screens enable antiviral drug target identification. Nat Chem Biol. 2016;12:361.
106. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. Nat Biotechnol. 2015;33:661.
107. Wallace J, Hu R, Mosbruger TL, Dahlem TJ, Stephens WZ, Rao DS, Round JL, O'Connell RM. Genome-wide CRISPR-Cas9 screen identifies MicroRNAs that regulate myeloid leukemia cell growth. PLoS One. 2016;11(4):e0153689.
108. Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera MDC, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol. 2013;32:267.
109. Morgens DW, Deans RM, Li A, Bassik MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. Nat Biotechnol. 2016;34:634.
110. Lin A, Giuliano CJ, Sayles NM, Sheltzer JM. CRISPR/Cas9 mutagenesis invalidates a putative cancer dependency targeted in on-going clinical trials. eLife. 2017;6:e24179.
111. Castanotto D, Rossi JJ. The promises and pitfalls of RNA-interference-based therapeutics. Nature. 2009;457(7228):426–33.
112. Tiemann K, Rossi JJ. RNAi-based therapeutics–current status, challenges and prospects. EMBO Molecular Medicine. 2009;1(3):142–51.
113. Jackson AL, Burchard J, Schelter J, Chau BN, Cleary M, Lim L, Linsley PS. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA. 2006;12(7):1179–87.
114. Sigoillot FD, Lyman S, Huckins JF, Adamson B, Chung E, Quattrochi B, King RW. A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. Nat Methods. 2012;9:363.
115. Echeverri CJ, Beachy PA, Baum B, Boutros M, Buchholz F, Chanda SK, Downward J, Ellenberg J, Fraser AG, Hacohen N, et al. Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat Methods. 2006;3:777.
116. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816–21.
117. Hess GT, Frésard L, Han K, Lee CH, Li A, Cimprich KA, Montgomery SB, Bassik MC. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat Methods. 2016;13:1036.
118. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533:420.
119. Conaway JW. Introduction to theme "chromatin, epigenetics, and transcription". Annu Rev Biochem. 2012;81(1):61–4.
120. Gilbert Luke A, Larson Matthew H, Morsut L, Liu Z, Brar Gloria A, Torres Sandra E, Stern-Ginossar N, Brandman O, Whitehead Evan H, Doudna Jennifer A, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154(2):442–51.
121. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA–guided activation of endogenous human genes. Nat Methods. 2013;10:977.
122. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2014;517:583.
123. Horsthemke B, Buiting K. Chapter 8 Genomic Imprinting and Imprinting Defects in Humans. In: *Advances in Genetics*, vol. 61: Academic Press; 2008. p. 225–46.
124. Zovkic IB, Guzman-Karlsson MC, Sweatt JD. Epigenetic regulation of memory formation and maintenance. Learn Mem. 2013;20(2):61–74.
125. Kungulovski G, Jeltsch A. Epigenome editing: state of the art, concepts, and perspectives. Trends Genet. 2016;32(2):101–13.
126. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R. Editing DNA Methylation in the Mammalian Genome. Cell. 2016; **167**(1):233–47. e217
127. Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R. Functional annotation of native enhancers with a Cas9–histone demethylase fusion. Nat Methods. 2015;12:401.
128. Hilton IB, Gersbach CA. Enabling functional genomics with genome engineering. Genome Res. 2015;25(10):1442–55.
129. Chen B, Gilbert Luke A, Cimini Beth A, Schnitzbauer J, Zhang W, Li G-W, Park J, Blackburn Elizabeth H, Weissman Jonathan S, Qi Lei S, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell. 2013;155(7):1479–91.
130. Ma H, Tu L-C, Naseri A, Huisman M, Zhang S, Grunwald D, Pederson T. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nat Biotechnol. 2016;34:528.
131. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31:827.
132. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013;31(9):822–6.
133. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, Cheng AW, Trevino AE, Konermann S, Chen S, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol. 2014;32:670.
134. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotechnol. 2013;31:839.
135. Christie KA, Courtney DG, DeDionisio LA, Shern CC, De Majumdar S, Mairs LC, Nesbit MA, Moore CBT. Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. Sci Rep. 2017;7(1):16174.
136. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. Science. 2016; 351(6268):84–8.

137. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature. 2016;529:490.
138. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature. 2017;550(7676):407–10.
139. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520:186.
140. Zetsche B, Gootenberg Jonathan S, Abudayyeh Omar O, Slaymaker Ian M, Makarova Kira S, Essletzbichler P, Volz Sara E, Joung J, van der Oost J, Regev A, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163(3):759–71.
141. Kim D, Kim J, Hur JK, Been KW, Yoon S-H, Kim J-S. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat Biotechnol. 2016;34:863.
142. Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, Lopez JM, McCaw ZR, Aryee MJ, Joung JK. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat Biotechnol. 2016;34:869.
143. Glass Z, Lee M, Li Y, Xu Q. Engineering the delivery system for CRISPR-based genome editing. Trends Biotechnol. 2018;36(2):173–85.
144. Fanta CH. Asthma. N Engl J Med. 2009;360(10):1002–14.
145. Hans B, Stanley S. Prevalence of asthma-like symptoms in young children. Pediatr Pulmonol. 2007;42(8):723–8.
146. Moffatt MF. Genes in asthma: new genes and new ways. Curr Opin Allergy Clin Immunol. 2008;8(5):411–7.
147. Vercelli D. Discovering susceptibility genes for asthma and allergy. Nat Rev Immunol. 2008;8:169.
148. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, Bleecker ER. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. Journal of Allergy and Clinical Immunology. 2010;125(2):328–35. e311
149. Sleiman PMA, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SAG, Wang K, Rafaels NM, Michel S, Bonnelykke K, et al. Variants of DENND1B associated with asthma in children. N Engl J Med. 2010;362(1):36–44.
150. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk JB, Willis-Owen SAG, Klanderman B, Lasky-Su J, et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. Am J Hum Genet. 2009;84(5):581–93.
151. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson WOCM. A large-scale, consortium-based Genomewide association study of asthma. N Engl J Med. 2010; 363(13):1211–21.
152. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, Himes BE, Levin AM, Mathias RA, Hancock DB, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse north American populations. Nat Genet. 2011;43(9):887–92.
153. Ono JG, Worgall TS, Worgall S. Airway reactivity and sphingolipids—implications for childhood asthma. Molecular and Cellular Pediatrics. 2015;2:13.
154. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, den Dekker HT, Husby A, Sevelsted A, Faura-Tellez G, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet. 2013;46:51.
155. Bochkov YA, Watters K, Ashraf S, Griggs TF, Devries MK, Jackson DJ, Palmenberg AC, Gern JE. Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. Proc Natl Acad Sci U S A. 2015;112(17):5485–90.
156. Hawkins GA, Tantisira K, Meyers DA, Ampleford EJ, Moore WC, Klanderman B, Liggett SB, Peters SP, Weiss ST, Bleecker ER. Sequence, haplotype, and association analysis of ADRβ2 in a multiethnic asthma case-control study. Am J Respir Crit Care Med. 2006;174(10):1101–9.
157. Himes BE, Jiang X, Wagner P, Hu R, Wang Q, Klanderman B, Whitaker RM, Duan Q, Lasky-Su J, Nikolos C, et al. RNA-Seq Transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells. PLoS One. 2014;9(6):e99625.
158. Weiss JS, Møller HU, Lisch W, Kinoshita S, Aldave AJ, Belin MW, Kivelä T, Busin M, Munier FL, Seitz B, et al. The IC3D classification of the corneal dystrophies. Cornea. 2008;27(Suppl 2):S1–83.
159. Broadgate S, Yu J, Downes SM, Halford S. Unravelling the genetics of inherited retinal dystrophies: past, present and future. Prog Retin Eye Res. 2017;59:53–96.
160. Moore C, Christie K, Marshall J, Nesbit M. Personalised genome editing – the future for corneal dystrophies. Prog Retin Eye Res. 2018;1
161. Xue K, Oldani M, Jolly JK, Edwards TL, Groppe M, Downes SM, MacLaren RE. Correlation of optical coherence tomography and autofluorescence in the outer retina and choroid of patients with Choroideremia. Invest Ophthalmol Vis Sci. 2016;57(8):3674–84.
162. Jacobson SG, Cideciyan AV, Roman AJ, Sumaroka A, Schwartz SB, Heon E, Hauswirth WW. Improvement and decline in vision with gene therapy in childhood blindness. N Engl J Med. 2015;372(20):1920–6.
163. Ghazi NG, Abboud EB, Nowilaty SR, Alkuraya H, Alhommadi A, Cai H, Hou R, Deng W-T, Boye SL, Almaghamsi A, et al. Treatment of retinitis pigmentosa due to MERTK mutations by ocular subretinal injection of adeno-associated virus gene vector: results of a phase I trial. Hum Genet. 2016;135(3):327–43.
164. Parker MA, Choi D, Erker LR, Pennesi ME, Yang P, Chegarnov EN, Steinkamp PN, Schlechter CL, Dhaenens C-M, Mohand-Said S, et al. Test–retest variability of functional and structural parameters in patients with Stargardt disease participating in the SAR422459 gene therapy trial. Translational Vision Science & Technology. 2016;5(5):10.
165. Zallocchi M, Binley K, Lad Y, Ellis S, Widdowson P, Iqball S, Scripps V, Kelleher M, Loader J, Miskin J, et al. EIAV-based retinal gene therapy in the shaker1 mouse model for usher syndrome type 1B: development of UshStat. PLoS One. 2014;9(4):e94272.
166. Courtney DG, Moore JE, Atkinson SD, Maurizi E, Allen EHA, Pedrioli DML, McLean WHI, Nesbit MA, Moore CBT. CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting. Gene Ther. 2015;23:108.
167. Bakondi B, Lv W, Lu B, Jones MK, Tsai Y, Kim KJ, Levy R, Akhtar AA, Breunig JJ, Svendsen CN, et al. In vivo CRISPR/Cas9 gene editing corrects retinal dystrophy in the S334ter-3 rat model of autosomal dominant retinitis Pigmentosa. Mol Ther. 2016;24(3):556–63.
168. Baird RD, Caldas C. Genetic heterogeneity in breast cancer: the road to personalized medicine? BMC Med. 2013;11(1):151.
169. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.
170. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013;501:338.
171. Nowell P. The clonal evolution of tumor cell populations. Science. 1976; 194(4260):23–8.
172. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481:306.
173. Gerlinger M, McGranahan N, Dewhurst SM, Burrell RA, Tomlinson I, Swanton C. Cancer: evolution within a lifetime. Annu Rev Genet. 2014;48(1):215–36.
174. Harrod A, Fulton J, Nguyen VTM, Periyasamy M, Ramos-Garcia L, Lai CF, Metodieva G, de Giorgio A, Williams RL, Santos DB, et al. Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer. Oncogene. 2017;36(16):2286–96.
175. Dréan A, Williamson CT, Brough R, Brandsma I, Menon M, Konde A, Garcia-Murillas I, Pemberton HN, Frankum J, Rafiq R, et al. Modeling therapy resistance in BRCA1/2–mutant cancers. Mol Cancer Ther. 2017;16(9):2022–34.
176. Wang H, Sun W. CRISPR-mediated targeting of HER2 inhibits cell proliferation through a dominant negative mutation. Cancer Lett. 2017;385:137–43.
177. Schwarzenbach H, Hoon DSB, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011;11:426.
178. Openshaw MR, Page K, Fernandez-Garcia D, Guttery D, Shaw JA. The role of ctDNA detection and the potential of the liquid biopsy for breast cancer monitoring. Expert Rev Mol Diagn. 2016;16(7):751–5.
179. Shaw JA, Guttery DS, Hills A, Fernandez-Garcia D, Page K, Rosales BM, Goddard KS, Hastings RK, Luo J, Ogle O, et al. Mutation analysis of cell-free DNA and single circulating tumor cells in metastatic breast Cancer patients with high circulating tumor cell counts. Clin Cancer Res. 2017;23(1):88–96.
180. Catarino R, Ferreira MM, Rodrigues H, Coelho A, Nogal A, Sousa A, Medeiros R. Quantification of free circulating tumor DNA as a diagnostic marker for breast Cancer. DNA Cell Biol. 2008;27(8):415–21.
181. Yamamoto Y, Kosaka N, Tanaka M, Koizumi F, Kanai Y, Mizutani T, Murakami Y, Kuroda M, Miyajima A, Kato T, et al. MicroRNA-500 as a potential diagnostic marker for hepatocellular carcinoma. Biomarkers. 2009;14(7):529–38.
182. Pauline W, Carina R, Klaus P, Sabine K-B, Rainer K, Heidi S. Impact of platinum-based chemotherapy on circulating nucleic acid levels, protease

activities in blood and disseminated tumor cells in bone marrow of ovarian cancer patients. Int J Cancer. 2011;128(11):2572–80.
183. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: structure and dynamics. Phys Rep. 2006;424(4):175–308.
184. Nash DB. Harnessing the power of big data in healthcare. American Health & Drug Benefits. 2014;7(2):69–70.
185. Belle A, Thiagarajan R, Soroushmehr SMR, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. Biomed Res Int. 2015;2015:370194.
186. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. JMIR Med Inform. 2016;4(4):e38.
187. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010;42:504.
188. French Juliet D, Ghoussaini M, Edwards Stacey L, Meyer Kerstin B, Michailidou K, Ahmed S, Khan S, Maranian Mel J, O'Reilly M, Hillman Kristine M, et al. Functional variants at the 11q13 risk locus for breast Cancer regulate Cyclin D1 expression through long-range enhancers. Am J Hum Genet. 2013;92(4):489–503.
189. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008;322(5909):1845–8.
190. Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature. 2011;469:368.
191. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009;324(5924):218–23.
192. Reynoso MA, Juntawong P, Lancia M, Blanco FA, Bailey-Serres J, Zanetti ME: Translating Ribosome Affinity Purification (TRAP) Followed by RNA Sequencing Technology (TRAP-SEQ) for Quantitative Assessment of Plant Translatomes. In: *Plant Functional Genomics: Methods and Protocols*. Alonso JM, Stepanova AN. New York, NY: Springer New York; 2015: 185–207.
193. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. Nature. 2009;460:479.
194. Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C, Tuschl T. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods. 2008;44(1):3–12.
195. König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nature Structural &Amp; Mol Biol. 2010;17:909.
196. German MA, Luo S, Schroth G, Meyers BC, Green PJ. Construction of parallel analysis of RNA ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat Protoc. 2009;4:356.
197. German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, et al. Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol. 2008;26:941.
198. Pelechano V, Wei W, Jakob P, Steinmetz LM. Genome-wide identification of transcript start and end sites by transcript isoform sequencing. Nat Protoc. 2014;9:1740.
199. Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. Nature. 2013;497:127.
200. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP. Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci. 2011;108(27):11063–8.
201. Wan Y, Qu K, Ouyang Z, Chang HY. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. Nat Protoc. 2013;8:849.
202. Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, Lowe TM, Salama SR, Haussler D. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat Methods. 2010;7:995.
203. Sakurai M, Yano T, Kawabata H, Ueda H, Suzuki T. Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. Nat Chem Biol. 2010;6:733.
204. Meyer Kate D, Saletore Y, Zumbo P, Elemento O, Mason Christopher E, Jaffrey Samie R. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3'UTRs and near Stop Codons. Cell. 2012;149(7):1635–46.
205. Gu W, Lee H-C, Chaves D, Youngman Elaine M, Pazour Gregory J, Conte D Jr, Mello Craig C. CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as C.elegans piRNA Precursors. Cell. 2012;151(7):1488–500.
206. Affymetrix/Cold Spring Harbor Laboratory ETP. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. Nature. 2009;457:1028.
207. Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, et al. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res. 2006;16(1):123–31.
208. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, et al. A map of open chromatin in human pancreatic islets. Nat Genet. 2010;42:255.
209. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. Genome Res. 2007;17(6):877–85.
210. Ponts N, Harris EY, Prudhomme J, Wick I, Eckhardt-Ludka C, Hicks GR, Hardiman G, Lonardi S, Le Roch KG. Nucleosome landscape and control of transcription in the human malaria parasite. Genome Res. 2010;20(2):228–38.
211. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. Current Protocols in Molecular Biology. 2015;109(1):21.29.21–9.
212. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, et al. An oestrogen-receptor-α-bound human chromatin interactome. Nature. 2009;462:58.
213. Duan Z, Andronescu M, Schutz K, Lee C, Shendure J, Fields S, Noble WS, Anthony Blau C. A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. Methods. 2012;58(3):277–88.
214. Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Singh Sandhu K, Singh U, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet. 2006;38:1341.
215. Dostie J, Dekker J. Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc. 2007;2:988.
216. Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi–C: a comprehensive technique to capture the conformation of genomes. Methods. 2012;58(3):268–76.
217. Sanchez-Luque FJ, Richardson SR, Faulkner GJ. Retrotransposon Capture Sequencing (RC-Seq): A Targeted, High-Throughput Approach to Resolve Somatic L1 Retrotransposition in Humans. In: Garcia-Pérez JL, editor. *Transposons and Retrotransposons: Methods and Protocols*. New York, NY: Springer New York; 2016. p. 47–77.
218. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, et al. Somatic retrotransposition alters the genetic landscape of the human brain. Nature. 2011;479:534.
219. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods. 2009;6:767.
220. van Opijnen T, Camilli A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nature reviews Microbiology. 2013;11(7) https://doi.org/10.1038/nrmicro3033.
221. Klein Isaac A, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, Di Virgilio M, Bothmer A, Nussenzweig A, Robbiani Davide F, et al. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell. 2011;147(1):95–106.
222. Oliveira TY, Resch W, Jankovic M, Casellas R, Nussenzweig MC, Klein IA. Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. J Immunol Methods. 2012;375(1):176–81.
223. HHW V, van Doorn A. A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. Apidologie. 2006;37(4):421–51.
224. MJF B, Paxton RJ. The conservation of bees: a global perspective. Apidologie. 2009;40(3):410–6.
225. Linde B, Veerle M, Gamal A-A, Guy S. Lethal and sublethal side-effect assessment supports a more benign profile of spinetoram compared with spinosad in the bumblebee Bombus terrestris. Pest Manag Sci. 2011;67(5):541–7.
226. Thomson D. Detecting the effects of introduced species: a case study of competition between Apis and Bombus. Oikos. 2006;114(3):407–18.
227. Ellis JD, Munn PA. The worldwide health status of honey bees. Bee World. 2005;86(4):88–101.
228. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan P-L, Briese T, Hornig M, Geiser DM, et al. A metagenomic survey of microbes in honey bee Colony collapse disorder. Science. 2007;318(5848):283–7.
229. Anderson D, East IJ. The latest buzz about Colony collapse disorder. Science. 2008;319(5864):724–5.
230. Horvath P, Barrangou R. CRISPR/Cas, the immune system of Bacteria and Archaea. Science. 2010;327(5962):167–70.

231. The Honeybee Genome Sequencing C. Insights into social insects from the genome of the honeybee Apis mellifera. Nature. 2006;443(7114):931–49.
232. Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJ, Hasselmann M, Lozier JD, et al. The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. 2015;16(1):76.
233. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. N Engl J Med. 1995;332(3):133–8.
234. Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. Lancet. 2008;372(9643): 1107–19.
235. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature. 2007;448:470.
236. Verlaan DJ, Berlivet S, Hunninghake GM, Madore A-M, Larivière M, Moussette S, Grundberg E, Kwan T, Ouimet M, Ge B, et al. Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am J Hum Genet. 2009;85(3):377–93.
237. Miller M, Tam AB, Cho JY, Doherty TA, Pham A, Khorram N, Rosenthal P, Mueller JL, Hoffman HM, Suzukawa M, et al. ORMDL3 is an inducible lung epithelial gene regulating metalloproteases, chemokines, OAS, and ATF6. Proc Natl Acad Sci. 2012;109(41):16648–53.
238. Breslow DK, Collins SR, Bodenmiller B, Aebersold R, Simons K, Shevchenko A, Ejsing CS, Weissman JS. Orm family proteins mediate sphingolipid homeostasis. Nature. 2010;463(7284):1048–53.
239. Breslow DK, Weissman JS. Membranes in balance: mechanisms of Sphingolipid homeostasis. Mol Cell. 2010;40(2):267–79.
240. Worgall TS, Veerappan A, Sung B, Kim BI, Weiner E, Bholah R, Silver RB, Jiang X-C, Worgall S. Impaired Sphingolipid Synthesis in the Respiratory Tract Induces Airway Hyperreactivity. Science Translational Medicine. 2013;5(186): 186ra167.
241. Miller M, Rosenthal P, Beppu A, Mueller JL, Hoffman HM, Tam AB, Doherty TA, McGeough MD, Pena CA, Suzukawa M, et al. ORMDL3 transgenic mice have increased airway remodeling and airway responsiveness characteristic of asthma. J Immunol. 2014;192(8):3475–87.
242. Lopez J, Burtis CA, Bruns DE. Tietz fundamentals of clinical chemistry and molecular diagnostics, 7th ed.: Elsevier, Amsterdam, 1075 pp, ISBN 978-1-4557-4165-6. Indian J Clin Biochem. 2015;30(2):243.
243. Zivkovic AM, Wiest MM, Nguyen UT, Davis R, Watkins SM, German JB. Effects of sample handling and storage on quantitative lipid analysis in human serum. Metabolomics. 2009;5(4):507–16.
244. Dong J, Guo H, Yang R, Li H, Wang S, Zhang J, Chen W. Serum LDL- and HDL-cholesterol determined by ultracentrifugation and HPLC. J Lipid Res. 2011;52(2):383–8.
245. Hafiane A, Genest J. High density lipoproteins: measurement techniques and potential biomarkers of cardiovascular risk. BBA Clinical. 2015;3:175–88.
246. Mora S, Otvos JD, Rifai N, Rosenson RS, Buring JE, Ridker PM. Lipoprotein particle profiles by nuclear magnetic resonance compared with standard lipids and Apolipoproteins in predicting incident cardiovascular disease in women. Circulation. 2009;119(7):931–9.
247. Rosenson RS, Brewer HB, Chapman MJ, Fazio S, Hussain MM, Kontush A, Krauss RM, Otvos JD, Remaley AT, Schaefer EJ. HDL measures, particle heterogeneity, proposed nomenclature, and relation to atherosclerotic cardiovascular events. Clin Chem. 2011;57(3):392–410.
248. Caulfield MP, Li S, Lee G, Blanche PJ, Salameh WA, Benner WH, Reitz RE, Krauss RM. Direct determination of lipoprotein particle sizes and concentrations by ion mobility analysis. Clin Chem. 2008;54(8):1307–16.
249. Lavu M, Gundewar S, Lefer DJ. Gene therapy for ischemic heart disease. J Mol Cell Cardiol. 2011;50(5):742–50.
250. Ding Q, Strong A, Patel KM, Ng S-L, Gosis BS, Regan SN, Cowan CA, Rader DJ, Musunuru K. Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing: novelty and significance. Circ Res. 2014;115(5):488–92.
251. Musunuru K, Orho-Melander M, Caulfield MP, Li S, Salameh WA, Reitz RE, Berglund G, Hedblad B, Engström G, Williams PT, et al. Ion mobility analysis of lipoprotein subfractions identifies three independent axes of cardiovascular risk. Arterioscler Thromb Vasc Biol. 2009;29(11):1975–80.
252. Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, Etchin J, Lawton L, Sallan SE, Silverman LB, et al. An Oncogenic Super-Enhancer Formed Through Somatic Mutation of a Noncoding Intergenic Element. Science (New York, NY). 2014;346(6215):1373–7.

Discussion

7 Discussion

Gene therapy holds immense promise for the treatment of single gene diseases. The potential to correct disease at the level of the gene offers hope for these devastating diseases. However, past tragedies in the field of gene therapy, such as the untimely death of Jesse Gelsinger due to an overreaction to adenovirus and the death of 5 patients in a clinical trial for X-SCID due to insertional mutagenesis[1,2], have highlighted that we must enter this new age with great caution. Tremendous progress has been made in the field of gene therapy, such as the vast expansion of the toolkit and a greater appreciation for associated risks for these applications. However, there are still several pinnacle questions, such as genome-wide specificity, efficiency and delivery, that must be addressed to bring about the next frontier in gene therapy.

The work presented within this thesis establishes the groundwork for the development of a gene editing approach to treat the corneal dystrophies, a monogenic disease. However, the application of gene editing strategies to treat complex disease, in which a combination of alleles in addition to environmental factors can contribute to the risk of developing disease, is somewhat premature. Currently our understanding of complex disease, prevents the direct extension of this work to diseases dictated by complex genetics. Paper V discusses the immense progress made in both our understanding and interpretation of human disease since the completion of the human genome project. For mendelian monogenic diseases, such as the corneal dystrophies, focus at the level of the DNA has been sufficient to identify the disease-causing variants. However, for complex disease a greater appreciation of the vast number of potential interactions will be absolutely critical in determining the genetic factors behind these common diseases and how their interactions result in disease. The fusion of a catalytically dead Cas9 to various protein effectors, such as a KRAB effector domain or chromatin-modifying enzymes, has enabled the generation of tools capable of interrogating the complexities of genome. Thus, irrespective of whether the CRISPR/Cas9 system fulfils its potential in the gene

therapy landscape, it has become an indispensable technology to further our understanding of human disease.

7.1 Role of TGFBIp in *TGFBI* corneal dystrophies

The pathomechanism of *TGFBI* corneal dystrophies has not been fully elucidated. Several links have been made between the susceptibility of mutant TGFBIp and oxidative stress. Granular corneal dystrophy type II (GCD2) corneal fibroblasts were shown to be in oxidative stress and more susceptible to oxidative damage than their wild-type counterparts[3]. In addition, their membrane potential and mitochondrial activity was significantly reduced[4]. When mutant TGFBIp was exposed to superoxide species it was less stable and produced markedly more amyloid fibrils than wild-type TGFBIp[5]. These independent studies all indicate mutant TGFBIp has an increased susceptibility to oxidative stress. The cornea is continually exposed to ultraviolet light, with 90% of incident UV-B radiation known to be absorbed by the cornea[6]. Incidentally, UV-B radiation is known to be the main driver of superoxide species, resulting in oxidative stress[7]. It is conceivable that the susceptibility of mutant TGFBIp to the oxidative stress caused by the continual exposure of the cornea to UV light could have an important role in the pathomechanism of *TGFBI* corneal dystrophies.

Furthermore, proteolytic degradation of mutant TGFBIp has been shown to be impaired. The rate of protein clearance from GCD2 corneal fibroblasts was measured following treatment with a protein synthesis inhibitor[8]. Wild-type TGFBIp was cleared rapidly, whereas mutant TGFBIp remained after 1 hour. This effect was again observed when mutant TGFBIp was introduced into wild-type corneal fibroblasts. Together this indicates defective extracellular excretion or degradation of mutant TGFBIp in respect to wild-type TGFBIp. In addition, abnormal turnover of mutant TGFBIp from corneas associated with R124C, R124H and R124L was observed in comparison to wild-type TGFBIp from normal corneas[9]. Interestingly, all 3

missense mutations studied were shown to have abnormal turnover, however differential protein processing was observed for each mutation. As these 3 missense mutations all affect the same residue this may indicate how such strikingly different corneal dystrophy phenotypes are observed for each mutation. TGFBIp was found to be significantly increased in corneas with R124C, R124H and R124L, compared to wild-type corneas, in addition it was found that TGFBIp co-localised with the pathogenic deposits[10]. Again indicative that degradation of mutant TGFBIp is impaired. This was confirmed in an additional report that investigated the proteomic composition of R124C deposits, causative of lattice corneal dystrophy, to R555W deposits, causative of granular corneal dystrophy[7]. All deposits were shown to contain TGFBIp as previously reported, however the remaining constituents varied between LCD and GCD, indicating they undergo different proteolytic processing suggesting a possible reasoning for the phenotypic heterogeneity exhibited by these missense mutations.

Missense mutations within *TGFBI* have not been shown cause adverse phenotypes in other tissues in body. The cornea exists as the only transparent connective tissue in the body, its fundamental function is to maintain transparency[11]. *TGFBI* corneal dystrophies result in a disease phenotype due to the accumulation of mutant proteins in the corneal stroma that impair its function to remain transparent. Taken together a potential mechanism of disease exists whereby; under oxidative stress the normal degradation pathways implemented to eliminate TGFBIp are not sufficient, leading to abnormal proteolytic processing of the mutant protein which results in an accumulation of mutant protein and thus the formation of these blinding deposits.

It has been extensively reported that when patients who have an underlying *TGFBI* mutation, but a seemingly quiet cornea i.e. do not exhibit symptoms associated with corneal dystrophy, receive laser eye surgery they see a rapid and severe accumulation of opacities in their cornea[12–

[15]. In Paper III, the loss of TGFBIp is shown to negatively affect the wound healing process in a regenerating zebrafish tail-fin. This result is important in explaining why patients see a sudden emergence of deposits following laser eye surgery. In addition, it contributes to the understanding of the pathomechanism of *TGFBI* corneal dystrophies. Tissue damage, such as a tail-fin amputation or laser eye surgery initiates wound healing pathways, such as the TGFβ signalling pathway, which results in an upregulation of *TGFBI* expression. Thus, there is an increased production of mutant TGFBIp in the cornea which is unable to be degraded. Therefore, triggers the accumulation of opacities which impair the function of the transparent cornea. Importantly, this result provides insight into the concentration of mutant TGFBIp required to result in deposit formation. Prior to laser eye surgery these patients did not show symptoms of corneal dystrophies, indicative that when mutant TGFBIp is present at very low levels deposits do not form. This provides insight into the levels of mutant allele disruption that may be required to restore a normal phenotype in corneal dystrophy patients. Perhaps the presence of mutant TGFBIp in the cornea below a certain threshold does not result in disease, however passing this threshold results in deposit accumulation. Furthermore, it has implications on prospective gene therapy strategies for the *TGFBI* corneal dystrophies, the obvious role of TGFBIp in wound repair suggests complete knockout of the *TGFBI* gene would be inadvisable, as it would result in a cornea unable to mount a sufficient response to potential damage.

Zebrafish have become a popular model to study wound healing due to their capacity to regenerate tissues and the ease of genetic manipulation. As zebrafish tail-fins are relatively simple, symmetric structures they offer an ideal tissue to assess a gene's role in wound healing. Within this thesis *tgfbi* knockdown in a zebrafish tail-fin was shown to impair tail-fin regeneration. While this is an interesting finding, the ambiguity of how TGFBIp results in the *TGFBI* corneal dystrophies still persists. A clear mechanism of how disease manifests and how

wound repair plays a role in this process are outstanding. To date no corneal epithelial wound healing studies in zebrafish have been published, however the ability of the zebrafish corneal endothelium to regenerate has been demonstrated[16,17]. Following confirmation of the regenerative capacity of the zebrafish corneal epithelium this finding could be validated by comparing the corneal regeneration rates following corneal injury in a wild-type zebrafish and a *tgfbi* deficient zebrafish.

Recently Poulsen *et al* performed a proteomic comparison of wild-type mice and mice deficient in *Tgfbi* (*Tgfbi*-/-)[18]. However, they found minimal difference in the protein composition of the two corneas. The results from the regenerating tail-fin assay indicate that the predominant role of TGFBIp is in wound repair. Thus, future work may encompass inducing corneal injury in both wild-type and *Tgfbi*-/- mice and assessing the proteomic composition of the injured corneas. Superficial corneal injury could be induced by using a trephine to mark a wound size and employing rotating burr to remove the epithelium and basement membrane within the mark. The rate of wound closure can then be monitored by staining the wound with fluorescein and imaging the mice *in vivo* at different timepoints. Corneas can then be dissected at an appropriate timepoint and subjected to mass-spectrometry analysis to compare the proteomic profile of uninjured and injured corneas. This would provide insight into the pathways involved in corneal wound repair and how complete removal of TGFBIp impacts upon this.

7.2 Mutation-dependent allele-specific editing of the *TGFBI* corneal dystrophies

The demonstration of the critical role of TGFBIp in wound repair in Paper III highlights the need to maintain a functional copy of *TGFBI* in the cornea. Repair the *TGFBI* missense would be a potential treatment strategy, Taketani *et al* demonstrated this is patient-derived primary corneal keratocytes. However, the efficiency of homology directed repair is currently too low to be applicable for local administration. Currently, due to the low efficiencies, homologous

recombination is only applicable *ex vivo*[19]. Patient cells are CRISPR treated *ex vivo* and clones harbouring the desired edit are expanded and these clonal cells can be reintroduced into the body. This is approach is particularly effective in diseases of the blood, such as sickle cell anaemia[20]. For corneal gene editing, a limbal biopsy would be taken, which could contain limbal epithelial stem cells (LESCs). The LESCs would be edited ex vivo and reintroduced into the patient's cornea. However, in order for the clonal edited cells to repopulate the cornea, resident stem cells in the corneal would have to be completely removed. This would induce limbal stem cell deficiency (LSCD) which is an extremely severe condition. The risk that the graft of clonally edited LESCs would not take is too high, thus this would not be a feasible treatment strategy for the corneal dystrophies. However, an approach that selectively disrupts the mutant allele and leaves the wild-type allele intact would be a viable treatment option.

Disruption of the PAM sequence is known to impair Cas9 function[21–23], therefore utilising the generation of a SNP-derived PAM to achieve allele-specific editing of a missense mutation offers a promising treatment strategy. Previous reports indicated that in cases where the missense mutation generates a novel PAM stringent allele-specificity can be achieved[24–26]. However, mutational analysis of the 62 missense mutations associated with the *TGFBI* corneal dystrophies revealed <1/3 generate a novel PAM. Critically, the 5 most prevalent mutations (R124H, R124C, R124L, R555Q and R555W) do not generate a novel *S.pyogenes* PAM but have a PAM nearby. Thus an alternative approach to PAM-specific allele-specific editing, termed guide-specific allele-specific editing, in which the missense mutation is incorporated into the guide sequence, was explored for these prevalent mutations. Paper II reports the inadequacies of this guide-specific approach in comparison to the PAM-specific approach. The guide-specific approach appeared to confer differing specificities depending on the position of the missense mutation within the guide sequence, with PAM-proximal mutations offering superior specificity. These results reflect current literature that describes the tolerance of

S.pyogenes Cas9 to single base pair mismatches within the guide sequence[27–29]. In addition, these reports have implicated the region immediately adjacent to the PAM as critical in determining specificity validating the results observed in Paper II.

Smith *et al* explored a guide-specific approach to treat α1-antitrypsin (AAT) deficiency. They designed a guide incorporating the missense mutation at postion 5 within the guide sequence. They demonstrated clear discriminate between wild-type and mutant alleles in patient-derived iPSCs, albeit at very low efficiencies of ~1%[30]. Yamamoto *et al* sought to develop a guide-specific approach to treat QT syndrome. They designed a guide incorporating the missense mutation at position 1 in guide, however they opted for a nickase approach to improve genome wide specificity[31]. Again they observed very low efficiencies but successfully isolated two clones with a frameshifting mutation on only the mutant allele. Burnight *et al* developed an approach to selectively disrupt the mutant allele in the P23H mutation associated with retinitis pigmentosa[32]. They utilised *S.aureus* Cas9 and designed a guide incorporating the missense mutation at position 3 in the guide sequence. Within patient-derived iPSCs they achieved very high indel frequencies and report the majority of indels occur only with the mutant allele. Interesting, Li *et al* published a comprehensive report of allele-specific editing also targeting the P23H retinitis pigmentosa mutation *in vivo*[33]. They used engineered *S.pyogenes* variants *Sp*Cas9-KKH and *Sp*Cas9-VQR, recognising NNCAGT and NGA PAMs respectively [22,34]. Guide design incorporated the P23H mutation at position 4 or 12 within the guide sequence. However, neither nuclease variant was able to discriminate between the single base pair *in vivo*. Truncation of the guide sequence has been shown to improve specificity[35], truncation of the guide from 20bp to 17bp enabled the nuclease variants to distinguish between the two alleles. Finally, Gao *et al* investigated a guide-specific approach to treat autosomal dominant hearing loss caused by the M412K mutation in *TMC1*[36]. They investigated four different *S.pyogenes* Cas9 guides which incorporated the missense mutations at position , 10, 9 or 6 of the guide

sequence, the fourth guide was a truncated from of the guide containing the mutation at position 6. Using RNPs they achieved good allele discrimination in mouse primary fibroblasts (*Tmc*$^{M412K/+}$). *In vivo* they only observed 0.92% indel frequency which they report to predominantly occur with the mutant allele. Surprisingly, they found that 0.92%, which equated to around 1.8% disruption of the mutant allele, was enough to ameliorate hearing loss. These reports describe differing efficiencies based on the model and the Cas nuclease utilised. Indicating that application of a guide-specific approach requires individual assessment of the target sequence and cell type in which allele-specific editing is necessary.

Paper II relied solely on an *in vitro* readout of allele-specificity, using a dual-luciferase assay and an *in vitro* digest, and perhaps provided an overestimate of wild-type allele cleavage. In order to adequately demonstrate allele-specific cleavage *ex vivo* or *in vivo* a patient-derived cell line or mouse model for each mutation studied would have been required, which were not readily available. In lieu of these models two *in vitro* assays were used in parallel to assess allele-specificity. While both assays are widely accepted techniques they both have limitations affecting their reliability. In the case of the dual-luciferase assay three separate plasmids are co-transfected into AD293 cells; a plasmid expressing *firefly* luciferase with a target sequence of the gene of interest cloned into the 3'UTR of *firefly* luciferase, a plasmid expressing Cas9 and a sgRNA targeted to the gene of interest, and a plasmid expressing *renilla* luciferase. The ability for Cas9 to cleave the target site causes the plasmid to be linearised, RNA polymerase will not be able to reach the poly A tail, thus the mRNA will be degraded, resulting in a reduction in *firefly* luciferase activity. *Firefly* luciferase is normalised to *renilla* luciferase as an attempt to control for transfection efficiency. However, this assay relies on the assumption that all cells are transfected with the same copy number of each plasmid and that all plasmids are expressed at an equivalent level. The considerable variation in copy number and expression can greatly affect the reliability of this assay readout. The *in vitro* digest involves digesting a

DNA template containing the target site with a ribonucleoprotein (RNP) complex of Cas9 and a sgRNA targeted to the target site. The ratio of Cas9:sgRNA:target can be controlled for, making comparison of cleavage at different target sites more reliable. However, this assay is not carried out in a cellular context so may have little resemblance to how Cas9 will behave *in vivo*. Despite these limitations both assays indicated that a mutation dependent approach that relied upon a single base pair mismatch within the guide sequence could not confer stringent specificity. This finding is not surprising when considered in the context of genome-wide specificity, Cas9 is known to cleave at sites elsewhere in the genome that differ to the target site by several mismatches. The likelihood that Cas9 will be faced with a genomic off-target site equipped with an adjacent PAM and only a single base pair mismatch is extremely low. Thus, allele-specific editing presents a novel class of off-targets that are more challenging to address than genome-wide off-target sites. The ability to discriminate between alleles using the 20bp guide sequence instead of the 2bp PAM sequence would offer ~10 fold improved flexibility in identifying a suitable target site. Currently available CRISPR nucleases do not possess the degree of specificity required to tackle this class of off-target site. To overcome this future work will involve using directed evolution to engineer a Cas9 nuclease that has the capabilities to achieve single nucleotide resolution. A library of variant Cas9 nucleases will be generated and tested in a bacterial selection and counter selection system that will isolate Cas9 variants that cleave the on-target site but do not cleave an off-target site differing by a single base pair mismatch. These variants will then be tested in mammalian cells for their ability to achieve allele-specific editing with clinically relevant SNPs.

7.3 Mutation-independent allele-specific editing of the *TGFBI* corneal dystrophies

Paper II indicated the limitations associated with a mutation-dependent approach. While stringent allele-specificity can be achieved with a PAM-specific approach or a context dependent guide-specific approach it is not feasible, using currently available nucleases, in

cases where several mutations are associated with the disease. Paper II revealed that when employing *S.pyogenes* Cas9 20% of the known *TGFBI* missense mutations were not targetable by either approach, 30% generated a novel PAM while the remaining mutations were only targetable by a guide-specific approach. Thus, in Paper III an approach that removes the constraints of mutation-dependent targeting was investigated. As the specificity rules for PAM-generating SNPs are well established this mutation-independent approach maintains this mechanism to achieve specificity. It utilises commonly occurring SNPs that are associated with a PAM on the same allele as the disease causing mutation. Shin *et al* reported the use of the extensive knowledge of the huntington (*HTT*) gene haplotypes[37], encompassing the trinucleotide repeat (TNR) associated with Huntington's disease, to achieve allele-specific disruption of *HTT* [38]. For *HTT*, screening of only 10 SNPs is sufficient to provide haplotype information for 97% of the European Huntington's disease patients[37]. However, in the case of *TGFBI* missense mutations phase cannot be pre-determined, there is no subtle way to predict the phasing of these missense mutations.

Building on previous work we developed an approach for the *TGFBI* corneal dystrophies that would enable allele-specific targeting of the *TGFBI* locus of a patient harbouring any missense mutation, provided some heterozygosity exists in their haplotypes across the *TGFBI* locus. This approach was coined allele-specific, SNP-derived, *in cis*, personalised, CRISPR (ASNIP CRISPR). A major limitation of this approach is that it requires the design of common intronic guides that could be used in combination with the ASNIP guides. This would enable a dual-guide approach with minimal intervening distance that is capable of efficiently excising a portion of *TGFBI* coding sequence. The application of this modification would require complete assurance that the intronic region targeted does not contain any regulatory elements. As the diverse functions and actions of non-coding RNAs are still largely unknown this may be not be easily achieved[39]. Cas nucleases recognising different PAMs, such as *S.aureus* Cas9

(NNGRRT)[40] or *As*Cas12a (TTTN)[41] or indeed nucleases with engineered PAM specificities[34,42], could also be assessed for the identified common SNPs across the *TGFBI* locus. This would enable two allele-specific ASNIP guides in close proximity to be utilised to achieve an excision of the *TGFBI* coding region. The Cas nuclease toolkit is continually expanding, with a diverse repertoire of targetable PAMs the ASNIP approach holds vast potential.

Haplotype analysis of 24 SNPs within the *TGFBI* coding region that matched the ASNIP criteria was performed to determine the proportion of the East Asian population that could be targeted by this approach. Interestingly, the R124H patient with which phased sequencing was performed differed at one position to the predicted haplotype, indicating that recombination had occurred at this positon. This highlights that the haplotype predictions cannot be solely relied upon for determining at which site targetable SNPs occur. Furthermore, in this new era of personalised medicine where progress will be made with great caution whole genome sequencing (WGS) will undoubtedly be a requisite for any patient undergoing any form of gene editing therapy, in order to fully comprehend the success or failure of such therapies. As an extension to this whole-genome phased sequencing will allow the design of guides *in cis* with the mutation for autosomal disease but will also aid in the understanding of outcomes should unwanted off-targets effects or chromosomal translocations occur. Therefore, obligatory whole-genome phased sequencing would be essential to help ensure safe and successful treatment using ASNIP CRISPR.

7.4 Delivery of CRISPR/Cas9 components to the corneal layers

Due to the role of the cornea as a structural barrier to prevent entry of foreign substances into the eye, delivery to the corneal layers has proven to be quite a challenge. To-date, topical application of delivery agents has achieved minimal success, even with viral agents[43]. AAV

was investigated for its ability to transduce the corneal layers *in vivo* following an intracameral injection. Excitingly, AAV-2/9 demonstrated the ability to transduce all corneal layers following an intracameral injection. In addition, packaging of CRISPR/Cas9 components into a dual-AAV system achieved 25.7% indel frequency in the whole cornea. This provides a potential vector that could be used to deliver gene editing reagents targeted to corneal dystrophies associated in any corneal layer. Identifying or developing a delivery vehicle capable of delivery to the corneal layers following topical application would revolutionise ocular gene therapy, encouraging uptake by both clinicians and patients. Further work may entail directed evolution of a AAV vector that can transduce corneal cells following topical application. A library of mutant AAV vectors would be produced, which would be topically applied to a cornea *ex vivo.* AAV vectors can then be isolated from a specific layer or from the posterior side of the cornea, indicating that they have successfully transduced the desired region of the cornea following topical application.

Basche *et al* published a report just this month demonstrating that a systemic injection of AAV-2/9 results in transduction of the corneal epithelium, stroma and endothelium[44]. Furthermore AAV-2/9 appeared to transduce the LESCs, evidenced by sustained centripetal streaks from the limbus. This report validates our observation and importantly indicates that AAV-2/9 has the ability to transduce the LESCs. However, this report also highlights that delivery vehicles are capable of transporting from other areas of the body to the eye. This raises considerable concerns about utilising viral vectors for gene therapy. AAV has become perhaps the most widely used delivery vector, due to its low immunogenicity. However, commonly utilised AAV vectors result in sustained expression of the viral transgene. Thus, when utilised to deliver CRISPR/Cas9 components there will be continual production of Cas9:sgRNA, increasing the likelihood of off-target cleavage. Gao *et al* report that assessment of off-target cleavage in primary fibroblasts following plasmid nucleofection resulted in cleavage at 9/10 off-target sites

with efficiencies ranging from 0.68-8.1%, while off-target cleavage using RNPs produced indel at only 1/10 sites at an efficiency of 1.2%[36]. This ability of RNPs to achieve higher genome-wide specificity than plasmid is well accepted[45,46]. This is due to continual expression of CRISPR/Cas9 components by the plasmid compared to limited exposure time of the RNPs. Therefore an *in vivo* delivery system such as, as nanoparticles or cationic lipids, with the ability to deliver RNPs to the target cell type would be provide a much safer delivery vehicle [47,48]. This is an important consideration for gene editing in all contexts, but most critically for allele-specific editing. The wild-type allele exists as an almost perfect off-target site, with only a single base pair mismatch. It is extremely unlikely that Cas9 will be faced with such an off-target elsewhere in the genome. However, for allele-specific editing this off-target will always occur and it presents the most hazardous off-target site as cleavage here will unintentionally ablate the functional protein. Cas9 requires exquisite specificity to distinguish between the two alleles. As viral delivery will promote continual expression of the CRISPR components eventual cleavage at the wild-type allele will almost certainly occur. Delivery to the corneal layers and desired cell type, such as the LESCs, keratocytes or corneal endothelial cells, utilising a non-viral delivery vehicle will again aid in genome-wide specificity. Future work will entail assessing the delivery of CRISPR/Cas9 RNP complexes to the cornea using iTOP, a non-viral delivery agent shown to have achieved gene editing of both muscle and retinal cells *in vivo*[49]. iTOP (induced transduction by osmocytosis and propanebetaine) uses a combination of NaCl hypertonicity-induced micropinocytosis and a transduction compound (propanebetaine) to induce highly efficient transduction of proteins into cells. An Ai9 mouse model in which a *loxP*-flanked STOP cassette prevents transcription of a red fluorescent protein variant (tdTomato), will be used to assess delivery via these non-viral agents[50]. Initially, Cre recombinase will be used to monitor successful delivery to the cornea using non-viral agents, as its proficiency to cleave *loxP* sites is well established. If Cre successfully cleaves the

loxP sites the STOP cassette will be removed and tdTomato will be expressed. Cre in complex with these agents will be assessed in various routes of administration, including topical application, intrastromal injection and intracameral injection. Observation of tdTomato using the IVIS *in vivo* imager and confirmation of transduction via fluorescent microscopy will indicate successful delivery of Cre recombinase to the cornea using non-viral reagents. At the outset milder formulations will be used, if no tdTomato signal is observed a more concentrated buffer will be tested. Following successful delivery of Cre recombinase to the corneal layers, CRISPR RNPs will be complexed with the iTOP buffer. The ability of iTOP to deliver CRISPR RNPs targeted to the *loxP* sites to the cornea will then be assessed using the IVIS *in vivo* imager. Confirmation of gene disruption will be confirmed using targeted resequencing. Successful transduction of the corneal layers using a non-viral delivery agent would provide a promising alternative to viral delivery, in which the expression of the CRISPR components would be limited thus reducing the propensity of Cas9 to cleave the wild-type allele.

7.5 Conclusion

While this thesis presents considerable progression towards a personalised genome editing strategy to treat the corneal dystrophies, substantial work still remains to be done. The work presented within highlights the current limitations of CRISPR/Cas9 gene editing, such as a lack of global specificity and inadequate delivery vehicles. The remaining issues are not isolated to corneal gene therapy but extend to all gene therapy applications. Adequately addressing these outstanding hurdles will enable the safe translation of these therapies to the clinic. The genome-wide specificity of CRISPR/Cas9 persists as a major concern for the clinical use of this nuclease. The Cas nucleases are evolved for use in a bacterial genome which are considerably smaller in size, thus wild-type nucleases do not possess the on-target fidelity required to edit the mammalian genome. Critically, this has massive implications for allele-specific editing, as the non-target allele exists as the most challenging off-target site. The development of

nucleases capable of nuclease specificity will be paramount to the safe and successful translation of allele-specific genome editing applications to the clinic. Furthermore, viral vectors have been successful gene replacement vectors in gene therapy applications thus far, in which persisted expression is advantageous to continually produce the functional protein. However, persisted expression in the case of genome editing and particularly allele-specific genome editing will increase the likelihood of off-target cleavage. Thus, exploration of a non-viral alternative to delivery gene-editing reagents to the cornea is essential to ensure cleavage only occurs at the target site. CRISPR/Cas9 has revolutionised the field of molecular biology, providing an unrivalled tool to investigate gene function and human disease. Beyond this it holds huge promise for the treatment of devastating single gene diseases. While this is an immensely exciting time in the field of gene therapy, the progression to the clinic must not be made with haste. Appropriately addressing the current limitations of a lack of specificity and inadequate delivery vehicles will ensure the full potential of CRISPR nucleases is achieved.

7.6 Timeline of Achievements

May 2016	Award	Genetics Society Training Grant 2016 (£600)	
May 2016	Training	Hannover Medical School, Hannover, Germany, Prof Hildegard Büning	Trained in production and transduction of AAV vectors
August 2016	Award	Junior Scientist Travel Grant 2016 (£750)	
August 2016	Conference - Poster Presentation	Genome Engineering: The CRISPR-Cas Revolution, Cold Spring Harbour, USA	CRISPR/Cas9 gene editing in the cornea of a keratin 12-luciferse multi-target knock-in mouse model
October 2016	Conference - Oral Presentation	European Association for Vision and Eye Research, Nice, France	Regulating gene expression towards solving ocular surface diseases
December 2016	Training	Trinity College Dublin, Dublin, Ireland, Prof Pete Humphries	Trained to implement intracameral injections in mice *in vivo*
May 2017	Conference - Oral Presentation	ARVO Annual Meeting, Baltimore, USA	Toward genome editing to treat inherited autosomal dominant eye disease
October 2017	Conference - Poster Presentation	The CRISPR Forum, Dublin, Ireland	Development of a personalised allele-specific CRISPR system to treat autosomal dominant disorders
November 2017	Conference - Poster Presentation	The Genome Editing Congress, London, UK	Development of a personalised allele-specific CRISPR system to treat autosomal dominant disorders
November 2017	Publication	Scientific Reports	Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders
November 2017	Conference - Poster Presentation	European Alliance for Personalised Medicine Congress, Belfast, UK	Development of a personalised allele-specific CRISPR system to treat autosomal dominant disorders
January 2018	Publication	Progress in Retinal and Eye Research (PRER)	Personalised genome editing - The future for corneal dystrophies
March 2018	Training	Linköping University, Linköping, Sweden, Dr Lasse Jensen	Trained to implement a tail-fin regeneration in zebrafish *in vivo*

April 2018	Conference - Poster Presentation	ARVO Annual Meeting, Honolulu, Hawaii, USA	Development of a personalised allele-specific CRISPR system to treat autosomal dominant disorders
May 2018	Conference - Oral Presentation	The Genome Editing Congress, Boston, USA	Development of a personalised allele-specific CRISPR system to treat autosomal dominant disorders
August 2018	Publication	BMC Genomics	Gene editing in the context of an increasingly complex genome

7.7 References

1. Verma, I. M. A tumultuous year for gene therapy. *Mol. Ther.* **2,** 415–416 (2000).
2. Cavazzana, M., Six, E., Lagresle-Peyrou, C., André-Schmutz, I. & Hacein-Bey-Abina, S. Gene Therapy for X-Linked Severe Combined Immunodeficiency: Where Do We Stand? *Hum. Gene Ther.* (2016). doi:10.1089/hum.2015.137
3. Choi, S. Il *et al.* Decreased catalase expression and increased susceptibility to oxidative stress in primary cultured corneal fibroblasts from patients with granular corneal dystrophy type II. *Am. J. Pathol.* **175,** 248–261 (2009).
4. Kim, T. *et al.* Altered mitochondrial function in type 2 granular corneal dystrophy. *Am. J. Pathol.* **179,** 684–692 (2011).
5. Grothe, H. L. *et al.* Altered protein conformation and lower stability of the dystrophic transforming growth factor beta-induced protein mutants. *Mol. Vis.* **19,** 593–603 (2013).
6. Kolozsvári, L., Nógrádi, A., Hopp, B. & Bor, Z. UV absorbance of the human cornea in the 240- to 400-nm range. *Investig. Ophthalmol. Vis. Sci.* **43,** 2165–2168 (2002).
7. Courtney, D. G. *et al.* Protein composition of TGFBI-R124C- and TGFBI-R555W-associated aggregates suggests multiple mechanisms leading to lattice and granular corneal dystrophy. *Investig. Ophthalmol. Vis. Sci.* **56,** 4653–4661 (2015).
8. Choi, S. Il, Maeng, Y. S., Kim, K. S., Kim, T. I. & Kim, E. K. Autophagy is induced by raptor degradation via the ubiquitin/proteasome system in granular corneal dystrophy type 2. *Biochem. Biophys. Res. Commun.* **450,** 1505–1511 (2014).
9. Korvatska, E. *et al.* Amyloid and non-amyloid forms of 5q31-linked corneal dystrophy resulting from kerato-epithelin mutations at Arg-124 are associated with abnormal turnover of the protein. *J. Biol. Chem.* **275,** 11465–11469 (2000).
10. Korvatska, E. *et al.* On the role of kerato-epithelin in the pathogenesis of 5q31-linked

corneal dystrophies. *Investig. Ophthalmol. Vis. Sci.* **40,** 2213–2219 (1999).

11. García-Castellanos, R. *et al.* Structural and Functional Implications of Human Transforming Growth Factor β-Induced Protein, TGFBIp, in Corneal Dystrophies. *Structure* **25,** 1740–1750.e2 (2017).
12. Aldave, A. J. *et al.* A Clinical and Histopathologic Examination of Accelerated TGFBIp Deposition After LASIK in Combined Granular-Lattice Corneal Dystrophy. *Am. J. Ophthalmol.* **143,** 416–419 (2007).
13. Han, K. E. *et al.* Pathogenesis and treatments of TGFBI corneal dystrophies. *Progress in Retinal and Eye Research* **50,** 67–88 (2016).
14. Jun, R. M. *et al.* Avellino corneal dystrophy after LASIK. *Ophthalmology* **111,** 463–468 (2004).
15. Kim, T.-I. *et al.* Deposits of transforming growth factor-beta-induced protein in granular corneal dystrophy type II after LASIK. *Cornea* **27,** 28–32 (2008).
16. Heur, M., Jiao, S., Schindler, S. & Crump, J. G. Regenerative potential of the zebrafish corneal endothelium. *Exp. Eye Res.* **106,** 1–4 (2013).
17. Stepp, M. A. *et al.* Wounding the cornea to learn how it heals. *Experimental Eye Research* **121,** 178–193 (2014).
18. Poulsen, E. T. *et al.* Proteomic profiling of TGFBI-null mouse corneas reveals only minor changes in matrix composition supportive of TGFBI knockdown as therapy against TGFBI-linked corneal dystrophies. *FEBS J.* **285,** 101–114 (2018).
19. Taketani, Y. *et al.* Repair of the TGFBI gene in human corneal keratocytes derived from a granular corneal dystrophy patient via CRISPR/Cas9-induced homology-directed repair. *Sci. Rep.* **7,** (2017).
20. DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci. Transl. Med.* **8,** 360ra134-360ra134

(2016).

21. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573 (2014).

22. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481–485 (2015).

23. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33,** 187–198 (2015).

24. Courtney, D. G. *et al.* CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting. *Gene Ther.* **23,** 108–112 (2016).

25. Bakondi, B. *et al.* In vivo CRISPR/Cas9 gene editing corrects retinal dystrophy in the S334ter-3 rat model of autosomal dominant retinitis pigmentosa. *Mol. Ther.* **24,** 556–563 (2016).

26. Li, Y. *et al.* Exploiting the CRISPR/Cas9 PAM constraint for single-nucleotide resolution interventions. *PLoS One* **11,** (2016).

27. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31,** 839–843 (2013).

28. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31,** 822–826 (2013).

29. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).

30. Smith, C. *et al.* Efficient and allele-specific genome editing of disease loci in human iPSCs. *Mol. Ther.* (2015). doi:10.1038/mt.2014.226

31. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal

off-target effects. *Nat. Methods* (2014). doi:10.1038/nmeth.2857

32. Burnight, E. R. *et al.* Using CRISPR-Cas9 to Generate Gene-Corrected Autologous iPSCs for the Treatment of Inherited Retinal Degeneration. *Mol. Ther.* (2017). doi:10.1016/j.ymthe.2017.05.015

33. Li, P. *et al.* Allele-Specific CRISPR-Cas9 Genome Editing of the Single-Base P23H Mutation for Rhodopsin-Associated Dominant Retinitis Pigmentosa. *Cris. J.* **1,** 55–64 (2018).

34. Kleinstiver, B. P. *et al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33,** 1293–1298 (2015).

35. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32,** 279–284 (2014).

36. Gao, X. *et al.* Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553,** 217–221 (2018).

37. Lee, J. M. *et al.* Sequence-Level Analysis of the Major European Huntington Disease Haplotype. *Am. J. Hum. Genet.* (2015). doi:10.1016/j.ajhg.2015.07.017

38. Shin, J. W. *et al.* Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Hum. Mol. Genet.* ddw286 (2016). doi:10.1093/hmg/ddw286

39. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell* (2014). doi:10.1016/j.cell.2014.03.008

40. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520,** 186–191 (2015).

41. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759–771 (2015).

42. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523,** 481–485 (2015).

43. Moore, C. B. T., Christie, K. A., Marshall, J. & Nesbit, M. A. Personalised genome editing – The future for corneal dystrophies. *Progress in Retinal and Eye Research* **65,** 147–165 (2018).

44. Basche, M. *et al.* Sustained and widespread gene delivery to the corneal epithelium via in situ transduction of limbal epithelial stem cells using lentiviral and AAV vectors. *Hum. Gene Ther.* **44,** hum.2018.115 (2018).

45. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* **208,** 44–53 (2015).

46. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24,** 1012–1019 (2014).

47. Lee, K. *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* (2017). doi:10.1038/s41551-017-0137-2

48. Yin, H. *et al.* Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.* (2017). doi:10.1038/nbt.4005

49. D'Astolfo, D. S. *et al.* Efficient intracellular delivery of native proteins. *Cell* **161,** 674–690 (2015).

50. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13,** 133–140 (2010).