Mutation-independent allele-specific editing by CRISPR-Cas9, a novel approach to treat autosomal dominant disease

Kathleen A. Christie[1], Louise J. Robertson[1], Caroline Conway[1], Kevin Blighe[1], Larry A. DeDionisio[2], Connie Chao Shern[1,2], Amanda M. Kowalczyk[1], John Marshall[3], Doug Turnbull[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 Genetics, UCL Institute of Ophthalmology, London, EC1V 9EL, UK

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

* To whom correspondence should be addressed

Biomedical Sciences Research Institute, Ulster University, Coleraine, Northern Ireland BT52 1SA, United Kingdom.

Tel no: +44(0)2870124577. Email: tara.moore@ulster.ac.uk

Short title: Mutation-independent allele-specific CRISPR editing

Abstract

CRISPR-Cas9 provides a tool to treat autosomal dominant disease by NHEJ gene disruption of the mutant allele. In order to discriminate between wild-type and mutant alleles, *Sp*Cas9 must be able to detect a single nucleotide 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 numerous 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 CRISPR gene editing independent of the disease-causing mutation which is capable of achieving complete allele discrimination and we propose it as a targeting approach for autosomal dominant disease. Our approach utilises natural variants in the target region that contain a PAM on one allele which lies *in cis* with the causative mutation, removing the constraints of a mutation-dependent approach. Our innovative patient-specific guide design approach takes into account the patient's individual genetic make-up allowing on and off target activity to be assessed in a personalised manner.

Introduction

CRISPR-Cas9 holds immense promise for the treatment of single gene diseases, enabling sequence specific genome modifications to be induced to remove or correct the genetic cause [1–3]. The best characterised Cas nuclease, *S.pyogenes* Cas9 (*Sp*Cas9) in complex with a single guide RNA (sgRNA) searches the genome for a NGG protospacer adjacent motif (PAM). Once a NGG PAM is encountered *Sp*Cas9 determines whether the 20bp guide sequence in the sgRNA 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, lacks fidelity and can introduce insertions and deletions (indels) at the repair site. NEHJ-mediated indels introduced in the coding region of a gene can result in a frameshifting mutation leading to premature termination of translation or nonsense mediated decay of the mRNA and thus permanent disruption of the target gene. One important application of this technology is the potential to treat autosomal dominant disease by allele-specific NHEJ mediated gene disruption. Autosomal dominant disease is predominantly caused by single base-pair, missense mutations[4]. Therefore, to achieve allele-specific NHEJ mediated gene disruption of the mutant allele*, Sp*Cas9 must be able to discriminate between wild-type and mutant alleles which differ by only this 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 achieve reasonable allele discrimination with certain mutations[5–7]. However, successful application of this approach requires the mutation of interest to have a usable PAM in close proximity. The position of the missense mutation within the guide sequence, and critically within the first 8-12nt in the guide sequence, is known to limit this approach through reduced allele discrimination observed the more distal the mutation is from the PAM[8]. Similarly, although exploitation of a novel PAM has been shown to confer stringent allele-specificity, only a fraction of missense mutations will generate a novel PAM[9,10 8]. While both approaches can be efficiently utilised in the context of certain mutations, they both highlight the limitations of a mutation dependent approach.

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[11]. Transforming growth factor β-induced (*TGFBI*) has been implicated as the causative gene in some of the most commonly occurring 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, and has been shown to have a role in wound healing, angiogenesis, cancer and inflammatory diseases[12,13]. Despite the fact that TGFBIp is widely 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 >70 different disease-causing, predominantly missense mutations within *TGFBI* have been described. These mutations and the dystrophies associated with them are classified as epithelial-stromal *TGFBI* corneal dystrophies[14,15]. A very strong genotype-phenotype correlation exists between each missense mutation and the pattern of the mutant protein deposits that accumulate in the cornea. Autosomal dominant missense mutations are not amenable to conventional gene replacement therapy, as the production of mutant protein will persist. Repair of the *TGFBI* R124H missense mutation has been demonstrated in patient-derived primary corneal keratocytes[16]. However, template repair is considered a relatively infrequent event in most cell types[17]. Heterozygous nonsense mutations in *TGFBI* associated with a normal phenotype have been reported[18,19]. Indicating that *TGFBI* is haplosufficient and disruption of the mutant allele, whilst leaving the wild-type allele intact, 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[20]. However, as knockdown of mutant protein expression by siRNA is only transient and would

require continued application/injection into the eye, permanent disruption of the mutant allele would be an attractive alternative strategy. In the case of *TGFBI* corneal dystrophies, with over 70 missense mutations currently associated with disease, utilising either a guide-specific or PAM-specific approach would require the design of a different guide for each mutation. This is an insurmountable task, as greater than one third of these missense mutations cannot be targeted by either approach and not all of the remaining mutations will offer guides with good on-target and off-target profiles[8].

Using autosomal dominant *TGFBI* corneal dystrophies as a model, we present a mutation-independent allele-specific CRISPR editing approach that is capable of achieving stringent allele discrimination with wild-type *Sp*Cas9 and propose it as a targeting approach for autosomal dominant disease. Genetic variation has been shown to affect the target specificity of CRISPR by creating or abolishing prospective target sites[21,22]. This enables allele-specificity to be achieved in cases where phase cannot be pre-determined[23] utilising natural variants in the target region that are associated with a PAM that lie *in cis* with the causative mutation. Consequently, the targeting approach is no longer constrained by the location of the mutation. At the same time, employing common variants ensures that a pool of well tested guides can be used to treat the majority of affected individuals in a given population. Herein we present a workflow for gene editing of 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.

Results

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

The *TGFBI* gene covers ~35kb and contains 17 coding exons. Mutations within *TGFBI* occur in exons 4-16 but are clustered in hotspots found in exons 4, 11, 12 and 14. Previously, we analysed all missense mutations in *TGFBI* 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 [8]. Greater than a third were targetable by neither 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 [8].

Thus, the feasibility of an alternative mutation-independent strategy was explored. We proposed that allele-specificity could be achieved by targeting non-disease causing SNPs that lie *in cis* with the disease causing mutation and contain a suitable PAM on only one allele. To identify variants across the *TGFBI* locus suitable for this approach, SNPs were filtered to select those with a minor allele frequency (MAF) of > 0.1 across all of the individuals in the 1000 Genomes Project Phase 3 cohort. As TGFBI is highly conserved, SNPs with a MAF of >0.1 were only located in intronic regions. These SNPs were then examined to determine which contain a PAM on only one allele and 24 which fitted these criteria were identified. (Figure 1a and Supplementary Table 1).

Haplotype Analysis of identified SNPs across *TGFBI*

Granular corneal dystrophy type II (GCD2) (OMIM: 607541) caused by the R124H *TGFBI* mutation, is most prevalent in East Asia: in Korea, the reported prevalence is 1 in 870, while in China it rises to 1 in 400 [24]. Although we have identified 24 suitable SNPs, their

usefulness would be reduced if the PAM-associated alleles all lie together in the same linkage disequilibrium (LD) block. In order to determine whether the selected SNPs are suitable for allele discrimination in these populations, we performed haplotype analysis using the 1000 Genomes project phase 3 data for the East Asian cohort (EAS) (Supplementary Figure 1a,b) and in the sub-populations of Han Chinese in Beijing, China (CHB) (Supplementary Figure 1c,d) and Japanese in Tokyo, Japan (JPT) (Supplementary Figure 1e,f) (Highlighted in red on Supplementary Table 1) and determined the extent of the LD blocks. For both the EAS population and CHB sub-population LD blocks were located in intron 1-2, spanning intron 2-3 to intron 13-14, and spanning intron 13-14 to intron 14-15, with between three and four haplotypes described for each LD block (Supplementary Figure 1 a-d). The structure of the LD blocks in the JPT sub-population differed somewhat, with two LD blocks, spanning intron 1-2 to intron 10-11 (6 haplotypes) and intron 13-14 to intron 15-16 (3 haplotypes) (Supplementary Figure 1e,f).

Our mutation independent approach relies on variation across the target locus. The haplotypes identified in the large haploblocks (EAS-B2, CHB-B2, JPT-B1) were analysed to determine the percentage of the population in which our approach could be used to selectively disrupt only one allele. The fraction of homozygotes, in whom the two alleles cannot be distinguished, was calculated using the Hardy-Weinberg equation for multiple alleles. This showed that 66% of the EAS population and 67% of the CHB and 71% of the JPT sub-populations are heterozygous for these alleles and therefore potentially targetable. Critically, since this analysis was performed using only the largest haploblock in each population, the actual proportion of the population that is potentially targetable may be larger.

The distribution of the PAM-associated SNPs was then assessed in each allele and used to calculate the number of targetable SNPs for each possible heterozygous combination of alleles. This analysis (Supplementary Figure 2) reveals that, even when only considering the largest haploblock, the proposed approach has the potential to target, for at least one position, all heterozygous combinations across all populations investigated, indicating that the 24 SNPs identified could be used to treat the majority of East Asian patients.

Guide design based on patient haplotype

To validate this approach we performed phased sequencing of a Japanese patient harbouring a R124H *TGFBI* GCD2 mutation which allowed identification of SNPs associated with a PAM on only one allele that lie *in cis* with the patient's R124H mutation. (Figure 2, Supplementary Table 2). A range of guides targeted to the PAM located on the mutant allele were then designed (Figure 3a, Supplementary Table 3). 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. A PCR product containing either the allele associated with a PAM or the allele with no PAM present (Figure 3b), was incubated with ribonucleoprotein (RNP) complexes of Cas9 and sgRNA. Digestion products were then electrophoresed on an agarose gel and the intensity of the digested products revealed the *in vitro* specificity of each guide (Figure 3c). Of the 12 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. It appeared 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 [25,26], on the 'No PAM present' allele, only conferred partial discrimination at best. These results suggest that, in order to achieve stringent allele-specificity, the sequence of the non-target allele, *in trans* with the mutation,

should be NGC, NCG, NGT or NTG and an NGG PAM should be *in cis* with the mutation on the targeted allele

Allele-specificity of single guides in R124H lymphocyte cell line

To test the mutation-independent allele-specific approach, a proliferating lymphocyte cell line (LCL) was generated utilising peripheral blood mononuclear cells (PBMC) from the GCD2 patient harbouring a *TGFBI* R124H mutation. RNP complexes of SpCas9 and nine modified synthetic sgRNAs, previously tested by *in vitro* digestion, were individually nucleofected into the R124H LCLs. To determine the allele-specificity of each guide, targeted resequencing across the on-target region, where Cas9 is predicted to cleave, was performed. The target region for all 9 guides was PCR amplified and subjected to deep sequencing. Computational analysis was performed to determine whether indels had occurred and with which allele they were associated. For the guides screened, we found that all could efficiently distinguish between 'PAM associated' and 'No PAM present' alleles (Figure 4a). 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. This indicates that careful guide design and testing can achieve stringent allele-specificity in a mutation-independent manner. However, in contrast to the observations of the *in vitro* digestion of PCR products (Figure 3b), stringent allele-specificity was observed regardless of whether of a non-canonical PAM was present on the non-target allele. In addition, the guide targeting rs6860369 appeared inactive in the *in vitro* screen but was active in a cellular context. This indicates that the *in vitro* screen, while largely suitable as a means of selecting active guides, does not serve as a predictive tool for allele specificity in a cellular context. For 7 out of 9 guides tested the

predominant indels observed were 1 or 2bp insertions, which occurred 3 or 4bp upstream of the PAM (Figure 4b,c,d, Supplementary Figure 3).

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

The *in cis* PAM-associated approach presented here removes the limitations of a mutation-dependent approach. All of the non-disease-causing SNPs matching the guide selection criteria were located in intronic regions, thus indels introduced by single guides are unlikely to have therapeutic potential. To overcome this an *in cis*, dual-guide approach targeted to the mutant allele is required. Upon careful design, excision of the exon/s between the guides will result in a clearly defined and consistent frameshift that will result in premature termination of translation or 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* the region between the dual-guides must be excised at a high frequency. However, the minimum reduction of TGFBIp in the cornea required to achieve a therapeutic effect is unknown.

In some cases, the target SNPs described lie substantial distances apart, up to >18kb (Supplementary Table 4). As the efficiency of deletion drops with increasing intervening distance[27], additional guides were designed that lie closer to a particular PAM discriminatory guide yet still allow excision of exons (Supplementary Table 5). In contrast to the PAM discriminatory guides, the additional guides are not allele-specific, they were selected to target the intronic region of both alleles (Figure 5a). It was hypothesised that the PAM discriminatory guide will only cut the mutant allele while the common-intronic guide will cut both alleles. When both cuts are made on the mutant allele the chromosomal region between these cuts may be deleted. The wild-type allele will only be cut by the common-intronic guide

227 which, provided meticulous design has been applied to avoid important regulatory elements,
228 should have no functional effect.

229 Five dual gRNA combinations ranging in intervening size from 419bp to 4008bp were tested
230 by nucleofection into R124H patient derived LCLs (Supplementary Table 6). One pair (Dual 2)
231 included the previously validated R124H specific gRNA, which cuts beside the mutation SNP
232 within Exon 4 [8]. To confirm that the expected deletion was occurring, we performed PCR and
233 Sanger sequencing analysis on genomic DNA from cells nucleofected with each pair. PCR
234 amplification across the deletion site for all pairs revealed a band from DNA isolated from
235 treated cells corresponding to the expected size (Figure 5b). Sanger sequencing analysis of
236 each deletion PCR product using both the forward and reverse primer showed a wildtype
237 sequence trace until the gRNA cut site, followed by mixed trace, indicating that regions
238 outside of the deletion were unaffected (Figure 5c, Supplementary Figure 4). In some cases
239 (e.g. dual 4, Figure 5c), we were able to determine the precise deletion of the intervening
240 sequence as the trace after the cut site showed the expected truncation sequence or else
241 with a 1bp indel. Other traces were mixed after the cleavage site indicating that varied indel
242 events occurred alongside deletion which did not, however, extend beyond the predicted
243 Cas9 cleavage site. Importantly, we were also able to confirm allele specificity of the deletion
244 for many of the pairs when the strand recognised by the sgRNA placed the PAM on the
245 opposite side of the Cas9 cleavage site to the deletion. Thus, when the smaller, deletion band
246 was sequenced only the PAM-associated SNP allele and not the alternative non-PAM allele
247 was detectable in the trace (e.g. Figure 5c). Together these analyses confirm that the dual
248 gRNA approach for specifically editing the mutation allele in TGFBI affected patients produces
249 the expected deletions in genomic DNA.

Whole genome analysis of CRISPR off-targets

In addition to demonstrating on target allele specificity it is necessary to demonstrate that the selected guides do not cause deleterious alterations elsewhere in the genome. We performed an unbiased *in vitro* genome wide screen for potential off targets performed using Circularization for In vitro Reporting of Cleavage Effects by sequencing (CIRCLE-seq)[30]. After employing a stringent cut-off threshold of greater than or equal to 25% of the top hit by read count we still detected potential off target cleavage sites for all of the guides tested (median 31, range 13-161) (Supplementary Table 7, Supplementary Figure 5). The *TGFBI* on-target site was consistently detected with high read count, however it was not always the top hit (Supplementary Table 8). The CIRCLE-seq results contained many of the Benchling *in silico*-predicted off target sites at high read count but also many others that were not predicted. Off targets were detected in gene coding regions (in 4/8 guides), none of which were predicted by the *in silico* design programs used. CIRCLE-seq reproducibility was tested by performing a technical replicate analysis on one guide; where separate CIRCLE-seq libraries were prepared from the same DNA sample and sequenced on separate lanes on different days. There was a high correlation between technical replicates when results were ranked by read number ($r^2 = 0.83$).

To validate CIRCLE-seq results by an independent method we selected a number of off target hits across all guides and performed Sanger sequencing and TIDE analysis[31]. Targets to validate CIRCLE-seq results were selected as follows: 1. Top 5 off target hits by CIRCLE-seq read count, 2. All off-targets in gene-coding regions above the CIRCLE-seq read count threshold, 3. On target *TGFBI* location (if not included in 1.), 4. For two guides we selected the top 5 off targets detected *in silico* by Benchling. The total number of selected targets

was 62; of these, 50 successfully passed primer design criteria and 24 produced Sanger sequencing traces of sufficient quality for off target analysis by TIDE (Supplementary Tables 8, 9a and 9b). For the off targets that could be analysed by TIDE, 75% were validated as being true off targets (18/24). There were 10 off targets in gene coding regions across all guides that were detected by CIRCLE-seq and not by *in silico* design; seven failed validation, one was not validated and two did validate, with *in vitro* cutting efficiencies of 12% and 5% (Supplementary Table 9a).

Discussion

Conventionally, gene therapy refers to gene replacement strategies, whereby a functional copy of the defective gene is introduced to ameliorate the disease[32]. This approach has been widely adapted for the treatment of autosomal recessive retinal dystrophies[33]. However, gene replacement is not appropriate for the treatment of autosomal dominant diseases, such as the *TGFBI* corneal dystrophies, due to the persistent production of mutant protein from the still present mutant allele. Conversely, an approach that aims to correct or disrupt the mutant allele can offer a viable treatment strategy for autosomal dominant disease.

Previous reports have described alternative approaches to treat the corneal dystrophy caused by individual *TGFBI* mutations. Courtney *et al* demonstrated that allele-specific siRNAs targeted to a lattice corneal dystrophy (LCD1) (OMIM:122200) R124C mutation could achieve potent and specific knockdown of the mutant allele in patient-derived corneal epithelial cells[20]. However, as knockdown of mutant protein expression by siRNA is only transient continued application would be required. Alternatively, repair of the *TGFBI* R124H missense mutation by CRISPR-Cas9 mediated homology directed repair (HDR) has been demonstrated in patient-derived primary corneal keratocytes[16]. However, despite extensive efforts to

improve the efficiency of HDR, template repair is considered a relatively rare event, limiting *in vivo* use [34,35]. Furthermore, as HDR is restricted to the G2 and S phases of the cell cycle, non-dividing cells such as corneal endothelial cells are not targetable by this approach.

We have previously reported more than one third of the mutations associated with *TGFBI* corneal dystrophies are not targetable by either a guide- or PAM-specific approach[8]. Therefore, a strategy based on targeting each individual mutation would provide an incomplete approach to treat these dystrophies. While SpCas9 has been shown to tolerate single base-pair mismatches between the guide and target sequences [8,36–38], mutations within the PAM are much less tolerated and have been shown to impair the cleavage efficiency of Cas9[39–41]. The use of truncated guide RNAs or high-fidelity variants could improve specificity within the guide RNA[42–45]. However, we chose an approach, focused on allelic differences resulting in a PAM site on one allele, to derive more stringent allele-specificity that removes the requirement that the disease-causing mutation itself must create a novel PAM. Nucleases with altered PAM specificities[6,40,43,46–48] will broaden the targeting capacity and allow additional dual-guide combinations, highly specific for the mutant allele but in closer proximity to one another, to be designed. The method presented 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 where phase cannot be pre-determined.

In order to prevent expression of the disease-causing mutant protein further considerations for guide design are required. It is widely accepted that if 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[49,50].

Therefore, to induce NMD and selectively degrade the mutant allele guides that target exons early in the transcript will be most desirable. When this concept is applied to *TGFBI* a premature termination codon no later than 50-55 nucleotides from the 5' splice site in exon 15 will result in NMD. This is evident from an examination of the corneal dystrophy-causing mutations in *TGFBI* which are either missense mutations or in frame indels, with the exception of a frameshifting single base deletion at codon 626 reported by Munier *et al* [51]. 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.

If common intronic guides are required to increase deletion frequency, 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[52], 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 demonstrated 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[53]. While this would indicate that, provided they are highly allele-specific, single 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. Furthermore, while we have shown encouraging results for the use of a dual-cut to induce a therapeutically relevant deletion, Kosicki *et al* also reported that complex deletions and rearrangements may also be occurring, which would be undetectable by our PCR based assays.

The approach described within necessitates thorough genomic characterization of the target loci. In this new era of personalised medicine, where progress will be made with great caution, whole genome sequencing (WGS) will undoubtedly be a prerequisite for any patient undergoing gene editing therapies to allow the potential for success or failure of such therapies to be fully assessed. We would advocate that patients undergo whole-genome phased sequencing to enable the design of guides in cis with the mutation for autosomal disease. The phase information will also aid in the understanding of outcomes should unwanted off-targets effects, such as those detected by the in vitro and ex vivo assays used here, or chromosomal translocations occur. Other hurdles such as efficiency of delivery of the gene editing components to the target cells remain to be addressed, however the approach presented here offers a promising strategy to allele-specific gene therapy in a mutation independent manner for autosomal dominant disease.

Materials and Methods

Phased sequencing of R124H patient genome

Genomic DNA was extracted from 3ml 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[54] Phase III data in gzipped variant call format (VCF)[55] 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[56]. 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[57].

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[58] 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[59], 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 50bp insert of *TGFBI* sequence encompassing the target site and PAM for either the 'No PAM' or 'PAM present' allele was cloned into the MCS of psiTEST-LUC-Target (York Bioscience Ltd, York, UK). A 587bp double-stranded DNA template was prepared by amplifying a region of the template plasmid using the primers listed in Supplementary Table 10. 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 PBMC

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 10 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 10. PCR products were purified using the Wizard® PCR

437 Preps DNA Purification System (Promega)and subjected to TruSeq PCR free library
438 preparation. Samples then underwent paired end sequencing using an Illumina MiSeq
439 instrument as per the manufacturer's instructions. For genomic DNA samples, paired FASTQ
440 files first underwent read filtration and trimming with Trim Galore! V0.4.0
441 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (utilising Cutadapt
442 v1.15 and FastQC v0.11.5), using default parameters and --qual 20 --length 70 –paired.
443 Reads from human samples were then aligned to the reference genome GRCh38 / hg38 /
444 GCA_000001405.15 (downloaded from the UCSC), using BWA v0.7.12 (mem algorithm with
445 default settings) [60]. Aligned reads in SAM format were converted to BAM, sorted, and
446 indexed with SAMtools v1.3.1[56]. PCR and optical duplicates were marked with Picard v1.119
447 (https://broadinstitute.github.io/picard/) and then expunged with SAMtools view function
448 with parameter -F set to 0x400. Reads with MAPQ below 30 were also expunged using
449 SAMtools view with parameter -q set to 30. Output BAMs were then sorted and indexed
450 using SAMtools. Reads in each sample's BAM file were then split based on the SNP of
451 interest. This was achieved using SAMtools view to first extract reads overlapping the target
452 SNP region, and then dividing these into allele-specific reads by using the shell function grep
453 -e to extract reads containing each SNP of interest flanked by 3 bases in both the 3' and 5'
454 directions. Allele-specificity of the resulting reads was visually checked for each sample in
455 IGV v2.3.97 (http://www.broadinstitute.org/igv). The number of properly-paired reads in
456 each allele-specific BAM file with and without indels was then tabulated by using SAMtools
457 view in combination with the shell function awk to filter on the CIGAR string. For example:
458 Reads with indels: samtools view -f 0x02 Allele1.BAM | awk '$6 ~ "I|D"', reads with no
459 indels: samtools view -f 0x02 Allele2.BAM | awk '$6 !~ "I|D"'. Separately, for each allele-
460 specific BAM file, pindel v0.2.5b9 [61] 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.

PCR and Sanger sequencing of genomic deletions

Genomic DNA extracted from LCLs transfected with each pair of sgRNAs and untransfected control was amplified using primers >80bp outside the cleavage site for each dual pair (Supplementary Table 10) and run on 1% agarose gel to visualise deletions. Bands corresponding to the size predicted to result from deletion were gel extracted or PCR purified (Wizard SV Gel and PCR purification system, ProMega) and Sanger sequenced (DNA sequencing facility, University of Cambridge, UK) from both the forward and reverse primer. Sequencing traces were analysed using A plasmid Editor (ApE) and aligned using Benchling.

Whole genome off-target analysis

Whole genome CRISPR off-target analysis was performed in-vitro on gDNA from the PBMC EBV transformed cell line described above using CIRCLE-seq as previously described[30]. Briefly, for each guide, 25 µg DNA was sheared to 300bp by sonication. For each guide to be tested, 4 x 4µg aliquots of sheared DNA was subject to end repair, A-tailing and ligation to a uracil-containing stem loop adapter using the KAPA HTP Library Preparation Kit PCR Free (Roche) followed by circularization and enzymatic digestion of non-circularized DNA. Circularized DNA was pooled to obtain 125ng DNA for in vitro sgRNA guided Cas9 digestion and linearized DNA fragments were subject to sequencing library preparation followed by PCR amplification with barcoded universal primers NEBNext Multiplex Oligos for Illumina (NEB). All samples were

subject to 150bp paired-end sequencing on an Illumina MiSeq instrument as per the manufacturer's instructions.

CIRCLE-seq data analysis was performed with CIRCLEseq (v1.1)[30] using Python 2.7 in a dedicated Conda environment. BWA (v0.7.17) and SAMtools (v1.9) were also installed in this environment and utilised by the CIRCLEseq algorithm. The UCSC December 2013 release of GRCh38 / hg38 was used as the reference genome (download: http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/). The parameters in the config files passed to the CIRCLEseq 'all' algorithm were: 'window_size: 3; mapq_threshold: 50; start_threshold: 1; gap_threshold: 3; mismatch_threshold: 6; merged_analysis: False; variant_analysis: True'. Annotation of regions identified by CIRCLEseq was performed with ANNOVAR (release 2015-06-17) against 'refGene' and 'cytoBand' resources. Generation of Manhattan plots was performed in R Programming Language (v3.5.2) using custom scripts and the *qqman* package.

Validation of off-targets identified by CIRCLE-seq

The top 5 off target sites identified by CIRCLE-seq by read count and any off targets located in gene coding regions were selected for validation by Sanger sequencing and TIDE analysis[31]; an automated method to report on the mutation spectrum and efficiency of genome editing by sanger sequence trace decompostition. In total 62 sites across 8 tested guides were selected for validation. Genomic DNA extracted from the corresponding RNP transfected lymphoblastoid cell line ex vivo model described above was subject to PCR amplification, sanger sequencing of the target region, followed by TIDE analysis using the

online analysis tool (https://tide.nki.nl/). The same steps were completed for DNA extracted from the untreated lymphoblastoid cell line to act as a control for analysis purposes.

Acknowledgements

This work was supported by Avellino Labs USA.

We would like to thank Synthego, specifically Kevin Holden and Travis Maures, for their kind gift of sgRNA and their helpful discussions and contribution.

Author disclosure statement

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

References

1. Jinek, M, Chylinski, K, Fonfara, I, Hauer, M, Doudna, JA and Charpentier, E (2012). A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptice Bacterial Immunity. *Science* **337**: 816–822.

2. Cong, L, Ran, FA, Cox, D, Lin, S, Barretto, R, Habib, N, *et al.* (2013). Multiplex Genome Engineering Using CRISPR/Cas System. *Science (80-.).* **339**: 819–824.

3. Mali, P, Yang, L, Esvelt, KM, Aach, J, Guell, M, DiCarlo, JE, *et al.* (2013). RNA-guided human genome engineering via Cas9. *Science (80-.).* **339**: 823–826.

4. Guo, Y, Wei, X, Das, J, Grimson, A, Lipkin, SM, Clark, AG, *et al.* (2013). Dissecting disease inheritance modes in a three-dimensional protein network challenges the 'guilt-by-Association' principle. *Am. J. Hum. Genet*.doi:10.1016/j.ajhg.2013.05.022.

5. Gao, X, Tao, Y, Lamas, V, Huang, M, Yeh, WH, Pan, B, *et al.* (2018). Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**: 217–221.

6. Li, P, Kleinstiver, BP, Leon, MY, Prew, MS, Navarro-Gomez, D, Greenwald, SH, *et al.* (2018). Allele-Specific CRISPR-Cas9 Genome Editing of the Single-Base P23H Mutation for Rhodopsin-Associated Dominant Retinitis Pigmentosa. *Cris. J.* **1**: 55–64.

7. György, B, Lööv, C, Zaborowski, MP, Takeda, S, Kleinstiver, BP, Commins, C, *et al.* (2018). CRISPR/Cas9 mediated disruption of the Swedish APP allele as a therapeutic approach for early-onset Alzheimer's disease. *Mol. Ther. - Nucleic Acids* **11**: 429–440.

8. Christie, KA, Courtney, DG, Dedionisio, LA, Shern, CC, De Majumdar, S, Mairs, LC, *et al.* (2017). Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. *Sci. Rep.* **7**.

9. Courtney, DG, Moore, JE, Atkinson, SD, Maurizi, E, Allen, EHA, Pedrioli, DML, *et al.* (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.

10. Bakondi, B, Lv, W, Lu, B, Jones, MK, Tsai, Y, Kim, KJ, *et al.* (2016). 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.

11. Klintworth, GK (2009). Corneal dystrophies. *Orphanet J. rare Dis. Cornea Dystrophy Foundaion* **4**: 7.

12. Han, KE, Choi, S Il, Kim, TI, Maeng, YS, Stulting, RD, Ji, YW, *et al.* (2016). Pathogenesis and treatments of TGFBI corneal dystrophies. *Prog. Retin. Eye Res.* **50**: 67–88.

13. Lakshminarayanan, R, Vithana, EN, Chai, SM, Chaurasia, SS, Saraswathi, P, Venkatraman, A, *et al.* (2011). A novel mutation in transforming growth factor-beta induced protein (TGFβIp) reveals secondary structure perturbation in lattice corneal

dystrophy. *Br. J. Ophthalmol.*doi:10.1136/bjophthalmol-2011-300651.

14. Weiss, JS, Møller, HU, Aldave, AJ, Seitz, B, Bredrup, C, Kivelä, T, *et al.* (2015). IC3D Classification of Corneal Dystrophies—Edition 2. *Cornea* **34**: 117–159.

15. Weiss, JS, Moller, HU, Lisch, W, Kinoshita, S, Aldave, AJ, Belin, MW, *et al.* (2008). The IC3D Classification of the Corneal Dystrophies. *Cornea* **27**: S1–S42.

16. Taketani, Y, Kitamoto, K, Sakisaka, T, Kimakura, M, Toyono, T, Yamagami, S, *et al.* (2017). 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**.

17. Richardson, CD, Kazane, KR, Feng, SJ, Zelin, E, Bray, NL, Schäfer, AJ, *et al.* (2018). CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* **50**: 1132–1139.

18. Sakimoto, T, Kanno, H, Shoji, J, Kashima, Y, Nakagawa, S, Miwa, S, *et al.* (2003). A novel nonsense mutation with a compound heterozygous mutation in TGFBI gene in lattice corneal dystrophy type I. *Jpn. J. Ophthalmol.* **47**: 13–17.

19. Song, JS, Lim, DH, Chung, ES, Chung, TY and Ki, CS (2015). Mutation analysis of the TGFBI gene in consecutive korean patients with corneal dystrophies. *Ann. Lab. Med.* **35**: 336–340.

20. Courtney, DG, Atkinson, SD, Moore, JE, Maurizi, E, Serafini, C, Pellegrini, G, *et al.* (2014). Development of allele-specific gene-silencing siRNAs for TGFBI Arg124Cys in lattice corneal dystrophy type i. *Investig. Ophthalmol. Vis. Sci.* **55**: 977–985.

21. Lessard, S, Francioli, L, Alfoldi, J, Tardif, J-C, Ellinor, PT, MacArthur, DG, *et al.* (2017).

572 Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at
573 therapeutically implicated loci. *Proc. Natl. Acad. Sci.* **114**: E11257–E11266.

574 22. Scott, DA and Zhang, F (2017). Implications of human genetic variation in CRISPR-based
575 therapeutic genome editing. *Nat. Med.* **23**: 1095–1101.

576 23. Shin, JW, Kim, K-H, Chao, MJ, Atwal, RS, Gillis, T, MacDonald, ME, *et al.* (2016).
577 Permanent inactivation of Huntington's disease mutation by personalized allele-
578 specific CRISPR/Cas9. *Hum. Mol. Genet.*: ddw286doi:10.1093/hmg/ddw286.

579 24. Chao-Shern, C, Me, R, Dedionisio, LA, Ke, BL, Nesbit, MA, Marshall, J, *et al.* (2018). Post-
580 LASIK exacerbation of granular corneal dystrophy type 2 in members of a Chinese
581 family. *Eye* **32**: 39–43.

582 25. Hsu, PD, Scott, DA, Weinstein, JA, Ran, FA, Konermann, S, Agarwala, V, *et al.* (2013).
583 DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**: 827–832.

584 26. Zhang, Y, Ge, X, Yang, F, Zhang, L, Zheng, J, Tan, X, *et al.* (2014). Comparison of non-
585 canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.* **4**.

586 27. Canver, MC, Bauer, DE, Dass, A, Yien, YY, Chung, J, Masuda, T, *et al.* (2014).
587 Characterization of genomic deletion efficiency mediated by clustered regularly
588 interspaced palindromic repeats (CRISPR)/cas9 nuclease system in mammalian cells. *J.*
589 *Biol. Chem.* **289**: 21312–21324.

590 28. Richardson, CD, Ray, GJ, Bray, NL and Corn, JE (2016). Non-homologous DNA increases
591 gene disruption efficiency by altering DNA repair outcomes. *Nat. Commun.* **7**.

592 29. Ma, L and Chung, WK (2014). Quantitative analysis of copy number variants based on
593 real-time lightcycler PCR. *Curr. Protoc. Hum.*

*Genet.*doi:10.1002/0471142905.hg0721s80.

30. Tsai, SQ, Nguyen, NT, Malagon-Lopez, J, Topkar, V V, Aryee, MJ and Joung, JK (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nat. Methods* **14**: 607–614.

31. Brinkman, EK, Chen, T, Amendola, M and Van Steensel, B (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**.

32. Moore, CBT, Christie, KA, Marshall, J and Nesbit, MA (2018). Personalised genome editing – The future for corneal dystrophies. *Prog. Retin. Eye Res.* **65**: 147–165.

33. DiCarlo, JE, Mahajan, VB and Tsang, SH (2018). Gene therapy and genome surgery in the retina. *J Clin Invest* **128**: 2177–2188.

34. Paquet, D, Kwart, D, Chen, A, Sproul, A, Jacob, S, Teo, S, *et al.* (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*doi:10.1038/nature17664.

35. Cox, DBT, Platt, RJ and Zhang, F (2015). Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**: 121–131.

36. Fu, Y, Foden, JA, Khayter, C, Maeder, ML, Reyon, D, Joung, JK, *et al.* (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**: 822–826.

37. Cho, SW, Kim, S, Kim, Y, Kweon, J, Kim, HS, Bae, S, *et al.* (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**: 132–141.

615 38. Josephs, EA, Kocak, DD, Fitzgibbon, CJ, McMenemy, J, Gersbach, CA and Marszalek, PE
616 (2015). Structure and specificity of the RNA-guided endonuclease Cas9 during DNA
617 interrogation, target binding and cleavage. *Nucleic Acids Res.* **43**: 8924–8941.

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

620 40. Kleinstiver, BP, Prew, MS, Tsai, SQ, Topkar, V V., Nguyen, NT, Zheng, Z, *et al.* (2015).
621 Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**: 481–
622 485.

623 41. Tsai, SQ, Zheng, Z, Nguyen, NT, Liebers, M, Topkar, V V, Thapar, V, *et al.* (2014). GUIDE-
624 seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.
625 *Nat. Biotechnol.* **33**: 187–197.

626 42. Fu, Y, Sander, JD, Reyon, D, Cascio, VM and Joung, JK (2014). Improving CRISPR-Cas
627 nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**: 279–284.

628 43. Kleinstiver, BP, Pattanayak, V, Prew, MS, Tsai, SQ, Nguyen, NT, Zheng, Z, *et al.* (2016).
629 High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target
630 effects. *Nature* **529**: 490–495.

631 44. Slaymaker, IM, Gao, L, Zetsche, B, Scott, DA, Yan, WX and Zhang, F (2016). Rationally
632 engineered Cas9 nucleases with improved specificity. *Science (80-.).* **351**: 84–88.

633 45. Chen, JS, Dagdas, YS, Kleinstiver, BP, Welch, MM, Sousa, AA, Harrington, LB, *et al.*
634 (2017). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**:
635 407–410.

636 46. Zetsche, B, Gootenberg, JS, Abudayyeh, OO, Slaymaker, IM, Makarova, KS,

637 Essletzbichler, P, *et al.* (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2
638 CRISPR-Cas System. *Cell* **163**: 759–771.

639 47. Gao, L, Cox, DBT, Yan, WX, Manteiga, JC, Schneider, MW, Yamano, T, *et al.* (2017).
640 Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35**: 789–792.

641 48. Kleinstiver, BP, Prew, MS, Tsai, SQ, Nguyen, NT, Topkar, V V., Zheng, Z, *et al.* (2015).
642 Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying
643 PAM recognition. *Nat. Biotechnol.* **33**: 1293–1298.

644 49. Nagy, E and Maquat, LE (1998). A rule for termination-codon position within intron-
645 containing genes: When nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**:
646 198–199.

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

649 51. Munier, FL, Frueh, BE, Othenin-Girard, P, Uffer, S, Cousin, P, Wang, MX, *et al.* (2002).
650 BIGH3 mutation spectrum in corneal dystrophies. *Investig. Ophthalmol. Vis. Sci.* **43**:
651 949–954.

652 52. Monteys, AM, Ebanks, SA, Keiser, MS and Davidson, BL (2017). CRISPR/Cas9 Editing of
653 the Mutant Huntingtin Allele In Vitro and In Vivo. *Mol. Ther.* **25**: 12–23.

654 53. Kosicki, M, Tomberg, K and Bradley, A (2018). Repair of double-strand breaks induced
655 by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat.*
656 *Biotechnol.*doi:10.1038/nbt.4192.

657 54. 1000 genome project consortium (2010). A map of human genome variation from
658 population-scale sequencing. *Nature* **467**: 1061–73.

55. Danecek, P, Auton, A, Abecasis, G, Albers, CA, Banks, E, DePristo, MA, *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

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

57. Purcell, S, Neale, B, Todd-Brown, K, Thomas, L, Ferreira, MAR, Bender, D, *et al.* (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**: 559–575.

58. Barrett, JC, Fry, B, Maller, J and Daly, MJ (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**: 263–265.

59. Gabriel, SB, Schaffner, SF, Nguyen, H, Moore, JM, Roy, J, Blumenstiel, B, *et al.* (2002). The structure of haplotype blocks in the human genome. *Science (80-.).* **296**: 2225–2229.

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

61. Ye, K, Schulz, MH, Long, Q, Apweiler, R and Ning, Z (2009). Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**: 2865–2871.

Figure 1: Non-disease causing mutations within the *TGFBI* locus with a MAF of >0.1 are identified, these SNPs are then analysed to determine if they generate a novel *S.pyogenes* PAM (NGG) on only one allele. Allele-specific guides are designed, and prospective guides are analysed using *in silico* design programs.

Figure 2: 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 allele-specific gene-editing, 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 3: *In vitro* assessment of guide allele-specificity a) Based on the phased sequencing data 12 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 high on-target scores scores using Benchling an *in silico* off-target predictor tool. The *TGFBI* gene is shown in green. Location of the 12 guides and the R124H mutation are depicted by grey arrows. The R124H mutation is shown in red. b) Schematic to shown how the cleavage templates were generated. A 50bp region containing

the target site and PAM was cloned into the MCS of a reporter plasmid. Primers flanking this region were to generate a 587bp product with an offset target site. c) Initially *in vitro* digests were used to determine the allele-specificity of the 12 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.

Figure 4 a) LCLs were transfected with RNPs for each of the 12 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. **b)** Representative DNA repair outcomes for the 'PAM associated' allele for the rs7725702 guide shown in Figure 6a. Numbering indicates the frequency of the edit observed, with 1 being the most frequently observed indel. Insertions are shown in green and deletions are shown in red, SNP resulting in a PAM shown in yellow **c)** Respective schematic for the 'No PAM' allele for the rs7725702 guide shown in Figure 6a **d)** Frequency of indels for the rs7725702 guide shown for the 'PAM associated' and 'No PAM' allele.

Figure 5: Allele-specific dual gRNA editing in patient derived LCLs. a) Schematic of dual editing approach when using a common intronic gRNA (CI-gRNA) on one side. The wildtype allele has only the CI-gRNA PAM (green). The mutation-associated allele has a SNP associated PAM (purple) and a CI-gRNA PAM. When editing takes place, a double-stranded break (DSB) is induced and the region between the two cuts is excised on the mutation-associated allele, while, at most, only a small indel should occur in the intron of the wildtype allele leaving the exon intact. b) PCR products revealing deletion of the region lying between dual-guides. A

forward primer upstream of target site 1 and reverse primer downstream of target site 2 was used to amplify across the intended deletion. There is no band in the untreated lane (UT) when the primer sites are too far apart for PCR amplification and a PCR product of the expected size (red box) is only obtained in the treated cells (T) when the deletion brings them within amplification range. Water controls remain negative for amplification (H_2O). c) Example of Sanger sequencing chromatograms showing editing events when cells were transfected with dual pair 4 (creating a 419bp deletion). Sanger sequencing was performed on gel-extracted deletion bands PCR-product and sequenced from the forward (F) and reverse (R) primer. The reverse chromatogram is displayed as a reverse complement to aid visualization. The wildtype (WT) and R124H mutation-associated allele sequences are displayed across the top of the trace. gRNA target sequences and direction are indicated above and NGG PAM highlighted in red. The SNP nucleotide is underlined in bold. *denotes where the PAM-associated allele SNP matches the chromatogram. The predicted deletion sequence is shown below.

Figure 1

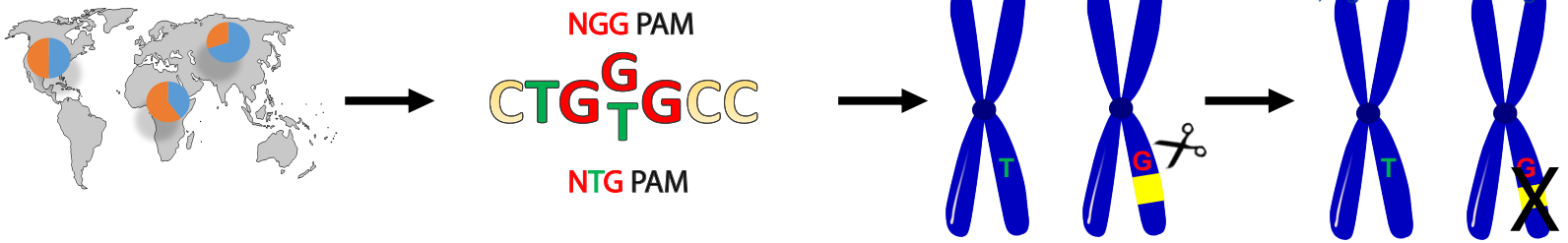
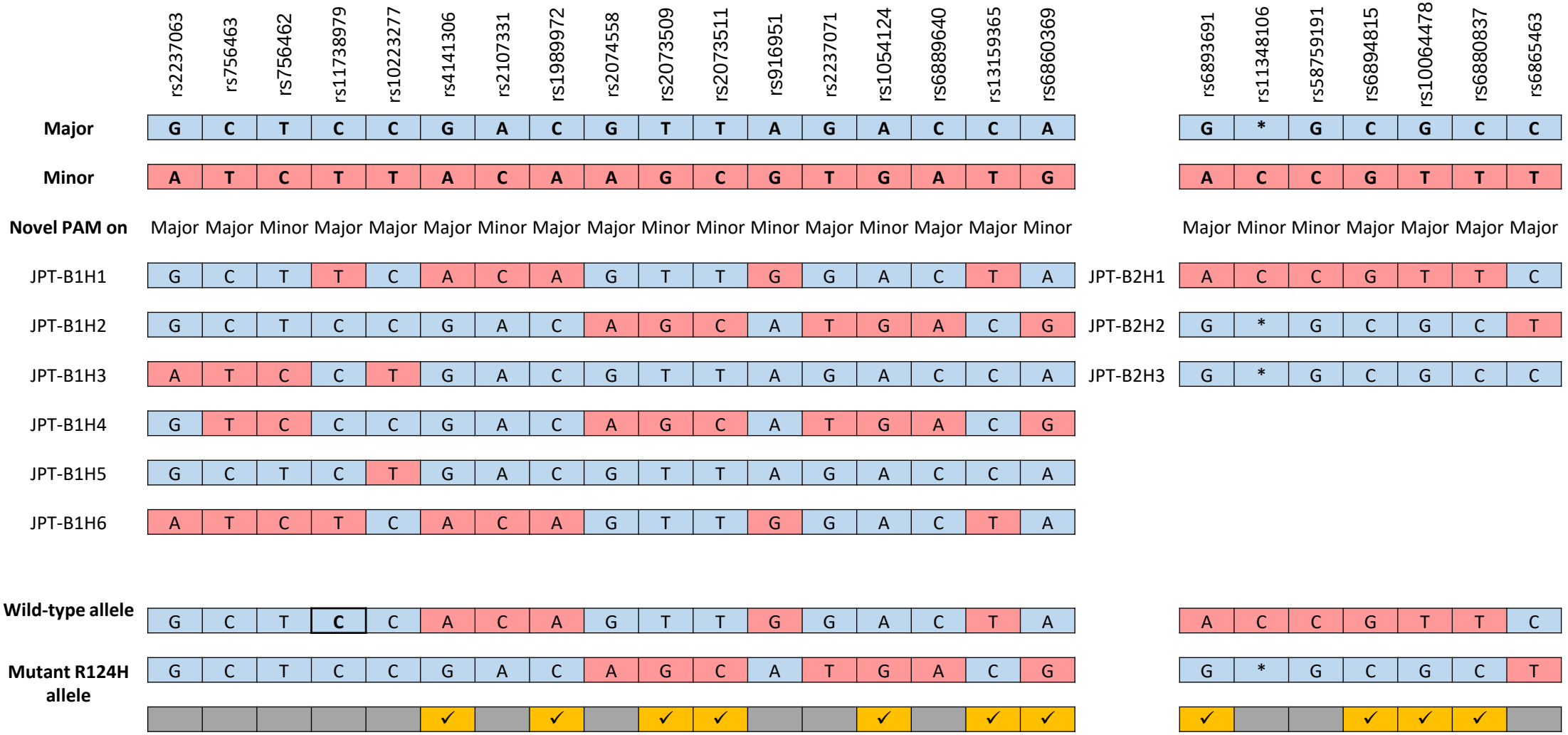


Figure 2

Click here to access/download;Figure;200212 Figures 2.pdf

2a



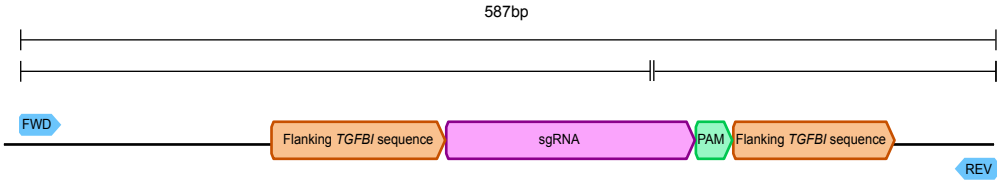
b

Figure 3

Click here to access/download;Figure;200212 Figures 3.pdf

3a

b



c

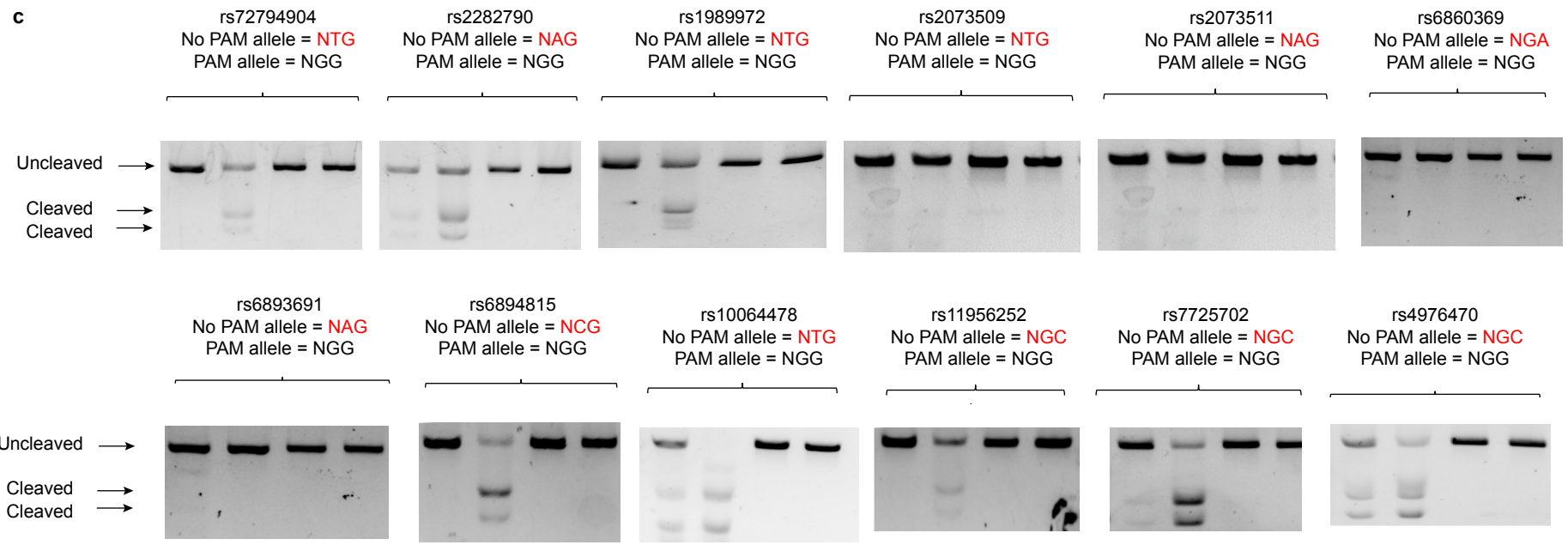
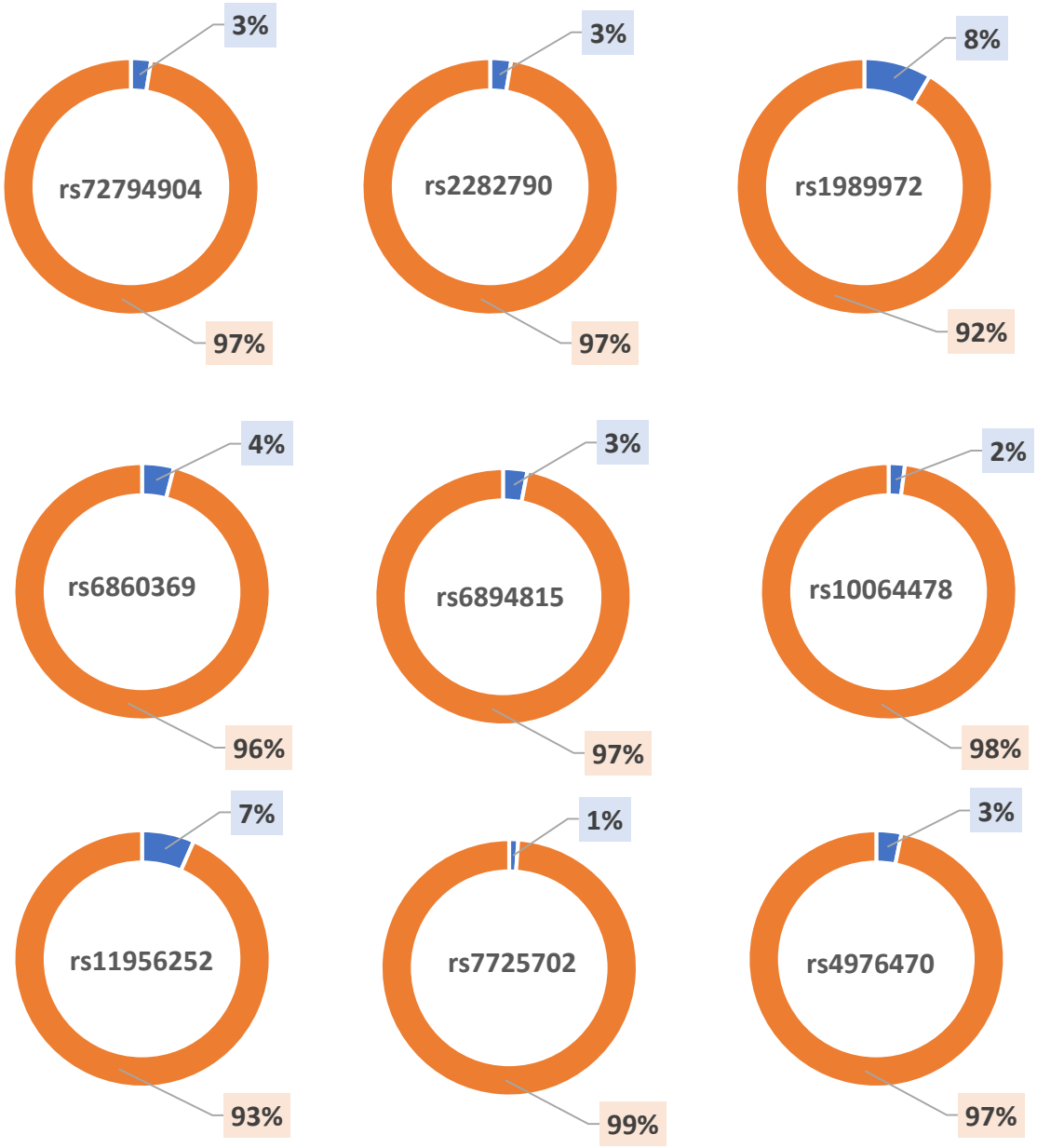


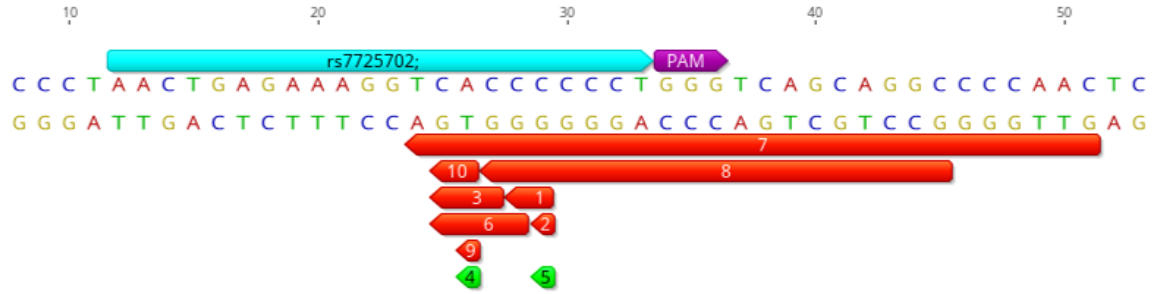
Figure 4

Click here to access/download;Figure;200212 Figures 4.pdf

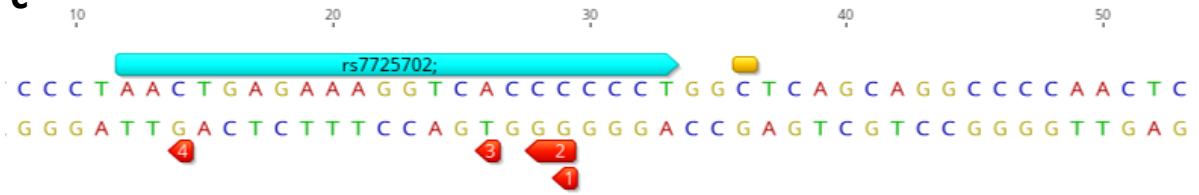
4a



b



c

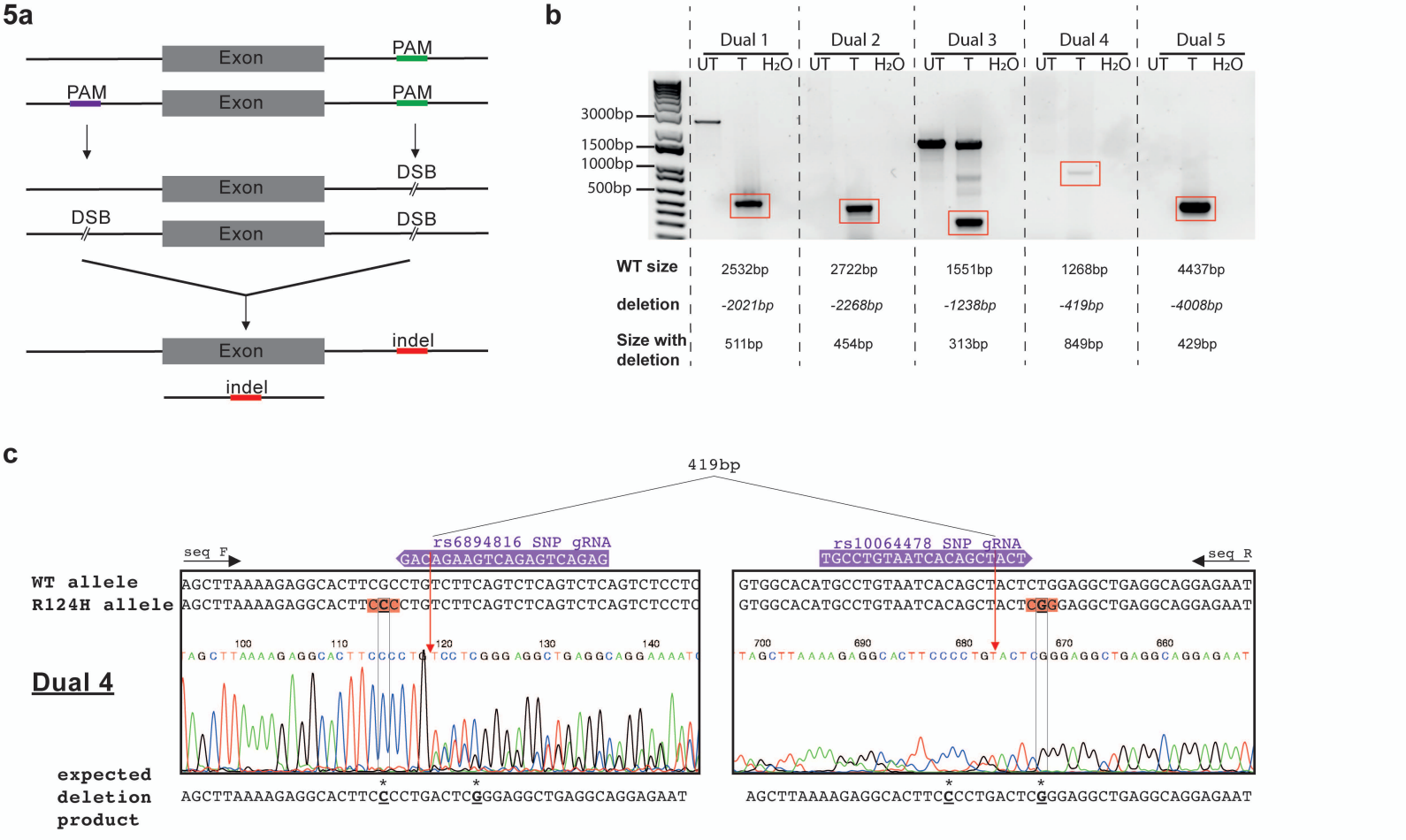


d

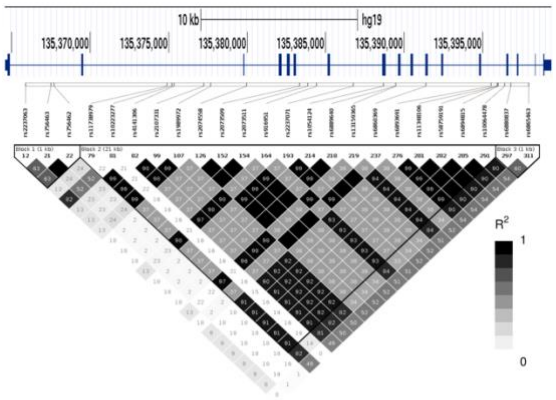
'PAM associated' allele			
Insertion	Deletion	Reads	Efficiency estimate %
-	CC	72573	19.65
-	C	67678	18.32
-	CAC	2219	0.60
A	-	2064	0.56
C	-	637	0.17
-	CACC	494	0.13
-	TCACCCCTGGGTCAGCAGGCCCCAAC	368	0.10
-	CACCCCTGGGTCAGCAGGC	338	0.09
-	A	171	0.05
-	CA	162	0.04
'No PAM' allele			
Insertion	Deletion	Reads	Efficiency estimate %
-	C	664	0.02
-	CC	357	0.02
-	A	95	0.01
-	C	74	0.01

Figure 5

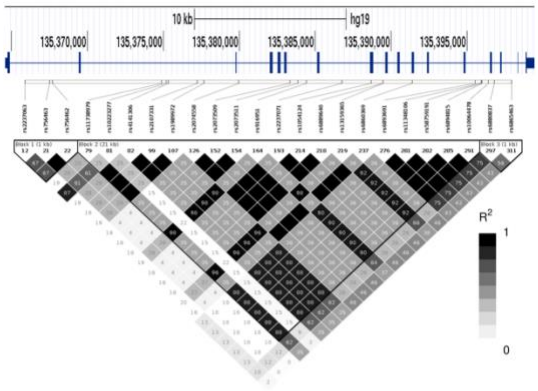
Click here to access/download;Figure;200212 Figures 5.pdf



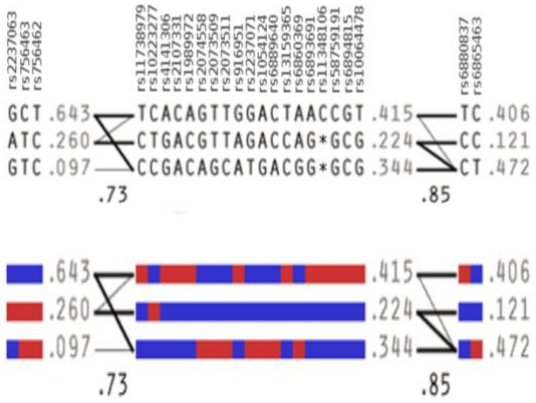
1a


10 kb
hg19
135,370,000
135,375,000
135,380,000
135,385,000
135,390,000
135,395,000
R2
1
0

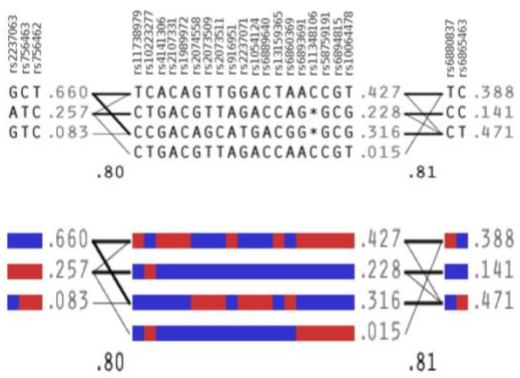
c


10 kb
hg19
135,370,000
135,375,000
135,380,000
135,385,000
135,390,000
135,395,000
R2
1
0

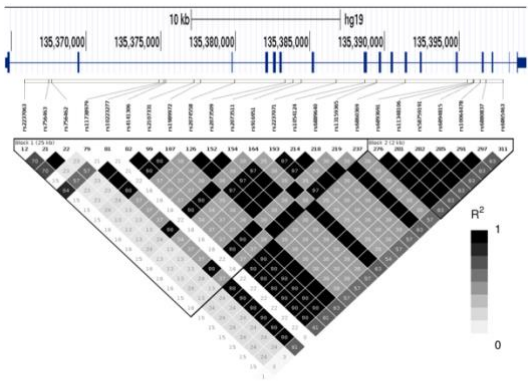
b


rs2237063
rs756463
rs756462
rs11738979
rs10223277
rs4141306
rs2107331
rs1989972
rs2074558
rs2073509
rs2073511
rs5916951
rs2237071
rs1054124
rs6889640
rs6860369
rs6893691
rs11348106
rs58759191
rs6894815
rs10064478
rs6880837
rs6865463
GCT .643
ATC .260
GTC .097
TCACAGTTGGACTAACCGT .415
CTGACGTTAGACCAG*GCG .224
CCGACAGCATGACGG*GCG .344
TC .406
CC .121
CT .472
.73
.85

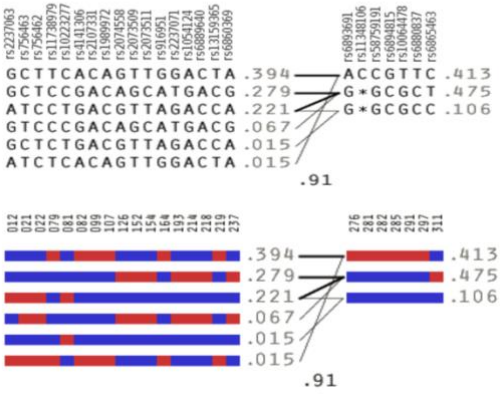
d


rs2237063
rs756463
rs756462
rs11738979
rs10223277
rs4141306
rs2107331
rs1989972
rs2074558
rs2073509
rs2073511
rs1054124
rs223701
rs13159365
rs6860369
rs6889640
rs11348106
rs58759191
rs6894815
rs10064478
rs6880837
rs6865463
GCT .660
ATC .257
GTC .083
TCACAGTTGGACTAACCGT .427
CTGACGTTAGACCAG*GCG .228
CCGACAGCATGACGG*GCG .316
CTGACGTTAGACCAACCGT .015
TC .388
CC .141
CT .471
.80
.81

e


10 kb
hg19
135,370,000
135,375,000
135,380,000
135,385,000
135,390,000
135,395,000
R2
1
0

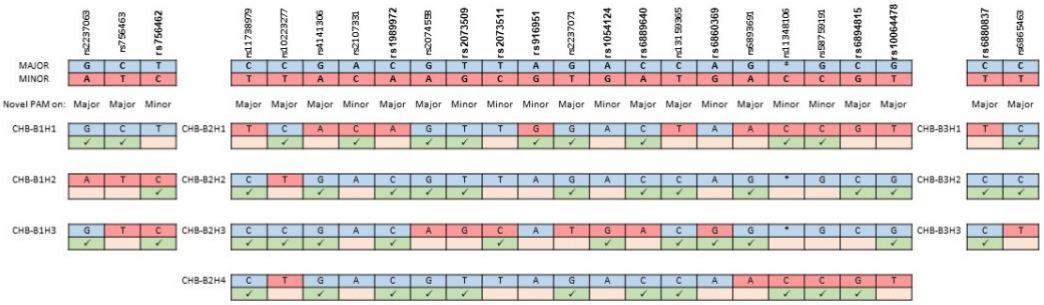
f


rs2237063
rs756463
rs756462
rs11738979
rs10223277
rs4141306
rs2107331
rs1989972
rs2074558
rs2073509
rs2073511
rs5916951
rs2237071
rs1054124
rs6889640
rs13159365
rs6860369
rs6893691
rs11348106
rs58759191
rs6894815
rs10064478
rs6880837
rs6865463
GCTTCACAGTTGGACTA .394
GCTCCGACAGCATGACG .279
ATCCTGACGTTAGACCA .221
GTCCCGACAGCATGACG .067
GCTCTGACGTTAGACCA .015
ATCTCACAGTTGGACTA .015
ACCGTTC .413
G*GCGCT .475
G*GCGCC .106
.91
012
021
022
079
081
082
099
107
126
152
154
164
193
214
218
219
237
276
281
282
285
291
297
311
.394
.279
.221
.067
.015
.015
.413
.475
.106
.91

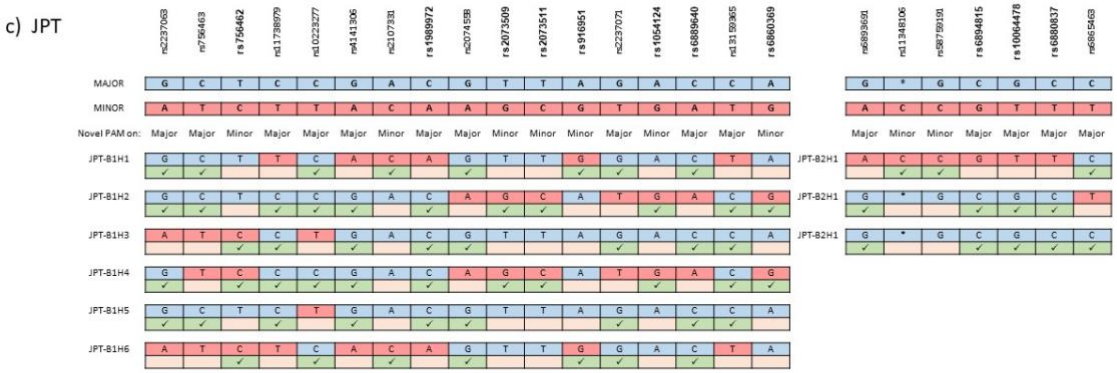
Supplementary Figure 1: Linkage disequilibrium of the identified allele-discriminatory SNPs in the coding region of the TGFBI gene – plots were generated using the 1000 Genomes Project phase 3 data for EAS, CHB and JPT populations, black indicates very strong LD, while a white indicates no LD a) LD plot 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 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.

a) EAS

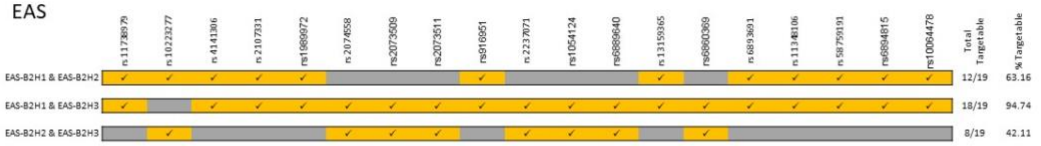
b) CHB



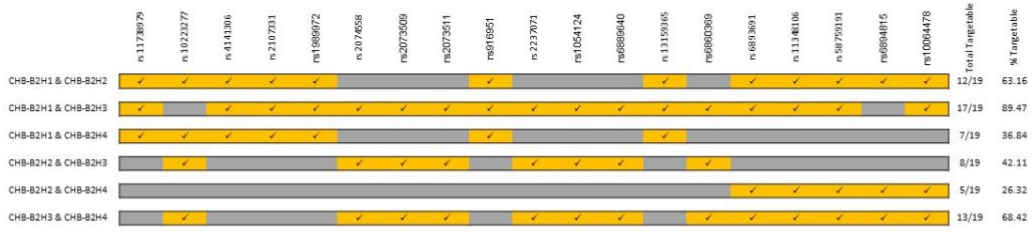
c) JPT



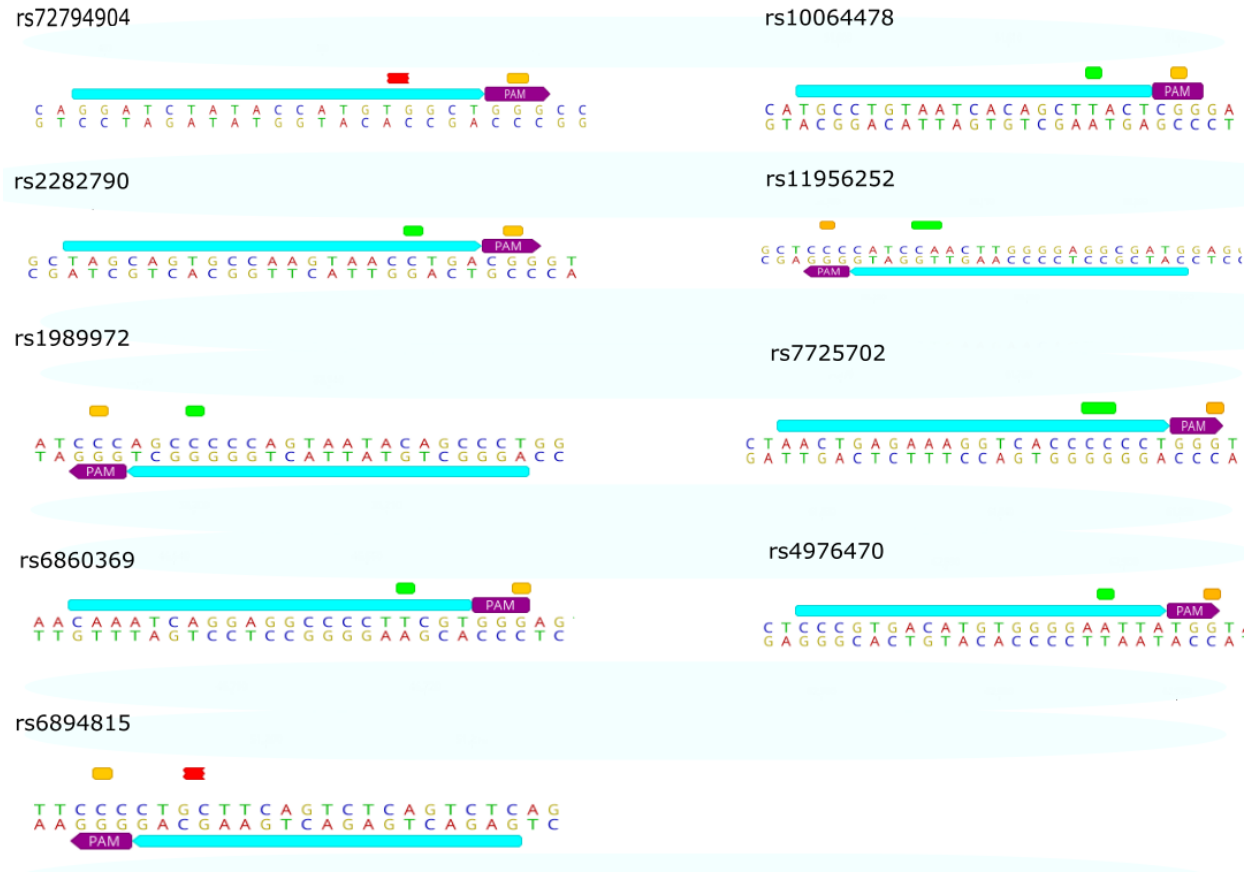
d) EAS



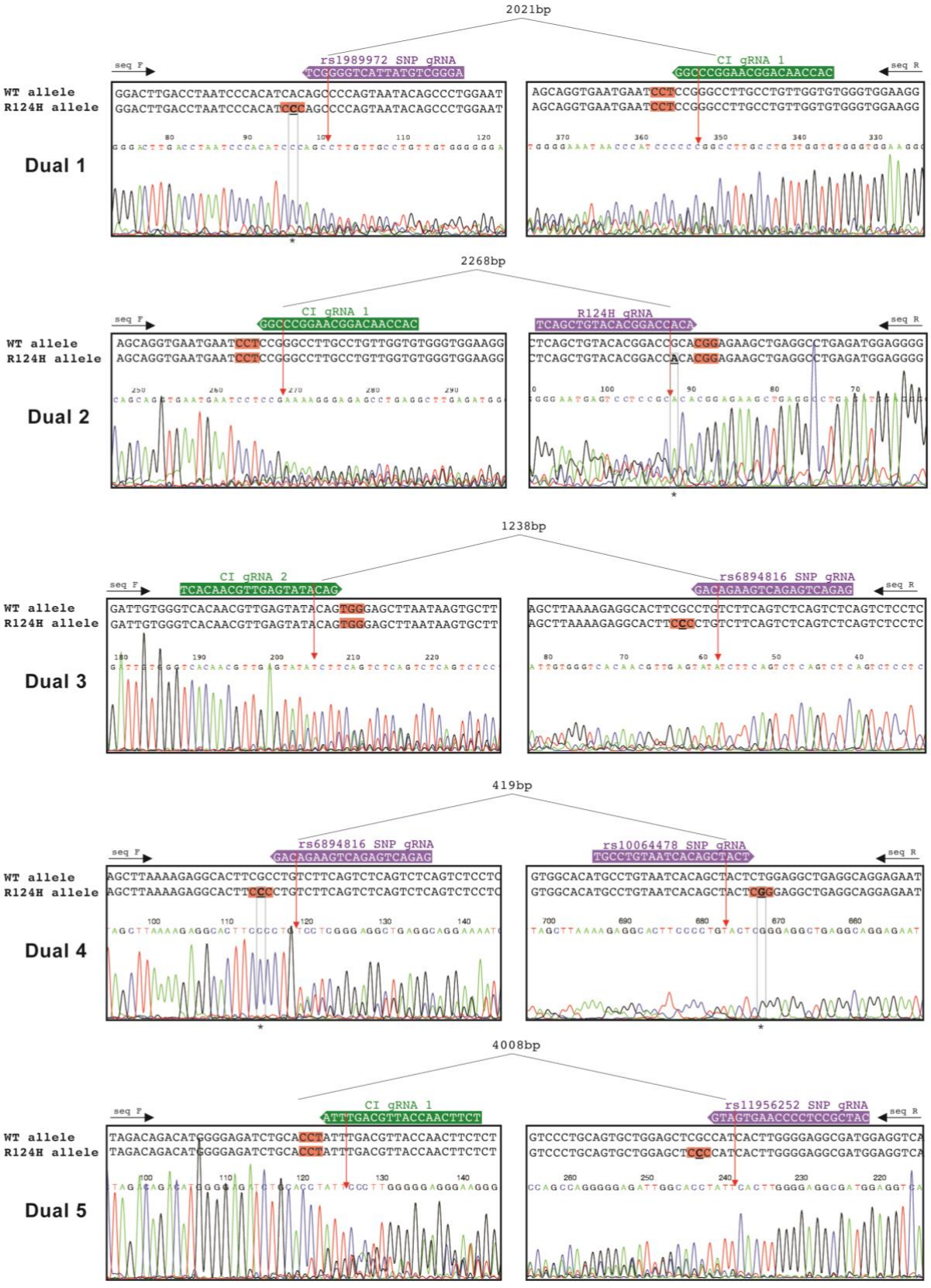
e) CHB



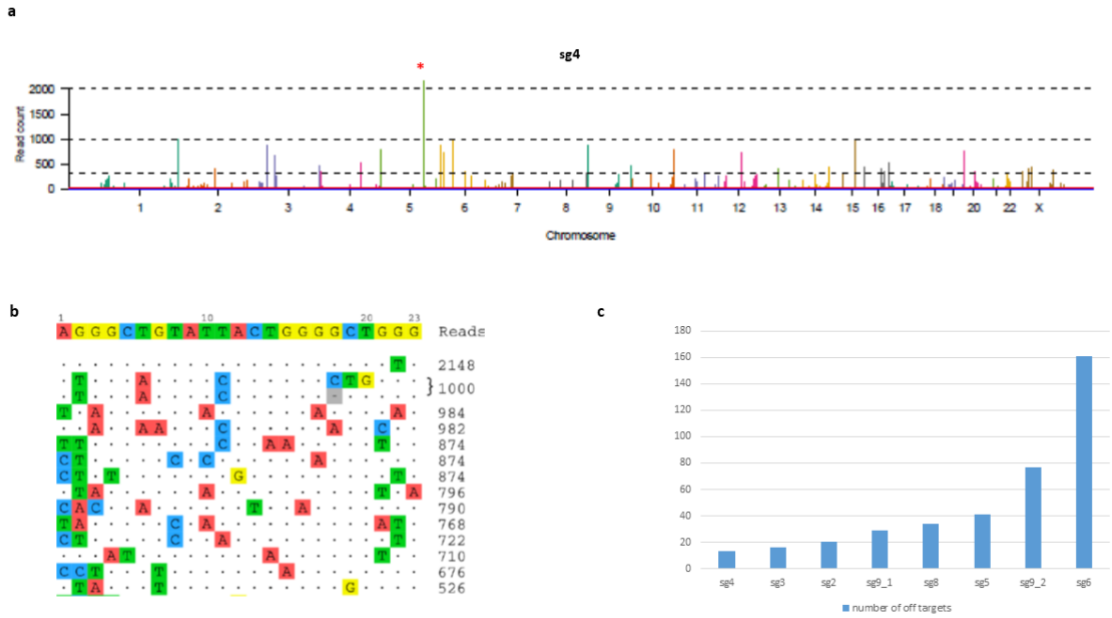
Supplementary Figure 2: Haplotype analysis to determine proportion of population targetable by the identified SNPs. **a,b,c)** Identified SNPs across each haplotype in the a) EAS population b) CHB population and c) JPT population were assessed to determine which have the PAM generating allele and which 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 allele-specific gene-editing. 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 3: Predominant indels of each of the ASNIP guides shown in Figure 4a. PAM is shown in purple, gRNA is shown in blue, SNP associated with a PAM on only one allele is shown in yellow, insertions depicted in green and deletions depicted in red.


2021bp
seq F
rs1989972 SNP qRNA
TCGGGGTCATTATGTCGGGA
CI qRNA 1
GGCCCGGAACGGACAACCAC
seq R
WT allele
R124H allele
Dual 1
2268bp
CI qRNA 1
GGCCCGGAACGGACAACCAC
R124H qRNA
TCAGCTGTACACGGACCACA
Dual 2
1238bp
CI qRNA 2
TCACAACGTTGAGTATACAG
rs6894816 SNP qRNA
GACAGAAGTCAGAGTCAGAG
Dual 3
419bp
rs6894816 SNP qRNA
GACAGAAGTCAGAGTCAGAG
rs10064478 SNP qRNA
TGCCTGTAATCACAGCTACT
Dual 4
4008bp
CI qRNA 1
ATTTGACGTTACCAACTTCT
rs11956252 SNP qRNA
GTAGTGAACCCCTCCGCTAC
Dual 5

Supplementary Figure 4: Sanger sequencing chromatograms showing editing events when cells were transfected with dual pairs. Sanger sequencing was performed on gel-extracted or PCR purified deletion bands PCR-product and sequenced from the forward (F) and reverse (R) primer. The reverse chromatogram is displayed as a reverse complement to aid visualization. The wildtype (WT) and R124H mutation-associated allele sequences are displayed across the top of the trace. gRNA target sequences and direction are indicated above and NGG PAM highlighted in red. The SNP nucleotide is underlined in bold. *denotes where the PAM-associated allele SNP matches the chromatogram. The predicted deletion sequence is shown below.



Supplementary Figure 5: Off-target cleavage by CRISPR/Cas9 single guides investigated using CIRCLE-seq. Off targets were identified for each sample and analysed for genome location, read count (a) and mismatch number to target sequence (b). A threshold of >25% reads of top hit was used to rank guides by off-target cleavage and identify the most specific guides (c)

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
TATATTAGCACAAATAAAGTAAAACAAGTGTTTTTTTTTTAGAAAGTAGTTCTATAAGGA	No
ACGATTTTGAGATTCACCCATGTTATAGCATGTTATGGATAATTTGTTCTTTTTTTTAGCT	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
ACCTAATACATCCACTCTTGTCGTTTTCCAGGAGGGCAGAGGAACCTCGCCTGCTTCACCAT	Yes
CACAGGTCAGGATCTATACCATGTGGGCTGTGCCTGTTTCCTCCAAGAGTTTCCAGATCGT	Yes
TTGAATCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCATGCCACTGCACTTCAGCCTGGG	Yes
GCAGAAATTTGGGCAAGAAATGGAGAAGAAGACACCAGACTGTGGTGGGCATGTGGGAAAC	No
TAAAATTATCTCATATAATCCTCCTAATAGTCCTGTAAGTAGGTATTATTCTTATTCTCAA	No
TAATCTTTTAGAGCATTCCTAGCTGGAAACTTTTCCCGGAAATCCATAAGTTAAGTGCTCC	No
TGATGAGATGTTTTTACCCCAAATCCAGCCGTGTCTTTCCCTGTTGTTTAGGATTATTAGA	No
ACTGATTTAGAGTGATCAACCTATGTCCCCAGAGTAAGTTACAGAAACTAAATTTTTAAAA	Yes
CATGCATGCACTGTTTTTTTTTTCCTACTGTTTTTTTTTTCCTTCTGCTTTCTTCCCTGC	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
GTTTGGGGGGGACTTGACCTAATCCCACATCCCAGCCCCAGTAATACAGCCCTGGAATTTAT	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
GATCTCTGCCTAACTGAGCTCACCTCTCCTCCCTCTCCTGACTGGTTAGATTTTCTAG	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
TTTGAGGCTGAATAATATTCCATTGCATATATATACCACTCACTTTGTTTATTCACTCATC	No
TTAATTTGGGAGGGGGGGGAATTGCTATAGTGTTTCCATTGTGGCTGCACTATGTCACATT	Yes
AGTTGCTCCACATCCTCTCTAATACTTGTTATTTTCTGGTTTTTGGTAATAACCATTCTAT	No
TTTACTCTGTTGACAGTGTCCTTTGATGCATAAACATCTTTAATTTAGATGAAGTTCACAA	No
ACCACACTGTTTTGATTATTGTAGCCTTGGCGTAAGTTTTAAAATTAGTAAGTATGAGTCC	No
GTAAGTATGAGTCCTTCAACTTTGTTCTTCTTTTCAAAATTGTTGTGGCTATTGGAAATT	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
GTCTAAACATTGTTGAGAAGAAGCAGTGGCAAAAAAAAAAAAAAGTTGAAAATACATTAGAG	No
ATTTTAACATCAGGAATGCCTGATTTTTTTTAAAAAAAAAACCTCCCAGCAAACTAGGAAT	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
TCTTGAGGTTCTAAGTACGGTATTTTTCTGAAAAAAAAAAAAAAAATGGTTTTACTGAAAG	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
ATGGGAGTGAATTTCTTTCTTTCTTTCTTTTTTTTTTGAGACAGAGTCTCATTCTGTCA	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 1: Table showing the 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 were 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 2: 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 3: Guide sequences of 12 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 4: Distance in base pairs between dual combinations consisting of guides designed based on the haplotype analysis

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 5: 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 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?
Dual 4 (5 & 6)	rs6894815 & rs10064478	GAGACTGAGACTGAAGACAG	TGCCTGTAATCACAGCTACT	419	Only remove intronic region	N/A
Dual 3 (C2 & 5)	CI-2 & rs6894815	TCACAACGTTGAGTATACAG	GAGACTGAGACTGAAGACAG	1238	125	41.66666667
Dual 1 (3 & C1)	rs1989972 & CI-1	AGGGCTGTATTACTGGGGCT	CACCAACAGGCAAGGCCCGG	2021	65	21.66666667
Dual 2 (C1 & H)	CI-1 & R124H	CACCAACAGGCAAGGCCCGG	TCAGCTGTACACGGACCACA	2268	Cut site in exon, difficult to predict	Unknown
Dual 5 (C4 & 7)	CI-4 & rs11956252	AGAAGTTGGTAACGTCAAAT	CATCGCCTCCCCAAGTGATG	4008	632	210.6666667

Supplementary Table 7: Number of off-targets detected for all guides investigated

Guide ID	Target SNP	Total off targets	Off targets above threshold (>25% of top hit)	Number of top 10 *in silico* targets identified by CIRCLE-seq	Mismatch range in off-targets above threshold	Rank position of on-target in CIRCLE-seq	Read count for top hit	Read count of on-target in CIRCLE-seq	On-target % cleavage *in vitro*	On-target % cleavage by deep seq	On-target % cleavage by TIDE analysis
sg2	rs72794904	155	20	6	3-6	18	1324	390	60	2.15	NA - F
sg3	rs2282790	221	16	8	4-6	1	1892	1892	55	0.51	8
sg4	rs1989972	305	13	7	4-6	1	2148	2148	25	3.68	12
sg5	rs6860369	426	41	10	3-6	8	1814	898	80	2.3	NA - F
sg6	rs6894815	1448	161	8	2-6	10	2492	1556	0	1.75	NA - F
sg8	rs11956252	1554	34	4	2-6	2	796	646	45	4.6	9
sg9_1	rs7725702	782	29	9	4-6	1	2536	2536	75	39.82	27
sg9_2	rs7725702	761	77	10	3-6	20	1318	690	75	39.82	27

Supplementary Table 8: Top off-target sites detected for each guide investigated

guide ID	PCR target ID	Target type	TIDE result	% Efficiency	Validation result
sg2	2.1	TOP HITS	F		FAIL
sg2	2.2	TOP HITS	F		FAIL
sg2	2.3	TOP HITS	Y	8.3	POSITIVE VALIDATION
sg2	2.4	TOP HITS	F		FAIL
sg2	2.5	TOP HITS	F		FAIL
sg2	2.6	TGFBI	F		FAIL
sg2	2.7	EXONIC	F		FAIL
sg3	3.1	TGFBI	Y	7.6	POSITIVE VALIDATION
sg3	3.2	TOP HITS	Y	5.1	POSITIVE VALIDATION
sg3	3.3	TOP HITS	Y	54.7	POSITIVE VALIDATION
sg3	3.4	TOP HITS	Y		FAIL
sg3	3.5	TOP HITS	F		FAIL
sg3	3.6	TOP HITS	Y	2.3	POSITIVE VALIDATION
sg4	4.1	TOP HITS -TGFBI	Y	11.6	POSTIVE VALIDATION
sg4	4.2	TOP HITS	Y	2.4	POSITIVE VALIDATION
sg4	4.3	TOP HITS	F		FAIL
sg4	4.4	TOP HITS	Y	0.5	NEGATIVE VALIDATION
sg4	4.5	TOP HITS	Y		FAIL
sg4	4.6	TOP HITS	F		FAIL
sg4	4.7	TOP HITS	F		FAIL
sg4	4.8	BENCHLING	F		FAIL
sg4	4.9	BENCHLING	F		FAIL
sg5	5.1	TOP HITS	Y	16.1	POSITIVE VALIDATION
sg5	5.2	TOP HITS	F		FAIL
sg5	5.3	TOP HITS	Y	0	NEGATIVE VALIDATION
sg5	5.4	TOP HITS	F		FAIL
sg5	5.5	TOP HITS	F		FAIL
sg5	5.6	TGFBI	F		FAIL
sg5	5.7	EXONIC	F		FAIL
sg5	5.8	EXONIC	Y	1.2	NEGATIVE VALIDATION
sg5	5.9	EXONIC	F		FAIL
sg6	6.1	TOP HITS	Y		FAIL
sg6	6.2	TOP HITS	Y	3	POSITIVE VALIDATION
sg6	6.3	TOP HITS	Y		FAIL
sg6	6.4	TOP HITS	Y	3.1	POSITIVE VALIDATION
sg6	6.5	TOP HITS	Y	4.2	POSITIVE VALIDATION
sg6	6.6	TGFBI	F		FAIL
sg6	6.7	BENCHLING	Y	16.7	POSITIVE VALIDATION
sg6	6.8	BENCHLING	Y	1.7	POSITIVE VALIDATION
sg6	6.9	BENCHLING	F		FAIL
sg6	6.10	BENCHLING	Y		FAIL
sg6	6.11	BENCHLING	F		FAIL
sg6	6.12	BENCHLING	Y	0.9	NEGATIVE VALIDATION
sg6	6.13	EXONIC	F		FAIL
sg6	6.14	EXONIC	F		FAIL
sg6	6.15	EXONIC	Y		FAIL
sg6	6.16	EXONIC	Y	10.7	POSITIVE VALIDATION
sg6	6.17	EXONIC	Y	4.6	POSITIVE VALIDATION
sg8	8.1	TOP HITS	F		FAIL
sg8	8.2	TOP HITS TGFBI	Y	9.3	POSITIVE VALIDATION
sg8	8.3	TOP HITS	F		FAIL
sg8	8.4	TOP HITS	Y	7.9	POSITIVE VALIDATION
sg8	8.5	TOP HITS	F		FAIL
sg8	8.6	EXONIC	F		FAIL
sg8	8.7	TOP HITS	F		FAIL
sg9	9.1	TGFBI	Y	26.7	POSITIVE VALIDATION
sg9	9.2	TOP HITS	Y	0.80%	NEGATIVE VALIDATION
sg9	9.3	TOP HITS	F		FAIL
sg9	9.4	TOP HITS	F		FAIL
sg9	9.5	TOP HITS	Y	0.40%	NEGATIVE VALIDATION
sg9	9.6	TOP HITS	F	F	FAIL
sg9	9.7	TOP HITS	F	F	FAIL

Supplementary Table 9a: Validation of detected off-target sites detected

A: results for all samples that passed through TIDE analysis					
TIDE results					
GUIDE id	Target/ Sample ID	TARGET SOURCE	gene region	CONCLUSION	% EFFICIENCY
sg2	2.3	TOP HITS	ncRNA_intronic	POSITIVE VALIDATION	8.3
sg3	3.1	TGFBI	upstream	POSITIVE VALIDATION	7.6
sg3	3.2	TOP HITS	intergenic	POSITIVE VALIDATION	5.1
sg3	3.3	TOP HITS	intergenic	POSITIVE VALIDATION	54.7
sg3	3.6	TOP HITS	intronic	POSITIVE VALIDATION	2.3
sg4	4.1	TOP HITS -TGFBI	intronic	POSITIVE VALIDATION	11.6
sg4	4.2	TOP HITS	intergenic	POSITIVE VALIDATION	2.4
sg4	4.4	TOP HITS	ncRNA_intronic	NEGATIVE VALIDATION	0.5
sg5	5.1	TOP HITS	intronic	POSITIVE VALIDATION	16.1
sg5	5.3	TOP HITS	intronic	NEGATIVE VALIDATION	0
sg5	5.8	EXONIC	exonic	NEGATIVE VALIDATION	1.2
sg6	6.2	TOP HITS	upstream_downstream	POSITIVE VALIDATION	3
sg6	6.4	TOP HITS	intronic	POSITIVE VALIDATION	3.1
sg6	6.5	TOP HITS	intronic	POSITIVE VALIDATION	4.2
sg6	6.7	BENCHLING	intergenic	POSITIVE VALIDATION	16.7
sg6	6.8	BENCHLING	intronic	POSITIVE VALIDATION	1.7
sg6	6.12	BENCHLING	intronic	NEGATIVE VALIDATION	0.9
sg6	6.16	EXONIC	exonic	POSITIVE VALIDATION	10.7
sg6	6.17	EXONIC	exonic	POSITIVE VALIDATION	4.6
sg8	8.2	TOP HITS TGFBI	intergenic	POSITIVE VALIDATION	9.3
sg8	8.4	TOP HITS	intergenic	POSITIVE VALIDATION	7.9
sg9	9.1	TGFBI	intergenic	POSITIVE VALIDATION	26.7
sg9	9.2	TOP HITS	intronic	NEGATIVE VALIDATION	0.80%
sg9	9.5	TOP HITS	intronic	NEGATIVE VALIDATION	0.40%

Supplementary Table 9b:

B: Summary of sample drop off through CIRCLE-seq validation steps by TIDE analysis.		
Summary table:		
	TOTAL TARGETS SELECTED	62
	Passed primer design	50
	produced PCR product for treated and control	46
	produced sanger sequencing result for treated and control	26
	TIDE fail	2
	TOTAL TIDE RESULTS	24
	positive off-target validations	18
	negative off-target validations	6

Supplementary Table 10: 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

eTOC

The well documented propensity of CRISPR/Cas9 to cleave at unintended off-target sites has impeded the progression of this promising tool to the clinic. In the case of autosomal dominant disease, the most perilous off-target site exists as the wild-type allele, which differs to the mutant allele by only a single base pair. This manuscript presents an innovative approach to selectively cleaving the mutant allele in a mutation-independent manner. Using corneal dystrophy as a model we show successful allele-specific editing of TGFBI and importantly offer a potential targeting strategy for all autosomal dominant disease, in which selective disruption of the mutant allele offers a viable treatment approach.

Graphical Abstract

Click here to access/download;Graphical Abstract;Graphical abstract-01.png