Fakultät für Medizin

Gene Editing in the STAT3 Hyper-IgE Syndrome

Andreas Christian Eberherr

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Carsten Schmidt-Weber

Prüfer*innen der Dissertation:
1. Prof. Dr. Ellen Renner
2. Prof. Dr. Benjamin Schusser
3. Prof. Dr. Eckhard Wolf

Die Dissertation wurde am 28.07.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 21.02.2023 angenommen.

Table of content

Abbreviations

A	**Adenine**
bp	Base pair
BER	Base excision repair
C	Cytosine
Cas9	CRISPR-associated protein 9
CCL2	Chemokine (C-C Motif) Ligand 2
CMV	Cytomegalovirus
CRISPR	Clustered regularly interspaced short palindromic repeat
DMSO	Dimethyl sulfoxide
DNase	Deoxyribonuclease
DSB	Double strand break
G	Guanine
GFP	Green fluorescent protein
HDR	Homology directed repair
HIES	Hyper-IgE syndrome
HTS	High-throughput sequencing
I	Inosine
InDel	Insertion/Deletion
MMR	Mismatch repair
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
qRT-PCR	Reverse transcription quantitative real-time PCR
R382Q	Human *STAT3* c.1145G>A/p.R382Q mutation
R382W	Human *STAT3* c.1144C>T/p.R382W mutation
RIG-I	Retinoic acid inducible gene 1
RNase	Ribonuclease
RNP	Ribonucleoprotein
SCC	Single cell clone
sgRNA	Single guideRNA

SOCS3	Suppressor of cytokine signaling 3
STAT3	Signal transducer and activator of transcription 3
STAT3-HIES	STAT3 hyper-IgE syndrome
T	Thymine
TBP	TATA-binding protein
WGS	Whole genome sequencing

Summary

Recent developments of gene editing techniques based on CRISPR/Cas9 allow for the programmable editing of DNA with unprecedented flexibility, specificity and efficiency. A CRISPR/Cas9 based gene editing technique called adenine base editing enables the direct conversion of adenine (A) to inosine (I) which is read and replicated as guanine (G) by polymerases. Thus, single base pairs can be converted from A·T to I/G·C enabling the correction of G·C to A·T mutations underlying homozygous and heterozygous monogenic diseases. In this work, we applied classical CRISPR/Cas9 editing and adenine base editing to heterozygous human *STAT3* mutations known to cause Signal Transducer and Activator of Transcription 3 Hyper-IgE Syndrome (STAT3-HIES).

STAT3-HIES is an inborn error of immunity caused by heterozygous mutations in the *STAT3* gene resulting in a dominant-negative effect on STAT3 signaling. Patients present with recurrent infections of skin and lung, eczema, skeletal manifestations, eosinophilia, reduced Th17 cell numbers and high levels of serum IgE. Especially the recurrent lung infections which lead to the formation of pneumatoceles and reduced lung function pose a threat to the patients. To date, there is no curative therapy for STAT3-HIES available. Therefore, treatment is mainly based on relief or prevention of symptoms. Hematopoietic stem cell transplantation (HSCT) is of limited benefit concerning the extra-hematopoietic symptoms and already existing lung damage.

We hypothesized that the targeted genetic editing of the underlying mutations in patients by adenine base editors (ABEs) might provide a causative treatment approach for STAT3-HIES. To confirm this hypothesis we developed gene editing systems for two of the most prevalent STAT3-HIES causing heterozygous *STAT3* mutations c.1144C>T;p.R382W (R382W) and c.1145G>A;p.R382Q (R382Q).

We show that ABE7.10 in combination with our guideRNA WRA-2 is able to correct the R382W mutation with high efficiency (22 - 44%) and specificity in primary patient fibroblasts ($STAT3^{+/R382W}$) as confirmed by Sanger sequencing, high-throughput sequencing (HTS), whole genome sequencing (WGS) and *in silico* analyses. We did not observe any off-target effects when we analyzed genomic integrity via Sanger sequencing, HTS and WGS. Analysis of STAT3 DNA binding activity via DNA binding ELISA (STAT3 TransAM) after interleukin 6 (IL-6) stimulation showed an increase in bulk treated primary patient fibroblasts and a significant increase in repaired single cell

clones (SCCs) when compared to untreated cells. We observed the same pattern when we analyzed STAT3 target gene expression of *Suppressor of Cytokine Signaling 3* (*SOCS3*) and *Chemokine C-C Motif Ligand 2* (*CCL2*) via reverse transcription quantitative real-time PCR (qRT-PCR). Further, *SOCS3* and *CCL2* expression of bulk treated patient cells and repaired SCCs was comparable to the levels we observed in primary fibroblasts of healthy donors.

For the R382Q mutation we developed an editing system for an allele specific knock-out to reduce the dominant-negative effect of the mutation via *S. pyogenes* Cas9 and our guideRNA QR-1. We observed a high knock-out efficiency of the affected allele (>70%) in bulk treated patient cells with only minimal editing of the wild type allele (<4%) despite the low difference of only one base pair (bp) between alleles. Another treatment approach, a combination of ABE7.10 and our guideRNA QRA-3, leads to a robust correction of the R382Q mutation (~20%) in bulk treated cells without undesired editing effects at the target region as determined via Sanger sequencing and *in silico* analysis.

Our results show that adenine base editing is able to correct heterozygous STAT3-HIES causing mutations in primary patient fibroblasts with high efficiency and specificity. An unimpaired genomic integrity and significant improvements of STAT3 signaling indicate a possible therapeutic value of the treatment approach.

Therefore, our proof-of-concept for a causative therapy approach of STAT3-HIES is an important first step for the treatment of STAT3-HIES and emphasizes the potential of adenine base editing in the treatment of monogenic diseases.

Zusammenfassung

Neue entwickelte CRISPR/Cas9 basierte Geneditierungsmethoden erlauben eine programmierbare Editierung von DNA mit bisher unerreichter Flexibilität, Spezifität und Effizienz. Eine dieser CRISPR/Cas9 basierten Geneditierungsmethoden, die Adenin-Baseneditierung, ermöglicht die direkte Umwandlung von Adenin (A) zu Inosin (I), das von Polymerasen als Guanin (G) gelesen und repliziert wird. Somit können einzelne Basenpaare von A·T zu I/G·C konvertiert werden, was die Korrektur von G·C zu A·T Mutationen ermöglicht, die homozygoten und heterozygoten monogenen Krankheiten zu Grunde liegen. In dieser Arbeit haben wir klassische CRISPR/Cas9 Editierung und Adenin-Baseneditierung im Kontext heterozygoter humaner *STAT3*-Mutationen, die dafür bekannt sind, dass sie zum Signal Transducer and Activator of Transcription 3 Hyper-IgE Syndrom (STAT3-HIES) führen, verwendet.

STAT3-HIES ist eine angeborene Immunerkrankung und wird durch heterozygote Mutationen in dem *STAT3*-Gen verursacht, die zu einem dominant-negativen Effekt auf den STAT3-Signalweg führen. Patienten zeigen wiederkehrende Infektionen von Haut und Lunge, Ekzeme, Skelettauffälligkeiten, Eosinophilie, reduzierte Th17-Zellzahlen und hohe Mengen von IgE im Serum. Besonders die wiederkehrenden Lungeninfektionen, die zur Bildung von Pneumatozelen und reduzierter Lungenfunktion führen, sind eine Gefahr für die Patienten. Bisher existiert keine kurative Therapie für STAT3-HIES, daher beschränkt sich die Behandlung hauptsächlich auf die Linderung oder Verhinderung von Symptomen. Hematopoetische Stammzelltransplantationen (HSCT) helfen wenig in Bezug auf extra-hematopoetische Symptome und bereits bestehende Lungenschäden.

Wir stellten die Hypothese auf, dass die gezielte genetische Editierung der zu Grunde liegenden Mutationen in Patienten durch Adenin-Baseneditoren (ABEs) einen ursächlichen Behandlungsansatz für STAT3-HIES darstellen könnte. Um diese Hypothese zu bestätigen entwickelten wir Geneditierungssysteme für die zwei bei unseren Patienten häufigsten STAT3-HIES verursachenden heterozygoten *STAT3*-Mutationen c.1144C>T;p.R382W (R382W) und c.1145G>A;p.R382Q (R382Q).

Wir zeigen, dass ABE7.10 in Kombination mit unserer guideRNA WRA-2 in der Lage ist die R382W-Mutation mit hoher Effizienz (22-44%) und Spezifität in primären Patientenfibroblasten ($STAT3^{+/R382W}$) zu korrigieren. Bestätigt wurde dies per Sanger-Sequenzierung, Hochdurchsatzsequenzierung (HTS), Gesamtgenomsequenzierung

(WGS) und *in silico* Analysen. Sanger-Sequenzierung, HTS und WGS zeigten keine Nebeneffekte auf die genomische Integrität. Die Analyse der STAT3-Bindeaktivität per DNA-binde ELISA (STAT3 TransAM) nach Interleukin 6 (IL-6) Stimulation zeigte einen Anstieg bei zum Großteil behandelten primären Patientenfibroblasten und einen signifikanten Anstieg bei reparierten Einzelzellklonen (SCCs) im Vergleich zu unbehandelten Zellen. Dasselbe Muster zeigte die Analyse der STAT3-Zielgenexpression der Gene *Suppressor of Cytokine Signaling 3* (*SOCS3*) und *Chemokine C-C Motif Ligand 2* (*CCL2*) mittels reverser Transkription und quantitativer Echtzeit-PCR (qRT-PCR). Weiterhin war die *SOCS3*- und *CCL2*-Expression in zum Großteil behandelten Patientenzellen und reparierten SCCs mit der von primären Fibroblasten gesunder Spender vergleichbar.

Für die R382Q-Mutation entwickelten wir ein Editierungssystem für einen allel-spezifischen Knock-out, um den dominant-negativen Effekt der Mutation durch *S. pyogenes* Cas9 und unsere guideRNA QR-1 zu reduzieren. Wir beobachteten eine hohe Knock-out Effizienz des betroffenen Allels (>70%) in zum Großteil behandelten Patientenzellen gleichzeitig zu einer nur minimalen Editierung des Wildtyp-Allels (<4%), trotz des nur kleinen Unterschiedes von einem Basenpaar (bp) zwischen den Allelen. Ein anderer Behandlungsansatz, eine Kombination von ABE7.10 und unserer guideRNA QRA-3, führte zu einer robusten Korrektur der R382Q-Mutation (~20%) in zum Großteil behandelten Zellen. Sanger-Sequenzierung und *in silico* Analyse zeigten keine unerwünschten Editierungseffekte an der Zielregion.

Unsere Ergebnisse legen nahe, dass Adenin-Baseneditierung verwendet werden kann, um heterozygote STAT3-HIES verursachende Mutationen in primären Patientenfibroblasten mit hoher Effizienz und Spezifität zu korrigieren. Eine unkompromittierte genomische Integrität und signifikante Verbesserungen des STAT3-Signalweges deuten auf einen möglichen therapeutischen Wert des Behandlungsansatzes hin.

Unser damit erbrachter Machbarkeitsnachweis für einen kausalen STAT3-HIES Therapieansatz ist ein wichtiger erster Schritt für die STAT3-HIES Behandlung und hebt das Potential von Adenin-Baseneditierung zur Behandlung von monogenischen Krankheiten hervor.

1. Introduction

1.1 Signal Transducer and Activator of Transcription 3 Hyper-IgE Syndrome

1.1.1 STAT3-HIES history, diagnosis, symptoms and treatment

The Signal Transducer and Activator of Transcription 3 Hyper-IgE Syndrome (STAT3-HIES) or autosomal dominant Hyper-IgE Syndrome (AD-HIES) was already described as early as 1966, although using a different name [1]. Later the disease was correlated with an elevated serum IgE level giving the disease the name Hyper-IgE Syndrome (HIES) [2]. It took over 30 years until a molecular cause for HIES was identified, heterozygous dominant-negative mutations in the *STAT3* gene [3-5]. Since then many *STAT3* mutations have been identified to cause STAT3-HIES[6]. Additionally, *STAT3* gain of function mutations have been identified causing a completely different clinical picture with early-onset polyautoimmunity, lymphoproliferation, and growth failure [7].
Another form of HIES is autosomal recessive HIES (AR-HIES), now called DOCK8 deficiency, which is caused by autosomal recessive mutations in the gene *Dedicator Of Cytokinesis 8* (*DOCK8*). Further, there are other inborn errors such as in *Phosphoglucomutase 3* (*PGM3*), *Serine Peptidase Inhibitor Kazal-type 5* (*SPINK5*) and *Tyrosine Kinase 2* (*TYK2*), Wiskott-Aldrich syndrome (WAS) and Omenn Syndrome (OS) that are accompanied with eczema and immunodeficiency [8-13]. All of which are stated as rare diseases in Orphanet, a portal founded and coordinated by the French Institute of Health and Medical Research (INSERM) [14].

Diagnosis of STAT3-HIES is based on characteristic symptoms such as elevated serum IgE levels, eczema and recurrent skin and lung infections and a genetically confirmed pathogenic heterozygous dominant-negative variant of the *STAT3* gene [15]. Symptoms include but are not limited to characteristic facial features, hyperextensibility of joints, retained primary teeth, minimal trauma bone fractures and scoliosis. Immunologically STAT3-HIES patients show a high serum IgE level, eosinophilia and reduced Th17 cell numbers [16]. Recurrent skin abscesses and lung

infections are mainly mediated by *Staphylococcus aureus* although other bacteria such as *Streptococcus pneumoniae*, *Haemophilus influenzae* or fungi such as *Aspergillus fumigatus* and *Candia albicans* have been described [17]. Healing of pneumonias is often impaired resulting in pneumatoceles and bronchiectasis. The resulting lung disease is responsible for a high morbidity and mortality in patients with STAT3-HIES [18, 19].

Treatment of STAT3-HIES consists mainly of preventive and symptomatic measures such as antibiotic and antimycotic therapy, immunoglobulin replacement therapy, inhalation of sodium chloride and additives or surgical intervention [19, 20]. Results of hematopoietic stem cell transplantation (HSCT) are promising as it is able to resolve the immunological aspects of STAT3-HIES although it is of limited benefit concerning the extra-hematopoietic symptoms and already existing lung damage [19, 20]. Further HSCT requires a matching donor, conditioning and confers a risk of graft-versus-host disease [20].

1.1.2 The *STAT3* gene, STAT3 signaling and *STAT3* mutations

The human signal transducer and activator of transcription (STAT) family comprises seven *STAT* genes, *STAT1*, *STAT2*, *STAT3*, *STAT4*, *STAT5A*, *STAT5B* and *STAT6*. As shown in Figure 1 STAT proteins share a certain structure consisting of a 4-helix bundle domain, a DNA binding domain, a Src homology domain 2 (SH2) domain and a Tyrosine phosphorylation site [21].

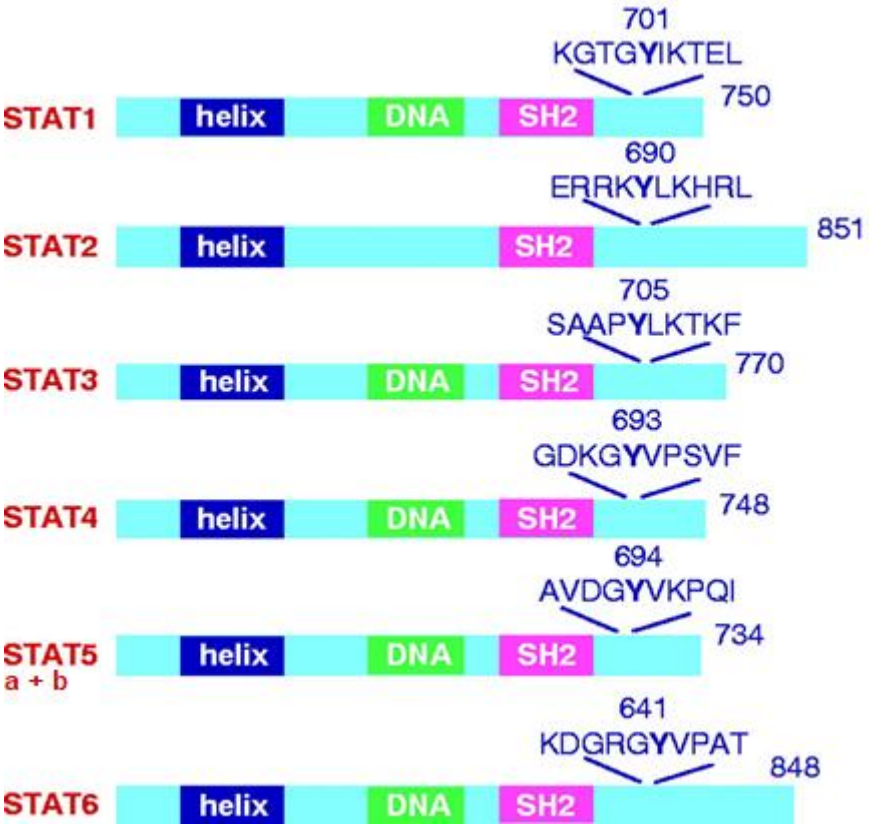

Figure 1: Schematic illustration of STAT family members

Helix: 4-helix bundle; DNA: DNA binding domain; SH2: Src homology domain 2
*adapted from Akira et al. [21]

The human *STAT3* gene is located on chromosome 17 on the reverse strand. The canonical *STAT3* transcript consists of 24 exons and 4921 nucleotides which result in a 770 amino acid protein (transcript STAT3-201 in Ensembl, flagged as MANE Select v0.93) (Figure 2) [22]. Four isoforms are known for the STAT3 protein namely STAT3α, STAT3β, STAT3γ and STAT3δ. While STAT3α and STAT3 β are generated via alternative splicing, STAT3γ and STAT3δ derive from proteolytic processing. Information on STAT3γ and STAT3δ other than a supposed role as negative regulator of STAT3 is lacking. STAT3α and STAT3β have been shown to have distinct properties and cellular roles [23]. Both isoforms are co-expressed in all cell types and levels of STAT3α are usually higher than levels of STAT3β [24]. STAT3 proteins have six domains (Figure 2). The aminoterminal domain mediates multiple functions such as cooperative DNA binding (as tetramer), nuclear translocation and protein-protein interactions [25]. The coiled-coil domain binds to importin-α3 and – α6 and thus mediates the translocation into the nucleus [26]. The DNA-binding domain binds to the STAT3 target sequence “TTCCSGGAA” (S = C or G) in the enhancer sequences of

STAT3 target genes [26]. The SH2 domain mediates the dimerization of two phosphorylated STAT3s via binding of the dimerization partner p-Tyr705 in a specific cavity in the SH2 domain [25]. Further the SH2 domain is also necessary for STAT3 to recognize activated docking sites of membrane receptors such as gp130 [26]. The transactivation domain harbors the Tyr705 phosphorylation site which is recognized by SH2 domains and a second phosphorylation site (Ser727) in the case of STAT3α which modulates STAT3 transcriptional activity [27].

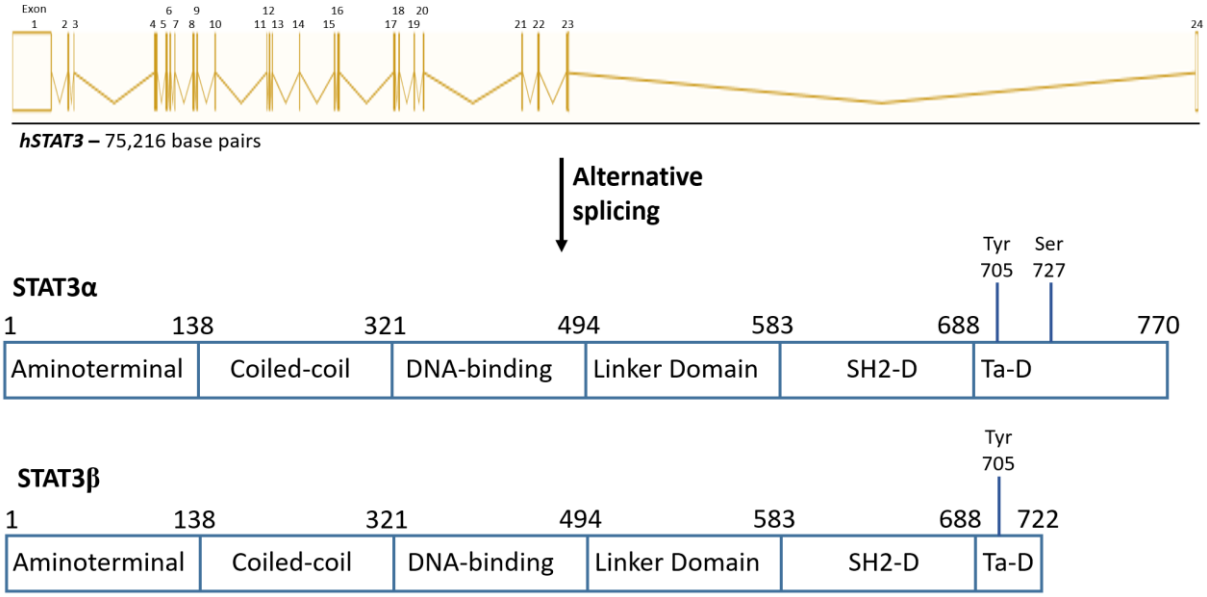

Figure 2: Human *STAT3* genomic structure and STAT3 domains and isoforms

Human *STAT3* consists of 24 exons and is processed to two separate transcripts via alternative splicing resulting in the two STAT3 isoforms STAT3α and STAT3β.
SH2-D = Src Homolgy 2 domain, Ta-D = Transactivation domain
*Adapted from Sgrignani *et al.* [25]

The canonical STAT3 signaling pathway (Figure 3A) consists of cytokine binding to their respective receptors (e.g. IL-6 and the IL-6 receptor/two glycoprotein 130 (gp130) monomers). Recruited Janus kinases (Jaks; Jak1, Jak2 and Tyk2) phosphorylate STAT3 docking regions on gp130 and enable binding of unphosphorylated STAT3 proteins. STAT3 proteins are phosphorylated (Tyr 705) via Jaks and are now able to dimerize via SH2 domain and Phospho-Tyr705 binding. The STAT3 dimers then translocate to the nucleus via importin-α3 and – α6 binding to the coiled-coil domain. In the nucleus STAT3 dimers bind to the enhancer regions of target genes via the DNA-binding domain and regulate target gene expression [28-31]. In case of heterozygous dominant-negative *STAT3* mutations such as R382W and R382Q in the DNA-binding

domain of STAT3 the dimers including mutated STAT3 are unable to bind to DNA leading to a dominant-negative effect on STAT3 signaling (Figure 3B) [3, 4].

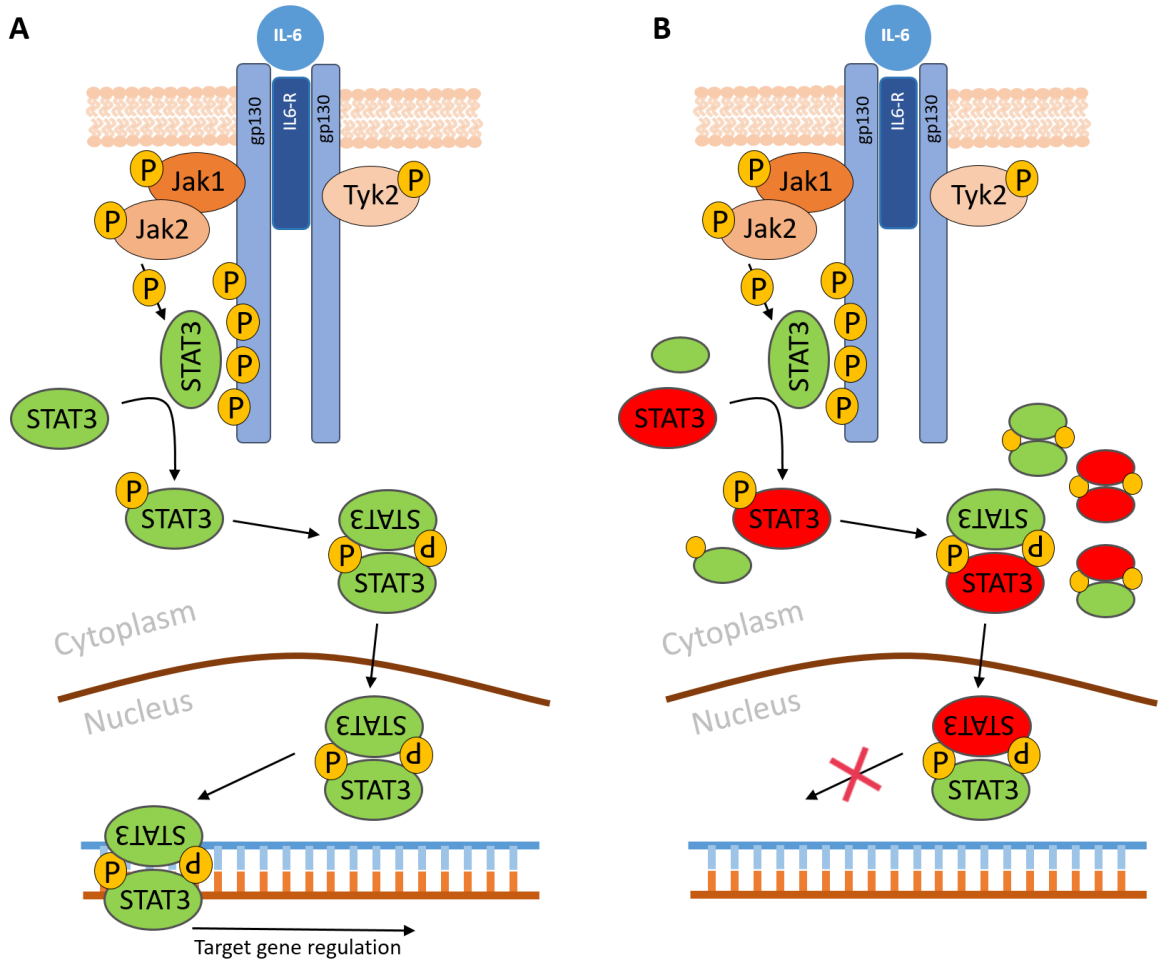

Figure 3: Canonical STAT3 signaling

A: IL-6 or similar ligands are bound by IL6-R and induces binding of two GP130 monomers. Recruited Janus kinases (JAK1, JAK2 and Tyk2) phosphorylate docking regions for STAT3 on gp130. STAT3 is phosphorylated by JAKs and is now able to dimerize. After translocation into the nucleus target genes are regulated. B: Heterozygous mutations in the STAT3 DNA-binding domain impair target gene regulation.
*adapted from Akira *et al.*, O'Shea *et al*., O'Shea *et al*. and Ma *et al*. [28-31]

Interestingly, mice with a heterozygous knock-out of *STAT3* show no phenotype [32] and a patient hemizygous for *STAT3* due to a splice site mutation in Intron 12 did not show the full scope of symptoms typical for STAT3-HIES such as recurrent bacterial infections, connective tissue abnormalities or chronic mucocutaneous candidiasis [33].

STAT3 is an important part of many biological processes such as growth, survival, apoptosis, stress and differentiation [34]. The importance of STAT3 as central transcription factor becomes even more apparent by the plethora of ligands which are able to induce STAT3 activation. STAT3 activating ligands include but are not limited to the overlapping IL-6 and IL-12 family ligands (IL-6, IL-11, IL-27, IL-31, LIF, OSM, CNTF, cardiotrophin-1, IL-23, IL-35, IL-12), the IFN family ligands (IL-10, IL-19, IL-20, IL-22, IL-24, IL-26, IFNα, IFNβ, IFNγ), the IL-2 family ligands (IL-2, IL-7, IL-9, IL-15, IL-21) and several others such as IL-5, CSF3/G-CSF, CSF1, leptin, EGF, FGF, IGF and IL-8 [3, 31, 35-37]. This also explains the many diverse symptoms of patients affected by *STAT3* mutations (1.1.1). Of the close to 100 known *STAT3* mutations the majority of STAT3-HIES causing ones is located in either the DNA-binding domain or the SH2 domain (Figure 4) [3-6, 38, 39]. Of all these *STAT3* mutations mutations such as R382W and R382Q represent hot-spot mutations [4, 5].

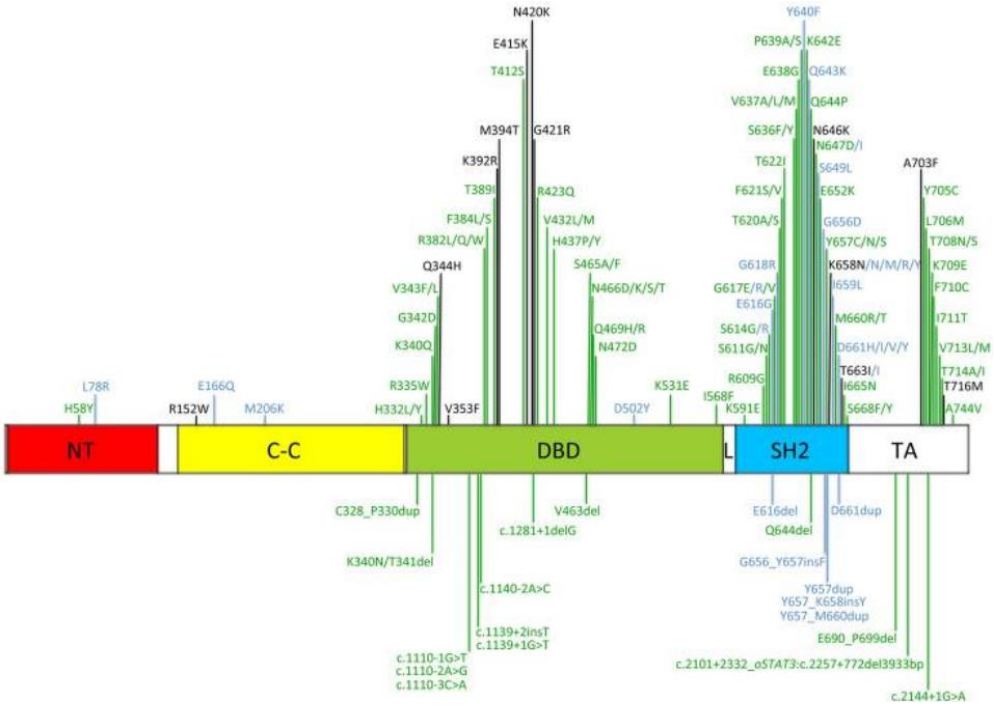

Figure 4: Locations of disease causing *STAT3* mutations

Mutations shown above the figure represent missense mutations and the resulting amino acid changes, while mutations shown below the figure are deletions, duplications or intronic mutations represented in nucleotide nomenclature. Dominant-negative mutations found in Hyper-IgE syndrome are indicated in green, those germline mutations reported in *STAT3* gain-of-function (GOF) are indicated in black, and the somatic, GOF mutations reported in malignancy, including LGL leukemia, are reported in blue. N-terminal domain = NT; coiled-coil domain = C-C; DNA-binding domain = DBD; linker domain = L; Src-homology 2 domain = SH2; transactivation domain = TA
*adapted from Vogel *et al.* [38]

1.2 CRISPR/Cas9 based Gene editing

1.2.1 The CRISPR/Cas9 system

In 2013 a research group showed that a bacterial defence system called clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated protein 9 (Cas9) system can be used to induce programmable and precise cleavage of DNA in human and mouse cells [40] . Prior to this milestone in the gene editing field other teams showed how the CRISPR/Cas systems provide adaptive immunity against viruses in bacteria and archaea and how the systems can be programmed to cut specific DNA regions [41, 42] .

CRISPR/Cas systems have been divided into three types. While type I and III systems are dependent on a complex of several proteins to function, the type II systems need only one protein to cleave DNA, the Cas9 protein [43].

Cas9 cleaves DNA via its two nuclease domains HNH and RuvC (Figure 5) and is guided to the target region by an RNA duplex of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA contains a target specific sequence, the protospacer, and a short sequence complementary to the tracrRNA. The tracrRNA consists of a short crRNA compatible sequence as well as a scaffold region which binds to the Cas9 protein. Both RNAs can be combined to a single guideRNA (sgRNA) facilitating guideRNA design and handling [42]. Additionally to the sgRNA and Cas9 protein a Cas9 dependent protospacer adjacent motif (PAM) adjacent to the protospacer sequence is needed at the target site for Cas9 to induce cleavage (e.g. NGG for *S. pyogenes* Cas9 or NNGRRT for *S. aureus* Cas9). Bacteria or archaea use the PAM to distinguish between self versus non-self [44].

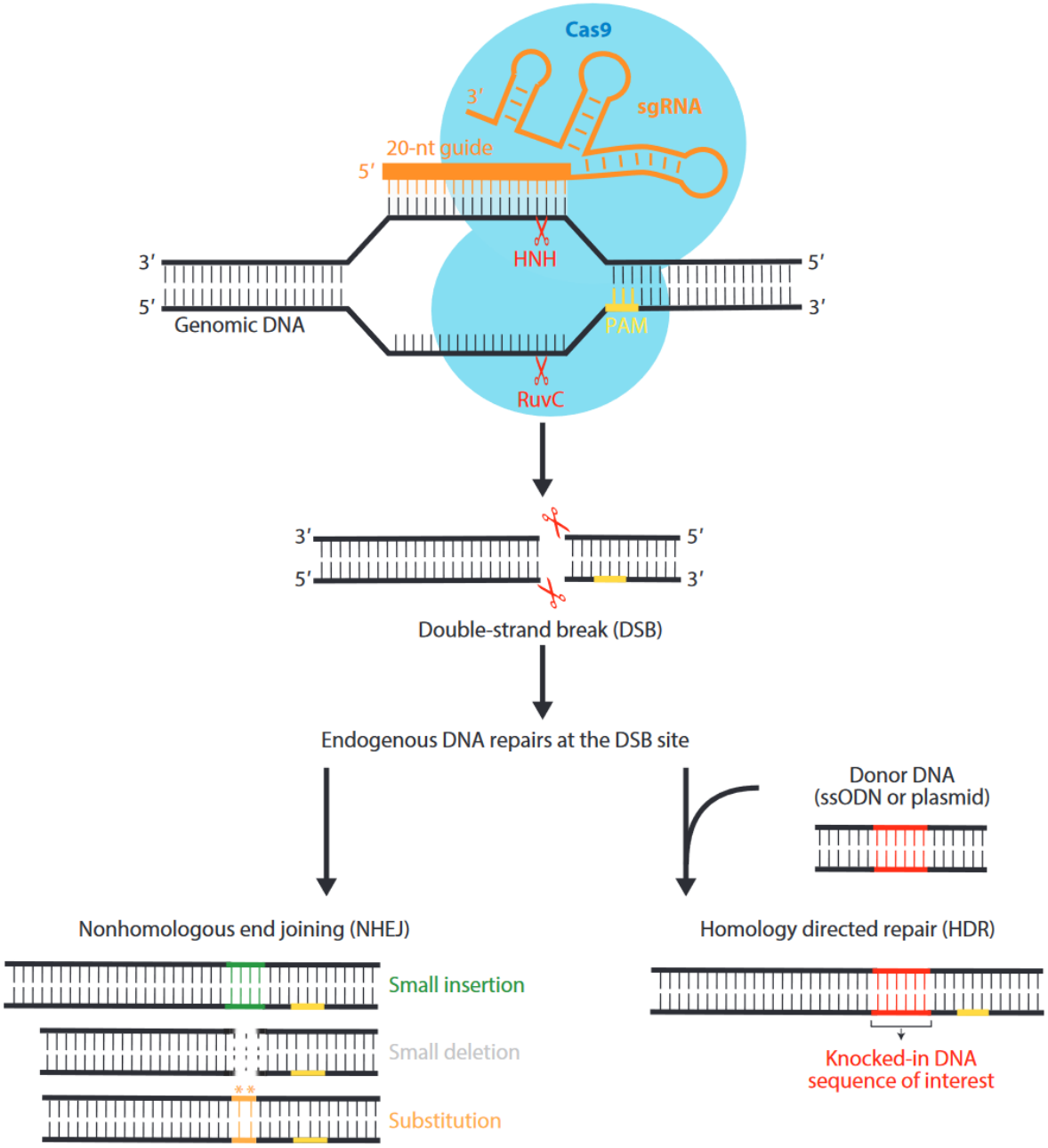

Figure 5: Cas9 DNA cutting process and possible editing outcomes

Schematic representation of the Streptococcus pyogenes Ca9 cutting process. A 20 nucleotide protospacer sequence is hybridized to the target sequence and both DNA strands are cleaved by the HNH and RuvC nuclease domains. The cellular repair processes NHEJ leads to small insertions, small deletions or substitutions. In the presence of donor DNA as repair template a precise knock-in via the HDR process is possible.
Adapted from Jiang *et al.* [45]

The CRISPR/Cas9 system allows for a programmable and specific editing of a target sequence. The editing results in the formation of double strand breaks (DSBs) and subsequent repair via non homologous end joining (NHEJ), homology directed repair

(HDR) and other cellular repair mechanisms [46]. As NHEJ is the main mechanism of repair the resulting insertions and deletions (InDels) usually cause a frame shift and consequently a premature stop codon terminating transcription of the target gene via nonsense mediated decay [47].
Further, if a DNA template is provided, the CRISPR/Cas9 mediated induction of DNA double strand breaks at the target site can be used to specifically introduce homozygous and/or heterozygous mutations with the use of the eukaryotic cell DNA repair pathway HDR. In this case the sequence of the template is copied into the target site flawlessly, although at low frequencies. An optimized approach of this strategy is the CORRECT method [48, 49].
In contrast to creating DSBs, Cas9 nickases with one impaired nuclease domain (HNH or RuvC) will only induce DNA nicks on one DNA strand [50].

The advantages of the CRISPR/Cas9 systems regarding the usage in gene editing are a high specificity, a high efficiency of InDel induction and the ability to perform small and large (1 bp to >30 000 bp) knock-ins. The disadvantages are the dependence on double strand breaks, the low efficiency of template mediated knock-ins and a dependency on certain cell cycle stages and thus mitosis for the HDR mediated repair to occur. [49, 51, 52]

1.2.2 CRISPR/Cas9 in therapy

The remarkable ability to induce targeted double strand breaks and thus insertions and deletions after their repair via the cellular main repair pathway NHEJ and to perform knock-ins via template mediated HDR made CRISPR to an important tool for the editing of DNA biochemically, *in vitro* and *in vivo* [53-56]. The use of CRISPR/Cas9 culminated in the ability to curatively treat to date only symptomatically treatable diseases.
Among the first diseases for which clinical trials involving CRISPR/Cas9 were started were Leber Congenital Amaurosis 10 (LCA10), Sickle Cell Disease (SCD), ß-Thalassemia and Transthyretin Amyloidosis. These diseases have a relatively simple mechanism of genetic pathology or treatment in common, explaining the choice as first targets. LCA10 is caused by autosomal recessive mutations in *CEP290*, SCD

and ß-Thalassemia by mutations in *HBB* and Transthyretin Amyloidosis by mutations in *TTR*. Treatments for these diseases are based on the induction of double strand breaks (DSB) by CRISPR/Cas9. In LCA10 and Transthyretin Amyloidosis the mutations in the genes *CEP290* and *TTR* are targeted directly. In SCD and ß-thalassemia a gene expressing a negative regulator of a substitution gene for the mutated gene (*BCL11A* is a negative regulator of fetal hemoglobin, which can substitute adult hemoglobin) is targeted. [57-59]

Treatments for other diseases are in development. This includes Duchenne muscular dystrophy, dystrophic Epidermolysis Bullosa, Parkinson, cholesterol management, cancer and others [60-64].

1.2.3 Adenine base editing

Some of the treatment approaches above were enabled by the development of adenine base editing a specialized version of CRISPR/Cas9 based editing for the direct conversion of adenines (A) to inosines (I) which are read and replicated as guanines (G). Adenine base editors (ABEs) including ABE7.10 which was used in this project were published in 2017. ABEs consist of a fusion protein of a *Streptococcus pyogenes* Cas9 with a via point mutation inactivated RuvC1 nuclease domain (a Cas9 nickase) and a hetero dimer of a wild type *E. coli* adenine deaminase, TadA, and a laboratory evolved TadA* which accepts single stranded DNA as substrate. As shown in Figure 6 the Cas9-TadA/TadA* fusion protein is guided to the target region by binding of Cas9 to the PAM and subsequent hybridization of the guideRNA and the target sequence. The opposing strand containing the target A becomes single stranded and thus accessible for the TadA/TadA* dimer which induces deamination and conversion to I. The still active HNH domain of Cas9 nicks the non-target strand and tricking the cellular repair machinery into favouring the converted (I containing) strand as repair template leading to an exchange of the thymine (T) opposite to the target A to cytosine (C). The resulting I·C base pair is read as G·C and will be replaced by G·C after replication. [65]

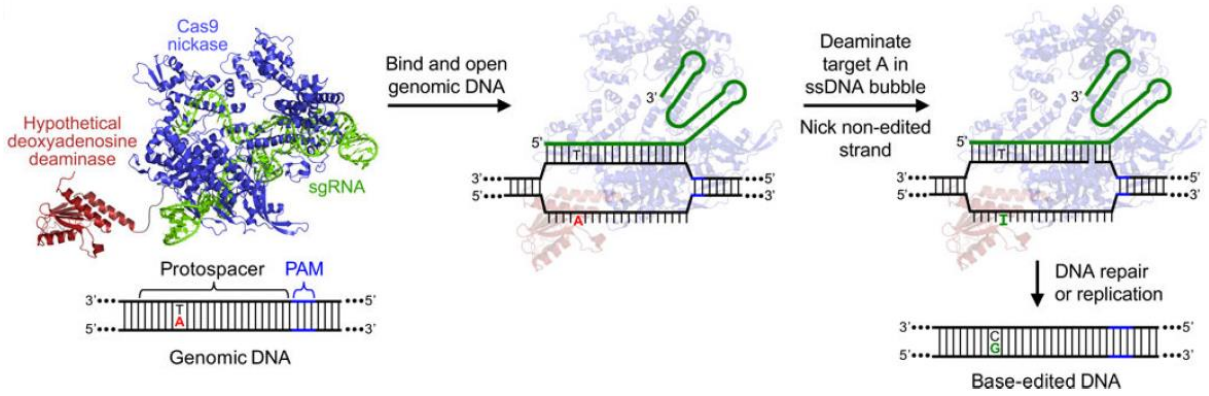

Figure 6: Adenine base editing process

ABE7.10 consists of a Cas9 nickase (blue) complexed with a guideRNA (green) and a laboratory evolved adenine deaminase (TadA/TadA*) (red). In the nucleus the fusion protein-guideRNA complex binds to the target region and unwinds the DNA. This allows the deaminase to convert any As on the now single stranded target region to Is. Further the non-target strand is nicked by the HNH domain of the Cas9 enzyme to increase acceptance of the edited strand by cellular repair mechanisms.
*Adapted from Gaudelli *et al.* [65]

For ABE7.10 to be able to perform the direct conversion of the target A to an I, the A needs to be positioned in the editing window. This window ranges from position 4 to 7 of the protospacer sequence when counted from the distal end of the PAM. As seen in Figure 7 protospacer positions 4 to 7 show high editing levels with a maximum at position 5 if using ABE7.10. Low levels of editing were detected for positions 3, 8, 9 and 10. [65]

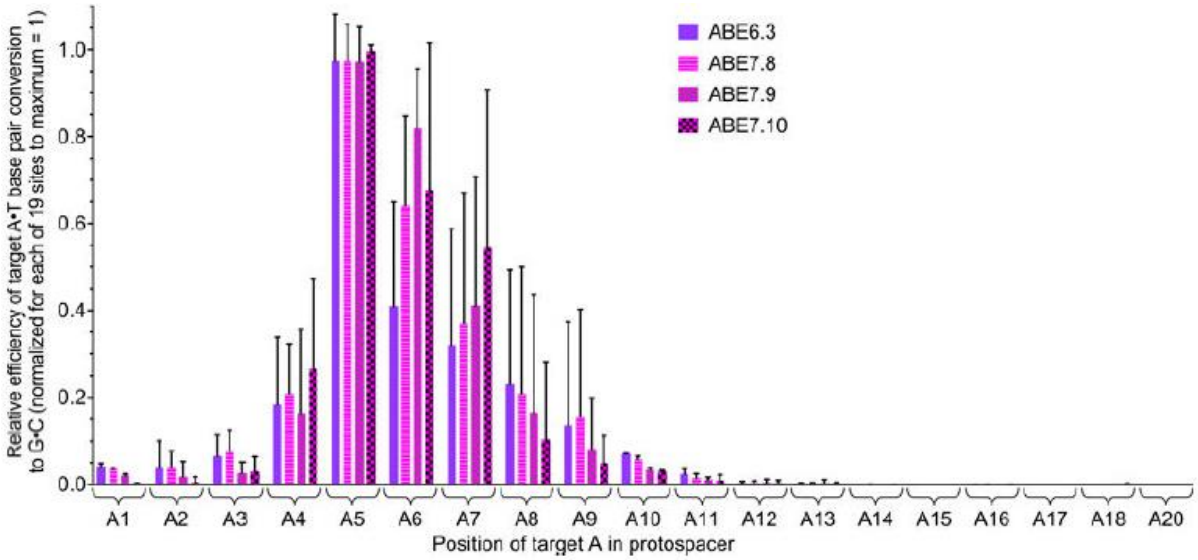

Figure 7: Activity window of ABEs

Relative A•T to G•C base editing efficiencies in HEK293T cells of late-stage ABEs at protospacer positions 1–18 and 20 across all 19 human genomic DNA sites tested. Values are normalized so that the maximum observed efficiency at each of the 19 sites for each ABE is 1. Values and error bars show the mean and s.d. across 19 sites, each with three independent biological replicates performed on different days.
Adapted from Gaudelli *et al.* [65]

Adenine base editing has been shown to be substantially more efficient in introducing single nucleotide conversions from A•T to G•C than the HDR mediated CORRECT method (Figure 8) [65].

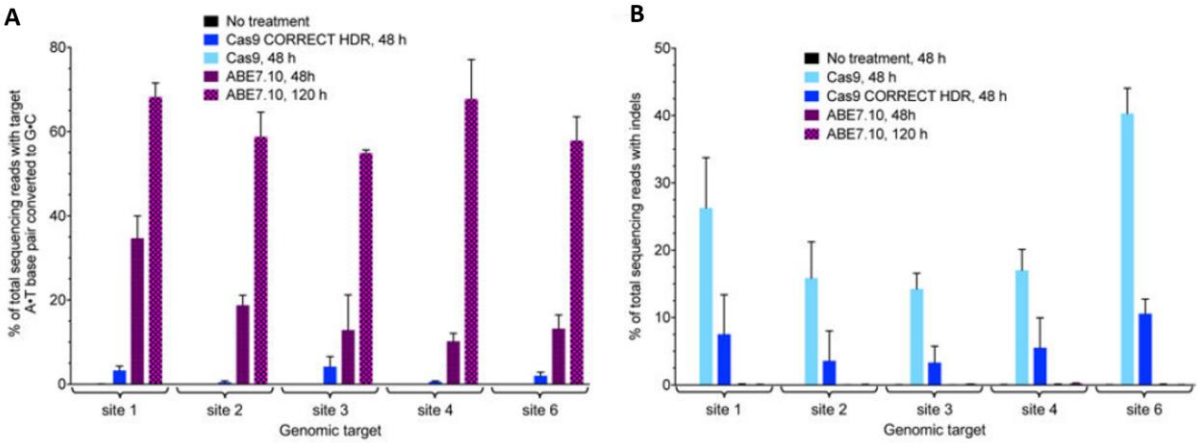

Figure 8: Comparison of ABE7.10 base editing and Cas9 HDR

A: A•T to G•C base editing efficiencies in HEK293T cells treated either with ABE7.10, or with Cas9 nuclease and an ssDNA donor template (following the CORRECT HDR method) targeted to five human genomic DNA sites. B: InDel formation in HEK293T cells treated as described in (A).
Adapted from Gaudelli *et al.* [65]

Approximately half of all known pathogenic single nucleotide polymorphisms (SNPs) are C to T changes (G to A on the opposing DNA strand). The ability to reverse these changes via an efficient A to I/G (or T to C on the opposing DNA strand) base editing tool allows for the development of treatment approaches for these mutations. [65]

The advantages of the ABE editing system compared to classical CRISPR/Cas9 knock-outs or template based repair approaches include the avoidance of DSBs and resulting unwanted by-products. As the ABE editing system does not employ the error prone NHEJ or cell cycle stage dependent HDR repair mechanism but base excision repair (BER) or mismatch repair (MMR) which are independent of cell cycle stage it allows editing in non-dividing cells. Further, ABEs allow the generation of fully repaired cells with high efficiency, independency of a repair template and increased DNA editing specificity due to deaminase substrate requirements (ABEs convert nearly exclusively As). The disadvantages include increased size of the effector enzyme, since a deaminase in the form of a laboratory evolved *E.coli* TadA dimer is attached to the Cas9 enzyme. Further, a decreased flexibility regarding the target as the target base or bases need to be positioned inside the editing window with a certain distance to the PAM, the limitation of the editing region to the size of the editing window (<10 bp) and a certain degree of random RNA editing by the constitutively active deaminase. [65-68]

ABE7.10 was continually refined to variants with higher editing efficiency such as ABEmax, which was optimized for expression and nuclear transport in mammalian cells [69]. The optimization included a different form of codon optimization (GenScript instead of the IDT codon usage in ABE7.10), the change of the used monopartite SV40 NLS (S**PKKKRKV**E) to a bipartite NLS (**KR**TAD GSEFE S**PKKKRKV**E) and the addition of a second bipartite NLS at the N-terminus [69-71].
The variant ABEmaxAW combines higher editing rates compared to ABE7.10 and low unwanted RNA editing and was the result of point mutations introduced into the two TadA monomers of ABE7.10 [66].
ABEs can not only be used to correct point mutations, but also to perform knock-outs without DSBs and in non-dividing cells. This can be achieved by ABE mediated mutation of either the start codon or splice sites of the gene of interest. Of the start codon sequence, ATG, A and T (= A on the opposing strand) can be converted via ABE but need a PAM with the appropriate distance away. If splice sites are mutated, exons will be skipped leading to a potential frameshift if the bp number of the skipped exon was not a multiple of three. A frameshift will usually lead to a premature stop codon which induces non-sense mediated decay of the transcript effectively abolishing expression of the gene of interest. [72-74]

1.2.4 Challenges of CRISPR/Cas9 based editing

While CRISPR/Cas9 mediated gene editing is a very promising technology for the future treatment of many diseases several challenges and limitations are known.
This includes off-target effects at regions with only few mismatches to the target region and DSB dependent effects such as large deletions [75-77]. Several approaches try to reduce these off-target effects by increasing the specificity of the editing system. Examples are the utilization of the mismatch sensitive seed region (first 10 bases from the PAM/3'-end) [75] or the even more sensitive core region (bases 4-7 from the PAM) of guideRNAs [78] during guideRNA design and exchange of the Cas9 enzyme with a high-fidelity version such as HiFi-Cas9 [79]. Further, the introduction of additional artificial mismatches in the guideRNA sequence reduces off-target binding although at the cost of on-target binding [77, 80]. Also truncation of guideRNAs [81, 82] and the addition of short inactive guideRNAs (CRISPR GUARDs) to shield off-target regions

from being edited [83, 84] were reported. Another problem for CRISPR/Cas9 based editing systems is host immunity to Cas9 and other parts of the editing system [85] as well as a potential positive selection of p53 compromised cells during the editing process [86].

Additionally, base editors showed a certain amount of unspecific RNA editing although refinements of the ABEs resulted in systems with substantially lowered RNA editing [66, 67, 87]. Special care needs to be taken if using *in vitro* transcribed guideRNAs as these are able to trigger a retinoic acid–inducible gene I (RIG-I) mediated immune response if the 5'-triphosphate is not removed [88]. In the case of plasmids as chosen vector for the editing system, plasmid integration can be a concern and needs to be checked after the editing process [89] .

One major hurdle in CRISPR/Cas9 based gene editing is the delivery of the editing system into the target cells. While there are several ways of delivery such as viral mediated delivery, Lipid-based nanoparticles (LNPs) mediated delivery and only *ex vivo* usable methods such as microinjection and electroporation, the method needs to be carefully optimized depending on cell type, an *ex/in vivo* setting and other factors [90]. Especially cell types such as primary low-passage number human cells are hard to transfect requiring electroporation which is a very harsh transfection method resulting in comparably low rates of surviving cells [91]. Electroporation leads to an increase in the permeability of cell membranes via the application of external electric fields allowing for the delivery of cargo into cells [92].

Another challenge of CRISPR/Cas9 based editing is the validation of the editing. Validating via a form of sequencing is most common. Sanger sequencing in combination with *in silico* analyses via programs such as TIDE and EditR allows for rough estimations of how many cells were edited [93, 94]. But this method suffers from the relatively poor resolution of Sanger sequencing. Validation via WGS allows for the detection of unknown off-target effects and an estimation of editing efficiency although WGS sensitivity is relatively low even with an already increased coverage of 70x in our case. The extremely sensitive HTS allows for a high coverage of target regions and easy subsequent analysis via programs such as CRISPResso 2 but requires to know the regions of interest before performing HTS [95]. This information needs to be acquired by *in silico* prediction via programs such as CRISPOR or other techniques such as Digenome-seq, GUIDE-seq or EndoV-seq to find potential off-target sites [96-99].

2. Aim of the study

STAT3-HIES is an inborn error of immunity caused by heterozygous dominant-negative *STAT3* mutations. Up to date no curative therapy is available. The aim of the here presented study was to assess if gene editing mediated correction of the disease-causing *STAT3* mutations might enable a form of curative treatment.

The first objective was to choose a DNA editing system that would be able to correct the most common *STAT3* mutations in STAT3-HIES patients' cells with high specificity and efficiency.

We planned to confirm functionality of this editing system via target region PCR template based assays first and then move on to an electroporation or lipofection mediated editing of primary control and patient fibroblasts.

After the correction of the mutation had been achieved and was confirmed by Sanger sequencing and an assay to quantify editing efficiencies, the goal was to optimize the editing system to a level sufficient for subsequent analyses such as a safety evaluation or functional analyses. The optimization was planned to include changes of effector enzymes, delivery protocols and methods as well as potential guideRNA adjustments and modifications.

The safety evaluation would consist of a confirmation of genomic integrity after treatment via a combination of high coverage whole genome sequencing to look for any abundant undesired effects and high-throughput sequencing of predicted off-target sites for a more sensitive analysis.

The functional analysis of STAT3 signaling in a patient cell model should confirm any beneficial effects of a correction of the *STAT3* mutation. To measure STAT3 signaling we wanted to employ STAT3 DNA binding ELISAs and/or quantitative real-time PCR of STAT3 target genes.

In the case of an improved STAT3 signaling after treatment we planned to adapt the developed editing system to other *STAT3* mutations and additional cell types as a last objective.

3. Results

3.1 A treatment approach for the *STAT3* R382W mutation

3.1.1 R382W treatment design considerations

The human *STAT3* c.1144C>T/p.R382W mutation in Exon 13 (R382W) results in an amino acid change from arginine to tryptophan in codon 382 and is one of the most common STAT3-HIES causing mutations. It is located in the DNA binding domain of STAT3 and influences STAT3 DNA binding negatively in $STAT3^{+/R382W}$ and $STAT3^{R382W/R382W}$ dimers.

In a first step to find a suitable treatment approach only *S. pyogenes* Cas9 mediated editing was considered as this was the most used, studied and versatile Cas9 enzyme. Several approaches employing *S. pygenes* Cas9 were taken into account:

- The specific knock-out of the affected allele to alleviate the dominant-negative effect and increase levels of functional STAT3 dimers
- The correction of the mutation via induction of a DSB close to the mutation site and subsequent template/HDR mediated repair of the mutation site
- The direct conversion of the mutated nucleotide adenine (A) to inosine (I) and consequentially guanine (G) on the + strand via the adenine base editor ABE7.10

With the use of *S. pyogenes* Cas9 the PAM was restricted to “NGG” enabling the guideRNAs shown in Figure 9 for R382W mutation specific editing. A guideRNA with its seed region (positions 1 – 10 counted from the PAM) positioned over the mutation site allowing for editing with high specificity was not available. Therefore, the treatment options with the classical CRISPR/Cas9 editing for an allele specific knock-out or template/HDR mediated repair were not considered further as high specificity was a requirement for any approach due to the low difference of only 1 bp of wild type and R382W allele.

Interestingly 3 guideRNAs positioned the ABE editing window over the mutation site and the mutation was represented as A (wild type G) making adenine base editing an option for the correction of the mutation. ABE7.10 mediated editing offers valuable advantages such as expected high repair efficiency rates compared to HDR mediated

approaches, no DSBs and no expected off-target effects for the healthy allele. ABE7.10-mediated repair depends on a NGG PAM and requires the target A (or As) to be at positions 4 - 7 of the protospacer and PAM sequence (PAM counted as 21-23). As described in 1.2.3, positions 8 and 9 can also show low levels of editing. Therefore guideRNAs need to be placed in such a way that no bystander As are at positions 4 – 9.

Figure 9: Available guideRNAs positioned over the mutation site

All guideRNAs with an NGG PAM spanning the R382W mutation at position 0.
+ strand = black; - strand = blue; PAMs = red; R382W mutation = bold and underlined; grey box = guideRNA seed region; ABE editing window = red box

WRA-2 was chosen as the most promising guideRNA, since the target A lies in the editing window of ABE7.10 and no other As are close to the editing window to prevent undesired editing of the A at position 3 (Figure 9). Conversion of the bystander A would result in an amino acid change from serine (TCC) to prolin (CCC) and should therefore be avoided.

The chosen editing strategy is summarized in Figure 10. With this editing system the nickaseCas9-ABE7.10-WRA-2 complex is expected to first bind to the AGG PAM, followed by the attempt to hybridize the WRA-2 guideRNA to the target region. In case of the R382W affected allele WRA-2 will hybridize flawlessly resulting in the target DNA strand to become single stranded and thus editable by ABE7.10. This should result in the direct conversion of the target A to I which is read and replicated as G by DNA and RNA polymerases. Due to the heterozygosity of the disease-causing mutation potential editing of the wild type allele needs to be taken into account. While hybridization of the guideRNA to the wild type allele is already discouraged by the mismatch at the mutation site, no A is present in the editing window of ABE7.10. We speculated that

this will prevent any editing of the healthy allele aside from very rare double strand breaks due to DNA nicks induced by the nickaseCas9 of the editing system.

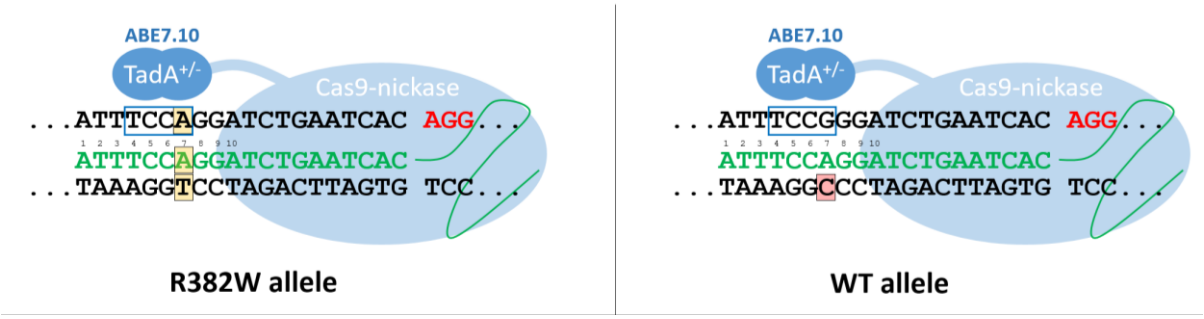

Figure 10: Binding of the editing complex to the target region

The ribonucleoprotein complex consisting of a Cas9 nickase, a dimer of wild type and mutated TadA monomers and the guideRNA WRA-2 binds to the R382W *STAT3* allele (left side) by flawless hybridization of the guideRNA. The target A is present in the ABE7.10 editing window at position 7. The editing complex is discouraged from editing the healthy *STAT3* allele (right side) by the mismatch of the guideRNA WRA-2 at position 7 and by the lack of an editable A in the editing window. Blue framed box = ABE7.10 editing window; yellow boxes = mutation site; red box = mismatch; red letters = PAM; green letters = guideRNA WRA-2

3.1.2 guideRNA quality control and validation

The WRA-2 guideRNA was produced by *in vitro* transcription with the EnGen sgRNA synthesis kit *s. pyogenes* from NEB and purified via column RNA isolation or phenol/chloroform isolation. To prevent a RIG-I mediated immune response, *in vitro* transcribed guideRNAs were treated with calf intestinal phosphatase (CIP) to remove the 5’ triphosphate.

As quality control produced guideRNAs were subjected to gel electrophoresis. The guideRNAs showed no signs of degradation and showed the correct length, when run on an agarose gel (Figure 11). NanoDrop analyses showed a high RNA purity for both isolation methods (260/280 and 260/230 ratios, data not shown). This analysis was performed for every batch of newly synthesized guideRNA.

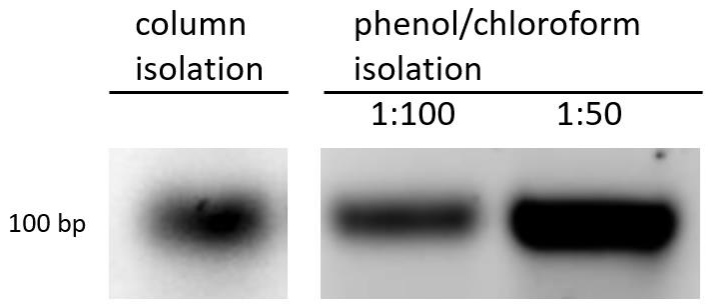

Figure 11: Quality control of purified guideRNA WRA-2

Agarose gel electrophoresis shows no degradation (lack of smearing below bands) of the produced guideRNA.

To validate produced WRA-2 guideRNA it was used for a Cas9 *in vitro* assay which confirmed the ability of guideRNAs complexed with Cas9 protein to cleave PCR generated amplicons of the target region. The WRA-2 guideRNA showed the ability to induce DNA double strand breaks in target region amplicons. As seen in Figure 12 the PCR amplicons (680 bp) were cleaved resulting in two bands at ~400 and ~300 bp. This confirmed the functionality of the designed sgRNA.

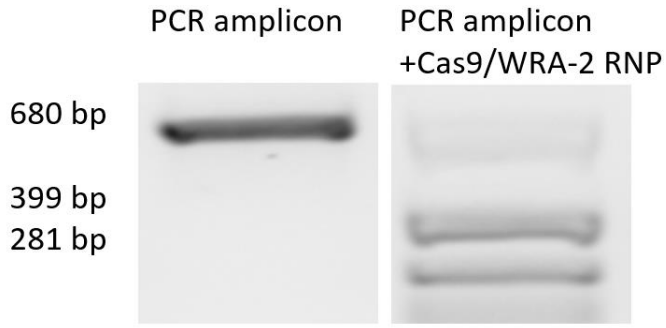

Figure 12: WRA-2 sgRNA induces cleavage in PCR amplicons of the target site

Cas9 *in vitro* assay with Cas9/WRA-2 RNP shows cleavage of 680 bp amplicons of the target region confirming functionality of the produced WRA-2 guideRNA.

3.1.3 Preparation of transfection substrates

To prepare for the repair of patient cells the guideRNA WRA-2 sequence was cloned into the plasmid pbs-U6 to enable expression in mammalian cells via a U6 promoter. Therefore, the WRA-2 sequence was ordered as DNA oligonucleotide and ligated into the vector. After transformation of the generated pbs-U6:WRA2 construct into *E. coli* and subsequent cultivation four colonies were picked to determine, if the insert was present in the vector. All four picked colonies showed the guideRNA WRA-2 insert when Sanger sequenced (Figure 13).

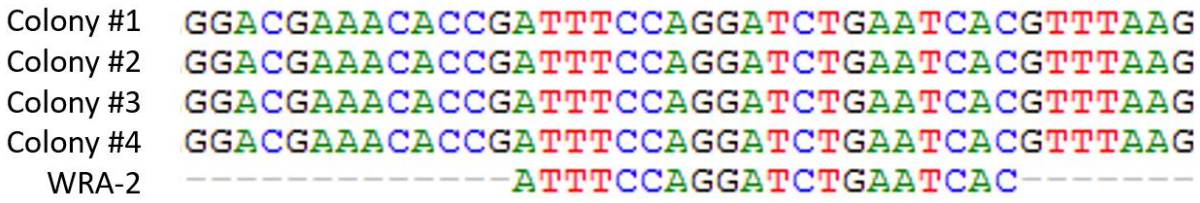

Figure 13: Successful integration of the WRA-2 sequence insert into pBS-U6:WRA2

All 4 picked DH5α *E. coli* colonies showed the inserted WRA-2 guideRNA sequence after transformation as indicated by the matching alignments.

Colony #1 was used for the next steps of the plasmid production. The complete Sanger sequence of the insertion site was aligned to the pBS-U6 vector sequence to check plasmid integrity in this region. As shown in Figure 14 no mutations were detected.

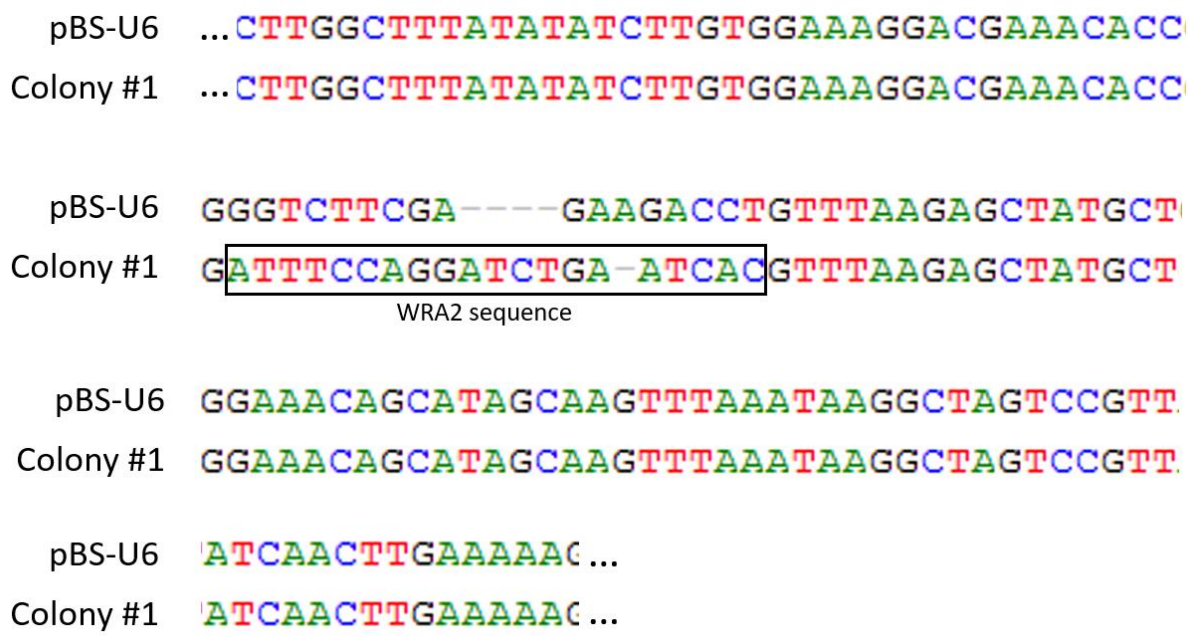

Figure 14: Colony #1 shows no mutations in the sequenced region

An alignment of the pBS-U6 sequence with the full Sanger sequenced region shows no mutations and indicates plasmid integrity.

In addition to the generated construct expressing the guideRNA WRA-2 via a U6 promoter we ordered the plasmid CMV-ABE7.10 encoding for ABE7.10 under CMV promoter control from addgene.

3.1.4 Establishment of a transfection protocol

Lipofection based transfection with the reagents JetMessenger (Polyplus), mRNA transfection kit (Miltenyi), Lipofectamine 2000, Lipofectamine 3000 and CRISPRmax (all Thermo Fisher) proved inefficient in primary *STAT3*$^{+/R382W}$ fibroblasts (data not shown). Therefore, we resorted to electroporation with the Lonza 4D Nucleofector X unit to enable efficient transfection of primary *STAT3*$^{+/R382W}$ fibroblasts.

Electroporation parameters for primary *STAT3*$^{+/R382W}$ fibroblasts were established by transfection of GFPmax plasmid (data not shown). To determine the required amounts of plasmid for the correction of R382W we electroporated varying amounts of the

pBS-U6:WRA2 plasmid and the CMV-ABE7.10 plasmid into *STAT3*$^{+/R382W}$ patient fibroblasts in a molecular weight ratio of 3:1 (guideRNA plasmid:ABE plasmid) which roughly translates to a molar ratio of 1:1. While 8 µg of total plasmid amount showed only low repair efficiency of the mutation 48 h after the electroporation, we observed 14% and 22% repair efficiency with 12 and 16 µg total plasmid amount as quantified by Sanger sequencing and subsequent EditR analysis (Figure 15). Since the viability of cells electroporated with 16 µg substrate was very low (phase contrast microscopy, Figure 15A), we settled for a total amount of 12 µg (9 µg CMV-ABE7.10 and 3 µg pBS-U6:WRA2).

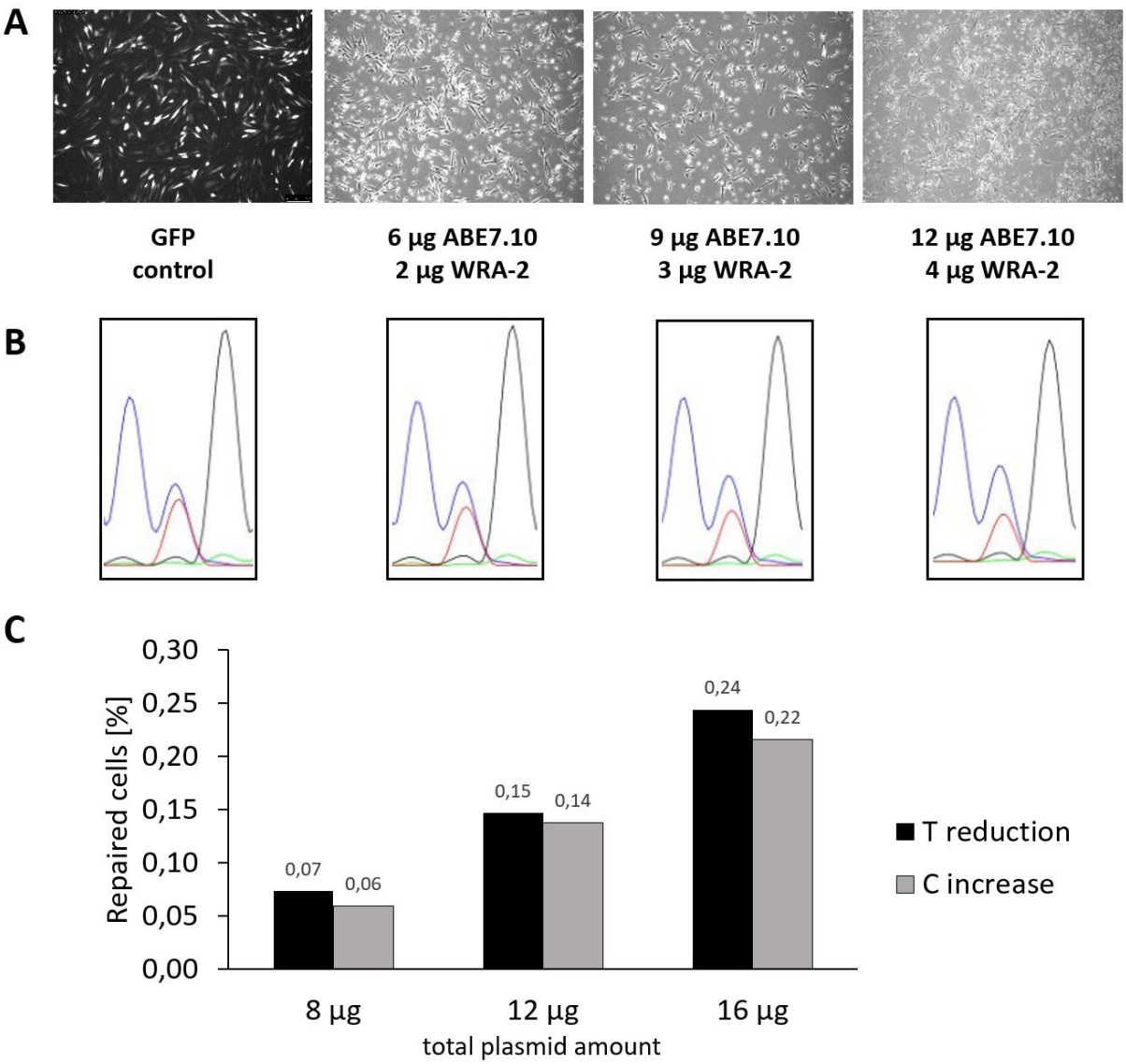

Figure 15: Titration of plasmid amounts

A: Fluorescence or phase contrast microscopy pictures of *STAT3*$^{+/R382W}$ fibroblasts 24 h after electroporation. B: Sanger chromatogram of the mutation site (double peaks) show a reduced T peak (red) and increased C peak (blue) indicating a successful correction (T to C) of the mutation. C: EditR analysis of the Sanger chromatogram to quantify the repair efficiency. n = 1

3.1.5 *Ex vivo* treatment of primary patient fibroblasts

We used the established electroporation conditions to treat three populations of primary *STAT3*$^{+/R382W}$ fibroblasts. We observed a repair efficiency (increase of wild type allele) of 29% ± 7% as determined by Sanger sequencing and subsequent EditR analysis. HTS and WGS showed a repair efficiency of 43.7% and 33.5% (Figure 16).

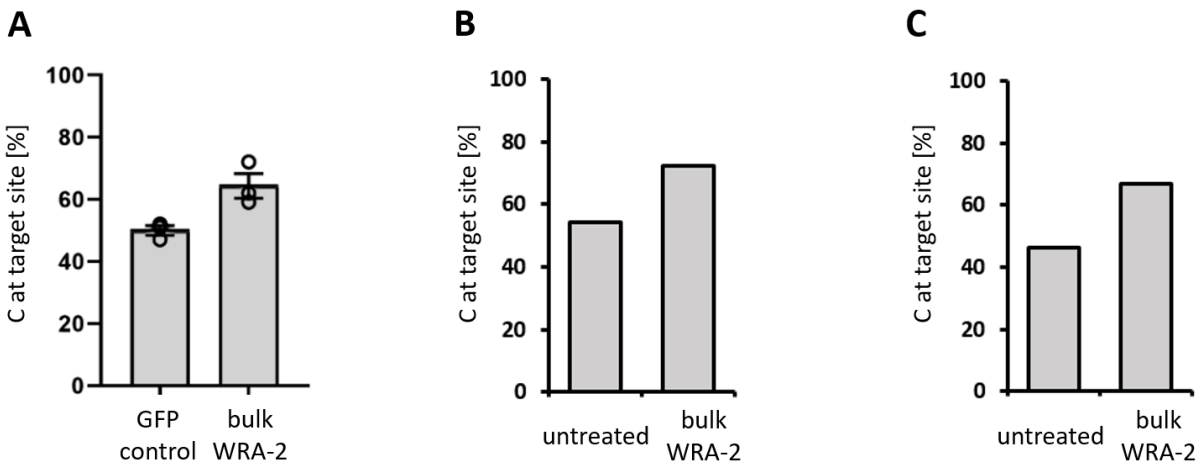

Figure 16: Correction at the mutation site in treated *STAT3*$^{+/R382W}$ fibroblasts

Analysis of A·T to G·C conversion at the target site in ABE7.10/WRA-2 treated *STAT3*$^{+/R382W}$ fibroblasts. A: Sanger sequencing and subsequent EditR analysis show a 29 ± 7% increase in the cytosine signal in ABE7.10/WRA-2 treated compared to GFP transfected (GFP control) *STAT3*$^{+/R382W}$ fibroblasts. Bars indicate mean values; dots individual experiments; error bars represent standard error of the mean (SEM); n = 3 B: Whole genome sequencing reveals a 33.5% increase of the cytosine signal of the bulk ABE/WRA-2 sample compared to untreated cells. n = 1 C: High-throughput sequencing shows a 43.7% increase of the cytosine signal of the bulk ABE/WRA-2 sample compared to untreated cells. n = 1
Adapted from Eberherr *et al.* [100]

To generate a 100% repaired cell population for further analyses we isolated single cell clones from a treated *STAT3*$^{+/R382W}$ fibroblast population. As seen in Figure 17 four of 20 single cell clones (numbers 5, 7, 18 and 20) showed no presence of the mutation and thus represented fully corrected cells. They were used for further experiments as repaired single cell clones #5, #7, #18 and #20.

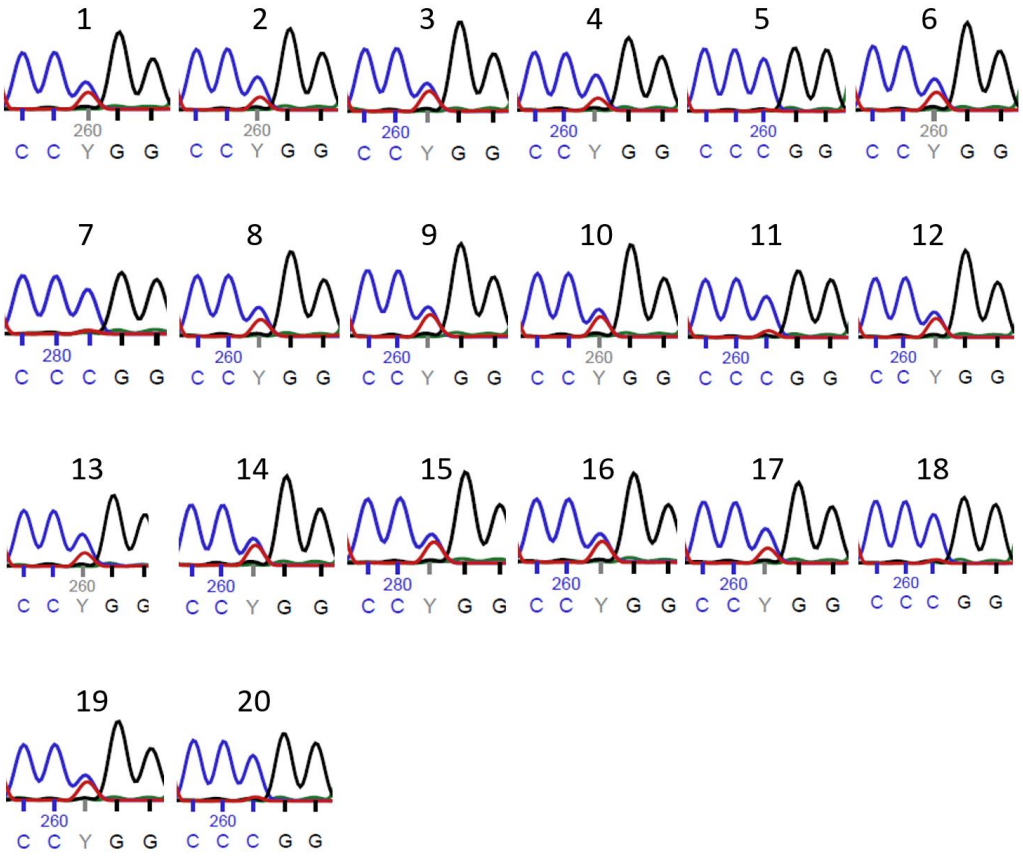

Figure 17: Sanger sequencing of single cell clones of treated fibroblasts

Primary *STAT3*$^{+/R382W}$ fibroblasts were seeded at low densities after treatment to obtain repaired single cell clones. Clones #5, #7, #18 and #20 show no signs of the R382W mutation in *STAT3*.

3.1.6 Safety evaluation

Since the genetic changes by CRISPR associated systems are known to cause side effects such as induced changes at off-target sites or plasmid integration, we performed a safety evaluation of the treatment approach. Off-target analysis via off-target *in silico* prediction and subsequent HTS of the most relevant predicted sites as well as WGS with increased coverage were used to check DNA integrity. Further we also analyzed the on-target site via HTS to see the effects on the healthy *STAT3* allele and to determine the product purity of the editing process.

In silico off-target prediction via the online tool CRISPOR resulted in the prediction of 188 off-targets for the guideRNA WRA-2 of which all except the *STAT3* wild type allele

had at least 3 mismatches. Off-targets evaluated as exonic by CRISPOR (in or close to exons) except the *STAT3* target locus showed at least four mismatches. While editing at the off-target sites was very unlikely due to at least three mismatches to the WRA-2 guideRNA, we checked the exonic and thus most relevant off-target sites regardless. We detected no off-target editing at the predicted exonic off-target sites via HTS (data shown in Eberherr *et al.* [100] – Supplementary Table 4). Analysis of bulk treated and untreated patient fibroblasts via WGS to detect abundant and guideRNA independent editing revealed similar patterns of single nucleotide variant (SNV) counts and InDel counts in both populations (Figure 18).

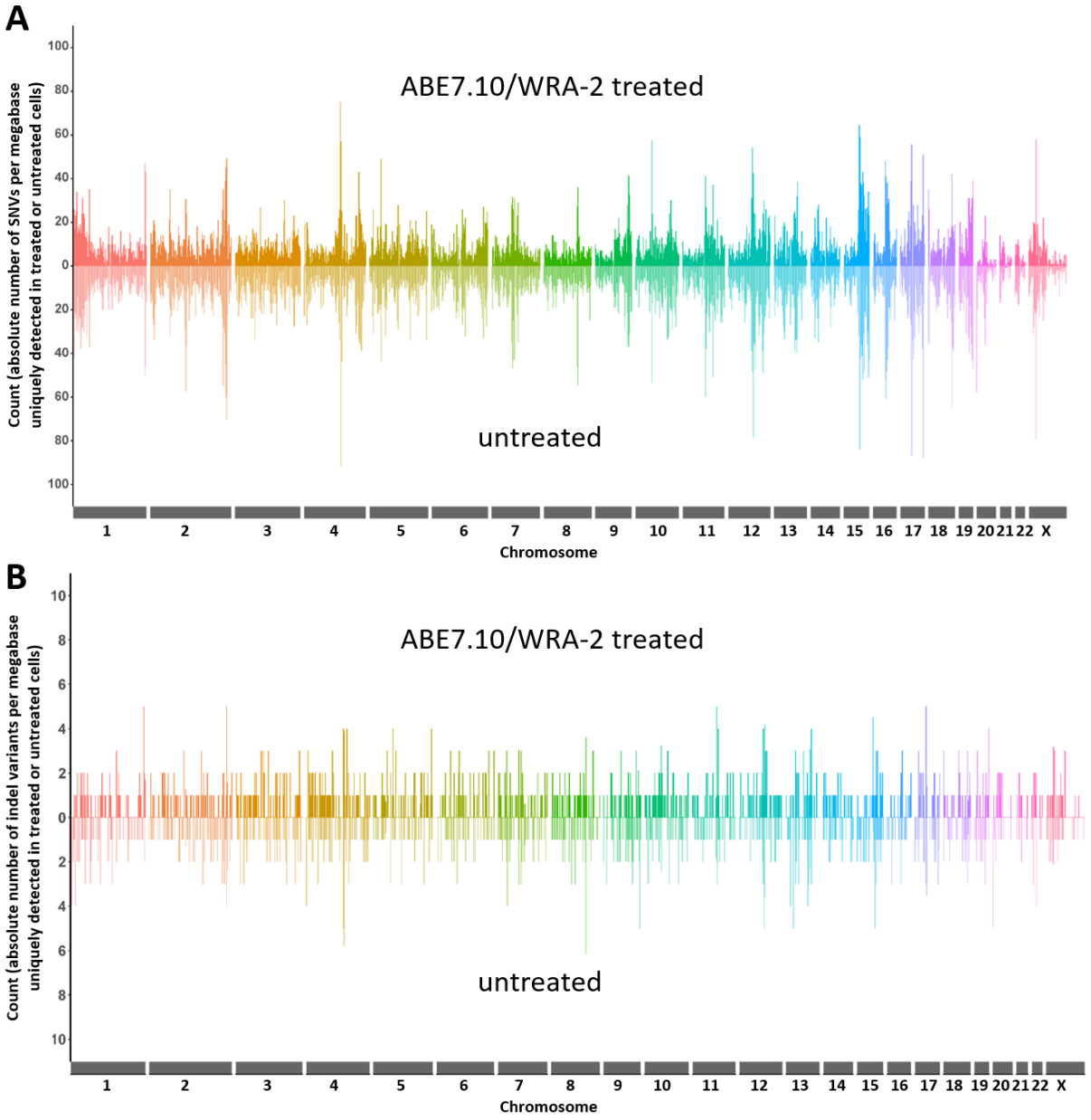

Figure 18: WGS data analysis – SNVs and InDels

A: Substantially symmetric patterns in genomic distribution of the absolute number of SNVs per megabase uniquely detected in untreated (pointing downward, light colors) and ABE7.10/WRA-2 treated (pointing upward, dark colors) *STAT3*$^{+/R382W}$ fibroblasts suggest a dependency on the sequencing content and complexity rather than on real variant patterns. Different colors indicate different chromosomes.
B: Substantially symmetric patterns in genomic distribution of the absolute number of InDel variants per megabase uniquely detected in untreated (pointing downward, light colors) and ABE7.10/WRA-2 treated (pointing upward, dark colors) *STAT3*$^{+/R382W}$ fibroblasts suggest a dependency on the sequencing content and complexity rather than on real variant patterns. Different colors indicate different chromosomes. SNVs = single-nucleotide variants; WGS = whole-genome sequencing
Adapted from Eberherr *et al.* [100]

Analysis of SNV distribution (Figure 19A) to check for increased frequencies of the ABE promoted conversion of A·T to G·C in the WGS data revealed higher counts of A·T to G·C SNVs in untreated patient cells compared to treated patient cells and the WGS control which consisted of two different tissues from the same donor run with identical WGS settings. Further, treated patient cells showed more C·G to A·T SNVs than the untreated cells and control. Since the ABE promoted conversion direction of A·T to G·C is underrepresented in treated patient cells, an ABE independent cause for the SNVs is suggested.

To check for guideRNA (and thus Cas9-nickase) dependent off-targets not included in the *in silico* prediction we performed a Levenshtein distance analysis. During this analysis the guideRNA sequence is matched to the DNA sequence 45 bp up- and downstream of detected SNVs. The minimal Levenshtein distance for all via WGS detected SNVs was four equalling four mismatches to the guideRNA sequence (Figure 19B). Since a relatively high number of four mismatches to the guideRNA sequence abolishes editing in CRISPR/Cas9 associated editing systems in the majority of cases, the result indicates a guideRNA independent cause of the SNVs such as naturally occurring mutations.

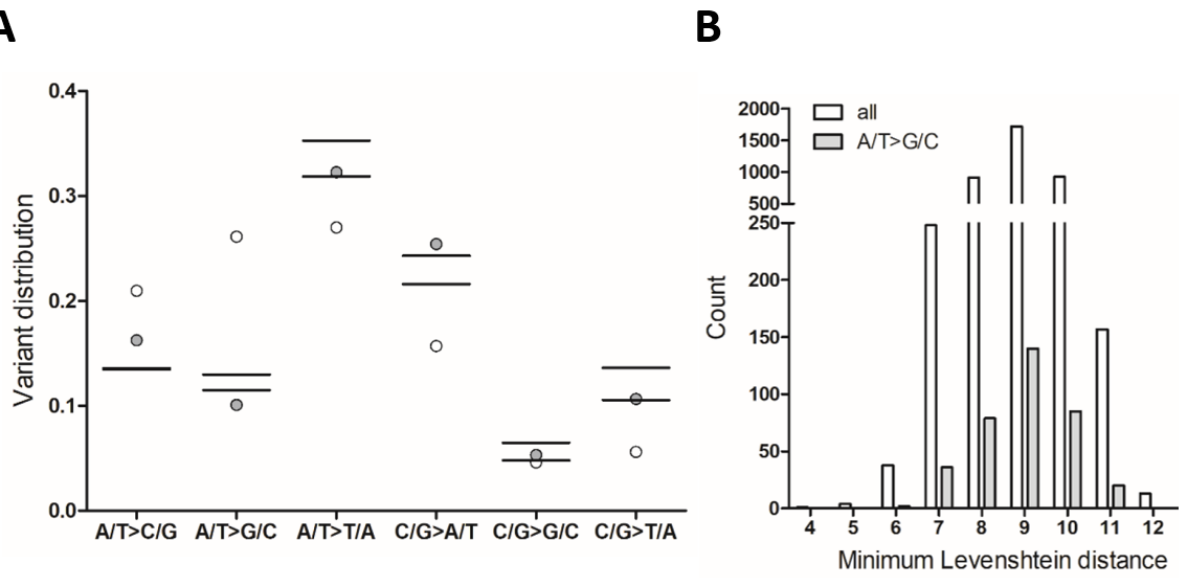

Figure 19: WGS data analysis – SNV distribution and Levenshtein distance

A: Relative distribution of SNVs of total genetic variants detected in WGS data of untreated (white dots) and ABE7.10/WRA-2 treated *STAT3*$^{+/R382W}$ fibroblasts (grey dots) compared with two control genomes from two different tissues of one donor analyzed on the same platform with equally prepared libraries (black lines) showed no increase in A·T to G·C conversions in treated cells. B: Levenshtein analysis of edits (SNVs, InDels) necessary to align the protospacer sequence of guideRNA WRA-2 to the genomic region 45 bp up- and downstream of each variant unique to the ABE7.10/WRA-2 treated *STAT3*$^{+/R382W}$ fibroblasts. The minimum Levenshtein distance of all detected SNVs (white bars) in ABE7.10/WRA-2 treated *STAT3*$^{+/R382W}$ fibroblasts was four, whereas the minimum Levenshtein distance of all A/T>G/C SNVs (grey bars) was six.
Adapted from Eberherr *et al.* [100]

Furthermore we found no signs of plasmid integration as all major structural variants detected were also present in untreated samples. Reads from both plasmids used during the electroporation were detected with a low coverage (pCMV-ABE7.10: 3.87x and pBS-U6-WRA2: 2.29x) suggesting that cellular plasmid amounts were diluted by cell division during the seven passages between electroporation and WGS.

The HTS analysis of the STAT3 target site (Table 1) showed 66.74% healthy allele (#1) in treated cells compared to 46.43% healthy allele in untreated cells. The allele affected by c.1144C>T (R382W, #2) was reduced to 23% in treated cells from 45,81% in untreated cells suggesting a robust repair efficiency of 43,7% (increase wild type allele) to 49,8% (decrease R382W allele). Further, low but increased amounts of several allele variants with additional point mutations around the editing site (#3 - #8)

were detected after treatment suggesting a possible effect of the treatment, although all allele variants were detected in untreated cells. Unintended changes were close to or in the ABE7.10 editing window except for the point mutation in #8 which lies in the PAM region but does not affect PAM functionality as the Cas9 specific PAM sequence 5'-NGG-3' is not changed. Seed region (10 first bases after the PAM) and nick-site did not show any unintended changes.

Aligned editing outcomes via HTS	Treated [%]	Untreated [%]
PAM WRA2 sgRNA 3'-GGACACTAAGTCTAGGACCTTTA-5'		
#1 CCTGTGATTCAGATCCCGGAAATTTAACATTCTGGGCACA	66,74	46,43
#2 CCTGTGATTCAGATCCTGGAAATTTAACATTCTGGGCACA	23,00	45,81
#3 CCTGTGATTCAGACCCTGGAAATTTAACATTCTGGGCACA	0,79	0,11
#4 CCTGTGATTCAGATGCCGGAAATTTAACATTCTGGGCACA	0,41	0,20
#5 CCTGTGATTCAGACCCCGGAAATTTAACATTCTGGGCACA	0,35	0,11
#6 CCTGTGATTCAGATCCCGGAAACTTAACATTCTGGGCACA	0,31	0,10
#7 CCTGTGATTCAGATCCTGAAAATTTAACATTCTGGGCACA	0,22	0,03
#8 CCCGTGATTCAGATCCCGGAAATTTAACATTCTGGGCACA	0,20	0,11

Table 1: HTS of the *STAT3* target site after bulk treatment

HTS of the target site in treated and untreated *STAT3*$^{+/R382W}$ primary fibroblasts. Treated cells show a higher amount of the healthy allele (#1) and a reduced amount of the R382W *STAT3* allele (#2). An allele with a mutation from T to C (#3, **C**, green) close to the R382W mutation site (#3, **T**, red) shows substantially higher amounts in treated cells (0.79% treated and 0.11% untreated) and might reflect undesired editing by the repair system. Black line indicates nick-site; blue box indicates seed region; grey box shows ABE7.10 editing window
Adapted from Eberherr *et al.* [100]

3.1.7 Analyses of STAT3 function in treated patient fibroblasts

The molecular cause for autosomal dominant STAT3-HIES in patients are heterozygous mutations in the *STAT3* gene. These mutations include point mutations such as p.R382W and p.R382Q which are among the most frequent mutations causing STAT3-HIES. To be able to evaluate if a repair of the mutation leads to improved STAT3 signaling, we utilized western blots of nuclear phosphorylated STAT3, a STAT3

DNA binding ELISA (TransAM) and measurements of STAT3 target gene expression (*CCL2 and SOCS3*) via reverse transcription quantitative real-time PCR.

To validate our set of functional analyses we analyzed the effect of a heterozygous R382W mutation on STAT3 function and compared healthy control fibroblasts (HDFa from ATCC) to *STAT3*$^{+/R382W}$ fibroblasts (Figure 20) after IL-6 stimulation. STAT3 function was impaired regarding amounts of phosphorylated (Y705) STAT3 in the nucleus (Figure 20A), STAT3 DNA binding capacity (Figure 20B) and target gene expression (Figure 20C and D).

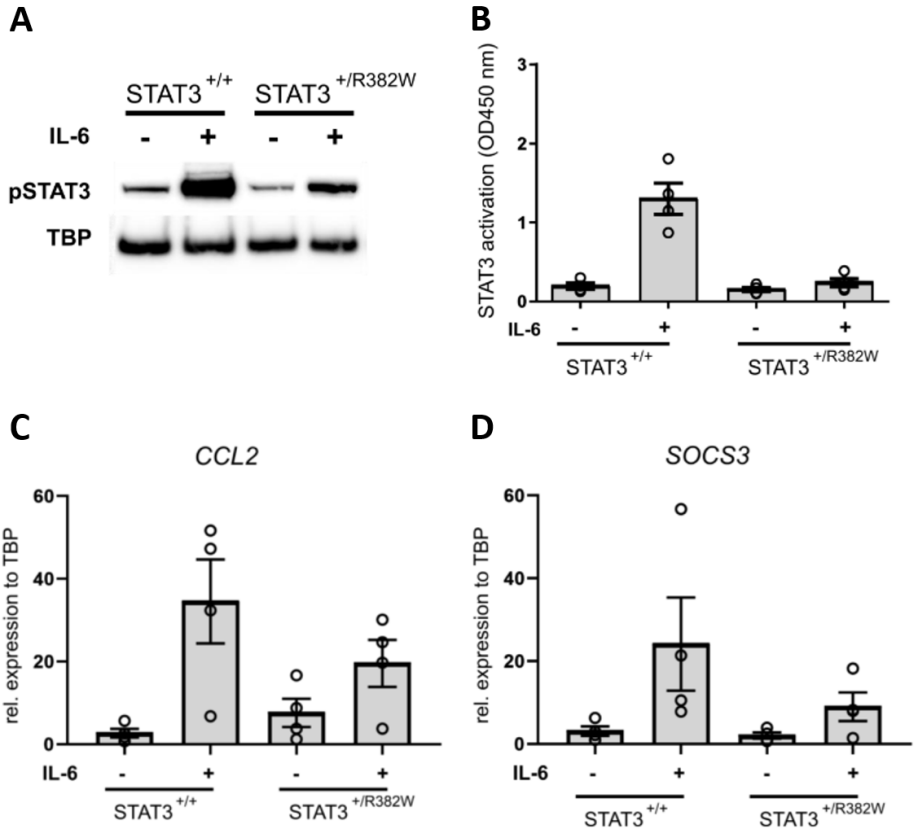

Figure 20: Primary *STAT3*$^{+/R382W}$ fibroblasts show impaired STAT3 signaling

A: Phospho(Y705)-STAT3 western blot of nuclear extracts of healthy control primary fibroblasts and primary *STAT3*$^{+/R382W}$ fibroblasts after IL-6 stimulation shows reduced amounts of phosphorylated STAT3. B: TransAM of healthy control primary fibroblasts and primary *STAT3*$^{+/R382W}$ fibroblasts shows reduced amounts of STAT3 DNA binding activity. C and D: STAT3 target gene expression of *CCL2* (C) and *SOCS3* (D) after IL-6 stimulation show reduced expression in *STAT3*$^{+/R382W}$ fibroblasts.
Bars indicate mean values; dots individual experiments; error bars represent standard error of the mean (SEM); n = 4 (B, C, D)
Adapted from Eberherr *et al.* [100]

To assess if treatment with ABE7.10 and WRA-2 guideRNA improves STAT3 signaling, analyses were performed on untreated, non-target control treated, bulk treated and repaired single cell clone (SCC) samples (Figure 21). Western blots of nuclear extracts after IL-6 stimulation showed an increased amount of phosphorylated STAT3 in repaired SCCs and suggest an increased nuclear retention through improved DNA binding capacity of STAT3 dimers (Figure 21A). STAT3 DNA binding activity measured via TransAM after IL-6 stimulation showed increased levels of activity in bulk treated patient cells and significantly increased activity in repaired SCCs (Figure 21B). *SOCS3* and *CCL2* expression after IL-6 stimulation were significantly increased in repaired SCCs (Figure 21C and D). All together these results suggest improvement of STAT3 signaling after treatment with the ABE7.10 base editor and WRA-2 guideRNA.

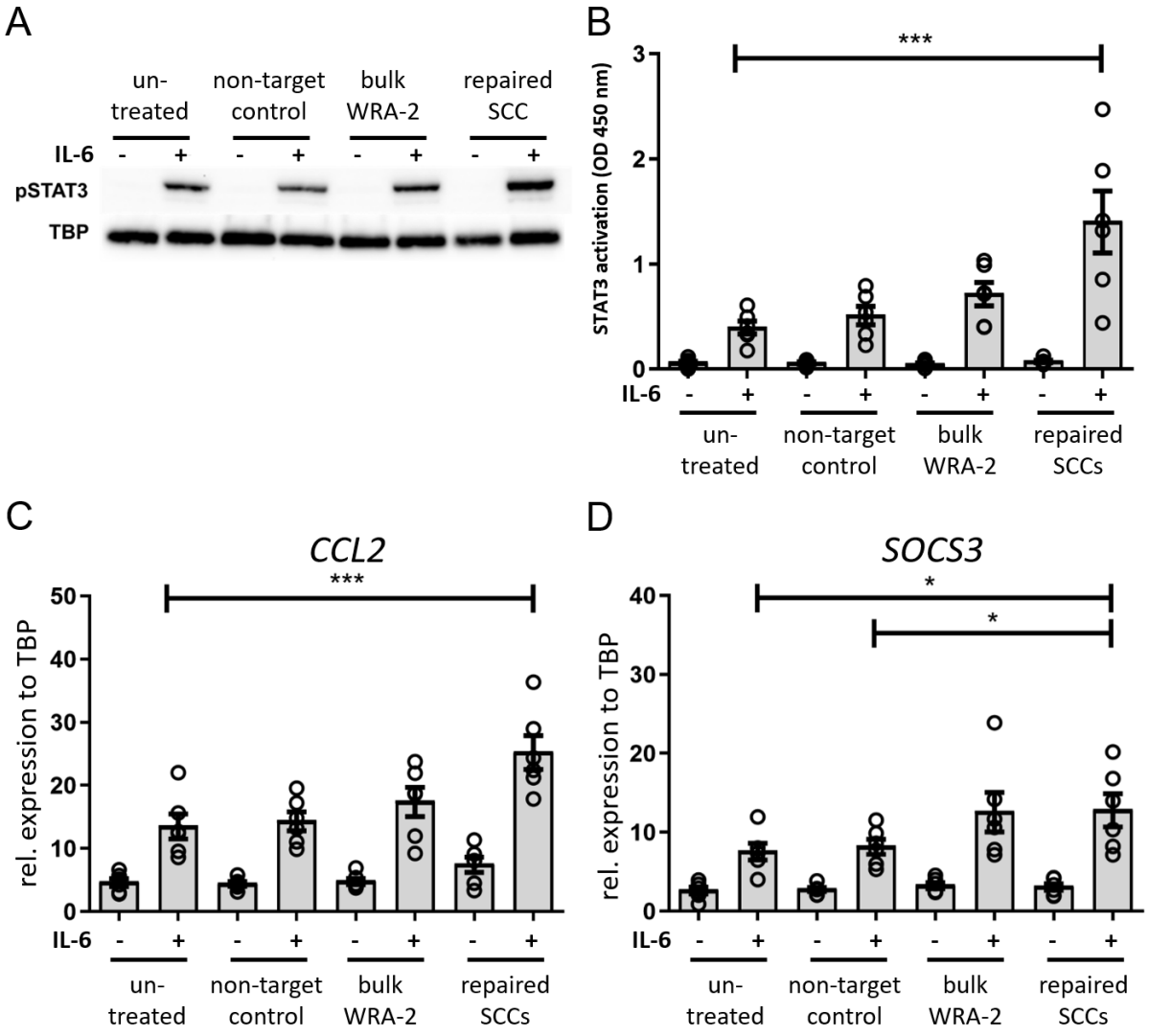

Figure 21: Functional analysis of IL-6 stimulated *STAT3*$^{+/R382W}$ fibroblasts

A: Western blot analysis of nuclear extracts shows increased nuclear translocation of STAT3 phosphorylated at Tyr705. B: TransAM DNA-binding assay shows significantly increased STAT3 DNA-binding activity in repaired SCCs. C and D: Reverse transcription quantitative real-time PCRs show significantly increased expression of STAT3 target genes *CCL2* (C) and *SOCS3* (D) in repaired SCCs compared to untreated *STAT3*$^{+/R382W}$ fibroblasts. Target gene expression was assessed relative to the housekeeping gene TATA-box-binding protein (TBP).
- = unstimulated; + = IL-6 stimulated; IL-6 = interleukin-6; bars indicate mean values; dots individual experiments; error bars represent SEM; significant differences *p <0.05, ***p <0.001 assessed by two-way ANOVA and Bonferroni post-test; n = 6 (B, C, D)
Adapted from Eberherr *et al.* [100]

To relate STAT3 signaling improvements to a wild type control, we compared target gene expression of *CCL2* and *SOCS3* after IL-6 stimulation between primary fibroblasts from several healthy donors (control) and untreated patient cells, bulk treated patient cells and repaired SCCs (Figure 22). Patient cell gene expression of

CCL2 and *SOCS3* was more comparable to healthy donor controls after treatment. Thus we suggest that the treatment leads at least to a partial rescue of STAT3 signaling.

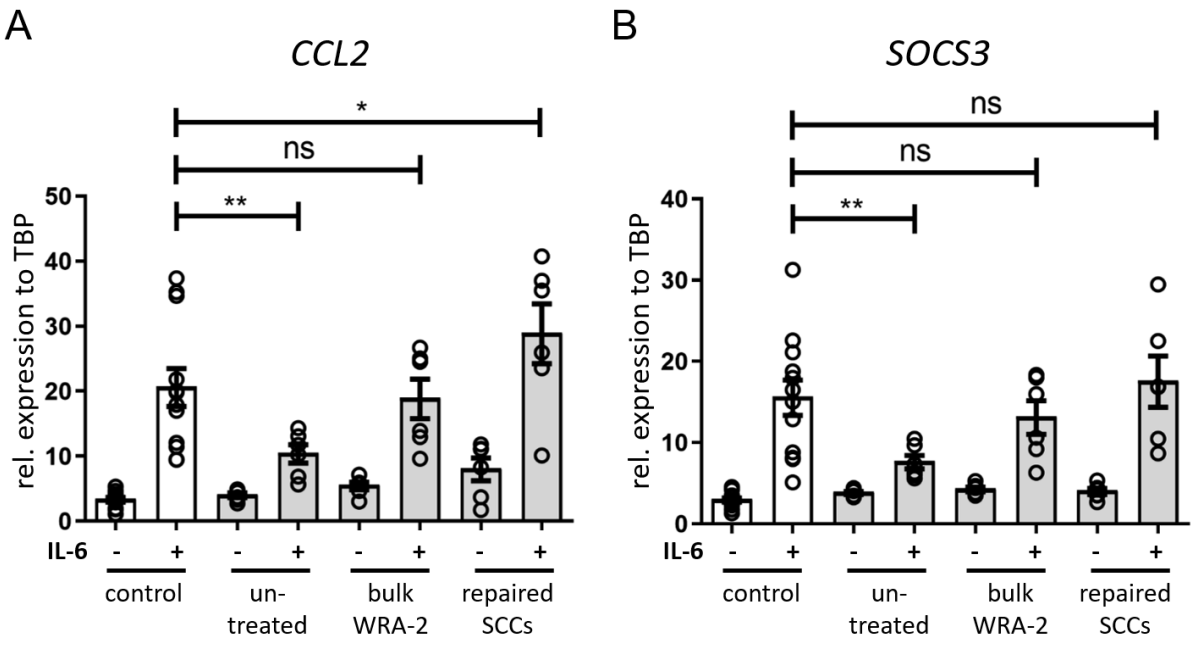

Figure 22: Target gene expression of healthy controls and treated patient cells

A and B: Reverse transcription quantitative real-time PCRs of STAT3 target genes *CCL2* (A) and *SOCS3* (B) show comparable expression in healthy primary control fibroblasts and primary treated *STAT3*$^{+/R382W}$ fibroblasts. Target gene expression was assessed relative to the housekeeping gene TATA-box-binding protein (TBP).
- = unstimulated, + = IL-6 stimulated; IL-6 = interleukin-6; bars indicate mean values; dots individual experiments; error bars represent SEM; significant differences *p <0.05, ***p <0.001 assessed by two-way ANOVA and Bonferroni post-test; n = 6 or 12 (control)
Adapted from Eberherr *et al.* [100]

Since STAT3 is known to promote proliferation we speculated that repaired fibroblasts might proliferate faster than not repaired cells. A proliferative advantage of treated cells would lead to an increasing repaired cell population in a treated organism and might influence the efficacy of the therapeutic benefit of a possible future treatment for STAT3-HIES. To evaluate any possible proliferative advantage of repaired cells, we cultured bulk repaired fibroblasts (initially 46.6% repair efficiency) over 10 passages and used Sanger sequencing to determine if the amount of repaired cells increased with increasing cell culture time. In this experiment ABEmaxAW, a refined version of ABE7.10 with high editing efficiency and a lack of unspecific RNA editing reported for other ABE variants, was used to correct the R382W STAT3 mutation. As shown in

Figure 23 the repaired cells showed no proliferative advantage over unrepaired patient cells after 10 passages.

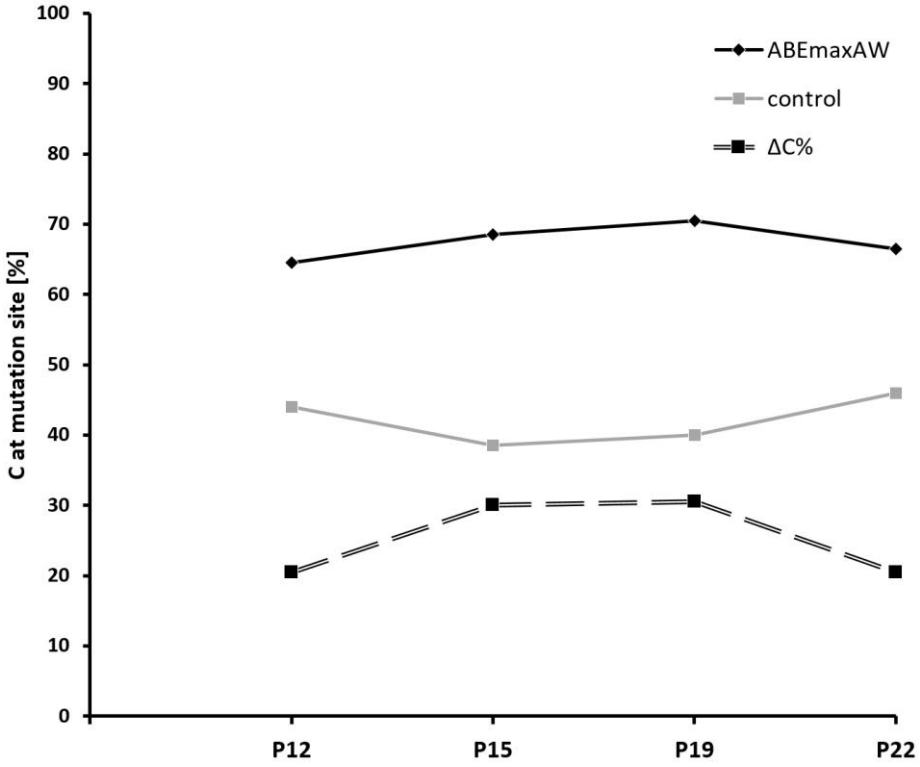

Figure 23: Repaired *STAT3*$^{+/R382W}$ fibroblasts show no proliferative advantage

Primary *STAT3*$^{+/R382W}$ fibroblasts were bulk treated (repair efficiency: 47% C increase) with ABEmaxAW and WRA-2 guideRNA plasmid or control plasmid. Repaired fibroblasts showed no proliferative advantage compared to control treated fibroblasts over a culture period of 10 passages. Black line = ABEmaxAW/WRA-2 treated, grey line = control treated, black dashed double line = ΔC% = C% (WRA) – C% (control)

3.1.8 Improvements to the editing system

The repair efficiency of 29±7% (Sanger sequencing and EditR) or 43.7% (HTS) indicates, that only a minority of the treatment surviving *STAT3*$^{+/R382W}$ fibroblasts is actually repaired. Therefore, we wanted to improve the repair efficiency, as this would facilitate a detection of any treatment effects investigated.

A new version of ABE7.10, ABEmax, was subjected to codon optimization and a change of the nuclear localization signal increasing the editing efficiency. Thus, we used ABEmax to correct primary *STAT3*$^{+/R382W}$ fibroblasts by electroporating the plasmid pCMV_ABEmax (addgene) encoding ABEmax under CMV promoter control and the previously constructed pBS-U6:WRA2 plasmid. As a sample intrinsic

electroporation control, we added 500 ng of GFPmax plasmid (Lonza) to the electroporation substrates.

We observed a substantial increase in the levels of editing achieved in cells treated with ABEmax (60-66%) compared to cells treated with ABE7.10 (34-36%) seven days after electroporation (Figure 24). While the electroporation efficiencies were high in ABE7.10 and ABEmax samples, we observed lower cell viability in ABEmax treated cells (Figure 24 A).

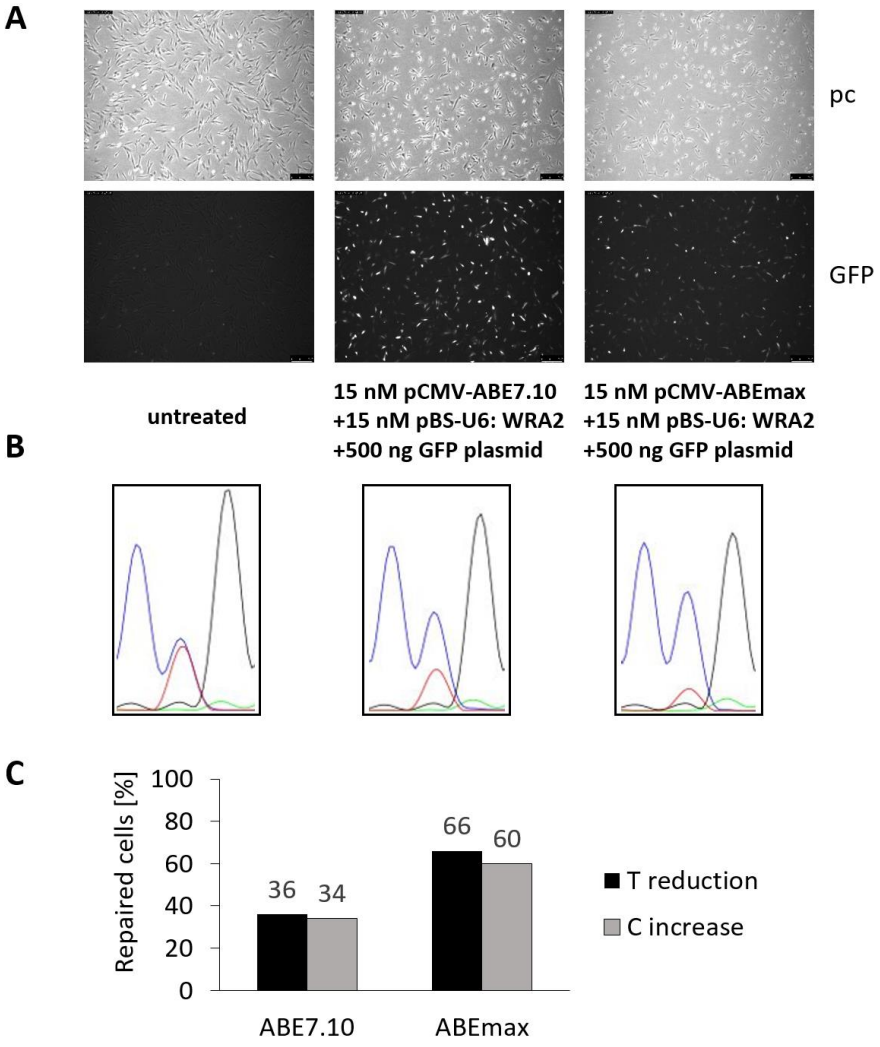

Figure 24: ABEmax improves editing efficiency

A: Cells 24 h after treatment show reduced viability of electroporated cells and a high electroporation efficiency. The lowest viability was observed in ABEmax treated cells. pc = phase contrast; GFP = green fluorescent protein. B: Sanger chromatogram of DNA isolated 7 days after treatment. C: EditR analysis of the obtained Sanger sequencing data. ABEmax treated cells show a higher repair efficiency.

Another approach to improve the editing efficiency was to test several different electroporation programs of the 4D Nucleofector namely CZ-167, ED-130, DT-130, DS-130, CA-201, CA-137, EN-150. After electroporation of 500 ng GFPmax plasmid (Lonza) into primary *STAT3*$^{+/R382W}$ fibroblasts, we observed the highest transfection efficiency with program CZ-167 instead of DT-130, although at the cost of reduced cell viability (Figure 25).

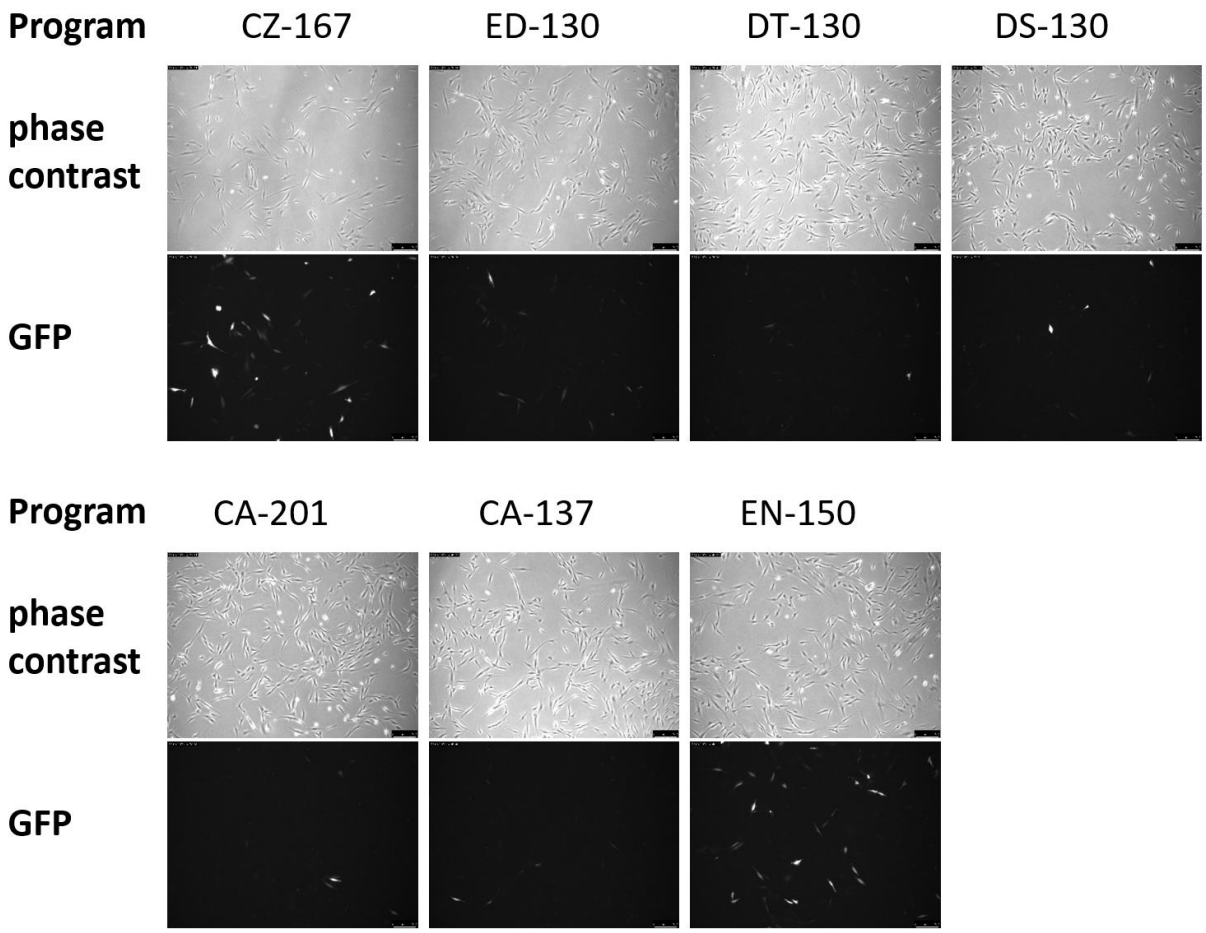

Figure 25: Programs CZ-167 and EN-150 show the highest transfection rates

Seven different electroporation programs (CZ-167, ED-130, DT-130, DS-130, CA-201, CA-137, EN-150) of the 4D Nucleofector X Unit (Lonza) were tested by transfection of 500 ng GFPmax plasmid (Lonza) into *STAT3*$^{+/R382W}$ primary fibroblasts. Programs CZ-167 and EN-150 showed the highest transfection rates as seen by the number of GFP expressing cells 48 h after electroporation.

While previous experiments focused on increasing transfection efficiency, we also wanted to increase flexibility of the editing system by using different substrates. Since RNA meditated editing proved difficult (results not shown) we used ribonucleoprotein (RNP) mediated editing. To find optimal electroporation settings we tested 8 different

programs und two buffers for the electroporation of 600 nM Cas9/WRA-2 RNP into *STAT3*$^{+/R382W}$ primary fibroblasts. As shown in Table 2, program CZ-167 and buffer P2 showed the highest editing rates (39.5%) and were used in any further experiments involving the electroporation of RNPs into primary fibroblasts.

Sample	program	kit	efficiency
#1	CZ-167	P2	39.5%
#2	EN-150	P2	20.5%
#3	DT-130	P2	12.2%
#4	CM-138	P2	9.9%
#5	DS-130	P2	11.1%
#6	CA-201	P2	12.3%
#7	CA-137	P2	0%
#8	ED-130	P2	12.5%
#9	CZ-167	P3	5.3%
#10	EN-150	P3	0.5%
#11	DT-130	P3	0.5%
#12	CM-138	P3	0.4%
#13	DS-130	P3	0.2%
#14	CA-201	P3	17.5%
#15	CA-137	P3	0%
#16	ED-130	P3	1.2%

Table 2: Program CZ-167 and buffer P2 show the highest transfection rate

8 different electroporation programs (CZ-167, ED-130, DT-130, DS-130, CA-201, CA-137, EN-150) of the 4D Nucleofector X Unit (Lonza) and 2 electroporation buffers (P2 and P3, Lonza) were tested by transfection of 600 nM Cas9/WRA-2 RNP into *STAT3*$^{+/R382W}$ primary fibroblasts. Editing rates were determined by Sanger sequencing and subsequent TIDE analysis. n = 1

In our next step we wanted to repair the R382W mutation via the electroporation of ABE7.10/WRA-2 RNPs into affected fibroblasts. Therefore, we ordered the pET42b-ABE7.10 plasmid from addgene for expression of ABE7.10 protein in bacteria and produced and purified the protein in collaboration with Arie Geerlof from the

Helmholtz Center Munich. The obtained ABE7.10 protein was used in two subsequent electroporations with 600 nM ABE7.10/WRA-2 RNP, electroporation program CZ-167, buffer P2 and primary *STAT3*$^{+/R382W}$ fibroblasts (Figure 26). The repair efficiency observed was 38 – 44.6% suggesting that RNPs are a viable alternative to plasmid based correction of the R382W mutation.

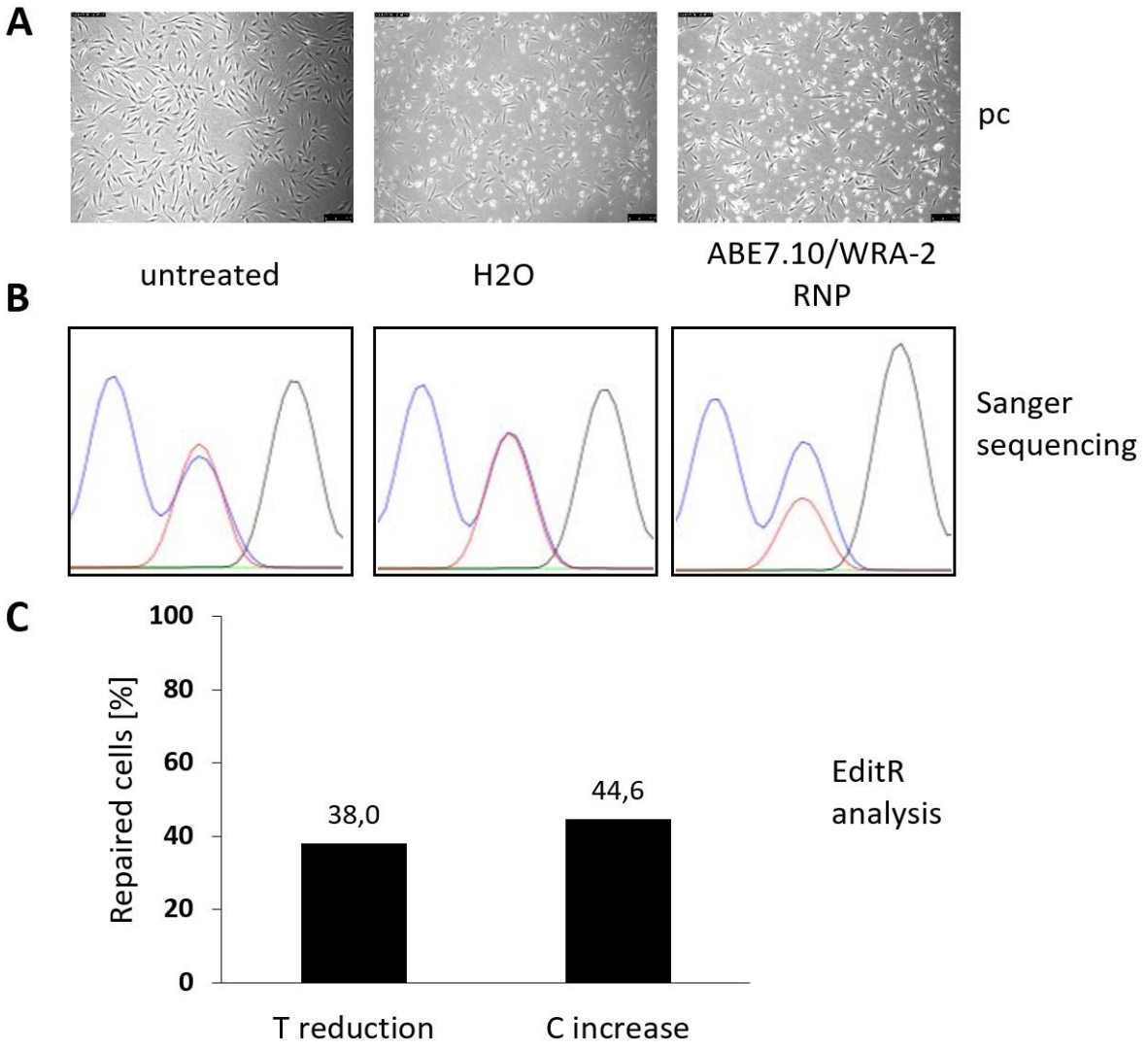

Figure 26: ABE7.10/WRA-2 RNPs show robust repair efficiency

A: Cells 24 h after treatment show reduced viability of electroporated cells. ABE7.10/WRA-2 RNP treated cells show similar viability as H_2O treated cells.
B: Sanger chromatograms prepared 24 h after treatment.
C: EditR analysis of the obtained Sanger sequencing data reveals an editing efficiency of 38 – 44.6%. n = 1

3.2 A treatment approach for the *STAT3* R382Q mutation

3.2.1 R382Q treatment design considerations

For the treatment of the heterozygous *STAT3* hot-spot mutation c.1145G>A/p.R382Q *STAT3* (R382Q) mutation two strategies were considered:

- The specific knock-out of the affected allele to increase levels of functional STAT3 dimers via reduction of the dominant-negative effect of the mutation
- The direct conversion of the mutated base pair (bp) to inosine (I) which is read an replicated as G (wild type) via adenine base editing

While the repair of the mutation via ABE showed very promising results for the c.1144C>T/p.R382W *STAT3* mutation (R382W), in this case the target A is located on the antisense strand and very close to three bystander As that might be edited unintentionally (Figure 27). Furthermore, a conversion of the AAA (lysine) codon to GAA (glutamine), GGA or GGG (both glycine) might not be confined to the R382Q allele as the triple A is also present on the healthy allele and the mismatch at the mutation site might not be enough to prevent editing.

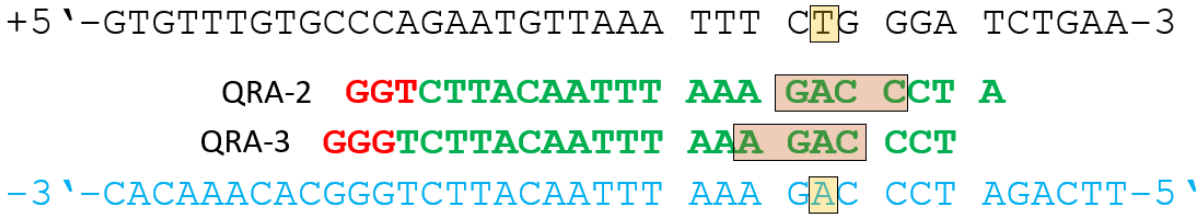

Figure 27: guideRNA designs to correct the *STAT3* R382Q mutation via ABE

Two guideRNAs (QRA-2 and QRA-3) enable ABE mediated correction of the *STAT3* R382Q mutation (yellow boxes). The editing windows (red box) clearly contain the target A but also contain or are very close to three bystander As that might be edited unintentionally. + strand = black; - strand = blue; protospacer sequence = green; PAM = red; R382W mutation = yellow box; editing window = red box

Since mice with a heterozygous knock-out of *STAT3* show no phenotype and a patient hemizygous for *STAT3* presented with only some of the manifestations typical for STAT3-HIES (1.1.2), we speculated that a heterozygous knock-out of the R382Q affected *STAT3* allele might be able to improve STAT3 signaling. With this and the

potential editing of the three bystander As on the healthy allele in mind we focused on the treatment approach utilizing the specific knock-out of the R382Q allele.

We started with the design of an editing system which would allow us the introduction of DSBs and consequently InDels specifically in the R382Q affected *STAT3* allele. InDels in coding sequences have a high probability of inducing frameshifts and consequently abolish expression by nonsense mediated decay of the resulting mRNAs. Since the healthy allele differs from the R382Q allele by only one nucleotide, the healthy allele is likely to be susceptible to off-target editing. To reduce any unintentional editing of the healthy allele we planned to increase the specificity of the editing system.

3.2.2 guideRNA design for a specific knock-out of the affected allele

In order to make the editing as specific as possible we looked for a *S. pyogenes* Cas9 guideRNA (PAM = NGG) which incorporates the mutation site in its seed region (positions 1-10 counted from the PAM) as mismatches in the seed region are known to reduce cleavage efficiency. Of all possible guideRNAs (Figure 28) only guideRNA #4 (QR1) positions the mutation site in the seed region and thus should confer the highest specificity for the affected allele.

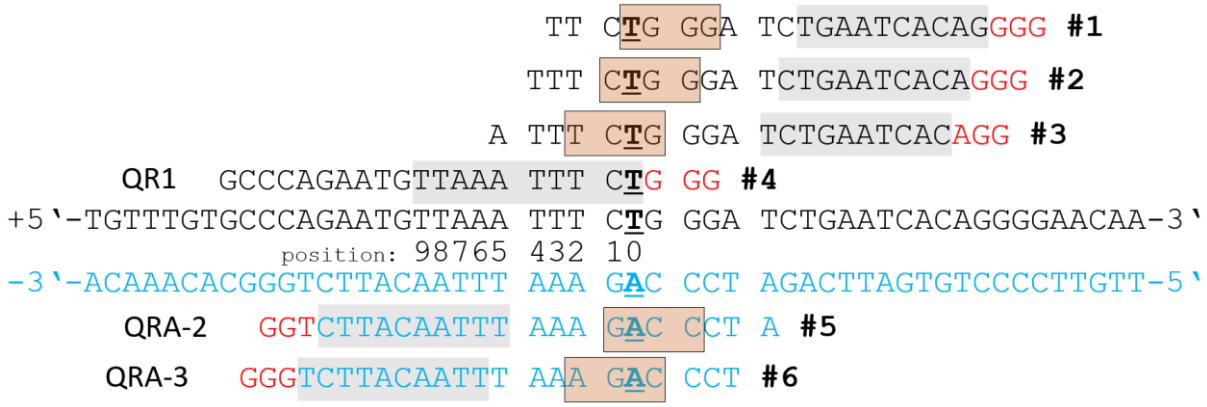

Figure 28: *STAT3* R382Q mutation specific guideRNAs for *S.p.* Cas9

Six different guideRNAs incorporate the mutation site in the protospacer sequence and thus confer specificity for the R382Q mutation. Only #4 positions the mutation site in the seed region of the guideRNA. Bold and underlined = R382Q mutation; grey boxes = seed regions; red boxes = ABE editing windows; red = PAMs

Since a difference of one nucleotide is likely to not confer a high enough specificity, we considered several strategies to increase the specificity further:

- Exchange of the *s.pyogenes* Cas9 for a high fidelity version
- Introduction of artificial mismatches to the healthy allele
- The addition of a short “inactive” guideRNA to shield the healthy allele from being edited (CRISPR GUARD)
- Truncation of the guideRNAs

An overview of the guideRNA dependent changes can be seen in Figure 29. All approaches were tested via Cas9 *in vitro* assays.

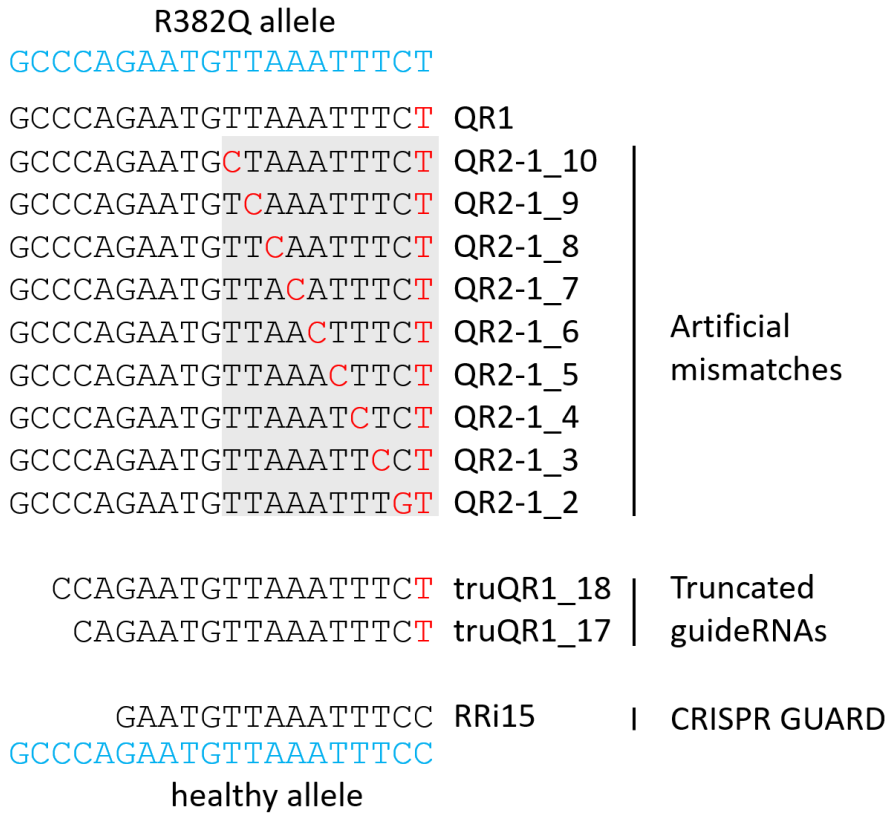

Figure 29: Strategies to increase guideRNA specificity

Specificity of guideRNAs can be increased by artificial mismatches, truncation or the shielding of the wild type allele (CRISPR GUARD). Red = Mismatches of guideRNAs to the healthy allele; grey box = seed region; blue = patient DNA

3.2.3 Specific knock-out guideRNA production and validation

We synthesized the guideRNAs as mentioned above and validated the guideRNAs via Cas9 *in vitro* assays. We speculated that an editing system with sufficient specificity would be able to discriminate between the healthy and the R382Q allele due to the one bp difference. In a first experiment we tested the guideRNA QR1 (one mismatch at position 1 to the healthy allele) with standard Cas9 or high fidelity Cas9 (both from IDT) and tried to cut PCR fragments amplified from patient DNA (*STAT3*$^{+/R382Q}$) or a healthy control (*STAT3*$^{+/+}$). None of the Cas9/QR1 or HF-Cas9/QR1 did show discrimination of the healthy allele as all PCR fragments were cut (Figure 30).

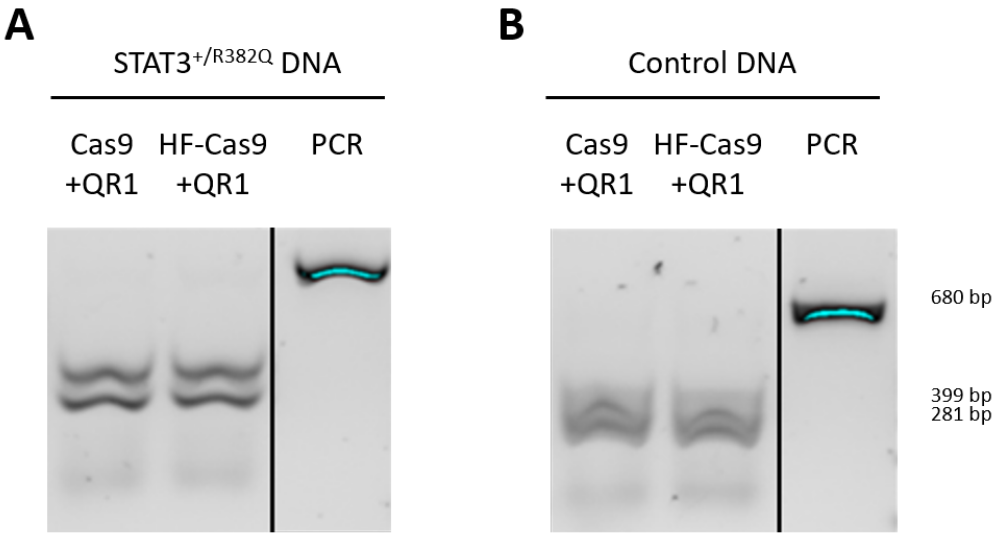

Figure 30: High Fidelity (HF) Cas9 – *in vitro* assay

A Cas9 in vitro assay (10 pmol RNPs : 1 pmol PCR fragments) performed with *s.p.* Cas9 and HF-Cas9, guideRNA QR1 and patient or control DNA shows no increased specificity of HF-Cas9. PCR fragments (680 bp) of the target region (patient and control DNA) are cleaved into 2 fragments (399 and 281 bp) regardless of the Cas9 variant used. A: Cas9 *in vitro* assay with *STAT3*$^{+/R382Q}$ DNA. B: Cas9 *in vitro* assay with control DNA

While also the use of truncated guideRNAs resulted in no discrimination (data not shown) we found varying degrees of healthy allele discrimination once we used guideRNAs containing a second artificial mismatch (Figure 31). Interestingly the specificity of the guideRNA seemed to increase if the mismatch was positioned closer to the PAM (Figure 31A). We attributed this effect to the mismatches starting to be positioned in the core region of guideRNAs. The core region is an especially mismatch

sensitive part of the seed region and ranges from position 4 to 7 counted from the PAM. Further the decrease in distance of the artificial mismatch to the already existing mismatch (only for the healthy allele) might destabilize the guideRNA hybridization and consequently increase specificity additionally.

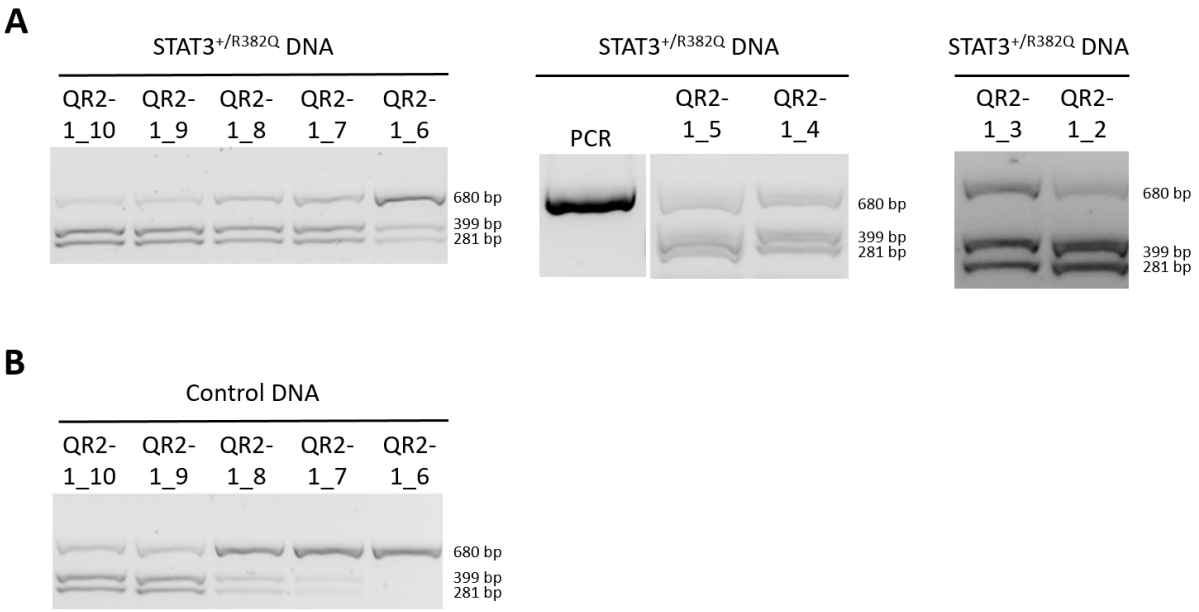

Figure 31: Artificial mismatches increase guideRNA specificity

guideRNAs with an added artificial mismatch show increased specificity as shown by Cas9 *in vitro* assays of the guideRNAs. A: Cas9 *in vitro* assay with *STAT3*$^{+/R382Q}$ DNA. B: Cas9 *in vitro* assay with control DNA.

A final approach to increase specificity and thus reduce editing of the wild type allele was the use of a combination of Cas9/QR1 and Cas9/RRi15 complexes. As shown in 3.2.2 (Figure 29) QR1 binds perfectly to the R382Q allele while RRi15 binds perfectly to the wild type allele. The guideRNA RRi15 is unable to induce Cas9 mediated cleavage due to its short length and should thus be able to shield the wild type allele from being edited by the Cas9/QR1 complex (CRISPR GUARD).

We tested QR1 and QR1+RRi15 with control DNA (*STAT3*$^{+/+}$) and observed that RRi15 abolishes the cutting of the healthy allele by QR1/Cas9 RNPs (Figure 32). Therefore we decided to test the concepts of the artificial mismatch and the CRISPR GUARD in primary patient fibroblasts (*STAT3*$^{+/R382Q}$).

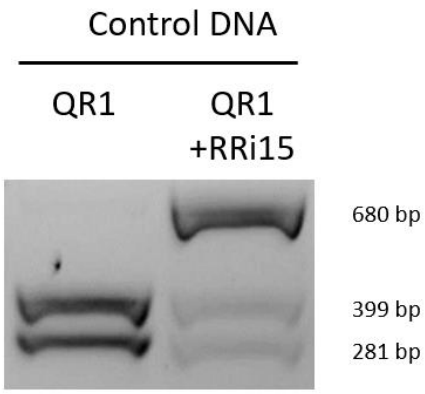

Figure 32: guideRNA RRi15 increases specificity *in vitro*

The due to the reduced length (15 bp) inactive guideRNA RRi15 with 100% homology to the healthy *STAT3* allele increases specificity. A Cas9 *in vitro* assay reveals a strong decrease of control DNA cleavage when using RRi15 in combination with QR1.

3.2.4 *Ex vivo* treatment of primary patient fibroblasts

To test the most promising approaches in a cellular setting we electroporated primary *STAT3*$^{+/R382Q}$ fibroblasts or healthy control fibroblasts (adult human dermal fibroblasts, HDFa, from ATCC) with Cas9 complexed with one of the seven guideRNAs QR2-1_10 to QR2-1_4 or QR1 as reference. QR2-1_3 and QR2-1_2 were only tested in *STAT3*$^{+/R382Q}$ fibroblasts. Sanger sequencing and subsequent *in silico* efficiency analysis using TIDE after the electroporations showed a very low level of editing in all guideRNAs possessing an artificial mismatch (Figure 33) while the QR1 reference showed an editing efficiency of 39.1% in *STAT3*$^{+/R382Q}$ cells, the editing in healthy controls was 8.6% indicating an already very pronounced specific targeting of the R382Q *STAT3* allele.

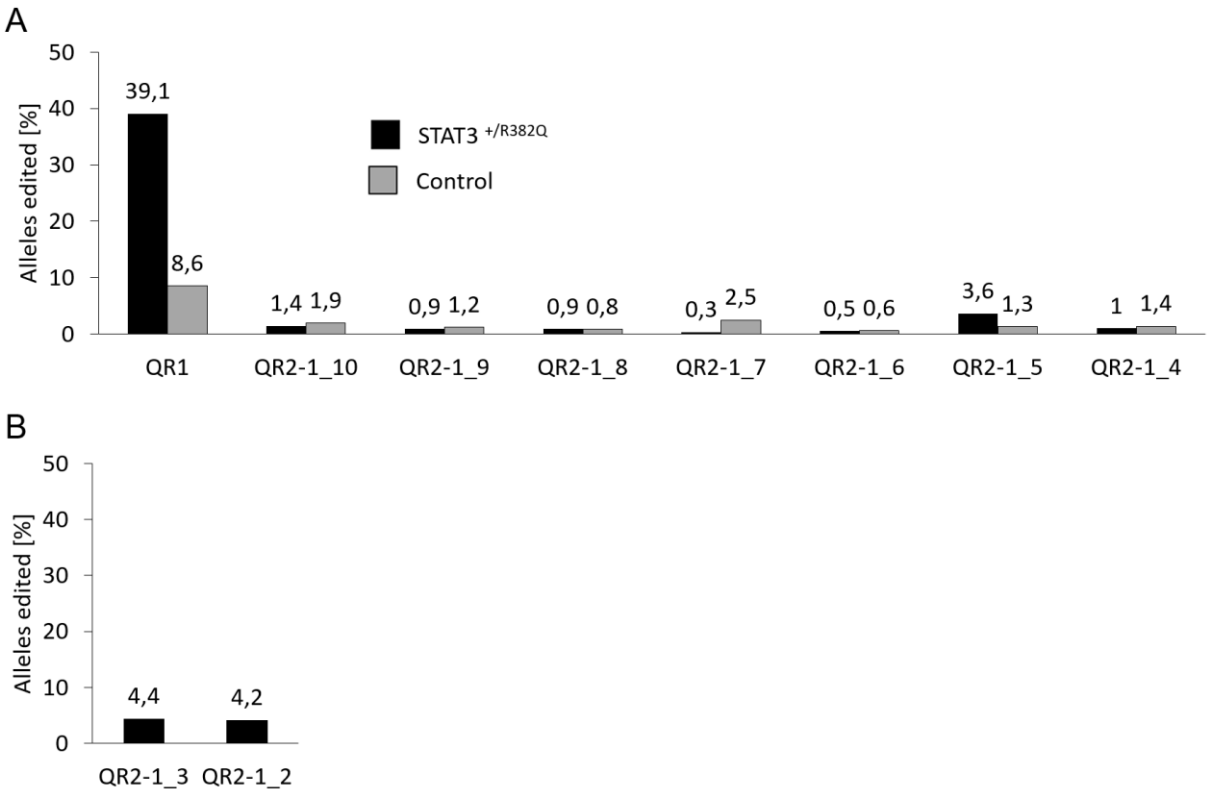

Figure 33: Low editing in patient fibroblasts with artificial mismatch guideRNAs

A: Electroporation of RNPs consisting of *s.p.* Cas9 and different guideRNAs into patient fibroblasts or healthy control fibroblasts show very low editing levels if artificial mismatches were introduced into the guideRNA. B: Editing levels with guideRNAs QR2-1_3 and QR2-1_2 in patient cells are similarly low as editing levels with other artificial mismatch harboring guideRNAs. n = 1

Since the increase in specificity of the guideRNAs achieved by introduction of artificial mismatches reduced the editing efficiency of the guideRNAs too much, we focused on the approach of shielding the healthy allele from editing via a CRISPR GUARD guideRNA which was already successful in the Cas9 *in vitro* assay. First we treated the primary control fibroblast line HDFa with Cas9/QR1 RNPs or Cas9/QR1 and Cas9/RRi15 RNPs and performed Sanger sequencing and subsequent TIDE analysis. In contrast to previous results (Figure 33) we observed lower levels of editing with QR1 this time with 1.4% (Figure 34) and comparably low editing levels with a combination of QR1 and RRi15 RNPs.

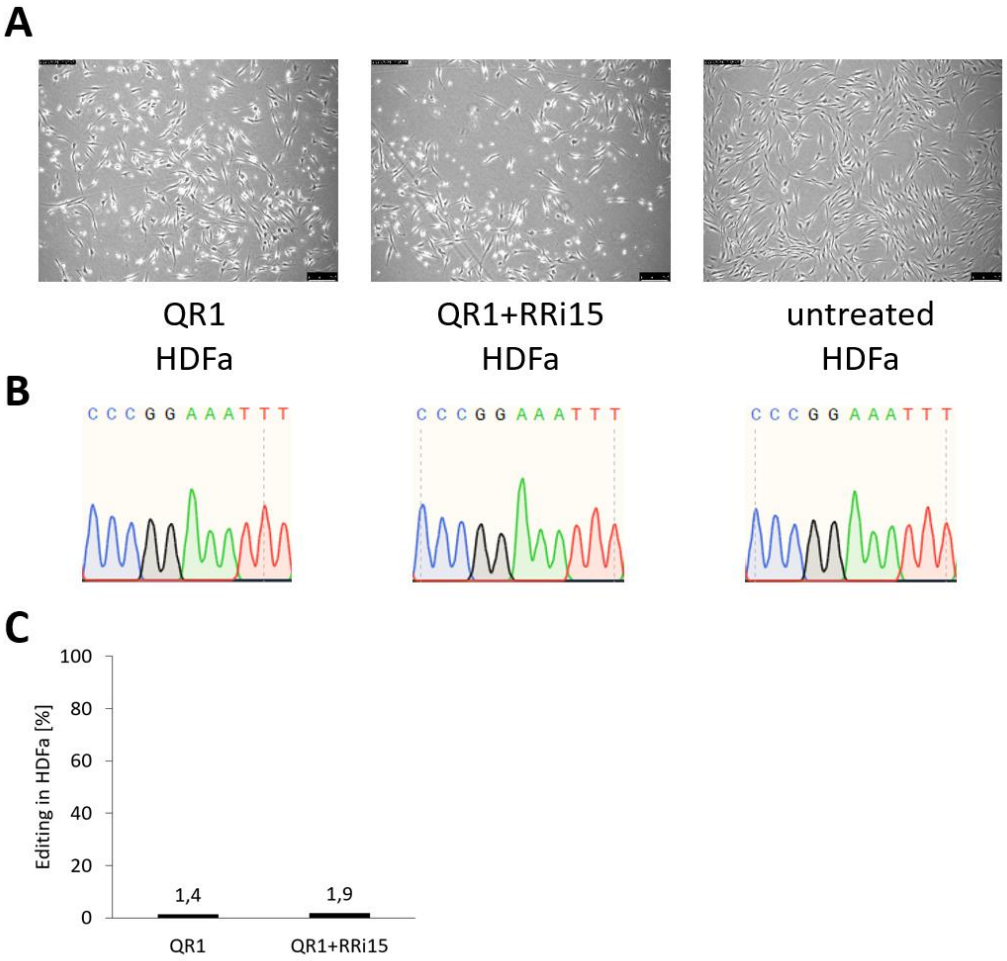

Figure 34: Low editing of QR1 or QR1/RRi15 in healthy control cells (HDFa)

Electroporation of RNPs consisting of *s.p.* Cas9 and the QR1 guideRNA or a combination of QR1 and RRi15 result in low editing levels in healthy control fibroblasts (HDFa). A: Phase contrast microscopy after electroporation. B: Sanger sequencing of the target region. C: Analysis of editing levels via TIDE. n = 1

The next step was to perform the same electroporation with QR1 and QR1/RRi15 RNPs in *STAT3*$^{+/R382Q}$ fibroblasts. With analyzed cell morphology and viability via phase contrast microscopy (Figure 35A) and found both to be comparable to the treated control cells (HDFa). Therefore we attributed the cell death to the electroporation and not the editing process. We observed much higher editing in the treated *STAT3*$^{+/R382Q}$ fibroblasts compared to the control cells as revealed by the aberrant peaks in the chromatogram of the Sanger sequencing (Figure 35B). TIDE analysis determined 38.1% editing with Cas9/QR1 and 47.1% with Cas9/QR1 and Cas9/RRi15 (Figure 35C). Surprisingly the addition of the RRi15 CRISPR GUARD seemed to increase the editing efficiency.

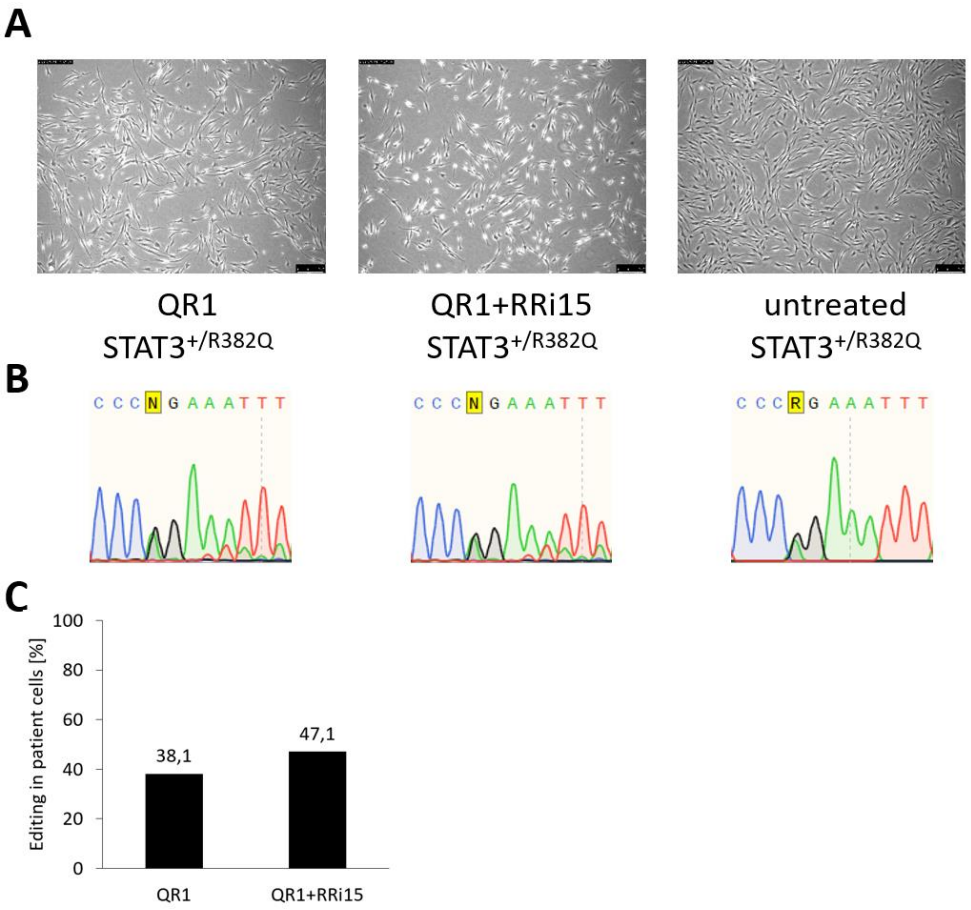

Figure 35: High editing levels of QR1 or QR1/RRi15 in primary patient fibroblasts

Electroporation of RNPs consisting of *s.p.* Cas9 and QR1 or a combination of QR1 and RRi15 results in high editing levels in primary patient fibroblasts (*STAT3*$^{+/R382Q}$). A: Phase contrast microscopy after electroporation. B: Sanger sequencing of the target region. C: Analysis of editing levels via TIDE. n = 1

To analyse the distribution of the editing on the healthy and the R382Q allele we performed HTS of the target region (Figure 36). We observed a drop of healthy allele reads from 41.03% in untreated patient cells to 39.58% in Cas9/QR1 treated and 37.82% in Cas9/QR1+RRi15 treated cells. Unexpectedly the CRISPR GUARD seemed to increase editing of the healthy allele instead of reducing it. Nevertheless, the editing rates of the healthy allele were low with 3.5% for the Cas9/QR1 and 7.8% for the Cas9/QR1+RRi15 sample. The editing efficiency of the R382Q allele on the other hand was extremely high with 78.8% for the Cas9/QR1 and 95.7% for the Cas9/QR1+RRi15 sample.

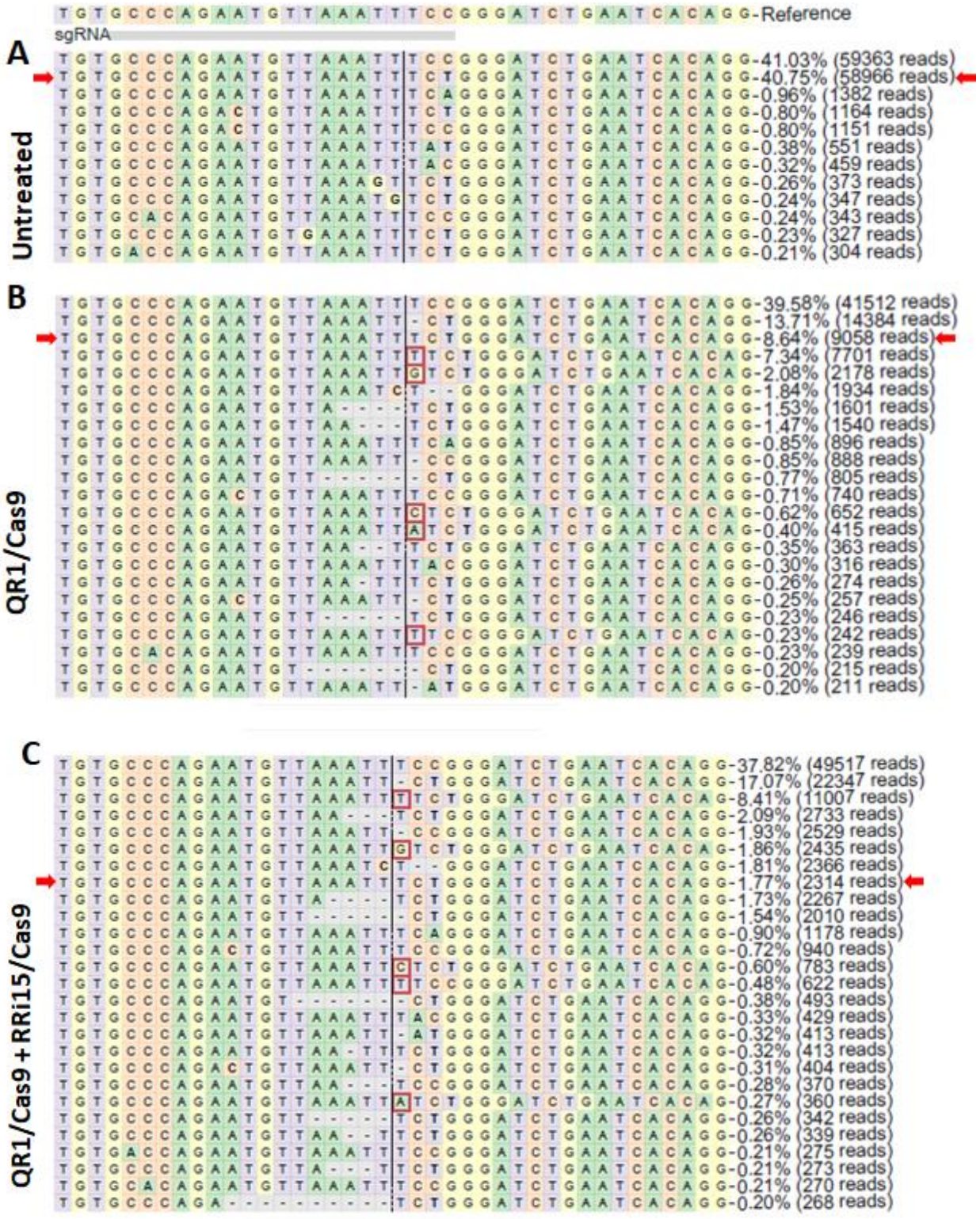

Figure 36: HTS of Cas9/QR1 or Cas9/QR1+RRi15 treated cells

STAT3$^{+/R382Q}$ primary fibroblasts were treated with Cas9/QR1 or Cas9/QR1+RRi15 RNPs and editing outcomes were analyzed via HTS and CRISPResso2. A: Untreated. B: Cas9/QR1 RNP treated. C: Cas9/QR1+ RRi15 treated. HTS cut-off was set to 0.2%, quality filter was set to 10 (phred33 scale). Red arrows = R382Q allele; red squares = insertions; bold letters = substitutions; dashes = deletions; black line = predicted cleavage site

To estimate the true knock-out percentage we removed all reads of events which were already present in the untreated analysis (rare substitutions), a result of editing of the healthy allele and all reads which lead to a frameshift and consequently to nonsense mediated decay of the protein. The left over events (Figure 37) were deletions resulting in no frameshift and are thus likely to still produce STAT3 protein. We added the percentages of the in-frame deletions to the percentages of the R382Q allele and calculated the true knock-out percentage with this number. This resulted in a true knock-out percentage of 73.3% for the Cas9/QR1 treated sample after the following calculation: (40.75 – (8.64+1.47+0.77))/40.75. For the Cas9/QR1+RRi15 treated sample we calculated an estimated knock-out percentage of 85.9% via the following: (40.75-(1.77+2.09+1.54+0.21))/40.75. We concluded that both treatments would induce a knock-out in the vast majority of cells and therefore refrained from picking isolated single cell clones to increase the treatment effect for functional analysis.

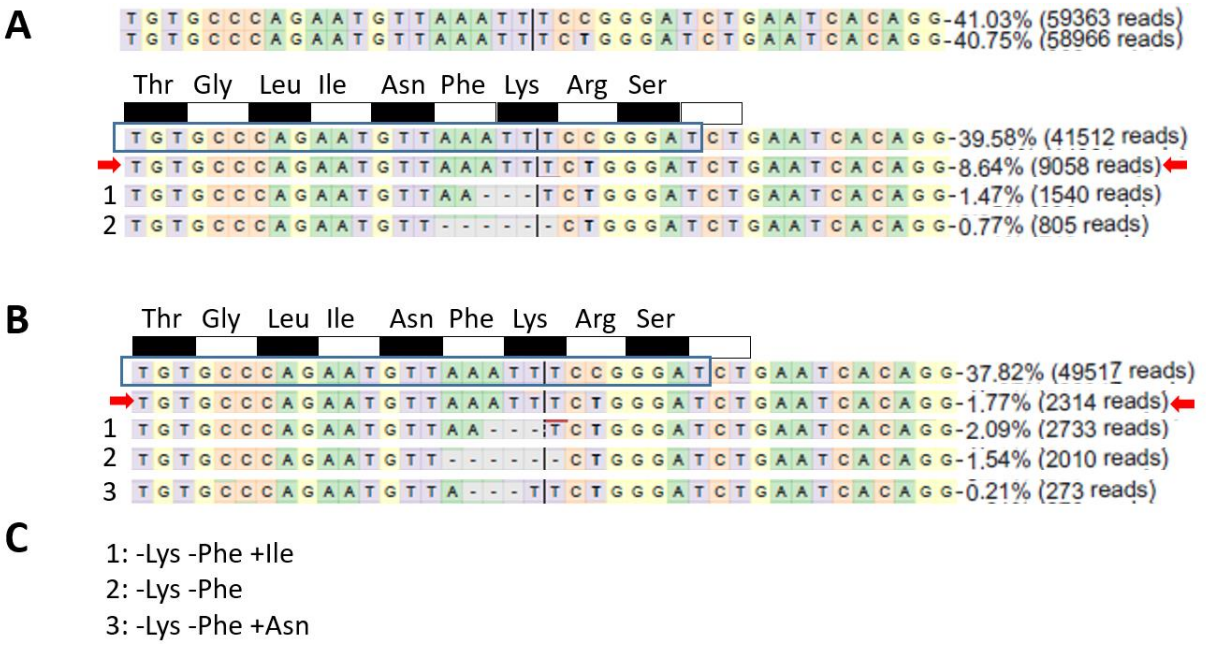

Figure 37: Unfavourable editing outcomes

STAT3$^{+/R382Q}$ primary fibroblasts were treated with Cas9/QR1 or Cas9/QR1+RRi15 RNPs and editing outcomes were analyzed. All editing events on the healthy allele or events leading to frameshifts were eliminated to estimate the true knock-out percentage. A: HTS of the Cas9/QR1 treated sample after elimination. B: HTS of the Cas9/QR1+RRi15 treated sample after elimination. C: Results of the events regarding the amino acid sequence. Red arrows = R382Q allele; black and white bars = codons; blue frames = part of exon 13

3.2.5 Off-target prediction for Cas9/QR1

Before any functional analysis of the Cas9/QR1 RNP or Cas9/QR1+RRi15 RNP treated patient fibroblasts we performed an *in silico* prediction with the online tool CRISPOR to check for any potential off-targets influencing the functional analysis. QR1 was predicted to cut at no off-target positions classified as exonic with less than 4 mismatches aside from the healthy *STAT3* allele (Table 3) which we had checked for editing via HTS. Due to the prediction we concluded that off-target editing is likely negligible when using Cas9/QR1 or Cas9/QR1+RRi15 RNP treated cells for functional analyses.

Locus	Mismatches
exon:STAT3	1
exon:SMC4/TRIM59/RP11-432B6.3	4
exon:VWA5B2	4
exon:RWDD1	4
exon:ESYT3	4
exon:JADE1/SCLT1	4
exon:RAB17	4

Table 3: CRISPOR off-target prediction of the QR1 guideRNA

CRISPOR *in silico* exonic off-target prediction for the QR1 guideRNA. Only six exonic off-targets with four or less mismatches were predicted aside from the healthy *STAT3* allele.

3.2.6 Analyses of STAT3 function in treated patient fibroblasts

Knowing that a large percentage of patient fibroblasts treated with Cas9/QR1 or Cas9/QR1+RRi15 RNPs (see above; 73.3 and 85.9%) had their R382Q affected *STAT3* allele knocked-out, we went on to test if the treatment would increase STAT3 target gene expression similarly to the repair of patient fibroblasts affected by the R382W mutation (3.1.7). We stimulated treated and untreated cells for 1 h with IL-6 to

activate STAT3 signaling and measured STAT3 target gene expression of *SOCS3* and *CCL2* by reverse transcription quantitative real-time PCR. We observed only minor increases of *SOCS3* expression and minimal increases in *CCL2* expression (Figure 38).

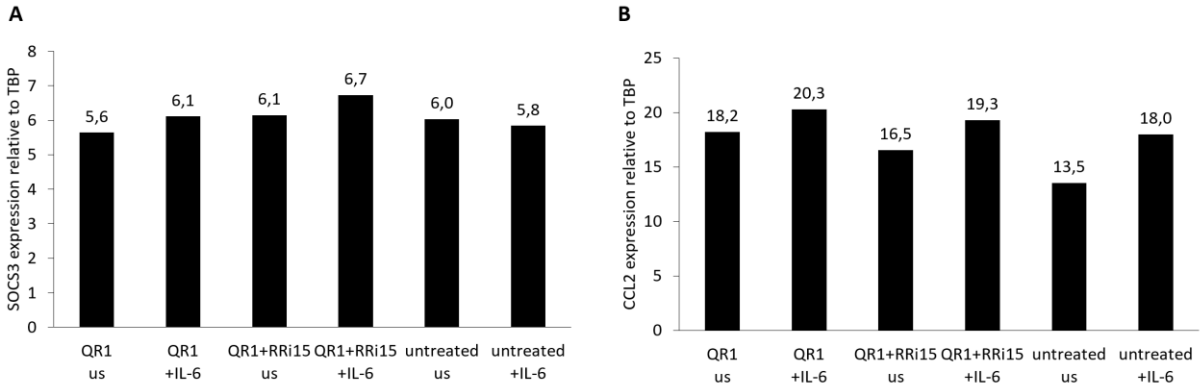

Figure 38: Treatment shows no effect on *SOCS3* and *CCL2* expression

Patient fibroblasts (*STAT3*$^{+/R382Q}$) treated with Cas9/QR1 or Cas9/QR1+RRi15 were stimulated with IL-6 to induce expression of STAT3 target genes *SOCS3* and *CCL2*. A: *SOCS3* expression relative to TBP. B: *CCL2* expression relative to TBP.

3.2.7 ABE mediated repair of the R382Q mutation

We tested the ABE mediated repair of the R382Q mutation before continuing with the Cas9/QR1 or Cas9/QR1+ RRi15 treatment. Therefore, we *in vitro* transcribed the guideRNAs QRA-2 and QRA-3 which were not chosen initially for treatment due to three bystander As close to the ABE7.10 editing window (3.2.1). To confirm functionality we performed Cas9 *in vitro* assays. As seen in Figure 39 both guideRNAs showed cleavage of PCR amplicons of the target region.

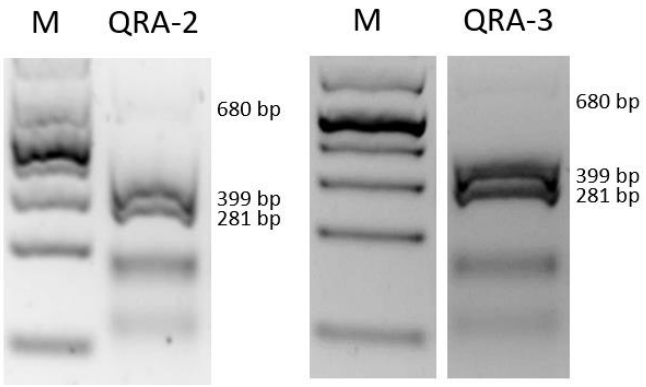

Figure 39: Cas9 *in vitro* assays for QRA-2 and QRA-3

guideRNAs QRA-2 and QRA-3 were validated via the ability to cut PCR fragments (680 bp) of the target region when complexed with Cas9 protein. Both guideRNAs cut the PCR fragments resulting in two smaller bands of 399 and 281 bp.

We continued with the electroporation of Cas9/QRA-2 or ABE7.10/QRA-2 RNPs into *STAT3*$^{+/R382Q}$ fibroblasts. We found no or only very few aberrant peaks in the chromatogram of the target region after Sanger sequencing (Figure 40A) of two independent electroporations with Cas9/QRA.2 RNPs. Further we found no editing activity of the ABE7.10/QRA-2 RNPs (Figure 40B).

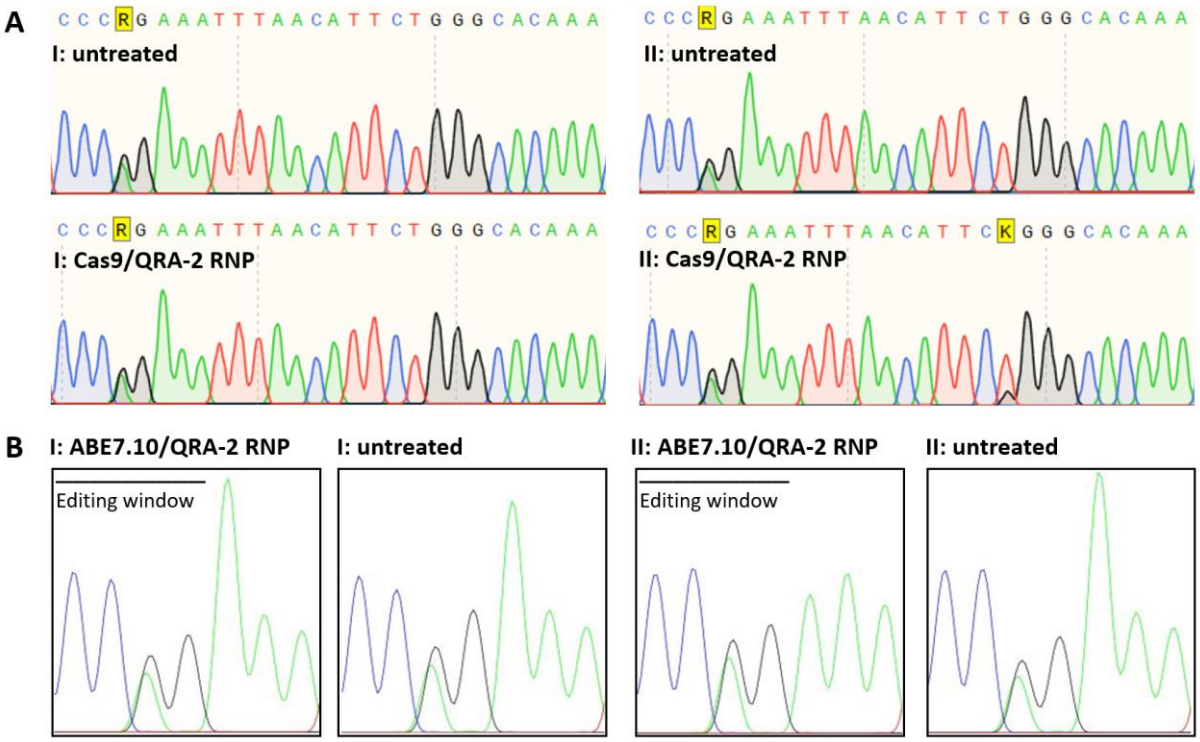

Figure 40: Cas9/QRA-2 or ABE7.10/QRA-2 RNPs transfected patient fibroblasts

Cas9/QRA-2 or ABE7.10/QRA-2 RNPs were electroporated into primary *STAT3*$^{+/R382Q}$ fibroblasts. A: Sanger sequencing of the target region in Cas9/QRA-2 treated cells. B: Sanger sequencing of the ABE7.10 editing window in ABE7.10/QRA-2 treated cells. I = 1st experiment; II = 2nd experiment

We continued with the test of guideRNA QRA-3 for the induction of DSBs and the correction of the R382Q mutation. We conducted two independent experiments during which we electroporated Cas9/QRA-3 or ABE7.10/QRA-3 RNPs into *STAT3*$^{+/R382Q}$ fibroblasts. Interestingly we found aberrant peaks in the chromatogram of Cas9/QRA-3 treated cells (Figure 41A) indicating the induction of double strand breaks. Further, we also found a reduction of the peak representing the mutation in ABE7.10/QRA-3 RNP treated cells indicating that a certain percentage of cells had their heterozygous R382Q *STAT3* mutation repaired (Figure 41B). Additionally, no editing of the bystander A in the ABE7.10 editing window (Figure 41B) was visible in the chromatogram.

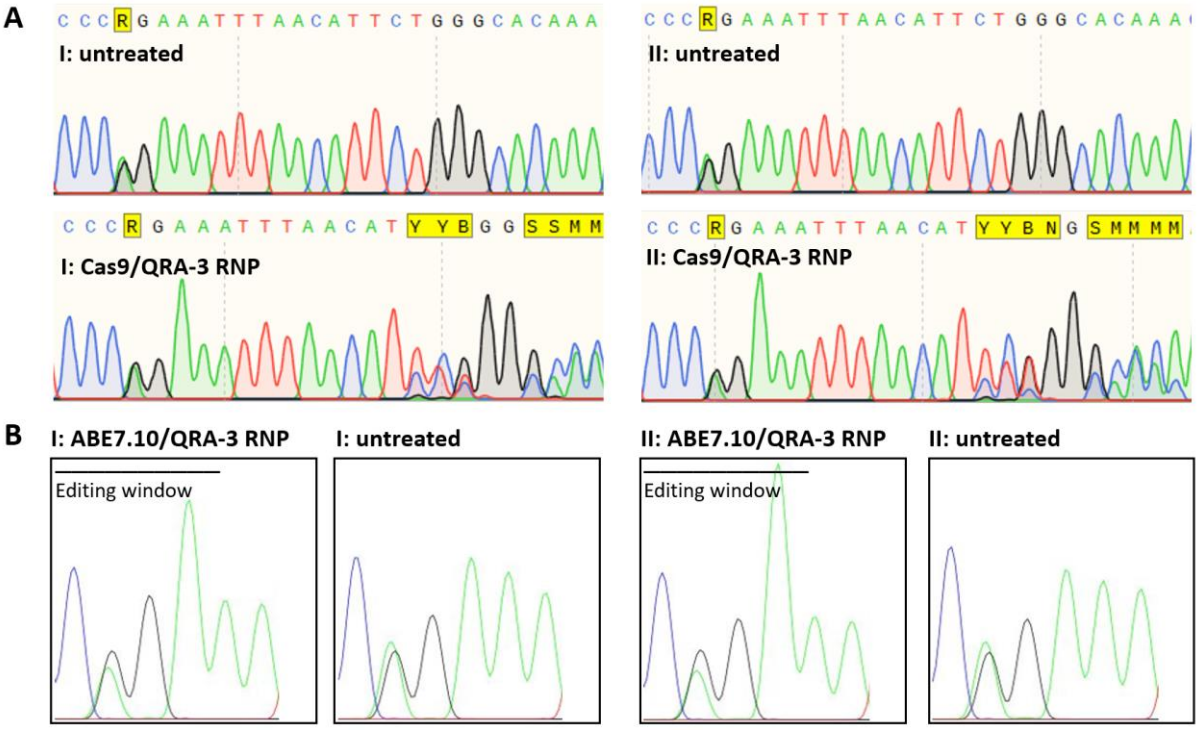

Figure 41: Cas9/QRA-3 or ABE7.10/ QRA-3 RNPs transfected patient fibroblasts

Cas9/QRA-3 or ABE7.10/QRA-3 RNPs were electroporated into primary *STAT3*$^{+/R382Q}$ fibroblasts. A: Sanger sequencing of the target region in Cas9/QRA-3 treated cells B. Sanger sequencing of the ABE7.10 editing window in ABE7.10/QRA-3 treated cells. A reduction of the green peak (of the double peak) representing the mutation (A) compared to untreated cells indicates that a part of the patient cells were repaired. I = 1st experiment; II = 2nd experiment

We quantified the results of the Sanger sequencing of QRA-2 and QRA-3 dependent treatments with TIDE and EditR (Figure 42). The guideRNA QRA-2 seemed to be inactive in cells despite showing the ability to cleave PCR amplicons *in vitro*. Interestingly, we observed a robust activity for guideRNA QRA-3 with 38.7% to 47.9% cutting efficiency and an ABE7.10 mediated repair efficiency of roughly 20%. With this result we speculate that the use of the ABE7.10/QRA-3 editing system might be the better treatment approach compared to the specific knock-out of the affected allele as the main reason against it - the editing of bystander As - seemed to not actually take place. Further, advantages of an ABE mediated repair of the mutation such as the lack of DSBs, the expected higher efficiency in STAT3 signaling improvement and less editing of the healthy *STAT3* allele remain. Nevertheless, a thorough investigation of ABE7.10/QRA-3 treated cells via HTS and functional experiments would need to be conducted to confirm this hypothesis.

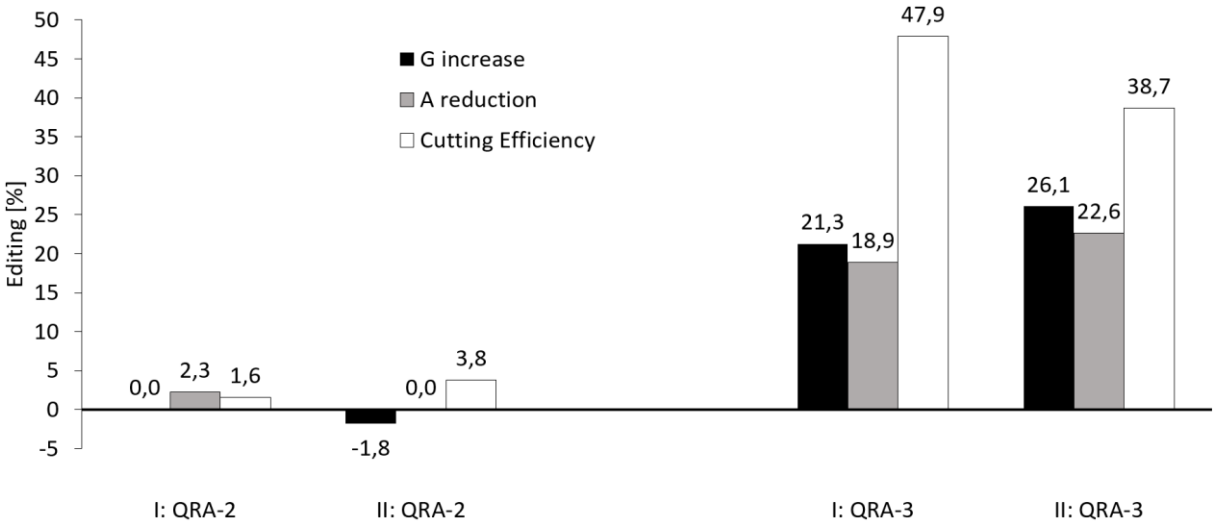

Figure 42: ABE7.10/QRA-2 and ABE7.10/QRA-3 RNP editing patient fibroblasts

Sanger sequencing of the target region and subsequent analyses via TIDE and EditR show close to zero editing for QRA-2 and a robust cutting and repair efficiency for QRA-3.

4. Discussion

With the discovery and development of *Streptococcus pyogenes* Cas9 as a programmable DNA editing tool in mammalian cells [42] and the subsequent generation of the first adenine base editor [65] a means to edit DNA with unprecedented precision was created. Since these tools had already been used to correct pathogenic mutations [65] we hypothesized that gene editing might be able to correct the heterozygous mutations underlying autosomal dominant STAT3-HIES and provide a curative therapy approach in the future. Since the *STAT3* R382W and R382Q heterozygous point mutations accounted for the majority of our STAT3-HIES patients, we focused to find a treatment approach for these first.

4.1 Efficiency of gene editing in primary fibroblasts

We achieved an editing efficiency of ~ 30% via plasmid encoded ABE7.10 and WRA-2 guideRNA (3.1.5) combined with the isolation of single cell clones was sufficient to provide a proof-of-concept for the treatment of the *STAT3* R382W mutation, higher efficiencies could be achieved with the exchange of ABE7.10 for ABEmax, a refined version of the ABE7.10 system (3.1.8.). The changes to the editing system included the replacement of the SV40 NLS in ABE7.10 with a bis-bpNLS and a different form of codon optimization. When we used ABEmax and guideRNA WRA-2 in primary patient fibroblasts, a repair efficiency of ~ 65% was measured indicating a substantial increase in editing efficiency compared to ABE7.10. Further changes to increase the editing efficiency in primary *STAT3*$^{+/R382W}$ fibroblasts included the change of the electroporation program from DT-130 (Lonza 4D Nucleofector – X Unit) to CZ-167 (3.1.8). The increase of editing efficiency due to these changes might lead to significantly improved STAT3 signaling in bulk treated cells (compared to untreated cells) which did not improve significantly in 3.1.7 likely due to too low amounts of repaired cells.

We observed that optimization of the substrate encoding (DNA, RNA, protein) and amount, type of effector enzyme and mode of delivery contributes to a higher editing

efficiency. Nevertheless, a major part of the editing efficiency is determined by the protospacer sequence specific efficiency of the guideRNA. This becomes apparent when comparing guideRNAs QRA-2 and QRA-3. Although the guideRNAs differ only by one bp, the cutting efficiency of QRA-2 (no editing) is substantially lower than the efficiency of QRA-3 (43.3%). This highlights the importance of testing several guideRNAs for their efficiency if more than one guideRNA can be used for the desired effect.

4.2 guideRNA synthesis and quality

We used a plasmid encoding guideRNAs under U6 promoter control or guideRNAs that were *in vitro* transcribed via a T7 promoter. Both promoters (U6 and T7) have the disadvantage that a 5' G is added to the 5' end of the guideRNA after synthesis if none is already present potentially leading to a mismatch at position 21. Nevertheless, we observed efficient editing despite the addition of mismatch producing 5' Gs. guideRNAs with an added 5' G used in this study are WRA-2 (20 – 65% repair efficiency of *STAT3* R382W) and QRA-3 (20% repair efficiency of *STAT3* R382Q). guideRNAs with no mismatch at position 21 are QR-1 (up to 95% cutting efficiency) and QRA-2 (2.7% cutting efficiency).
In regards to their efficiency especially the difference between the guideRNAs QRA-2 (~0% repair efficiency) and QRA-3 (~20% repair efficiency) was striking. In light of this result, a test of the two guideRNAs WRA-3 and WRA-4 (3.1.1) which are theoretically able to repair the *STAT3* R382W mutation via ABE besides WRA-2 seems promising. While the guideRNAs were initially excluded from the study due to a bystander A close to the editing window of ABE7.10 they might result in a higher repair efficiency than WRA-2.

In vitro transcribed guideRNAs were measured via NanoDrop to ensure sufficient purity (260/280 nm and 260/230 nm ratios) while RNA degradation was ruled out by gel electrophoresis (sharp bands of roughly the correct size). As a first measurement of efficiency we tested our guideRNAs via a Cas9 *in vitro* assay. While the Cas9 *in vitro* assays shown in this study were valuable to prove that the designed and manufactured guideRNAs were able to cleave the intended target region, they failed to predict a

sufficient activity in patient fibroblasts. The molar ratio of RNP to template used was 10 pmol : 1 pmol resulting in cleavage of the target template after incubation at 37 °C for 1 h. This was also true for the guideRNAs designed to increase specificity for the R382Q mutation. Nevertheless, many of these guideRNAs did not show editing when electroporated into patient fibroblasts (chapter 3.2.3). This indicates that the electroporation scenario was only insufficiently simulated by the used conditions making an adjustment of the protocol (reduction of RNP amount, increase of template amount or reduction of incubation time) necessary, if a usage of the assay as efficiency prediction for cellular editing additionally to the guideRNA quality control function is desired.

4.3 Genomic safety of the editing systems

The requirements for an editing system to treat the STAT3-HIES causing heterozygous *STAT3* R382W and R382Q mutations were an extremely high specificity in combination with a high efficiency. The extremely high specificity was necessary to prevent editing of off-target sites on the one hand and to prevent editing of the *STAT3* wild type allele which differs by only one bp from the mutated *STAT3* allele. A high enough efficiency was necessary for functional analysis of bulk treated cell populations, since untreated cells would always reduce any treatment effects in such a sample.

Classical CRISPR/Cas9 approaches such as the CORRECT method [48] rely on *Streptococcus pyogenes* Cas9 induced DSBs and subsequently by NHEJ generated InDels for a targeted knock-out and a DNA template and HDR to perform knock-ins or sequence substitutions. Undesired effects of classical CRISPR/Cas9 editing include the induction of a p53-mediated DNA damage response due to the generated DSBs [101, 102] and unexpected large chromosomal deletions or genomic rearrangements [76].

We used the ABE7.10 editing system in this study to achieve a functional STAT3 signaling improvement after repair of the R382W mutation. The system consists of a single guideRNA complexed with a fusion protein of a Cas9 nickase and an *E. coli.* TadA/TadA* heterodimer for the adenine base editing of A·T to G·C base pairs. The

TadA/TadA* heterodimer functions as deaminase and was laboratory evolved to accept DNA as target substrate [65]. Since the ABE7.10 editing system does not employ the NHEJ or HDR repair mechanism but rather base excision repair (BER) or mismatch repair (MMR) and generates DSBs in only minimal amounts, none of the mentioned classical CRISPR/Cas9 problems were reported for ABE7.10 [65, 68]. When a research group investigated guideRNA dependent off-target sites of Cas9 and ABE7.10 for target sites which were edited with on average 54% efficiency (InDels for Cas9 and substitutions for ABE7.10) by both editings systems, ABE7.10 was less prone to off-target genome modifications [65]. Nevertheless, several undesired effects of ABE7.10 mediated editing were reported. These effects include transcriptome-wide off-target deamination and unexpected nucleotide conversions [66, 67, 103-105].

As first step to ensure genomic safety of our guideRNAs WRA-2 and QR1 we predicted off-target sites via CRISPOR and found exonic off-targets to show at least four mismatches (except the *STAT3* target locus). Four mismatches to the guideRNA sequence abolish editing in CRISPR/Cas9 associated gene editing systems in the vast majority of cases [75, 77].
To further evaluate the safety of ABE7.10 mediated repair of the *STAT3* R382W mutation on a genomic level, we performed WGS and HTS of predicted exonic off-target sites (3.1.6). While genome wide detection of rare events was technically not possible despite an increased WGS coverage of 70x, we did not detect an increase in genome-wide A to G conversions or InDels in treated cells. Further we did not detect any plasmid integration as has been reported before [106].
The HTS of predicted off-target sites did not reveal any off-target editing. This was expected since the exonic off-targets sites all showed four mismatches to the guideRNA sequence resulting in a very low probability for hybridization. Further, HTS of the on-target site showed potential side effects of the treatment. Specifically, certain single nucleotide variants were detected that do not represent the wildtype or R382W allele (3.1.6, Table 1, #3 - #8). Several potential explanations for these variants come to mind:

1) All variants are the consequence of natural mutations in this region. This hypothesis is supported by the fact that all variants were already detected in the untreated sample albeit at lower frequencies. Further with a maximal frequency

of 0.79% for one variant (#3) the general frequency of the variants is low. Since cells had to be cultivated for several passages to acquire enough genomic DNA for subsequent analyses such as WGS and HTS, certain differences between untreated and treated cells are to be expected.

2) The variants were caused by the treatment with ABE7.10. Given the positioning of variants #3 - #7 close to the canonical editing window of the ABE7.10 protein this might be a possible explanation. This seems even more likely since variants #3, #5 and #6 are the result of A to G conversions which is the intended conversion mediated by ABE7.10. Variant #4 is the result of a G to C conversion, which is not a conversion generally mediated by ABE7.10 as shown by product purity analyses of the laboratory which created the ABE7.10 [65]. The same holds true for variant #7 which is the result of a C to T conversion. Variant #8 is relatively far away from the editing window of ABE7.10 in the PAM of the WRA-2 sgRNA. Although the conversion is from A to G and thus the ABE7.10 mediated conversion, an influence of ABE7.10 is unlikely as the PAM region is either double stranded DNA or not accessible for the ABE7.10 deaminase due to the bound Cas9-nickase during the editing process (the ABE7.10 deaminase requires single stranded DNA as substrate). Since ABE7.10 creates a DNA nick three nucleotides from the PAM, a failed repair by the cell could cause certain variants. But, as all variants are at least three nucleotides away from the nick site, this explanation is unlikely.

3) The variants were enriched due to a positive selection process of the cells in which variants in the region of the sgRNA conferred protection against the editing system via mismatches to the sgRNA and thus prevent DNA damage. While this scenario could explain variants in the sgRNA region becoming enriched after treatment, it would be likely for mutations to accumulate in the seed region as this would confer better protection.

Taken together further investigation on the origin of the variants would be necessary to determine their cause. Studies employing ABEmax instead of ABE7.10 might show an increased base editing efficiency and thus likely increased amounts of side-effects at the target site, if the variants are caused by ABE7.10. The analysis of additional cell

populations might also help to determine the cause of the detected variants. Ideally these cell populations should be analyzed directly after treatment without longer cell culture periods to minimize the amount of naturally occurring mutations.

4.4 Functional assays to assess STAT3 signaling in human primary fibroblasts

In STAT3-HIES the mutated STAT3 protein causes a dominant-negative effect on STAT3 signaling. To confirm a benefit for treated primary patient fibroblasts (*STAT3*$^{+/R382W}$), we analysed STAT3 signaling via STAT3 DNA binding ELISA (TransAM) and reverse transcription quantitative real-time PCR of STAT3 target genes. The Analysis via TransAM confirmed that an increased amount of STAT3 dimers in nuclear extracts of stimulated patient fibroblasts bind to the DNA binding sequence motif of STAT3 after treatment. This indicates a reduction of the *STAT3* R382W mutation mediated negative effect on STAT3 DNA binding. Additionally, reverse transcription quantitative real-time PCR analysis confirmed that the expression of the STAT3 target genes *SOCS3* and *CCL2* increased in stimulated patient fibroblasts after treatment. When we compared STAT3 target gene expression of bulk treated *STAT3*$^{+/R382W}$ fibroblasts and repaired single cell clones to several healthy donors, we found expression levels to be more comparable after treatment (3.1.7; Figure 22). We concluded that the treatment resulted in a functional improvement of STAT3 signaling in patient cells.

In contrast, a functional analysis of *STAT3*$^{+/R382Q}$ primary patient fibroblasts after a highly efficient and specific knock-out of the affected allele did not show increased *SOCS3* and *CCL2* expression after IL-6 stimulation (3.2.6). Although, the experiment was only conducted once and further investigations would be needed to assess the effects of a specific knock-out. A first interpretation of this result would be that the heterozygous knock-out of the affected *STAT3* allele while increasing the amounts of functional STAT3 dimers still ends in a form of STAT3 haploinsufficiency in which absolute levels of functional STAT3 dimer are still reduced to critical levels (50% and below) and IL-6 signaling is similarly affected as in cells with a heterozygous STAT3

R382Q mutation (~25% functional dimer). Two research groups suggest such a STAT3 signaling model [33, 107]. Interestingly, one of the groups described that in a STAT3 haploinsufficiency patient symptoms such as recurrent bacterial infections and the connective tissue abnormalities of HIES were absent although symptoms common for STAT3-HIES patients such as high IgE and aspergillosis were present [33]. Therefore, *STAT3*$^{+/R382Q}$ primary fibroblasts treated with the ABE7.10/QRA-3 editing system (3.2.7) might show a different response to the treatment as adenine base editing has the advantage that cells with two functional *STAT3* alleles are generated which will likely have wild type levels of functional STAT3 dimer. Nevertheless, additional replicates and further investigations are needed to confirm the result and hypothesis.

4.5 Therapeutic applicability of the R382W and R382Q approach in STAT3-HIES

The results of ABE7.10 mediated treatment of *STAT3*$^{+/R382W}$ and *STAT3*$^{+/R382Q}$ primary fibroblasts are encouraging in the way that editing in the cells takes place with robust efficiency while keeping genomic integrity. Further, a significant improvement of STAT3 signaling is observed after treatment of *STAT3*$^{+/R382W}$ fibroblasts. In regards to therapeutic applicability this indicates that if it is possible to get the editing system into the disease relevant cells with a high enough efficiency an improvement of STAT3 signaling will likley occur. Additionally, we did not observe a proliferative advantage of repaired patient fibroblasts compared to unrepaired patient fibroblasts in a cell culture setting (3.1.8). In order to further develop the treatment approach many aspects need to be addressed.

While fibroblasts may be important in the pathology of STAT3-HIES especially in the impaired wound healing [108], other cell types which bring the most benefit for the patient if repaired need to be identified. As hematopoietic stem cell transplantation (HSCT) for STAT3-HIES patients is reported to improve symptoms [20] a gene editing mediated targeting of hematopoietic stem cells (HSCs) might be beneficial for the hematopoietic symptoms of patients. Furthermore, alveolar type 2 cells are known to produce pulmonary surfactant in dependency on STAT3 signaling [109] which is

important for the alveolar tension and plays a role in host defence against pulmonary infections [110]. Thus, targeted repair of HSCs, alveolar type 2 cells or their progenitor or descendant cells might improve patients' defence versus pulmonary infections.

Once the target cell types are determined, a mode of gene editing as in *in vivo* or *ex vivo* needs to be chosen. The editing system used in this study uses electroporation as delivery method and is therefore only suited for *ex vivo* treatment. An *ex vivo* delivery of editing systems into autologous HSCs via electroporation mediated delivery of RNPs already has been shown by the gene editing company CRISPR Therapeutics with their drug CTX001 and the corresponding clinical trial [14]. A similar approach for STAT3-HIES patients might be possible. For *in vivo* approaches a change of the delivery system is necessary. Nanoparticle mediated transfection approaches [111] or virus mediated delivery [60] might be suitable delivery method for these approaches. Additionally, the editing system needs to be customized to ensure a high enough efficiency with the chosen delivery and editing method. This customization might involve the exchange of ABE7.10 for ABEmaxAW or other refined adenine base editing systems, the use of a certain form of substrate (DNA, RNA or RNP) and delivery method.
Also screenings for cytotoxicity, genotoxicity and immunogenicity need to be performed. The editing system we used lead to pronounced cytotoxicity as electroporation is a very harsh transfection method. A high cytotoxicity is problematic as a certain selection process of the treated cells might occur which could for example lead to an increased number of cells lacking an appropriate DNA damage response [102].

The search for genotoxic off-target effects is challenging as effects can be extremely rare and are thus undetectable by WGS. More sensitive analyses such as HTS have the need of an *a priori* knowledge of the off-target regions. Thus, we used *in silico* predicted off-target regions for this analysis. The gene editing company Editas medicine which was cofounded by one of the ABE7.10 creators developed a *Staphylococcus aureus* Cas9 based gene editing drug for the genetic disease Leber congenital amaurosis type 10 [57]. In order to ensure the genomic safety of their drug they performed additional assays to screen for DNA off-targets such as Digenome-Seq

and GUIDE-Seq as well as the check of all predicted off-target sites including intronic and intergenic predicted off-targets via HTS [112].

Digenome-Seq is an *in vitro* assay employing human genomic DNA and next-generation sequencing (NGS) to detect any genomic position at which DSBs are created by Cas9 and thus the off-target sites [97, 113]. Further GUIDE-seq is a cellular assay to detect DSB repair machinery in cells after DSB inducing treatments to find actual intracellular off-target sites [98]. A similar principle to the ABE adapted Digenome-seq is used by the method EndoV-Seq [99]. EndoV is an enzyme that will nick inosine containing DNA strands and thus create a detectable DSB together with the DNA nick induced by ABE7.10 or other base editors. With this ABE off-target detection method potential off-targets for ABE/WRA-2 or ABE/QRA-3 editing systems could be identified. An ABE adapted method similar to GUIDE-seq for the intracellular screening of *bona fide* off-targets has not yet been developed. Since EndoV-Seq likely overestimates the amount of potential off-targets as it operates on naked genomic DNA and therefore editing inhibiting parameters such as certain epigenetic modifications and chromatin structures are eliminated [99] an EndoV-Seq screening includes more off-targets than are a reality in cellular editing. Thus, an analysis of predicted off-target sites and biochemically confirmed off-target sites via HTS after editing might detect all guideRNA dependent off-target sites.

Immunotoxicity of editing systems is expected as all Cas9 based editing systems contain elements of prokaryotes which are also able to infect human cells and thus require an immune response. *Streptococcus pyogenes* Cas9 and *Staphylococcus aureus* Cas9 are the most commonly used Cas9 variants. Humans harbor pre-existing adaptive immune responses to both Cas9 orthologs [114]. As *S. pyogenes* Cas9 is also part of our editing system, especially RNP mediated editing *in vivo* will likely trigger an immune response. Additionally, also RNA encoded editing systems can trigger an immune response via the RIG-I pathway [88]. Therefore it is important to take these immune responses into account and avoid them for example by removing 5’-triphosphates from synthesized guideRNAs. Immunotoxicity in the context of STAT3-HIES might be less of a problem as the editing system is either degraded or diluted in *ex vivo* approaches or immunotoxicity might be reduced due to an already impaired immune system, intensive conditioning and/or immunosuppressive treatment in *in vivo* approaches [115]. When monocyte-related dendritic Cells are confronted with

adeno-associated virus (AAV) capsid, a popular delivery method for genetic agents, IL-6 secretion is activated [116]. As STAT3 mediated IL-6 signaling is impaired in STAT3-HIES patients (1.1.2) this might indicate a natural susceptibility to AAV mediated treatments.

Another important aspect of the therapeutic applicability of our editing system is the assessment of STAT3 function after treatment. We could show improved STAT3 signaling in treated patient fibroblasts as measured by STAT3 DNA binding ELISA (TransAM) and reverse transcription quantitative real-time PCR of target genes. Further, we found no proliferative advantage of treated fibroblasts compared to untreated cells during cell culture. While these results are an important proof-of-principle for the treatment of STAT3-HIES via ABE mediated gene editing, additional higher order functional analyses are needed to evaluate potential beneficial effects of a treatment on different cell types, whole tissues and processes such as scarring, immune cell generation, Th17 differentiation, angiogenesis and most importantly on the defence against infections.

4.6 Conclusion

In this study a treatment approach for STAT3-HIES patients affected by heterozygous *STAT3* R382W and R382Q mutations, which are known as hotspot mutations and the most common STAT3-HIES causing mutations, was developed.
The approach targeting the *STAT3* R382W mutation consisted of our single guideRNA WRA-2 and the adenine base editor ABE7.10 encoded on plasmids and delivered via electroporation. The ABE7.10/WRA-2 editing system was able to repair the *STAT3* R382W mutation in primary patient fibroblasts with robust efficiency which allowed us to isolate repaired single cell clones. We did not detect any DNA off-target effects when investigating bulk treated cells via WGS as well as predicted exonic off-target sites via HTS. Bulk treated cells and repaired single cell clones showed improved STAT3 signaling after treatment as shown by STAT3 DNA binding ELISA (TransAM) and reverse transcription quantitative real-time PCR of STAT3 target genes *SOCS3* and *CCL2*. We published these results in Eberherr *et al.* 2021 [100]. Additionally, we improved the efficiency of the editing system via optimized electroporation conditions

and the use of refined adenine base editors such as ABEmax and ABEmaxAW. We observed that ABEmaxAW treated cells showed efficient correction of the mutation and that repaired fibroblasts do not have a proliferative advantage when cultured over 10 passages. Also we were able to show that a repair of the *STAT3* R382W mutation was also possible efficiently via RNP mediated editing (WRA-2 guideRNA and ABE7.10 protein).
The initial treatment approach for the *STAT3* R382Q mutation to knock-out the affected allele was found to be highly efficient and specific. However, a functional benefit for STAT3 signaling of the treatment as measured via reverse transcription quantitative real-time PCR of STAT3 target genes was not observed. A switch of the treatment approach to guideRNA QRA-3 and ABE7.10 protein mediated repair of the mutation resulted in a moderate repair efficiency and unexpectedly no editing of the bystander A in the editing window as assessed by Sanger sequencing. Therefore, adenine base editing mediated repair is likely the better choice for the treatment of the *STAT3* R382Q mutation although no functional analyses have been performed yet.

In conclusion, we were able to provide an important proof-of-principle that adenine base editing is able to repair STAT3-HIES causing mutations and that the correction leads to improved STAT3 signaling. With only symptomatic treatment available for the disease this is a first step towards a causative and permanent therapy approach.

4.7 Outlook

As we successfully targeted two STAT3-HIES causing mutations (*STAT3* R382W and R382Q) in this study it is likely that the editing system could be adapted to other STAT3-HIES causing mutations. An exchange of the *S. pyogenes* Cas9 (PAM: NGG) for a less PAM restrictive variant such as xCas9 (PAM: NG) would greatly improve PAM compatibility and likely enable the correction of several additional STAT3-HIES causing mutations.

With the immense speed of the development of new DNA editing tools, many of them were not tested in this study although they might improve the editing system. These tools include ABE refinements such as ABE8s [117] and SECURE-ABE [118], Cas9

alternatives such as CasX [119], xCas9 [120], Sniper-Cas9 [121] and Cas-NG [122] and new generation editing systems such as prime editing [123]. Especially prime editing, an editing system capable of correcting mutations over a length of over 40 bp without DSBs, would be a powerful tool as it might allow for the correction of several closely clustered *STAT3* mutations with only one editing system [123].

In regards to the delivery system the electroporation used in this study proved efficient for fibroblasts but is only usable *ex vivo*. Therefore different delivery methods such as lipid nanoparticles (LNPs) or AAVs could provide more versatility [60]. This also holds true for the form in which the editing system is encoded and delivered as. Plasmid, RNA and RNP based editing systems each have their advantages and disadvantages. Therefore it would be important to investigate whether or not the form of encoding changes parameters such as efficiency, genotoxicity, cytotoxicity and immunogenicity.

Since HSCT showed promising results in the treatment of STAT3-HIES [20], developing an *in vivo* gene editing therapy in which underlying *STAT3* mutations in HSCs are repaired via HSC mobilization and subsequent intravenous injection of gene editing agents similar to a system already reported by a research group [124] might provide an effective treatment option. This approach would abolish the need to undergo HSCT and thus remove the problem of finding a matching donor, undergoing conditioning prior to HSCT and risk of graft-versus-host disease.

At last, any developed editing systems for STAT3-HIES causing mutations could likely be transferred to *STAT3* gain of function mutations such as *STAT3* mutations in solid and hematopoietic cancers [125] and thus be of benefit to a broad spectrum of patients.

5. Material and Methods

5.1 Materials

5.1.1 Cells

Cells or cell line	Company
One Shot Mach1 T1 Phage-Resistant Chemically Competent *E. coli*	Thermo Fisher Scientific, USA
Human Dermal Fibroblasts adult (HDFa)	ATCC, USA
STAT3 +/R382W primary fibroblasts	Isolated from skin sample
STAT3 +/R382Q primary fibroblasts	Isolated from skin sample

Table 4: Cells or cell lines and origin

5.1.2 Media

Media	Company
DMEM 1 g/L glucose, L-glutamine, sodium pyruvate and 3.7 g/L $NAHCO_3$	Pan Biotech, Germany
CryoStor CS10 Freeze Media	Biolife Solutions, USA
FBS superior	Biochrom, USA

Table 5: Media and companies

5.1.3 Kits

Kit	Company
Amaxa P2 Primary Cell 4D-Nucleofector X Kit L	Lonza, Switzerland
AmpliTaq Gold 360 Master Mix	Thermo Fisher Scientific, USA
Bio-Rad Protein Assay Kit II	Bio-Rad, USA
DNeasy Blood & Tissue Kit	Qiagen, Germany
EndoFree Plasmid Maxi Kit	Qiagen, Germany
EnGen sgRNA synthesis kit, *S. Pyogenes*	New England Biolabs, USA

High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems
innuPREP RNA Mini Kit 2.0	analytik jena, Germany
Nuclear Extract Kit	Active Motif, USA
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel, Germany
QIAprep Spin Miniprep Kit	Qiagen, Germany
RNA Clean & Concentrator-25	Zymo Research, USA
STAT3 TransAM	Active Motif, USA
Super Signal West Femto Maximum Sensitivity Substrate Kit	Thermo Scientific, USA
T4 DNA Ligase kit	fisher scientific, USA
TruSeq DNA PCR-Free Kit	Illumina, USA

Table 6: Kits and companies

5.1.4 Enzymes

Enzyme	**Source**
Alt-R S.p. Cas9 Nuclease V3	Integrated DNA Technologies, USA
BbsI restriction enzyme	New England Biolabs, USA
Quick CIP	New England Biolabs, USA
Proteinase K	Qiagen, Germany
ABE7.10 protein	Arie Geerlof of the Helmholtz Zentrum München

Table 7: Enzymes and origins

5.1.5 Chemicals and reagents

Reagent	**Company**
Boric acid, >99.8%	Carl Roth GmbH, Germany
Chloroform/Trichlormethan, 1l	Carl Roth GmbH, Germany
cOmplete Lysis-M, EDTA-free	Roche, Switzerland
Dimethyl sulfoxide research grade	SERVA Electrophoresis GmbH, Germany
EDTA – Solution pH 8.0 (0.5 M)	PanReac AppliChem ITW Reagents, Germany, Spain and Italy

GelRed Nucleic Acid Gel Stain	Biotium, USA
Gibco Ampicillin Sodium Salt	Thermo Fisher Scientific, USA
Gibco Antibiotic-Antimycotic (100x)	Thermo Fisher Scientific, USA
Gibco DPBS, -$CaCl_2$, -$MgCl_2$	Thermo Fisher Scientific, USA
Gibco TrypLE Express	Thermo Fisher Scientific, USA
Interleukin-6 human, IL-6, recombinant, expressed in *E. coli*, lyophilized powder	Sigma-Aldrich, USA
LB Broth (Luria/Miller)	Carl Roth, Germany
Methanol	Carl Roth GmbH, Germany
Page Ruler Prestained Protein Ladder, 10 to 180 kDa	Thermo Fisher Scientific, USA
Powdered milk	Carl Roth GmbH, Germany
Roti-Aqua-P/C/I	Carl Roth GmbH, Germany
Sodium acetate buffer solution, for molecular biology, 3 M	Sigma, USA
Sodium chloride	Carl Roth GmbH, Germany
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad, USA
TRIS	Carl Roth GmbH, Germany
TWEEN 20	Sigma-Aldrich, USA
UltraPure Agarose	Thermo Fisher Scientific, USA

Table 8: Chemicals, reagents and companies

5.1.6 Solutions and buffers

Blocking solution (10%): 100 ml TBST, 10 g powdered milk

TBE (1x; tris borate EDTA): 0.13 M tris, 45 mM boric acid, 2.5 mM EDTA

TBST (1x; tris-buffered saline, 0.1% Tween 20): 20 mM tris, 150 mM NaCl, 0.1% Tween 20

5.1.7 Primers and oligonucleotides

PCR Primers

Primer	Description	Sequence
hSTAT3 12F	Sanger sequencing, forward	GAGGTGTTATGTTGCGCTGATC
hSTAT3 14R	Sanger sequencing, reverse	CTGTTCATGTCACTTTGGCCTG
TBP-fwd	qRT-PCR, forward	CAGCCTGCCACCTTACGCT
TBP-rvs	qRT-PCR, reverse	GCCATAAGGCATCATTGGACTAA
CCL2-fwd	qRT-PCR, forward	GCCTCCAGCATGAAAGTCTC
CCL2-rvs	qRT-PCR, reverse	AGGTGACTGGGGCATTGAT
SOCS3-fwd	qRT-PCR, forward	GCCACCTACTGAACCCTCCT
SOCS3-rvs	qRT-PCR, reverse	ACGGTCTTCCGACAGAGATG

Table 9: PCR primer sequences

High-throughput sequencing primers

Adapter sequence forward: ACACTCTTTCCCTACACGACGCTCTTCCGATCT

Adapter sequence reverse: GACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Primer name	Orientation	Sequence (5'>3')
NGS-hSTAT3-fwd	Forward	AAAGAAATGCCCAGGAGCAC
NGS-hSTAT3-rvs	Reverse	CTTACATCACAATTGGCTCGGC
NGS-hPREP-fwd	Forward	ACATGGGGATTTGCTCTATGTAA
NGS-hPREP-rvs	Reverse	TGAAGGGTAATTCACCAAGCTA
NGS-hLINC00933-fwd	Forward	GACTGACCTTTACCTGGCAGA
NGS-hLINC00933-rvs	Reverse	AATGACCACGACCATGGGTT
NGS-hGNL1-fwd	Forward	CGTGAGTCAGGGCACCTTAG
NGS-hGNL1-rvs	Reverse	TCTCAGCTGGGACCTTTTCTC
NGS-hALDH2-fwd	Forward	ATGGGATACTGTATGTAAAGCCC
NGS-hALDH2-rvs	Reverse	AGGATGGTGACCACCAGATTC
NGS-hARPC5-fwd	Forward	ACAGGAAAAGACATATTTTGGTTGA
NGS-hARPC5-rvs	Reverse	CACTGCCTGACTCTTGGTGT
NGS-hSTK32C_201-fwd	Forward	CCCAGGGAGACTCAGACTCAT
NGS-hSTK32C_201-rvs	Reverse	TTTTGCACCCAAGTCTCAAGG

NGS-hSH2D7-fwd	Forward	CTACTCTTAGTGACCACTTGGCT
NGS-hSH2D7-rvs	Reverse	AAGGGCTCTTGCTCCTGTCTC
NGS-hCUEDC1-fwd	Forward	ATTTGAACTGCACATGGAGGG
NGS-hCUEDC1-rvs	Reverse	GTGTTGTTTGTACACAGCCCC
NGS-hDDIT3-fwd	Forward	CATGGAGCTTGTTCCAGCCA
NGS-hDDIT3-rvs	Reverse	TCCTGGTTCTCCCTTGGTCT
NGS-hSNX19P3+1-fwd	Forward	ATTCCTGAAGCAATTGTGTGC
NGS-hSNX19P3+1-rvs	Reverse	GTTCAGGATCTTTTTCTGGGGC
NGS-hKLHL29-fwd	Forward	TGTCAGTGTTGAGCCAGTCTTG
NGS-hKLHL29-rvs	Reverse	CTCTGAGGCTGTGTGTGAGAG

Table 10: High-throughput sequencing primers

Oligonucleotides

Oligo	Description	Sequence
WRA-2 oligo 1	Cloned into pBS-U6 after hybridization with oligo 2	CACCGATTTCCAGGATCTGAATCAC
WRA-2 oligo 2	Cloned into pBS-U6 after hybridization with oligo 1	AAACGTGATTCAGATCCTGGAAATC

Table 11: Oligonucleotide sequences

guideRNA synthesis templates for the EnGen sgRNA Synthesis Kit (NEB)

WRA-2 synthesis template	TTCTAATACGACTCACTATAGATTTCCAGGATCTGAATCACGTTTTAGAGCTAGA
QRA-2 synthesis template	TTCTAATACGACTCACTATAGATCCCAGAAATTTAACATTCGTTTTAGAGCTAGA
QRA-3 synthesis template	TTCTAATACGACTCACTATAGTCCCAGAAATTTAACATTCTGTTTTAGAGCTAGA
QR1 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTTAAATTTCTGTTTTAGAGCTAGA
QR2-1_10 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGCTAAATTTCTGTTTTAGAGCTAGA
QR2-1_9 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTCAAATTTCTGTTTTAGAGCTAGA

QR2-1_8 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTTCAATTTCTGTTTTAGAGCTAGA
QR2-1_7 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTTACATTTCTGTTTTAGAGCTAGA
QR2-1_6 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTTAACTTTCTGTTTTAGAGCTAGA
QR2-1_5 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTTAAACTTCTGTTTTAGAGCTAGA
QR2-1_4 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTTAAATCTCTGTTTTAGAGCTAGA
QR2-1_3 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTTAAATTCCTGTTTTAGAGCTAGA
QR2-1_2 synthesis template	TTCTAATACGACTCACTATAGCCCAGAATGTTAAATTTGTGTTTTAGAGCTAGA
truQR1_18 synthesis template	TTCTAATACGACTCACTATAGCCAGAATGTTAAATTTCTGTTTTAGAGCTAGA
truQR1_17 synthesis template	TTCTAATACGACTCACTATAGCAGAATGTTAAATTTCTGTTTTAGAGCTAGA
RRi15 synthesis template	TTCTAATACGACTCACTATAGAATGTTAAATTTCCGTTTTAGAGCTAGA

Table 12: guideRNA synthesis templates

5.1.8 Protospacer sequences of single guideRNAs

guideRNA	**Protospacer sequence**
WRA-2	ATTTCCAGGATCTGAATCAC
WRA-3	TTTCCAGGATCTGAATCACA
WRA-4	TTCCAGGATCTGAATCACAG
QRA-2	ATCCCAGAAATTTAACATTC
QRA-3	TCCCAGAAATTTAACATTCT
QR1	GCCCAGAATGTTAAATTTCT
QR2-1_10	GCCCAGAATGCTAAATTTCT
QR2-1_9	GCCCAGAATGTCAAATTTCT
QR2-1_8	GCCCAGAATGTTCAATTTCT

QR2-1_7	GCCCAGAATGTTACATTTCT
QR2-1_6	GCCCAGAATGTTAACTTTCT
QR2-1_5	GCCCAGAATGTTAAACTTCT
QR2-1_4	GCCCAGAATGTTAAATCTCT
QR2-1_3	GCCCAGAATGTTAAATTCCT
QR2-1_2	GCCCAGAATGTTAAATTTGT
truQR1_18	CCAGAATGTTAAATTTCT
truQR1_17	CAGAATGTTAAATTTCT
RRi15	GAATGTTAAATTTCC

Table 13: Protospacer sequences of the used guideRNAs

5.1.9 Antibodies

Antibody	Dilution	Company
Phospho-Stat3 (Tyr705) (3E2) Mouse mAb	1:1000	Cell Signaling Technology, USA
TBP (D5C9H) XP Rabbit mAb	1:1000	Cell Signaling Technology, USA
Stabilized Goat Anti-Rabbit/Anti-Mouse IgG, Peroxidase Conjugated	1:2500	Thermo Fisher Scientific, USA

Table 14: Antibodies

5.1.10 Plasmids

Plasmid	Source
pCMV-ABE7.10	addgene (#102919), USA
pCMV_ABEmax	addgene (#112095), USA
pCMV-ABEmaxAW	addgene (#125647), USA
pBS-U6:chimaericRNA	Kind gift from Florian Giesert of the Helmholtz Center Munich
pBS-U6:WRA2	WRA-2 protospacer sequence was cloned into the plasmid pBS-U6:chimaericRNA

pmaxGFP	Lonza, Switzerland

Table 15: Plasmids and origins

5.1.11 Laboratory equipment and consumables

Item	Company
2720 Thermal Cycler	Applied Biosystems, USA
4D-Nulceofector Core and X Unit	Lonza, Switzerland
Cell culture flasks, surface, filter cap, T-25/T-75/T-175	Sarstedt, Germany
Cellstar cell culture multiwall plates, 6/12/24/48/96 wells, PS, sterile	Greiner Bio-One, Germany
Cellstar serological pipette, graduated, sterile, 2/5/10/25/50 ml	Greiner Bio-One, Germany
CFX Connect Real-Time PCR Detection System	Bio-Rad, USA
Cryo.s, 2ml, PP, round bottom cryoconservation vials	Greiner Bio-One, Germany
ENVAIR safety cabinet eco safe Comfort Plus	ENVAIR Deutschland GmbH, Germany
Eppendorf New Brunswick Excella E24R Incubator Shaker	Eppendorf, Germany
Eppendorf Research plus mechanical pipettes 10/100/1000 µl	Eppendorf, Germany
Falcon Conical Centrifuge Tubes 15 and 50 ml	Thermo Fisher Scientific, USA
Heracell 150i CO2 Incubator	Thermo Fisher Scientific, USA
Heraeus Fresco 21 Centrifuge	Thermo Fisher Scientific, USA
Heraeus Multifuge X1R Centrifuge	Thermo Fisher Scientific, USA
INCU-Line IL 56, digital incubator	VWR, USA
Intas ChemoCam imaging device	Intas Science Imaging Instruments GmbH, Germany
Inverted Laboratory Microscope Leica DM IL LED	Leica Microsystems, Germany
Julabo SW22 water bath	Julabo, Germany

Low Retention Filter Tips 10, 100 and 1250 µl, sterile	Greiner Bio-One, Germany
Mini ReadySub-Cell GT Cell	Bio-Rad, USA
NanoDrop 1000 Spectrophotometer	Thermo Fisher Scientific, USA
NovaSeq 6000 platform	Illumina, USA
NuPAGE 4-12% Bis-Tris protein gels	Thermo Fisher Scientific, USA
pE-300^{lite} illumination system	CoolLED, England
PowerPac Basic Power Supply	Bio-Rad, USA
PVDF Transfer Membrane, 0.45 µM	Thermo Fisher Scientific, USA
Qubit 3.0 Fluorometer	Thermo Fisher Scientific, USA
Reaction tubes, PP, Natural, 1.5 and 2 ml	Greiner Bio-One, Germany
SAFE 2020 Class II Biological Safety Cabinet	Thermo Fisher Scientific, USA
XCell II Blot Module	Thermo Fisher Scientific, USA
XCell SureLock Mini-Cell Electrophoresis System	Thermo Fisher Scientific, USA

Table 16: Laboratory equipment, consumables and companies

5.1.12 Software

Name	**Use**	**Source**
MS Excel 2013	Data management and calculations	www.microsoft.com
GraphPad prism V5.03	Statistical analysis and Figure creation	www.graphpad.com
LABIMAGE 1D	DNA band visualization	www.kapelanbio.com
SnapGene	Visualization of cloning steps and sequences	www.snapgene.com
CRISPOR	Off-target analysis	Concordet *et al.* [96]
TIDE	DSB induction efficiency	Brinkman *et al.* [94]
EditR	Base editing efficiency	Kluesner *et al.* [93]
CRISPResso 2	HTS analysis	Clement *et al.* [95]

Table 17: General software

Whole genome sequencing:

Name	Use	Source
GATK (version 4) BaseRecalibrator	base quality score recalibration	McKenna *et al.* [126]
UCSC RepeatMasker	Repeat rejection	Smit *et al.* [127]
BreakDancer	structural variant calling	Fan *et al.* [128]
Delly	structural variant calling	Rausch *et al.* [129]
CNVnator	structural variant calling	Abyzov *et al.* [130]
Lumpy-SV	structural variant calling	Layer *et al.* [131]
Manta	structural variant calling	Chen *et al.* [132]
Pindel	structural variant calling	Ye *et al.* [133]
Integrative Genomics Viewer	Genome browser	Robinson *et al.* [134]

Table 18: Whole genome sequencing software

5.2 Methods

5.2.1 Institutional Review Board (IRB) approval

Skin samples were taken by a physician after IRB approved informed consent of the patients and healthy controls as stated in Eberherr *et al.* [100].

5.2.2 Primary fibroblast cell culture

As stated in Eberherr *et al.* [100] fibroblasts were isolated from a skin sample of a STAT3-HIES patient heterozygous for the R382W or R382Q mutation (*STAT3*$^{+/R382W}$ and *STAT3*$^{+/R382Q}$ fibroblasts) and two healthy individuals (control). Normal human adult primary dermal fibroblasts (HDFa, ATCC PCS-201-012, *STAT3*$^{+/+}$) were purchased from ATCC. Fibroblasts were cultured in DMEM (PAN Biotech) with 10% FBS superior (Biochrom) and 1% Antibiotic-Antimycotic (ThermoFisher Scientific) at 37 °C and 5% CO_2. Single cell clones were isolated via seeding of a 5 cell/µl cell suspension in 96-well format with a 1:1 dilution of normal fibroblast medium and medium conditioned by culturing *STAT3*$^{+/R382W}$ fibroblasts for 48 h. 24 h after seeding

wells containing only one fibroblast were selected for further cultivation. Cells were detached with Gibco TrypLE express (Thermo Fisher Scientific).

5.2.3 DNA or RNA isolation and purification

DNA isolation from cultured cells was performed with the DNeasy Blood and tissue kit (Qiagen) and RNA isolation with innuPREP RNA Mini Kit 2.0 (analytik jena) according to the manufacturer's instructions.
DNA purification was done with the Nucleo Spin Gel and PCR Clean-up kit (Macherey-Nagel) and RNA purification for low concentrations with the RNA Clean and Concentrator-25 kit (Zymo Research) or via Phenol/Chloroform RNA isolation for high concentrations. In brief the reaction volume was adjusted to 180 µl with nuclease-free H_2O and 20 µl of 3 M sodium acetate pH 5.2 were added. After mixing thoroughly an equal volume of phenol/chloroform/isoamyl alcohol (ROTI Aqua-P/C/I, Carl Roth) was added and the aqueous phase was transferred into a new tube. The aqueous phase was then mixed with an equal volume of chloroform and mixed thoroughly. After a second extraction with chloroform the RNA was precipitated by addition of 2 volumes of ethanol and incubation at – 20 °C for at least 30 min. The RNA pellet was collected via centrifugation and washed with 500 µl of a cold 70% ethanol solution. At last, the pellet was resuspended in H_2O.

5.2.4 Polymerase chain reaction – PCR

PCRs were performed with the AmpliTaq Gold 360 Master Mix (Thermo Fisher Scientific). The reaction was assembled in a total volume of 50 µl as follows:
25 µl AmpliTaq Gold 360 Master Mix
100 – 200 ng genomic DNA template in 5 µl volume
1 µl of primermix (10 µM each)
19 µl of H2O

The PCRs were run with the following program:

95°C for 10 min

40x 95°C for 30 sec

55 °C for 30 sec

72°C for X sec → depending on amplicon size (1 min for 1000 bp)

72°C for 7 min

4°C for storage

5.2.5 Gel electrophoresis

Agarose gels were produced from 1x TBE buffer, agarose and Gel Red (biotium). Gels were run in TBE buffer at 100 V for 60 min. Analysis was done via an Intas ChemoCam imaging device (Intas Science Imaging Instruments).

5.2.6 DNA and RNA quantification

Quantification of DNA and RNA was done by analysis via NanoDrop (Thermo Fisher Scientific) according to the manufacturer's instructions. For estimation of exact DNA amounts (required for WGS and HTS) a Qubit 3.0 system (Thermo Fisher Scientific) was used according to the manufacturer's instructions.

5.2.7 Cloning

To enable expression of guideRNAs via the plasmid pBS-U6:chimaericRNA under U6 promoter control protospacer sequences were ordered as two complimentary oligos resulting in BbsI resctriction enzyme cut sites. The pBS-U6:chimaericRNA plasmid was digested with BbsI and the hybridized oligos were ligated into the plasmid via T4 DNA ligase (fisher scientific) according to the manufacturer's instructions. The ligation reaction was used for transformation into One Shot Mach1 T1 Phage-Resistant Chemically Competent *E. coli* (Thermo Fisher Scientific). 1 µl of ligation reaction was pipetted into one aliquot of *E. coli.* The cells were incubated on ice for 30 min and heat shocked at 42 °C for 20 sec. After incubation for 2 min on ice 950 µl LB medium were added to the cells followed by incubation at 37 °C for 1 h. 100 µl of the *E. coli* solution were plated on ampicillin containing (50 µg/ml) LB agar plates. Plates were incubated

at 37 °C over night. Colonies were picked the next day and checked for integration of the insert via the QIAprep Spin Miniprep kit (Qiagen) and subsequent Sanger sequencing.

5.2.8 sgRNA production

guideRNAs were synthesized with the EnGen sgRNA synthesis kit *Streptococcus pyogenes* (New England Biolabs) according to the manufacturer's instructions. The only exception was an increased incubation time at 37 °C for 16 – 18 h instead of 30 min. The synthesized guideRNAs were treated with DNase I and calf intestinal phosphatase (quick CIP from New England Biolabs) as recommended. guideRNAs were purified via column based isolation or phenol/chloroform isolation. guideRNA concentration was measured via NanoDrop and purity was confirmed via sufficient 260/230 and 260/280 readings. guideRNAs were subjected to gel electrophoresis to confirm the correct size and to check for signs of degradation. Finally all guideRNAs were validated by Cas9 *in vitro* assay.

5.2.9 Cas9 *in vitro* assay

Synthesized sgRNAs were subjected to Cas9 in vitro assays to confirm their ability to induce Cas9 mediated cleavage of amplicons of the target region. 10 pmol Cas9 were mixed with 10 pmol of Alt-R S.p. Cas9 Nuclease V3 (IDT), 2 µl of 10x NEB buffer 3.1 and sufficient nuclease-free H_2O for a total volume of 20 µl. After incubation for 10 min at room temperature to assemble guideRNA and Cas9 RNPs, 1 pmol of amplicons containing the guideRNA target region was added in a total volume of 10 µl including 1 µl of 10x NEB buffer 3.1. The 30 µl reaction was then incubated at 37 °C for 1 h. To terminate the reaction 1 µl of Proteinase K (Qiagen; 1.25 µg/ml) was added followed by incubation at 56 °C for 10 min. The digested DNA was then subjected to gel electrophoresis to determine the cutting efficiency of the guideRNAs.

5.2.10 Electroporation

Electroporation of fibroblasts was done with the 4D-Nucleofector X System (program DT-130 or CZ-167) and the P2 Primary Cell Kit L (Lonza) as recommended by the

manufacturer. For plasmid based editing 1.5 pmol of ABE expressing plasmid and 1.5 pmol of guideRNA expressing plasmid were used in a volume of 100 µl nucleofection solution. For RNP based electroporation 60 pmol of Cas9 (IDT) and 60-72 pmol of guideRNA were used in a 100 µl volume.

5.2.11 Fluorescence microscopy

An inverted Laboratory Microscope Leica DM IL LED (Leica) in combination with a pE-300^{lite} illumination system (CoolLED) was used for fluorescence analysis of GFP expressing cells according to the manufacturer's instructions. GFP pictures were taken with 40 ms exposure time and a gain setting of 1.

5.2.12 Sanger sequencing

As stated in Eberherr *et al.* [100] Sanger sequencing was performed with PCR fragments of the target regions and exonic WRA-2 guideRNA off-targets with four mismatches predicted by CRISPOR [96]. Primer sequences are listed in 5.1.7. Editing efficiencies were determined via *in silico* analysis using EditR [93] or TIDE [94]. Base editing efficiency was calculated from the percentage of A signal in the chromatogram in untreated cells vs. treated cells. Chromatograms were received from Eurofins Genomics GmbH.

5.2.13 High-throughput sequencing

As in Eberherr *et al.* [100] primers for the *STAT3* target region and exonic WRA-2 guideRNA off-targets with four mismatches predicted by CRISPOR were designed to amplify 450 to 500 bp long amplicons centered on the off-target regions. The adapter sequence 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' was attached to forward primers and 5'-GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' to reverse primers at the 5'-end. PCR amplicons were sent for HTS to Eurofins Genomics or Genewiz. Analysis of HTS data was performed with CRISPResso 2 [95].

5.2.14 Whole genome sequencing

As stated in Eberherr *et al.* [100] The genomic library was prepared from 1000 ng of genomic DNA with the TruSeq DNA PCR-Free Kit (Illumina). DNA was fragmented to an average length of 350 bp by sonication. Libraries were validated according to standard procedures and sequenced via 150 bp paired-end on a NovaSeq 6000 platform. Reads were aligned using the mem algorithm of bwa 0.7.5a and aligned to the hg19 reference with decoy sequences and masked PAR regions. Base quality scores are recalibrated using GATK (version 4) BaseRecalibrator with enlarged context size for single nucleotide variants (SNVs) and InDels of respectively four and eight base pairs (instead of the default values 2 and 3). Variants were called and inspected both with GATK and custom scripts. Data are annotated with custom in house scripts using refSeq genes.

WGS data of ABE7.10/WRA-2 treated and untreated *STAT3*$^{+/R382W}$ fibroblasts were analyzed and compared to search for treatment induced SNVs and InDel variants. A deliberately low threshold of three supporting reads to call a variant was set, requiring a minimum of twenty reads on the position to retain it. To avoid confounding errors we restricted only to the high confidence regions calculated by the Genome In A Bottle consortium [135]. We rejected repeats as marked in UCSC RepeatMasker [127] and low complexity regions. We required a minimum mapping quality of 30 in the Phred quality score and a minimum base quality of 20 for the genomic distribution analysis and a minimum mapping quality of 50 in the Phred quality score and a minimum base quality of 25 for all other analysis. The genome was divided into 1279 similarly sized chunks and searched for variants uniquely detected in the ABE7.10/WRA-2 treated sample and not in the untreated sample and vice versa. Structural variants were called using a set of different structural variant callers: BreakDancer, Delly, CNVnator, Lumpy-SV, Manta, Pindel focusing on insertions, deletions, inversions, duplications and translocations [128-133]. The variants of each caller were combined together when overlap was detected for two separate calls of the same variant class. Potential integration of plasmid DNA was assessed by checking the presence of the plasmid sequences in the WGS data. To test for integration of the detected plasmid sequences we analyzed read pairs in which one read mapped on the plasmid sequence and the other read on the human nuclear genome. We inspected the reads manually using the Integrative Genomics Viewer (IGV) [134]. Of each pair we scanned both the read on

the plasmid and the read on the nuclear genome for signs of integration, for example the sequence being disrupted from a certain point to the 3'- or 5'-end. Secondly, we inspected the nearby region to the matched read on the nuclear genome to detect signs of an insertion, for example reads whose alignment is disrupted from a certain position on. To avoid confounding effects, we compared simultaneously the reads from our WGS sample from both the alignments with and without plasmid sequence and we added a different genomic sample to avoid reference genome effects. Off-target sites predicted using CRISPOR [96] were analyzed using the WGS data. Further sites were tested by independently aligning the protospacer sequence of WRA-2 to the hg19 reference genome and evaluating variants in nearby sites (±45 bp). Variants uniquely detected in the ABE7.10/WRA-2 treated and not the untreated sample were analyzed by Levenshtein distance analysis for potential binding of the guideRNA WRA-2 in a sliding window starting 45 bp upstream and ending 45 bp downstream of each variant position as mentioned above.

5.2.15 Western blot

As stated in Eberherr *et al.* [100] Western blot analysis was performed using 10-15 µg of nuclear extracts. Following SDS-PAGE with NuPAGE 4-12% Bis-Tris protein gels (Thermo Fisher Scientific) blots were probed with Phospho-Stat3 (Tyr705) (3E2) Mouse mAb (1:1000) and TBP (D5C9H) XP Rabbit mAb (1:1000) (all Cell Signaling Technology), blocked with and developed with secondary antibodies Stabilized Goat Anti-Rabbit/Anti-Mouse IgG, Peroxidase Conjugated (Thermo Scientific) and Super Signal West Femto Maximum Sensitivity Substrate Kit (Thermo Scientific) and an Intas ChemoCam imaging device (Intas Science Imaging Instruments GmbH). Page Ruler Prestained Protein Ladder (Thermo Scientific) was used as a size standard. Quantification was performed using LabImage 1D (Intas Science Imaging GmbH).

5.2.16 STAT3 DNA binding ELISA (TransAM)

As stated in Eberherr *et al.* [100] Nuclear extracts of unstimulated and IL-6 (20 ng/ml for 20 min) (Biochrom) stimulated *STAT3*$^{+/R382W}$ fibroblasts were prepared with the Nuclear Extract Kit (Active Motif) according to the manufacturer's instructions. Nuclear extract protein concentrations were measured via Bradford assay (Bio-Rad) and

STAT3 DNA-binding activity was analyzed via STAT3 TransAM (Active Motif) according to the manufacturer's instructions.

5.2.17 Reverse transcription quantitative real-time PCR (qRT-PCR)

As stated in Eberherr *et al.* [100] RNA isolations from unstimulated and IL-6 (20 ng/ml for 60 min) (Biochrom) stimulated *STAT3*$^{+/R382W}$ fibroblasts were performed with the innuPREP RNA Mini Kit 2.0 (analytik jena) with DNase digestion according to the manual. cDNA was generated via the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions. Gene expression was analyzed by qRT-PCR using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) and a CFX Connect Real-Time PCR Detection System (Bio-Rad).

5.2.18 Statistical analysis

As stated in Eberherr *et al.* [100] Assessment of significance was done by two-way ANOVA to check for overall differences and with a Bonferroni's posttest comparing each sample to the untreated cell population using GraphPad Prism software (Version 5.03) if applicable.

6. References

1. Davis, S.D., J. Schaller, and R.J. Wedgwood, *Job's Syndrome. Recurrent, "cold", staphylococcal abscesses.* Lancet, 1966. **1**(7445): p. 1013-5.

2. Buckley, R.H., B.B. Wray, and E.Z. Belmaker, *Extreme hyperimmunoglobulinemia E and undue susceptibility to infection.* Pediatrics, 1972. **49**(1): p. 59-70.

3. Minegishi, Y., et al., *Dominant-negative mutations in the DNA-binding domain of STAT3 cause hyper-IgE syndrome.* Nature, 2007. **448**(7157): p. 1058-62.

4. Holland, S.M., et al., *STAT3 mutations in the hyper-IgE syndrome.* N Engl J Med, 2007. **357**(16): p. 1608-19.

5. Renner, E.D., et al., *STAT3 mutation in the original patient with Job's syndrome.* N Engl J Med, 2007. **357**(16): p. 1667-8.

6. Bocchini, C.E., et al., *Protein stabilization improves STAT3 function in autosomal dominant hyper-IgE syndrome.* Blood, 2016. **128**(26): p. 3061-3072.

7. Fabre, A., et al., *Clinical Aspects of STAT3 Gain-of-Function Germline Mutations: A Systematic Review.* J Allergy Clin Immunol Pract, 2019. **7**(6): p. 1958-1969 e9.

8. Renner, E.D., et al., *Autosomal recessive hyperimmunoglobulin E syndrome: a distinct disease entity.* J Pediatr, 2004. **144**(1): p. 93-9.

9. Zhang, Q., et al., *Combined immunodeficiency associated with DOCK8 mutations.* N Engl J Med, 2009. **361**(21): p. 2046-55.

10. Sassi, A., et al., *Hypomorphic homozygous mutations in phosphoglucomutase 3 (PGM3) impair immunity and increase serum IgE levels.* J Allergy Clin Immunol, 2014. **133**(5): p. 1410-9, 1419 e1-13.

11. Chavanas, S., et al., *Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome.* Nat Genet, 2000. **25**(2): p. 141-2.

12. Minegishi, Y., et al., *Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity.* Immunity, 2006. **25**(5): p. 745-55.

13. Stadler, P.C., et al., *Inborn Error of Immunity or Atopic Dermatitis: When to be Concerned and How to Investigate.* J Allergy Clin Immunol Pract, 2021. **9**(4): p. 1501-1507.

14. *Orphanet version 5.49.0. - Last update: 2021-10-25*. Available from: https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Expert=2314&lng=EN.

15. Hsu, A.P., et al., *STAT3 Hyper IgE Syndrome*, in *GeneReviews((R))*, M.P. Adam, et al., Editors. 1993: Seattle (WA).

16. Zhang, L.Y., et al., *Clinical features, STAT3 gene mutations and Th17 cell analysis in nine children with hyper-IgE syndrome in mainland China.* Scand J Immunol, 2013. **78**(3): p. 258-65.

17. Danion, F., et al., *Aspergillus fumigatus Infection in Humans With STAT3-Deficiency Is Associated With Defective Interferon-Gamma and Th17 Responses.* Front Immunol, 2020. **11**: p. 38.

18. Freeman, A.F. and S.M. Holland, *Clinical manifestations, etiology, and pathogenesis of the hyper-IgE syndromes.* Pediatr Res, 2009. **65**(5 Pt 2): p. 32R-37R.

19. Kroner, C., et al., *Lung disease in STAT3 hyper-IgE syndrome requires intense therapy.* Allergy, 2019. **74**(9): p. 1691-1702.

20. Harrison, S.C., et al., *Hematopoietic Stem Cell Transplantation Resolves the Immune Deficit Associated with STAT3-Dominant-Negative Hyper-IgE Syndrome.* J Clin Immunol, 2021. **41**(5): p. 934-943.

21. Akira, S., *Functional roles of STAT family proteins: lessons from knockout mice.* Stem Cells, 1999. **17**(3): p. 138-46.

22. Howe, K.L., et al., *Ensembl 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D884-D891.

23. Aigner, P., V. Just, and D. Stoiber, *STAT3 isoforms: Alternative fates in cancer?* Cytokine, 2019. **118**: p. 27-34.

24. Caldenhoven, E., et al., *STAT3beta, a splice variant of transcription factor STAT3, is a dominant negative regulator of transcription.* J Biol Chem, 1996. **271**(22): p. 13221-7.

25. Sgrignani, J., et al., *Structural Biology of STAT3 and Its Implications for Anticancer Therapies Development.* Int J Mol Sci, 2018. **19**(6).

26. Hirai, H., P. Karian, and N. Kikyo, *Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor.* Biochem J, 2011. **438**(1): p. 11-23.

27. Aznar, S., et al., *Simultaneous tyrosine and serine phosphorylation of STAT3 transcription factor is involved in Rho A GTPase oncogenic transformation.* Mol Biol Cell, 2001. **12**(10): p. 3282-94.

28. Akira, S., *Roles of STAT3 defined by tissue-specific gene targeting.* Oncogene, 2000. **19**(21): p. 2607-11.

29. O'Shea, J.J., S.M. Holland, and L.M. Staudt, *JAKs and STATs in immunity, immunodeficiency, and cancer.* N Engl J Med, 2013. **368**(2): p. 161-70.

30. O'Shea, J.J., et al., *The JAK-STAT pathway: impact on human disease and therapeutic intervention.* Annu Rev Med, 2015. **66**: p. 311-28.

31. Ma, J.H., L. Qin, and X. Li, *Role of STAT3 signaling pathway in breast cancer.* Cell Commun Signal, 2020. **18**(1): p. 33.

32. Takeda, K., et al., *Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality.* Proc Natl Acad Sci U S A, 1997. **94**(8): p. 3801-4.

33. Natarajan, M., et al., *Aspergillosis, eosinophilic esophagitis, and allergic rhinitis in signal transducer and activator of transcription 3 haploinsufficiency.* J Allergy Clin Immunol, 2018. **142**(3): p. 993-997 e3.

34. Levy, D.E. and J.E. Darnell, Jr., *Stats: transcriptional control and biological impact.* Nat Rev Mol Cell Biol, 2002. **3**(9): p. 651-62.

35. Sun, L., et al., *Interleukin 12 (IL-12) family cytokines: Role in immune pathogenesis and treatment of CNS autoimmune disease.* Cytokine, 2015. **75**(2): p. 249-55.

36. Metcalfe, R.D., T.L. Putoczki, and M.D.W. Griffin, *Structural Understanding of Interleukin 6 Family Cytokine Signaling and Targeted Therapies: Focus on Interleukin 11.* Front Immunol, 2020. **11**: p. 1424.

37. Powell, M.D., et al., *IL-12 signaling drives the differentiation and function of a TH1-derived TFH1-like cell population.* Sci Rep, 2019. **9**(1): p. 13991.

38. Vogel, T.P., J.D. Milner, and M.A. Cooper, *The Ying and Yang of STAT3 in Human Disease.* J Clin Immunol, 2015. **35**(7): p. 615-23.

39. Jiao, H., et al., *Novel and recurrent STAT3 mutations in hyper-IgE syndrome patients from different ethnic groups.* Mol Immunol, 2008. **46**(1): p. 202-6.

40. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

41. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.

42. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.

43. Makarova, K.S., et al., *Evolution and classification of the CRISPR-Cas systems.* Nat Rev Microbiol, 2011. **9**(6): p. 467-77.

44. Sashital, D.G., B. Wiedenheft, and J.A. Doudna, *Mechanism of foreign DNA selection in a bacterial adaptive immune system.* Mol Cell, 2012. **46**(5): p. 606-15.

45. Jiang, F. and J.A. Doudna, *CRISPR-Cas9 Structures and Mechanisms.* Annu Rev Biophys, 2017. **46**: p. 505-529.

46. Sansbury, B.M., A.M. Hewes, and E.B. Kmiec, *Understanding the diversity of genetic outcomes from CRISPR-Cas generated homology-directed repair.* Commun Biol, 2019. **2**: p. 458.

47. Popp, M.W. and L.E. Maquat, *Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine.* Cell, 2016. **165**(6): p. 1319-1322.

48. Paquet, D., et al., *Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9.* Nature, 2016. **533**(7601): p. 125-9.

49. Chu, V.T., et al., *Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.* Nat Biotechnol, 2015. **33**(5): p. 543-8.

50. Mali, P., et al., *CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.* Nat Biotechnol, 2013. **31**(9): p. 833-8.

51. Sakamoto, Y., et al., *Mitotic cells can repair DNA double-strand breaks via a homology-directed pathway.* J Radiat Res, 2021. **62**(1): p. 25-33.

52. He, X., et al., *Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair.* Nucleic Acids Res, 2016. **44**(9): p. e85.

53. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

54. Yang, L., et al., *CRISPR/Cas9-Directed Genome Editing of Cultured Cells.* Curr Protoc Mol Biol, 2014. **107**: p. 31 1 1-17.

55. Qin, W., et al., *Generating Mouse Models Using CRISPR-Cas9-Mediated Genome Editing.* Curr Protoc Mouse Biol, 2016. **6**(1): p. 39-66.

56. Karvelis, T., et al., *Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements.* Genome Biol, 2015. **16**: p. 253.

57. Maeder, M.L., et al., *Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10.* Nat Med, 2019. **25**(2): p. 229-233.

58. Gillmore, J.D., et al., *CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis.* N Engl J Med, 2021. **385**(6): p. 493-502.

59. Frangoul, H., et al., *CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia.* N Engl J Med, 2021. **384**(3): p. 252-260.

60. Moretti, A., et al., *Somatic gene editing ameliorates skeletal and cardiac muscle failure in pig and human models of Duchenne muscular dystrophy.* Nat Med, 2020. **26**(2): p. 207-214.

61. Osborn, M.J., et al., *Base Editor Correction of COL7A1 in Recessive Dystrophic Epidermolysis Bullosa Patient-Derived Fibroblasts and iPSCs.* J Invest Dermatol, 2020. **140**(2): p. 338-347 e5.

62. Chang, K.H., et al., *In vitro genome editing rescues parkinsonism phenotypes in induced pluripotent stem cells-derived dopaminergic neurons carrying LRRK2 p.G2019S mutation.* Stem Cell Res Ther, 2021. **12**(1): p. 508.

63. Musunuru, K., et al., *In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates.* Nature, 2021. **593**(7859): p. 429-434.

64. Razeghian, E., et al., *A deep insight into CRISPR/Cas9 application in CAR-T cell-based tumor immunotherapies.* Stem Cell Res Ther, 2021. **12**(1): p. 428.

65. Gaudelli, N.M., et al., *Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage.* Nature, 2017. **551**(7681): p. 464-471.

66. Rees, H.A., et al., *Analysis and minimization of cellular RNA editing by DNA adenine base editors.* Sci Adv, 2019. **5**(5): p. eaax5717.

67. Grunewald, J., et al., *Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors.* Nature, 2019. **569**(7756): p. 433-437.

68. Jeong, Y.K., B. Song, and S. Bae, *Current Status and Challenges of DNA Base Editing Tools.* Mol Ther, 2020. **28**(9): p. 1938-1952.

69. Koblan, L.W., et al., *Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction.* Nat Biotechnol, 2018. **36**(9): p. 843-846.

70. Wu, J., A.H. Corbett, and K.M. Berland, *The intracellular mobility of nuclear import receptors and NLS cargoes.* Biophys J, 2009. **96**(9): p. 3840-9.

71. Suzuki, K., et al., *In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration.* Nature, 2016. **540**(7631): p. 144-149.

72. Kluesner, M.G., et al., *CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary and immortalized cells.* Nat Commun, 2021. **12**(1): p. 2437.

73. Chen, S., et al., *CRISPR Start-Loss: A Novel and Practical Alternative for Gene Silencing through Base-Editing-Induced Start Codon Mutations.* Mol Ther Nucleic Acids, 2020. **21**: p. 1062-1073.

74. Wang, X., et al., *Efficient Gene Silencing by Adenine Base Editor-Mediated Start Codon Mutation.* Mol Ther, 2020. **28**(2): p. 431-440.

75. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.

76. Kosicki, M., K. Tomberg, and A. Bradley, *Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements.* Nat Biotechnol, 2018. **36**(8): p. 765-771.

77. Anderson, E.M., et al., *Systematic analysis of CRISPR-Cas9 mismatch tolerance reveals low levels of off-target activity.* J Biotechnol, 2015. **211**: p. 56-65.

78. Zheng, T., et al., *Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence.* Sci Rep, 2017. **7**: p. 40638.

79. Vakulskas, C.A., et al., *A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells.* Nat Med, 2018. **24**(8): p. 1216-1224.

80. Kellner, M.J., et al., *SHERLOCK: nucleic acid detection with CRISPR nucleases.* Nat Protoc, 2019. **14**(10): p. 2986-3012.

81. Fu, Y., D. Reyon, and J.K. Joung, *Targeted genome editing in human cells using CRISPR/Cas nucleases and truncated guide RNAs.* Methods Enzymol, 2014. **546**: p. 21-45.

82. Fu, Y., et al., *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* Nat Biotechnol, 2014. **32**(3): p. 279-284.

83. Coelho, M.A., et al., *CRISPR GUARD protects off-target sites from Cas9 nuclease activity using short guide RNAs.* Nat Commun, 2020. **11**(1): p. 4132.

84. Rose, J.C., et al., *Suppression of unwanted CRISPR-Cas9 editing by co-administration of catalytically inactivating truncated guide RNAs.* Nat Commun, 2020. **11**(1): p. 2697.

85. Crudele, J.M. and J.S. Chamberlain, *Cas9 immunity creates challenges for CRISPR gene editing therapies.* Nat Commun, 2018. **9**(1): p. 3497.

86. You, L., et al., *Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research.* Mol Ther Methods Clin Dev, 2019. **13**: p. 359-370.

87. Li, J., et al., *Structure-guided engineering of adenine base editor with minimized RNA off-targeting activity.* Nat Commun, 2021. **12**(1): p. 2287.

88. Wienert, B., et al., *In vitro-transcribed guide RNAs trigger an innate immune response via the RIG-I pathway.* PLoS Biol, 2018. **16**(7): p. e2005840.

89. Lino, C.A., et al., *Delivering CRISPR: a review of the challenges and approaches.* Drug Deliv, 2018. **25**(1): p. 1234-1257.

90. Yip, B.H., *Recent Advances in CRISPR/Cas9 Delivery Strategies.* Biomolecules, 2020. **10**(6).

91. Martufi, M., et al., *Single-Step, High-Efficiency CRISPR-Cas9 Genome Editing in Primary Human Disease-Derived Fibroblasts.* CRISPR J, 2019. **2**: p. 31-40.

92. Sherba, J.J., et al., *The effects of electroporation buffer composition on cell viability and electro-transfection efficiency.* Sci Rep, 2020. **10**(1): p. 3053.

93. Kluesner, M.G., et al., *EditR: A Method to Quantify Base Editing from Sanger Sequencing.* CRISPR J, 2018. **1**: p. 239-250.

94. Brinkman, E.K. and B. van Steensel, *Rapid Quantitative Evaluation of CRISPR Genome Editing by TIDE and TIDER.* Methods Mol Biol, 2019. **1961**: p. 29-44.

95. Clement, K., et al., *CRISPResso2 provides accurate and rapid genome editing sequence analysis.* Nat Biotechnol, 2019. **37**(3): p. 224-226.

96. Concordet, J.P. and M. Haeussler, *CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens.* Nucleic Acids Res, 2018. **46**(W1): p. W242-W245.

97. Kim, D., et al., *Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells.* Nat Methods, 2015. **12**(3): p. 237-43, 1 p following 243.

98. Tsai, S.Q., et al., *GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.* Nat Biotechnol, 2015. **33**(2): p. 187-197.

99. Liang, P., et al., *Genome-wide profiling of adenine base editor specificity by EndoV-seq.* Nat Commun, 2019. **10**(1): p. 67.

100. Eberherr, A.C., et al., *Rescue of STAT3 Function in Hyper-IgE Syndrome Using Adenine Base Editing.* CRISPR J, 2021. **4**(2): p. 178-190.

101. Haapaniemi, E., et al., *CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response.* Nat Med, 2018. **24**(7): p. 927-930.

102. Ihry, R.J., et al., *p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells.* Nat Med, 2018. **24**(7): p. 939-946.

103. Zhou, C., et al., *Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis.* Nature, 2019. **571**(7764): p. 275-278.

104. Kim, H.S., et al., *Adenine base editors catalyze cytosine conversions in human cells.* Nat Biotechnol, 2019. **37**(10): p. 1145-1148.

105. Jeong, Y.K., et al., *Adenine base editor engineering reduces editing of bystander cytosines.* Nat Biotechnol, 2021. **39**(11): p. 1426-1433.

106. K., K., *Broadening the GMO risk assessment in the EU for genome editing technologies in agriculture.* Environ Sci Eur, 2020.

107. Laurence, A.D. and H.H. Uhlig, *When half a glass of STAT3 is just not enough.* Blood, 2016. **128**(26): p. 3020-3021.

108. Dmitrieva, N.I., et al., *Impaired angiogenesis and extracellular matrix metabolism in autosomal-dominant hyper-IgE syndrome.* J Clin Invest, 2020. **130**(8): p. 4167-4181.

109. Matsuzaki, Y., et al., *STAT3 regulates ABCA3 expression and influences lamellar body formation in alveolar type II cells.* Am J Respir Cell Mol Biol, 2008. **38**(5): p. 551-8.

110. Han, S. and R.K. Mallampalli, *The Role of Surfactant in Lung Disease and Host Defense against Pulmonary Infections.* Ann Am Thorac Soc, 2015. **12**(5): p. 765-74.

111. Zhang, H., et al., *Aerosolizable Lipid Nanoparticles for Pulmonary Delivery of mRNA through Design of Experiments.* Pharmaceutics, 2020. **12**(11).

112. Stefanidakis, M., *Preclinical Assessment of In Vivo Gene Editing Efficiency, Specificity, and Tolerability of EDIT-101, an Investigational CRISPR Treatment for Leber Congenital Amaurosis 10 (LCA10)*, in *European Society of Gene & Cell Therapy (ESGCT) 26th Annual Congress*. 2019.

113. Kim, D., et al., *Genome-wide target specificity of CRISPR RNA-guided adenine base editors.* Nat Biotechnol, 2019. **37**(4): p. 430-435.

114. Charlesworth, C.T., et al., *Identification of preexisting adaptive immunity to Cas9 proteins in humans.* Nat Med, 2019. **25**(2): p. 249-254.

115. Wagner, D.L., L. Peter, and M. Schmueck-Henneresse, *Cas9-directed immune tolerance in humans-a model to evaluate regulatory T cells in gene therapy?* Gene Ther, 2021. **28**(9): p. 549-559.

116. Kuranda, K., et al., *Exposure to wild-type AAV drives distinct capsid immunity profiles in humans.* J Clin Invest, 2018. **128**(12): p. 5267-5279.

117. Gaudelli, N.M., et al., *Directed evolution of adenine base editors with increased activity and therapeutic application.* Nat Biotechnol, 2020. **38**(7): p. 892-900.

118. Grunewald, J., et al., *CRISPR DNA base editors with reduced RNA off-target and self-editing activities.* Nat Biotechnol, 2019. **37**(9): p. 1041-1048.

119. Liu, J.J., et al., *CasX enzymes comprise a distinct family of RNA-guided genome editors.* Nature, 2019. **566**(7743): p. 218-223.

120. Hu, J.H., et al., *Evolved Cas9 variants with broad PAM compatibility and high DNA specificity.* Nature, 2018. **556**(7699): p. 57-63.

121. Lee, J., et al., *Using Sniper-Cas9 to Minimize Off-target Effects of CRISPR-Cas9 Without the Loss of On-target Activity Via Directed Evolution.* J Vis Exp, 2019(144).

122. Nishimasu, H., et al., *Engineered CRISPR-Cas9 nuclease with expanded targeting space.* Science, 2018. **361**(6408): p. 1259-1262.

123. Anzalone, A.V., et al., *Search-and-replace genome editing without double-strand breaks or donor DNA.* Nature, 2019. **576**(7785): p. 149-157.

124. Wang, H., et al., *In vivo hematopoietic stem cell gene therapy ameliorates murine thalassemia intermedia.* J Clin Invest, 2019. **129**(2): p. 598-615.

125. Shahmarvand, N., et al., *Mutations in the signal transducer and activator of transcription family of genes in cancer.* Cancer Sci, 2018. **109**(4): p. 926-933.

126. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.

127. Smit; AFA; Hubley, R.G., P., *RepeatMasker Open-3.0*. 1996-2010.

128. Fan, X., et al., *BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping.* Curr Protoc Bioinformatics, 2014. **45**: p. 15 6 1-11.

129. Rausch, T., et al., *DELLY: structural variant discovery by integrated paired-end and split-read analysis.* Bioinformatics, 2012. **28**(18): p. i333-i339.

130. Abyzov, A., et al., *CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing.* Genome Res, 2011. **21**(6): p. 974-84.

131. Layer, R.M., et al., *LUMPY: a probabilistic framework for structural variant discovery.* Genome Biol, 2014. **15**(6): p. R84.

132. Chen, X., et al., *Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications.* Bioinformatics, 2016. **32**(8): p. 1220-2.

133. Ye, K., et al., *Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads.* Bioinformatics, 2009. **25**(21): p. 2865-71.

134. Robinson, J.T., et al., *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.

135. Zook, J.M., et al., *Extensive sequencing of seven human genomes to characterize benchmark reference materials.* Sci Data, 2016. **3**: p. 160025.

List of figures

List of tables

Acknowledgement

First I want to express my sincere gratitude to Prof. Dr. Ellen Renner who accepted me as her PhD student and gave me the possibility to work on this exciting project. Ellen, thank you for giving me room to explore on my own and on the same time supporting me whenever necessary! Thank you for your many efforts and your trust in me.

Also a big thankyou to my supervisor Dr. Beate Hagl. Thank you, Beate, for teaching me many many things and always helping me. Be it a friendly hint on how to improve the setup of an experiment, help with writing abstracts, putting together presentations or writing a paper.

Further, I want to thank Prof. Dr. Benjamin Schusser for reviewing my project plan, attending several thesis committee meetings and always being available for help and advice.

And of course a big thankyou to all the helping colleagues in the lab, namely Christine Wolf, Dr. André Maaske, Daniela Kreilinger, Michelle Plummer, Anica Lechner and Dr. Renate Effner. Christine, your westernblotting and qPCR skills are unparalleled, thank you for your hard work regarding our paper! André, thank you for your help with the iPSC part of the paper and being the perfect partner for ping-ponging ideas and concepts. Dani, thank you for your help in the lab and lending an ear for any problems. Michelle, thank you for your help with the DNA-extraction and Sanger sequencing. Anica, thank you for helping in the lab, especially tending to iPSCs! And Renate thank you for always helping when necessary and showing me how to work with the FACS.

Also thank you to our collaboration partners Anna Pertek and Ejona Rusha for teaching me how to culture iPSCs and performing iPSC assays for our Paper. Thank you to Miriam Kastlmeier and Carola Voss for performing the iPSC differentiation and subsequent analysis. Thank you Riccardo Berutti and Elisabeth Graf for performing the WGS and subsequent analysis! Also thank you to Florian Giesert, your help and many advice regarding gene editing were immensely helpful!

Also a big thank you to my co-students Daniele, Annabel and Alisa.

A final thank you to Isabella who always supported me, especially during the difficult parts of the thesis. Without you I would not have been able to do this.