Developing a Gene Editing Strategy to Target the COL1A1 Promoter: Significance for Fibrotic Disorders and Pioneering Beyond the Boundaries of CRISPRi System

Inaugural Dissertation

Zur

Erlangung des Doktorgrades

philosophiae doctor (PhD) in Health Sciences

der Medizinischen Fakultät

der Universität zu Köln

vorgelegt von

Karim Daliri

aus Shiraz, Iran

Hundt Druck, Köln

Köln, 2024

Betreuer*in: PD Dr. Kurt Paul Pfannkuche

Gutachter*in: Prof. Dr. Gerhard Sengle

Prof. Dr. Marcel Halbach

Datum der Mündlichen Prüfung: 13.05.2024

Acknowledgment

I would like to express my profound gratitude to Jürgen Hescheler, the director of the Physiology Center, for providing me with the opportunity to engage in scientific activities during my doctoral studies.

My deep appreciation goes to my supervisor, PD.Dr. Kurt Pfannkuche, who directed me into the fascinating realm of gene editing. His expertise and passion for the field have been a constant source of inspiration for me. His guidance and support have been instrumental in my academic journey.

I would also like to extend my thanks to all the exceptional colleagues whose support has greatly facilitated my work environment. Their dedication and teamwork have made a significant impact on my research.

Despite originating from the continent of Asia, s Iran and the city of Shiraz, my scientific journey to Cologne has been an enlightening experience. The transition was challenging, but the knowledge and skills I have gained are invaluable.

My heartfelt gratitude goes to my parents,wife, and sisters, who, despite the distance, have consistently supported me with their positivity. Their unwavering faith in my abilities has been a driving force behind my accomplishments.

As I continue my journey in the world of science, I am reminded of the words of the great Persian poet, Hafez Shirazi. His verse serves as a reminder of my roots and the wisdom of my heritage:

ساقی به نور باده برافروز جامِ ما
مطرب بگو که کارِ جهان شُد به کامِ ما

ما در پیاله عکس رخ یار دیده‌ایم
ای بی‌خبر ز لذتِ شُربِ مدامِ ما

هرگز نمیرد آن که دلش زنده شد به عشق
ثبت است بر جَریدهٔ عالم دوام ما

حافظ شیرازی

Table of Contents

Abbreviation

SEM-Scanning Electron Microscopy

RNA-seq-RNA-sequencing

DEG-Differential Expression Gene

PE-Prime Editing

BE-Base Editing

ABE-Adenine Base Editors

pegRNA-Prime editing guide RNA

CRISPR-Clustered Regularly Interspaced Short Palindromic Repeats

NHEJ-Non-Homologous End Joining

ZFN-Zinc Finger Nucleases

TALE-Transcription Activator like Effector

PAM-Protospacer Adjacent Motif

ECM-Extracellular Matrix

CBE-Cytosine Base Editor

crRNA-CRISPR RNA

tracrRNA-transactivating crRNA

sgRNA-single guide RNA

DSBs-Double Strand DNA Breaks

RT-Reverse transcriptase

Col-I-Collagen type I

GFP-Green Fluorescent Protein

PCR-Polymerase Chain Reaction

qPCR-quantitative Polymerase Chain Reaction

PBS-Phosphate Buffered Saline

TIDE-Tracking Indels Decomposition

pegRNA-Prime editing guide RNA

epegRNA-engineered prime editing guide RNA

NGS-Next-Generation Sequencing

LC-MS-Liquid Chromatography Mass Spectrometry

TEAB-Triethylammoniumbicarbonate

DTT-Dithiothreitol

KEGG-Kyoto Encyclopaedia of Genes and Genome

Abstract

CRISPR-Cas9 gene editing is a promising therapy for pathogenic mutations, but current editing approaches are mainly aimed at relatively small groups of patients with specific mutations. Here, we introduce a new application of gene editing to develop an antifibrotic strategy that could potentially be used to address a wide range of organs with fibrosis.

In our research on adenine base editing, we successfully targeted the CCAAT box inside the promoter of Col1a1gene promoter in fibroblasts to substitute a twin adenine (AA) with guanine (GG). This innovative modification led to reduction in collagen indicating the precision and efficiency of our editing approach.To assess the broader implications of this intervention; we employed high-throughput multi-omics technologies. Through transcriptomics (RNA-seq) and proteomics (tandem mass spectrometry), we aimed to evaluate our gene editing approach on the molecular signature. Notably, our investigation revealed that the reduction in Col1a1 expression influencing in PI3K-Akt signalling pathway which is linked with various organ fibrosis such as liver, kidney, lung and heart. This suggests the specificity and efficacy of our editing approach for possible antifibrosis goals.

By targeting CCAAT promoter region, our method presents a versatile alternative to the CRISPR interference (CRISPRi) system. This addresses its key drawbacks, such as reversible and transient outcomes, which are suboptimal especially in the context of therapeutic applications where permanent cure is desired. Col1a1 promoter editing may thus represent a permanent and optimal therapeutic strategy across diverse fibrotic disorders.

Zusammenfassung

Die CRISPR-Cas9-Geneditierung ist eine vielversprechende Therapie für pathogene Mutationen, aber die derzeitigen Ansätze zielen hauptsächlich auf relativ kleine Patientengruppen mit spezifischen Mutationen ab. In dieser Arbeit stellen wir eine neue Anwendung der Geneditierung vor, um eine antifibrotische Strategie zu entwickeln, die potenziell für eine breite Palette von Organen mit Fibrose eingesetzt werden könnte.

In unserer Forschung zur Adenin-Baseneditierung haben wir erfolgreich die CCAAT-Box im Promotor des Col1a1-Gens in Fibroblasten angesteuert, um ein doppeltes Adenin (AA) durch Guanin (GG) zu ersetzen. Diese innovative Modifikation führte zu einer Verringerung des Kollagens, was die Präzision und Effizienz unseres Editing-Ansatzes belegt. Um die breiteren Auswirkungen dieses Eingriffs zu beurteilen, haben wir genomweite Multi-Omics-Technologien eingesetzt. Mittels Transkriptomik (RNA-Seq) und Proteomik (Tandem-

Massenspektrometrie) wollten wir unseren Geneditierungsansatz auf die molekulare Signatur hin untersuchen.

Interessanterweise ergab unsere Untersuchung, dass die reduzierte Col1a1-Expression den PI3K-Akt-Signalweg beeinflusst, der mit verschiedenen Organfibrosen wie Leber-, Nieren-, Lungen- und Herzfibrose in Verbindung steht. Dies deutet auf die Spezifität und Wirksamkeit unseres Editing-Ansatzes für mögliche antifibrotische Ziele hin.

Mit der gezielten Bearbeitung der CCAAT-Promotorregion stellt unsere Methode eine vielseitige Alternative zum CRISPR-Interferenz(CRISPRi)-System dar. Sie behebt dessen entscheidende Nachteile, wie reversible und transiente Effekte, die im Kontext therapeutischer Anwendungen, wo eine dauerhafte Heilung angestrebt wird, suboptimal sind. Die Bearbeitung des Col1a1-Promotors könnte daher eine dauerhafte und optimale therapeutische Strategie für verschiedene fibrotische Erkrankungen darstellen.

1. Introduction

Historical Background The field of biology emerged from the study of living organisms and their behaviors and physiological characteristics. Early investigations revealed that cells are the fundamental units of life capable of independent reproduction, and genes play a crucial role in inherited traits. These findings led to the understanding that all living organisms possess a genome, which is inherited across generations and contains instructions for essential life processes, including physical appearance, reproduction, behavior, and interactions with the environment. Consequently, the pursuit of linking genotype (genetic makeup) to phenotype (observable traits) has been an ongoing endeavor.

In recent decades, significant breakthrough in biology have revolutionized our understanding of molecular processes in the living organisms. The development of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) technology has surpassed previous macroscopic biological comprehension and has profoundly transformed our perception of organisms, tissues, and cells. This novel gene-editing tool, CRISPR/Cas, has obtained considerable attention within the scientific community.

1.1. The Evolution of Genome Engineering Technologies

Gene targeting is considered the most reliable approach for analyzing the functionality of genes and their variants. This approach involves intentionally replacing or modifying genetic codes with new ones. Researchers initially showed that homologous recombination machinery could be used to introduce exogenous DNA into a specific locus in mice in the late 1980s, and later in human cells. However, the method was restricted by the low frequency of template integration into the genome, and the possibility of off-target insertions. As a result, rigorous screening procedures were required for proper clone selection, a time-consuming and labour-intensive process that limited the technique's effectiveness (Capecchi, 2022; Rajewsky et al., 1996; Rong & Golic, 2000). A significant advancement in genome engineering technology was the discovery that double-stranded DNA breaks (DSBs) can significantly stimulate cellular DNA repair mechanisms. When DSBs occur, they are usually repaired through one of two pathways: non-homologous end joining (NHEJ) or homologous recombination. NHEJ repairs the break by simply joining the broken ends of the DNA, often leading to small mutations (Indel) (Komor et al., 2018; Zaboikin et al., 2017). These two studies established the basis for

genome editing. Meganucleases were effective in specifically cleaving chromosomal DNA in cells due to their long DNA recognition sequences, which ranged from 12 to 40 base pairs in length. However, these enzymes were challenging to reprogram, which made them less practical for most genome editing applications (Kan et al., 2014; Silva et al., 2011). As a result, the focus shifted towards developing programmable tools that could target any specific region of the genome.

1.2. Engineering Gene Target Specificity with ZFNs and TALENs

A significant development occurred in the 1990s when Srinivasan Chandrasegaran, a biochemist at The Johns Hopkins University, discovered that FokI, a type IIS restriction enzyme, could be divided by a protease into two separate domains: a DNA-binding domain and a DNA-cutting nuclease domain (Chandrasegaran & Carroll, 2016; Mani et al., 2005). The discovery that FokI could be divided into separate DNA-binding and DNA-cutting domains suggested the possibility of creating a new sequence-specific nuclease. In 1996, Chandrasegaran and his research team were able to demonstrate in vitro that zinc finger nucleases (ZFNs) could cleave target DNA site-specifically by fusing the FokI nuclease domain to zinc finger proteins (L. Li et al., 1992; Mino et al., 2009). This discovery immediately suggested that a novel sequence-specific nuclease could be created by fusing the FokI nuclease domain to a DNA-binding protein.

Zinc finger proteins differ from meganucleases in that they recognize target DNA in a modular manner. Each zinc finger protein is composed of at least three zinc finger domains, and a single domain interacts with a 3-bp sequence, making them highly suitable for the development of sequence-specific DNA-binding proteins (Alwin et al., 2005; Petersen & Niemann, 2015). Subsequently, other researchers also applied custom-designed ZFNs for in vivo genome editing (Maeder et al., 2008; Urnov et al., 2005).

Although there were numerous successful illustrations of genome editing using ZFNs in cells and entire organisms, creating ZFNs with specific target specificities remained a challenging task. Many ZFNs were found to be cytotoxic, likely due to cleavage at numerous off-target sites that had high sequence homology with on-target sites (Cradick et al., 2011). Later on two new gene editing technologies were introduced. Transcription-activator-like effector (TALE) proteins derived from Xanthomonas (a genus of bacteria) interact with DNA through a code

that is different from that of zinc fingers. Each repeat domain in TALE proteins recognizes a single base, and it is possible to mix and match four different repeat domains to generate new DNA-binding proteins. By linking these to the FokI domain, a new set of programmable nucleases, known as TALENs, can be created. Importantly, TALENs exhibited minimal cytotoxicity in human cells. However, the emergence of CRISPR-Cas9 technology quickly displaced TALENs from their dominant position in the field of genome editing (Becker & Boch, 2021; Cradick et al., 2011).

1.3. RNA-guided Genome Editing with CRISPR–Cas9

The RNA-guided CRISPR endonuclease system was first discovered in Escherichia coli as descibed by its distinctive genomic structure. This system developed as an adaptive immune mechanism, which bacteria and archaea use through CRISPR-associated (Cas) genes to integrate foreign genetic material into the CRISPR locus (Daliri et al., 2024; Ishino et al., 2018). The integrated material is then transcribed into RNA templates, that guide the targeted destruction of mobile genetic elements. So far, three types of CRISPR systems have been identified, each with different mechanisms of action (Daliri et al., 2024; Makarova et al., 2011).

Unlike type II CRISPR Cas system, which utilizes a single endonuclease, Cas9, to identify and cleave target DNA, type I and III systems employ a collection of Cas genes to perform RNA processing, target recognition, and cleavage. A pair of non-coding RNAs, which include a guide-bearing and different crRNA (CRISPR RNA) as well as a necessary auxiliary transactivating crRNA (tracrRNA), directs the Cas9 endonuclease to the target DNA (Makarova et al., 2015) The crRNA is comprised of a 20-nt guide sequence, or spacer, that provides target specificity via Watson-Crick base-pairing with the target DNA (Chylinski et al., 2013).

The CRISPR-Cas system obtained from Streptococcus pyogenes is characterized by a target DNA sequence that is always followed by a 5'-NGG protospacer adjacent motif (PAM). The PAM can vary depending on the specific CRISPR system (Geng et al., 2016).It is possible to redirect Cas9 towards any target with an appropriate PAM by modifying the 20-nt guide sequence within the single guide RNA (sgRNA). Additionally, a chimeric, sgRNA can be created by artificially linking segments from the crRNA and tracrRNAs (Filippova et al., 2019; Jinek et al., 2012; Kiani et al., 2015).

1.3.1 Challenges and Issues Related to the Genome Engineering Technologies

Several investigations regarding Cas9 specificity have shown that although each base within the 20-nt guide sequences contributes to overall specificity, multiple mismatches between the sgRNA and its complementary DNA can still be tolerated to some degree. This means that Cas9 can cleave genomic regions that exhibit imperfect homology with the target 20-nt guide sequence, potentially resulting in off-target DSBs and NHEJ repair (Cho et al., 2014; Koo et al., 2015). When DSBs are repaired by NHEJ, non-specific indels insertions or deletions are created at the DSB site, which can cause frameshifts leading to gene dysfunction (Malzahn et al., 2017).

The efficiency of HDR -based gene editing approach that uses a homologous DNA template to repair a DSB in the target gene- is not the same across different mammalian cell types because the HDR pathway is only operational during particular cell cycle phases. Furthermore, the HDR is in constant competition with NHEJ for repairing DSBs, and NHEJ usually outcompetes HDR. As a result, enhancing HDR yields and/or reducing NHEJ rates through the development of novel approaches and tools have become among major challenging areas of research in this field (V. T. Chu et al., 2015; Devkota, 2018).

1.4 The New Era of Genome Editing

1.4.1 Base Editors

One CRISPR-based technology, which developed to address multiple challenges of creating targeted single-nucleotide alterations in a precise and efficient manner, is base editing (BE). Base editing is a distinctive method of gene editing that does not involve cleaving the nucleic acid backbone. Instead, it utilizes enzymes that modify DNA and are fused to a programmable DNA-targeting agent (Jeong et al., 2020). In 2016, introducing of the first cytosine base editors (CBEs) were reported. These CBEs were developed using natural single-stranded DNA deaminase domains that convert cytosine nucleotide into uracil. During DNA replication and repair uracil is considered as thymine resulting in a conversion of the original cytosine to a thymine (Komor et al., 2016) (Figure 1-1).

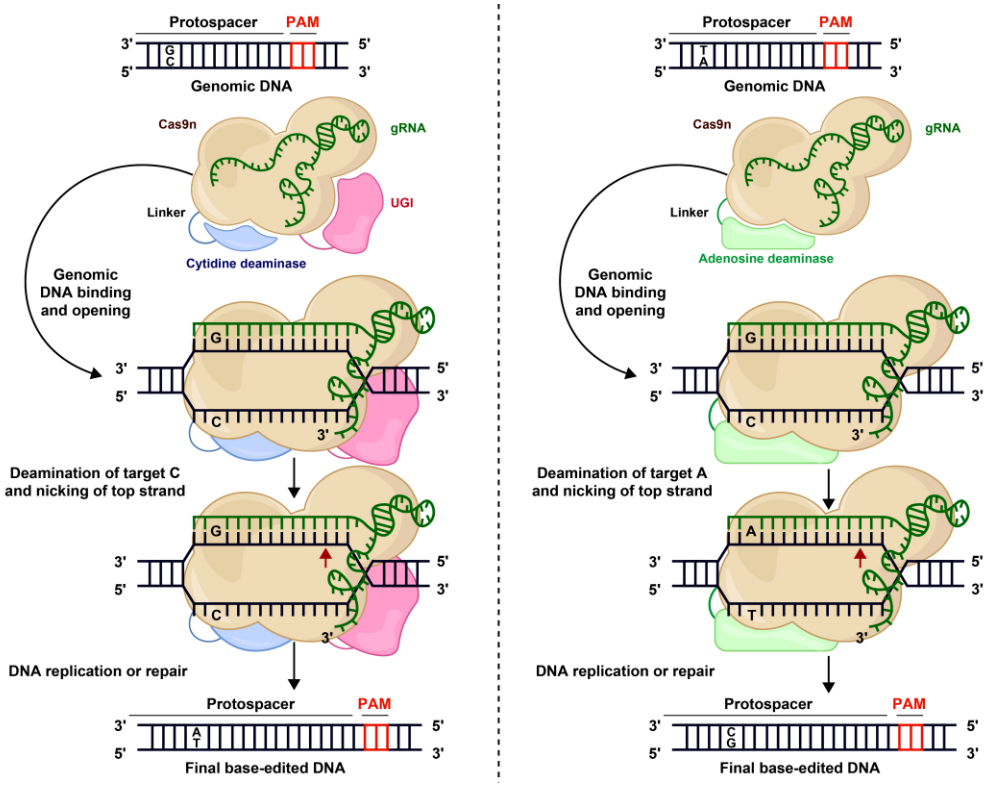


Figure1-1. Illustrating the mechanism of Adenine and Cytosine base editors. On the right, the adenine base editor (ABE) mechanism shows the conversion of Adenine (A) to Inosine (I), which is read as Guanine (G) during DNA replication. On the left, the cytosine base editor (CBE) mechanism depicts the conversion of Cytosine (C) to Uracil (U), which is read as Thymine (T) in the subsequent replication even

CBEs works by converting C•G base pairs into U•G base pairs and finally into T•A base pairs after DNA repair. To ensure the deaminase acts only on the intended genomic site, nuclease-impaired Cas9 is fused to single-strand cytidine deaminases. When the guide RNA and catalytically inactivated Cas9 engage with the target DNA, a single-stranded genomic DNA stretch (known as the R-loop) is created. Within this R-loop, the deaminase converts cytosine in editing window into uracil. The cellular DNA repair machinery can then resolve the resulting base mismatch to the desired state or revert to the initial state. However, using a Cas9 nickase to nick the non-deaminated strand directs the mismatch repair towards replacing that strand with a sequence based on the edited one (Anzalone et al., 2020).

Later, adenine base editors (ABEs) was introduced which convert A•T base pairs into G•C base pairs. Since there are no known natural deaminases that function on deoxyadenosine, developing ABEs required engineering a deoxyadenosine deaminase through laboratory evolution. The deoxyadenosine deaminase that has been utilized in all current ABEs is a variant of the laboratory-evolved enzyme (Gaudelli et al., 2017). Over 100 variations of base editors

have been developed since their launch, with modifications tailored to specific sequence contexts or designed to control the editing window, activity, or specificity of the editors (Porto et al., 2020).

Base editors have gained significant attention as new genome editing tools due to several advantages: (1) they are not depending on cellular HDR machinery, which means they can efficiently install programmed edits even in non-dividing cells (2) the editing results are highly precise (3) base editing does not require DNA template delivery and can be performed entirely with mRNA or ribonucleoprotein (RNP) agents and (4) SNPs (single nucleotide polymorphisms) are the most prevalent type of human disease-associated mutations and base editors have the potential to correct over 70% of disease-associated SNPs (Koblan et al., 2018, 2021).

The primary limitation of base editors is the requirement for precise location of the target within the optimal editing window while avoiding unwanted bystander edits. The optimal editing window for the original base editors is typically nucleotides 4-7 of the protospacer, with the first nucleotide being the farthest from the protospacer-adjacent motif (PAM) (Huang et al., 2021). Efficient base editing may not be possible if the protospacer cannot be appropriately positioned, as Cas9 necessitates a PAM to bind its target sequence. Numerous new Cas and deaminase variants have been described to overcome these limitations, with some having altered PAM specificities or editing windows (S. H. Chu et al., 2021; Tan et al., 2019).

Finally, in terms of translational research ex vivo base editing of cells followed by transplantation has been offered a promising therapeutic approach. For example, hematopoietic stem and progenitor cells (HSPCs) can be utilized for autologous bone marrow transplantation and can be effectively edited in most of defect alleles by base editors before transplantation (Zeng et al., 2019, 2020). CBEs and ABEs were found to enable editing of over 90% of three different sites in primary human T cells, while human HSPCs can be edited with an efficiency of greater than 80% at two different sites (Gaudelli et al., 2020; Webber et al., 2019). Base editing has also achieved an efficiency of over 80% in human induced pluripotent stem cells (iPSCs) (Nami et al., 2021).

1.4.2. Prime Editors

In 2019, prime editing (PE) as another revolutionary method for making precise DNA edits, without the need for DSBs was introduced by Anzlone et al. Prime editors consist of an engineered reverse transcriptase (RT) fused to modified Cas9 nickase generating a nick in the R-loop at the target DNA site. The prime editing guide RNA (pegRNA) has a 3' extension that binds to the nicked target DNA strand and this complex serves as a primer-template complex for RT, to add the desired sequence according to a template encoded in the pegRNA (Anzalone et al., 2019). Prime editors were demonstrated to be capable of effecting changes to any single base pair, in addition to causing deletions with a minimum of 80 nucleotides and insertions with a minimum of 44 nucleotides (J. Yan et al., 2020).

The two-components of PE system are modified Cas9 nickase and pegRNA that typically achieve editing outcomes ranging from 5% to 20% with less than 1% indels off-targets. To enhance editing efficiencies a second gRNA can be used for directing the Cas9 nickase to nick the non-edited strand without a 3' extension for RT. However, this editing approach (known as PE3) has the potential to increase indel formation in approximately 10% of alleles, despite increasing editing efficiencies by 1.5- to 4-fold (Jang et al., 2022). Later PE4 , PE5 and PEmax based on the presence or absence of a nicking gRNA ,using codon usage, nuclear localization signal (NLS) architectures, and Cas9 variants in order to increase PE efficiency were introduced (Y. Liu et al., 2021; Scholefield & Harrison, 2021).

1.5. The Collagen Superfamily

Collagens, which make up around 30% of the total protein mass in mammals, are considered the most abundant proteins in human body as well. Since the initial discovery of collagen I by Miller and Matukas in 1969, researchers have identified an additional 28 types of collagen (Miller & Matukas, 1969; Ricard-Blum, 2011). The development of innovative molecular biology tools has significantly accelerated the exploration and identification of these various collagen types (Exposito et al., 2010; M. K. Gordon & Hahn, 2010). Numerous papers have been published on the collagen family, shedding light on the structure and biological functions of collagens and addressing the question of what truly constitutes collagen. While the question "What is collagen, what is not?" may still be valid, these publications have provided valuable answers and fresh perspectives on the subject (Gay & Miller, 1983). In addition to the 28 different collagen types, the collagen family exhibits further diversity through the presence of multiple molecular isoforms within each type. For instance, collagen IV have various isoforms

such as IVa, and there are hybrid isoforms composed of α chains from two different collagen types (type V/XI molecules) indicating complexity and diversity of the collagen family (Birk & Brückner, 2010; Mienaltowski & Birk, 2014).

Collagen α chains vary in size, ranging from 662 to 3152 amino acids in humans. They can form either homotrimers (e.g., collagen II) when the three α chains are identical or heterotrimers (e.g., collagen IX) which they are different (Martel-Pelletier et al., 2008). The stability of the triple helix structure is mainly ensured by factors such as the presence of glycine every third residue, a notable amount of proline and hydroxyproline, and electrostatic interactions (Ricard-Blum et al., 2000).

1.5.1. Roles of Collagens

Fibrillar collagens have a significant structural role in shaping and providing mechanical properties to various tissues. They contribute to the molecular architecture and provide tensile strength in the skin, as well as resistance properties in ligaments. While some collagens were previously considered "minor," they play a crucial role in tissue integrity (Martel-Pelletier et al., 2008).

During fibrosis, there is an excessive accumulation of collagen within the extracellular matrix (ECM). Targeting fibrillogenesis, the process of collagen fibril formation, has emerged as a novel approach to mitigate fibrosis. This can be achieved by inhibiting the peptide-mediated interactions between collagen molecules, thereby preventing excessive collagen deposition (H. J. Chung et al., 2008). Collagens interacts with cells through multiple receptors, and their involvement in regulating cell growth, differentiation, and migration via receptor binding is extensively documented. Indeed, collagens are present within the ECM and they actively engage in interactions with cells through various types of receptors (Leitinger, 2011; Rosso et al., 2004).

1.5.2. Collagen Type I

Collagen type I (Col-I) is comprised of two genes: COL1A1, situated at 17q21.3-q22 on the long arm of the chromosome, and COL1A2, located at 17q21.3-22.1 within the same chromosome region. Col-I is available extensively in both healthy and diseased tissues. The intron sizes of these two genes exhibit significant variation, with COL1A1 potentially reaching up to 18 kb in length, while COL1A2 can extend up to 38 kb (Naomi et al., 2021).

Collagen has a crucial triple-helix structure that facilitates and sustains interactions between cells and the ECM. There are currently 28 known types of collagen, which are widely

distributed in various tissues such as the skin, bones, organ capsules, tendons, and cornea. The production of Col-I is a multifaceted process involves several steps, beginning with gene transcription in the cell nucleus, followed by the formation of heterotrimeric structures and fibrils (Engel & Bächinger, 2005). Indeed, The synthesis encompasses transcription and translation, post-translation modifications, the assembly of the triple helix structure, and ultimately, the secretion of Col-I (Naomi et al., 2021). The synthesis of collagen typically follows a general pathway, which involves the removal of the propeptide followed by the formation of lysyl-crosslinks (Ishikawa & Bächinger, 2013). This process contributes to the overall synthesis of collagen (Figure1-2).

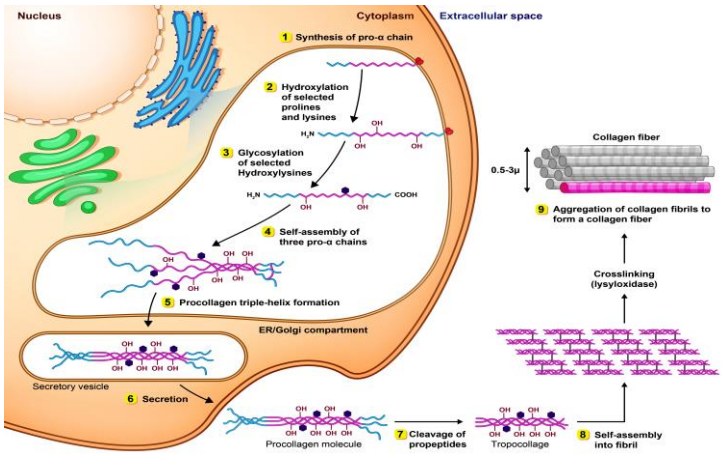


Figure 1-2. Illustrating the intricate processes involved in the production of Col-I. Each stage, from gene expression to protein assembly, is detailed to provide a comprehensive understanding of Col-I biosynthesis. Image modified from(Levine & Levine, 2011).

The regulation of transcription in Col-I is significantly influenced by several factors such as the cell type, growth factors, and cytokines. The presence of exons in the Col-I gene ranges from 3 to 117, which explains the diversity of collagen mRNAs. This variation mostly arises from the utilization of different initiation sites or alternative splicing of exons in the gene transcription process (Sorushanova et al., 2019). Subsequently, the mature mRNA is transported to the cytoplasm where it undergoes translation at the endoplasmic reticulum producing pre-procollagen. During translation, it is referred as the pro-polypeptide chain. With the aid of a signal recognition domain, this chain is transferred into the lumen of endoplasmic reticulum for post-translation modifications (Gistelinck et al., 2016; Jun Liu et al., 2007).

During the synthesis of a newly unfolded polypeptide chain, multiple post-translational modifications (PTMs) are added simultaneously. Specific enzymes that are involved in collagen-related processing including assembly and secretion of procollagen facilitate these

modifications, such as lysyl and prolyl hydroxylations as well as hydroxyl glycosylation. These enzymes form complexes with other chaperones and catalysts are responsible for protein folding (Koide & Nagata, 2005).The formation of collagen trimeric monomers, composed of three α chains, is influenced by various factors. N-linked carbohydrates and intra-chain disulfide bonds play critical roles in stabilizing the globular structure of the propeptide. Once procollagen is assembled, the helical structure is subsequently packaged into secretory vesicles within the Golgi apparatus. These vesicles are then released into the extracellular space. After secretion, the trimers of procollagen undergo further processing, which varies depending on the specific type of collagen (Koide & Nagata, 2005).

1.6. Collagen Functions in Physiological and Pathological Conditions

The extracellular matrix (ECM) refers to the non-cellular elements present in tissues and organs. The significance of ECM in cell biology, especially in the context of different diseases such as cancer pathogenesis, has gained significant attention as it has been documented that for example tumor cells and the ECM have a strong and reciprocal interaction (Crotti et al., 2017; Yue, 2014). Furthermore, the ECM plays a key functional role in various cellular processes such as cell morphology, cell migration, cell differentiation, and cell-cell interactions. Conversely, cells themselves actively modify the composition, structure, and mechanical properties of the ECM (Fane & Weeraratna, 2020).

In addition to polysaccharides, ECM predominantly consists of two other main types of molecules including proteoglycans and fibrous proteins, such as collagen and fibronectin. Proteoglycans occupy the extracellular interstitial space, while fibrous proteins serve as the primary structural components providing support to the ECM (Helm & Potts, 2012) (Figure 1-3).

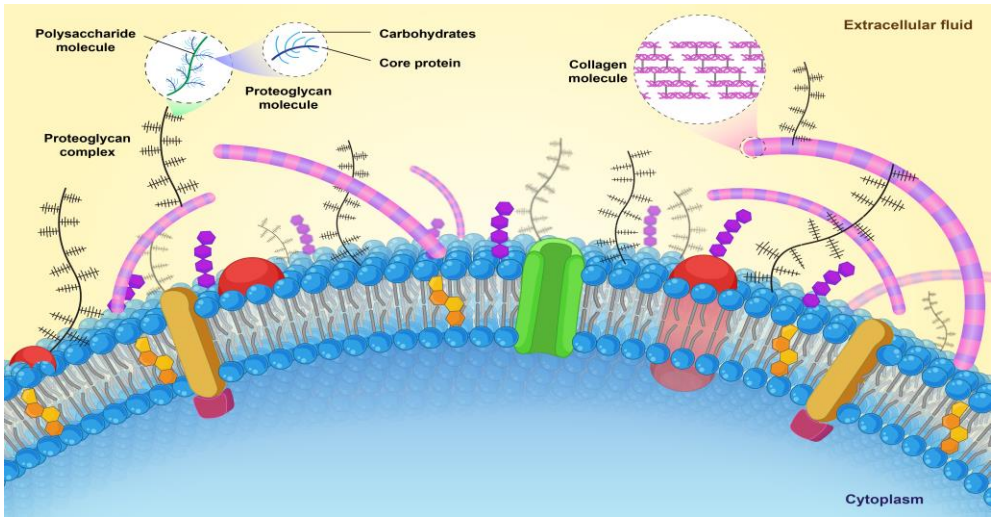


Figure 1-3. Graphical representation showing the composition of the extracellular matrix (ECM). The figure emphasizes the various predominant types of molecules that are integral to the ECM structure.

Collagens also act as a barrier separating the epithelium from the underlying tissue. The basement membrane serves as the initial obstacle in the spread of cancer cells. The ECM, which provides structural support, varies across tissues and changes over time due to factors like homeostasis, diseases, injuries, and aging. Various cells including fibroblasts, epithelial cells, macrophages, and cancer cells, continually modify the ECM network (Goldberg & Smith, 2004; Sila-Asna et al., 2007) via influencing on collagen production.

The balance and stability of normal tissues are maintained through tightly regulated interactions between cells and their surrounding environment. This controlled interaction ensures an optimal environment for normal physiological processes. However, in many types of fibrotic conditions- the last stage of organ failure- and tumors this tissue homeostasis is disrupted, leading to abnormal tissue organization. This disruption creates conditions that promote tumor growth and metastasis. The ECM has emerged as a key component in the tumor microenvironment and has been extensively studied for its role in tumor progression (Najafi et al., 2019).

Fibrosis can affect any organ or tissue, making it associated with a range of diseases. Chronic fibroproliferative diseases account for 45% of all deaths globally (Figure 1-4). This highlights the importance of addressing these diseases due to their significant impact on quality of life and the associated healthcare costs resulting from organ failure (Henderson et al., 2020; Pinzani, 2008). There is also an increasing demand for organ transplants despite limited availability, often leading to fatal outcomes (Watson & Dark, 2012).

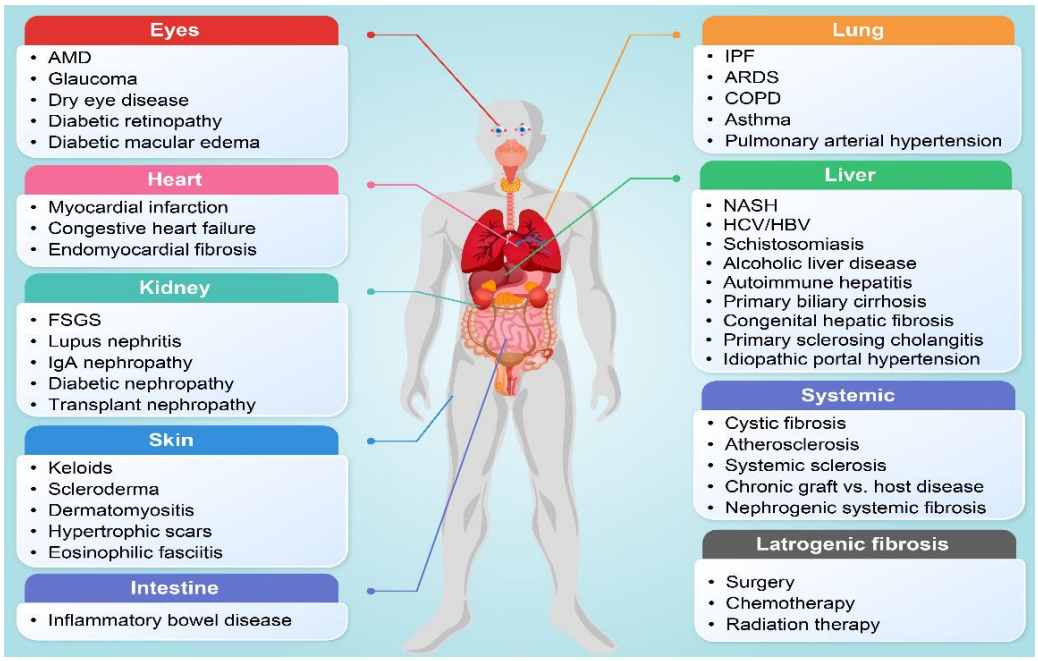


Figure 1-4. Graphical representation illustrating the range of diseases associated with collagen in human mortality rates. Age-related macular degeneration(AMD), Focal Segmental glomerulosclerosis(FSGS), Idiopathic pulmonary fibrosis (IPF), Acute respiratory dis distress syndrome (ARDS), Chronic obstructive pulmonary disease (COPD), Non-alcohol related steatohepatitis (NASH), Hepatitis C virus (HCV), Hepatitis B virus (HBV). Image modified from (Morten Asser Karsdal et al., 2014).

Furthermore, the severity and limited treatment options for these diseases, along with their high prevalence in most cases and vague status in certain fibrotic conditions, have recently captured the attention of major pharmaceutical companies in this field (McVicker & Bennett, 2017).

Fibrotic diseases share a common characteristic, which is the disruption of normal tissue remodelling. This disruption causes an abnormal buildup of ECM components, resulting in an ECM that has altered structural and signalling properties (Wynn, 2007) (Figure 1-5).

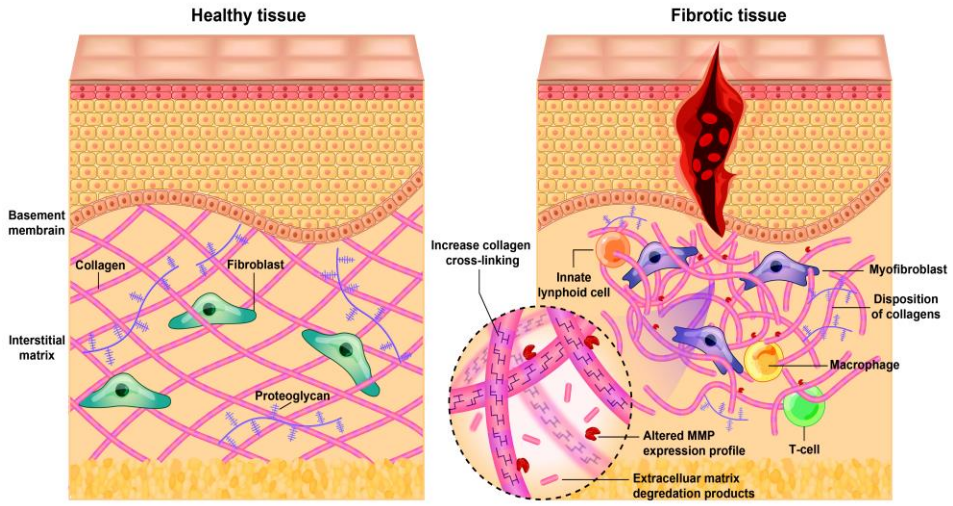


Figure 1-5. Displaying tissue remodelling. On the left side, normal tissue remodelling and ECM composition are depicted. Conversely, the right side illustrates the characteristic disruption found in fibro proliferative diseases, with an abnormal buildup of ECM components, leading to modified structural and signalling attributes.

For a long time, it was widely believed that fibrotic tissue lacked the capacity to reverse and was considered an irreversible scaffold. However, this notion has been proven incorrect, as fibrosis is now understood to be a dynamic and reversible condition resulting from an ongoing remodeling process (Henderson et al., 2020). Consequently, it can be effectively addressed through intervention. Currently, there are notable examples of interventions targeting fibrosis to reverse, including the use of antiviral therapy for chronic hepatitis B and the eradication of chronic hepatitis C using interferon-α-based. These interventions have shown promising results, particularly in cases of liver fibrosis (Henderson et al., 2020; Schuppan & Kim, 2013).

Historically, growth factors, cytokines, and various small molecules were predominantly acknowledged for their roles in intercellular, intracellular, and paracrine signalling (Schultz & Wysocki, 2009). Nevertheless, it has become increasingly clear that ECM also assumes significant importance in facilitating direct or indirect paracrine and even endocrine communication.

The ECM has the ability to regulate cell phenotype through two main mechanisms. Firstly, it serves as a reservoir of powerful signalling fragments, which can influence cell behaviour. Secondly, the ECM interacts with cells through specific receptors like integrins and proteoglycans, directly impacting cell phenotype (Cong Li et al., 2023). An instance of this is the overactivity of cardiac fibroblasts, which can lead to the excessive synthesis and accumulation of the ECM proteins within the myocardium, a condition known as cardiac fibrosis which has detrimental effects on the function of the heart (Fan et al., 2012; van

Nieuwenhoven & Turner, 2013). In addition to being recognized as the primary source of the ECM proteins, fibroblasts also produce various cytokines, peptides, and enzymes that serve important functions. For instance, matrix metalloproteinases (MMPs) and their inhibitors play a direct role in controlling the turnover of ECM components, thereby influencing the process of cardiac remodelling (Krenning et al., 2010). Ultimately, the progressive synthesis and build-up of fibrotic deposits can result in the disruption of myocytes interactions within the myocardium. This can lead to a significant decline in cardiac function (Moore-Morris et al., 2015; Ravassa et al., 2023).

1.7. Collagen and Fibrosis

Fibrosis is characterized by the excessive build-up of ECM proteins within tissues, signifying an abnormal and unregulated collagen production. This process significantly contributes to organ dysfunction in a wide range of diseases, including interstitial lung diseases, liver cirrhosis, progressive systemic sclerosis, diabetic nephropathy and so on (Bhogal et al., 2005).

The transition from healthy tissue to diseased tissue can lead to an increase in the stiffness of the ECM. This stiffening has been demonstrated to facilitate tumour cell migration and activate myofibroblasts resulting in excessive collagen deposition (Henderson et al., 2020). Activated myofibroblasts are the primary cellular source that are responsible for the excessive deposition of fibrotic ECM in various organs, including the lung, liver, kidney, skin and heart (Kurose & Mangmool, 2016; LeBleu et al., 2013; Mack & Yanagita, 2015; Phan, 2002).

Excessive accumulation of interstitial collagens in the alveolar structures is a characteristic feature of idiopathic pulmonary fibrosis (IPF), leading to impaired gas exchange. This condition is indicated by the continuous presence of extra interstitial collagens, particularly types I, III, and VI (King et al., 2011). Fibrotic kidney diseases also exhibit an increased risk of mortality due to elevated levels of microfilamentous interstitial collagen (Panizo et al., 2021). Similar patterns of interstitial collagen deposition are observed in skin, liver, and cardiac fibrosis (Gabrielli et al., 2009; Karsdal et al., 2020; Kurose & Mangmool, 2016).

In the context of cardiac fibrosis (Kurose, 2021), multiple cell types, including fibroblasts work together to create an environment conducive to fibrosis. Moreover, immune cells like macrophages, mast cells, and lymphocytes are recruited and become active in remodelling

hearts, and potentially stimulating fibroblast activation through the release of fibrogenic substances such as cytokines and growth factors. Furthermore, endothelial cells, pericytes, and vascular smooth muscle cells also have the ability to secrete molecular signals that can affect the behaviour of fibroblasts (Frangogiannis, 2021).

1.8. Collagen-Targeted Therapies

Several experimental evidences have demonstrated the effectiveness of pharmacological inhibitors that specifically target type I collagen in fibrotic context. Some of these compounds are under evaluation in different phases of clinical trials. Some of these inhibitors are:

1.8.1 The Inhibition of Transcription

Baicalein is a flavonoid compound known for its remarkable anti-inflammatory and anticancer properties. Studies have demonstrated that baicalein effectively inhibits the production of type I collagen induced by TGF-β1. Notably, baicalein showed a low toxicity property in normal lung fibroblast cells (Hu et al., 2017; Sun et al., 2020).

Histone deacetylation is a critical process in the transcription of COL1A1 gene. Experimental findings have revealed that phenylbutyrate effectively decreases the transcription of COL1A1 resulting in reduced levels of type I collagen protein in fibroblasts. Moreover, some studies have demonstrated that sodium phenylbutyrate holds promise as an anticancer agent in prostate cancer by the same approach (Rishikof et al., 2004; Shi et al., 2022a).

Furthermore, the compound C9 was introduced through a high throughput screening process that specifically targeted La-related protein 6 (LARP6). Its mechanism of action involves disrupting the interaction between LARP6 and type I collagen 5′SL RNA, resulting in the inhibition of collagen production. Notably, C9 showed antifibrotic activity with some acceptable adverse side effects during in vivo studies. These promising findings introduced C9 as a candidate for further investigation as a potential antifibrotic therapeutic agent (B. Stefanovic et al., 2021; L. Stefanovic & Stefanovic, 2019).

Moreover, ethyl 3,4-dihydroxybenzoate has been identified as a compound that effectively suppress some type of breast cancer cell by inhibiting collagen synthesis (Han et al., 2014). Another compound of interest is 2- pyridine-5-carboxylic acid, which has demonstrated efficacy in inhibiting collagen biosynthesis and inhibiting tumor growth (Shi et al., 2022a).

Finally, minoxidil has shown the ability to inhibit cancer cell migration by blocking collagen production resulting in reduced metastasis of cancer cells (Eisinger-Mathason et al., 2013; Qi & Xu, 2018).

1.8.2 Inhibition of HSP47

As a collagen-specific chaperone promoting tumor growth and invasion, HSP47 is an attractive therapeutic target for cancer treatment (Duarte & Bonatto, 2018; Ito et al., 2017). AK-778 is an HSP47 inhibitor that intervene the interaction between HSP47 and collagen and inhibit the HSP47 activity without affecting HSP47 synthesis while showing antifibrotic effects (Thomson et al., 2005).

As mentioned, type I collagen is frequently upregulated during tumorigenesis. The binding of type I collagen to its receptors on tumour cells promotes tumor cell proliferation, epithelial-mesenchymal transition and metastasis (Kauppila et al., 1998). Therefore, targeting type I collagen expression can influence on the efficacy of tumor therapies such as chemotherapy, radiotherapy and immunotherapy. It should be mentioned that angiogenesis is required for tumor growth as well as metastasis, and type I collagen function is pivotal for angiogenesis (Baldari et al., 2022; Shi et al., 2022b).

2. Materials

2.1. Materials

Name	Manufacturer	Catalogue Number
Plasmid DNA Maxiprep Kits	Invitrogen	K210006
Proteinase K	NEB	P8107S
DNA Ladder	NEB	N3231S
Gel Loading Dye, Purple (6X)	NEB	B7025S
OneTaq® DNA Polymerase	NEB	M0480S
PowerUp™ SYBR™ Green	ABI	A25742
Triton™ X-100	Invitrogen	85111
Lipofectamine 3000	Invitrogen	L3000001
Opti-MEM Reduced Serum Medium	Invitrogen	31985070
TransIT®-mRNA	Mirus Bio	MIR2250
PureLink™ PCR Purification	Invitrogen	K310001
DNeasy Blood & Tissue Kit	Qiagen	69581
peqGREEN DNA/RNA dye	Avantor	732-3196
RNeasy Mini Kit (50)	Qiagen	74004
ViaFect™	Promega	E4981
Hydroxyproline Assay Kit	Sigmaaldrich	MAK008
RNeasy Mini Kit (50)	Qiagen	74004

2.2. Common Reagents

Name	Manufacturer	Catalogue Number
Dulbecco's modified eagle medium (DMEM)	Invitrogen	41965039
FBS heat inactivated 944855K	Invitrogen	10500064
Penicillin/Streptomycin 100x	Invitrogen	15140122
MEM Non-essential amino acids 100x	Invitrogen	11140035
ProLong™ Gold Antifade Mountant	Invitrogen	P36930
Tris (1 M), pH 8.0, RNase-free	Invitrogen	AM9856
0.25% Trypsin-EDTA (1X)	Invitrogen	25300-056
0.05% Trypsin-EDTA (1X)	Invitrogen	25300-054
PBS: Dulbecco's (1x) (without Ca^{2+} and Mg^{2+})	Invitrogen	14190169
Trypan Blue Stain 0.4%	Invitrogen	15250-061
Dimethylsulfoxide (DMSO)	Applichem	2045215
Tris-acetate-EDTA (TAE)	Carl Roth	CL86.1
Propanol	Carl Roth	6752.1
UltraPure™ DNase/RNase-Free Distilled Water	Invitrogen	10977-035

2.3. Common Consumables

Name	Specification	Manufacturer	Catalogue Number
Bacteriological dishes	10 cm	Cornig	430166
Cell culture pipettes	5, 10 ,25	Greiner	606180 607180 760180
Cell strainers	40 µm	Becton Dickinson	352340
Multi-well culture plates	6,48,96 wells	Greiner	665180
Falcon tubes	15, 50 ml	BD Falcon™	1110502
Cell Culture Dish	60 mm	Greiner Bioone	628160
Cell Culture Dish	35 mm	Greiner Bioone	627160
Cell Culture Dish	35x10mm	Greiner Bioone	664160
Cover slips	16 mm	Carl Roth	LH23.1

3. Aims

In the landscape of genetic therapies, CRISPR-Cas9 gene editing stands out as a revolutionary approach with the high potential to revolutionize the treatment of a myriad of diseases. Fibrotic disorders, which are characterized by the excessive accumulation of ECM proteins (mostly collagen I) leading to organ dysfunction, is a significant therapeutic challenge. Current treatments are often limited in the effectiveness and mostly target symptoms rather than the underlying causes. Our research pivots on the innovative application of adenine base editing (ABE8), a derivative of the CRISPR-Cas9 system, focusing on the Col1a1 gene, a key player in fibrosis, to reduce Col1a1 gene expression via targeting basal promoter (CCAAT box) resulting in ablation of fibrotic signaling pathway in fibroblasts. This approach will aim to provide a permanent and broadly applicable strategy in various fibrotic conditions.

Objectives

-Development of an Antifibrotic Strategy Using Base Editing: At the forefront of our aims is the application of a CRISPR-Cas9 targeting the Col1a1 gene promoter(CCAAT box) to reduce collagen expression.

-Efficacy and Precision of Adenine Base Editing: A key objective is to evaluate the efficacy of adenine base editing in reducing collagen production within fibroblasts as the major source of collagen production in the body. This task involves the high precision and specificity of our approach, aiming for minimal off-target effects resulting in maximum therapeutic benefit.

-Signaling Pathway Assessment: Utilizing high-throughput multi-omics analysis, we aim to examine molecular changes following the gene editing process. This investigation is crucial in specificity of our goals, for its clinical applicability in fibrotic diseases.

By combining these goals, this thesis aims to open up novel application of ABE8. We plan to use the powerful base editing technology to create better, more accurate, and wide-ranging treatments.

4. Methods

4.1. Cell Culture

The fibroblast cells (NIH3T3), were cultured in DMEM medium (Gibco), supplemented with 10% fetal bovine serum and 1% Penicillin/Streptomycin. These cells were kept at 37°C and 5% CO2 with a routine check for any signs of contamination. The culture medium was changed every two days. Moreover, depending on the 80% confluency criterion; the cells were passaged every 4-5 days using trypsin.

During the passaging procedure, cells were first washed twice with PBS (-/-). They were then exposed to 0.05% trypsin/EDTA for 3 minutes in a 37°C incubator. Cell detachment was verified under the light microscopy. To stop the trypsin, an equal amount of 10% serum-enriched medium was added. After gentle agitation with a 1000µl pipette tip and centrifuging at 1000 RPM for 5 minutes, the supernatant was removed, and fresh medium was added. The resulting mixture was suitable for both passaging and cryopreservation in -80 °C.

For the cryopreservation process, a solution of 90% FBS and 10% Dimethysulfoxide (DMSO, Sigma-Aldrich) was used. Cells were then immersed in this freezing solution, and left to freeze at -80°C overnight. Following this, the cells were kept at -140°C in a nitrogen gas tank for long-term usage.

4.2. Genotyping the Col1a1 Promoter

Genomic DNA was extracted from wildtype fibroblasts (WTs) using the DNeasy Blood & Tissue Kit (QIAGEN, Cat. No. 69581). This kit was used as the manufacturer's instruction to ensure the optimal yield and purity of DNA from cells. Primers were designed with PrimerBLAST (NCBI, 2022) and ordered from Thermo Fisher Scientific, under the Invitrogen label.The DNA was then amplified through PCR using the high-fidelity Taq DNA Polymerase (Cat No.M0480S, OneTaq® DNA Polymerase, NEB). The primer sequences were:

Forward 5' GTCCCAGAAAGAAAGTACAAGGG 3'

Reverse 5' TGGAGAGCTGGGAGGAACC 3'

The sequence composition of the amplicon was subsequently decoded by sanger sequencing at Eurofins company, Germany.

4.2.1. gRNA Design and Plasmid

We designed the gRNAs using the CRISPRO algorithm with a guide length of 20 nucleotides.

4.3. Cloning Protocol

Once the gRNA design met the experimental requirements, the next critical phase was the synthesis and cloning process by Gene Script (Staal et al., 2019). The designed gRNA was first synthesized. In the post-synthesis phase, the gRNA was efficiently ligated into the 132777 plasmid (pU6-pegRNA-GG-acceptor, Addgene ID: 132777). This plasmid from Addgene is known for its reliable performance in various genetic experiments, providing the necessary framework for successful integration and expression of the introduced gRNA.

4.3.1. CloneEZ Cloning Technique

CloneEZ -an innovative cloning technique by obviating the need for conventional restriction sites and ligation- was applied by GenScript (GenScript Biotech, Netherlands) to clone gRNA. The process began with the amplification of the desired insert sequence via PCR, during which homologous regions were seamlessly added to the ends of the insert sequence (J.-D. Zhang et al., 2011)

Instead of employing multiple restriction enzymes, the vector was linearized, priming it for the subsequent annealing process. Quality checks, such as gel electrophoresis, were rigorously carried out to verify successful linearization, ensuring the integrity of the cloning process. Subsequently, the insert sequence and the linearized vector underwent an incubation process, allowing them to hybridize. Using the innate ability of DNA molecules to repair themselves, a continuous DNA sequence was formed without the requirement for external ligases, ensuring seamless integration of the insert sequence into the vector. Once the gRNA was effectively integrated into the pU6 plasmid, several validation assays was meticulously conducted (Tang et al., 2012). These included high-fidelity sequencing to confirm the accuracy of the cloned sequence and restriction digestion to verify the orientation and integrity of the gRNA insert within the plasmid. These comprehensive validation steps ensured the authenticity and functionality of the cloned gRNA construct for downstream applications.

It should mentioned that the ABE8 synthetic RNA was a generous gift from Professor. David Liu lab. Upon receipt of the synthetic ABE8, the proper storage was considered to maintain their activity (-80°C).

4.4. Fibroblast Cell Culture, Transfection, and Genomic Analysis

For transfection, TransIT®-mRNA Transfection Kit was used, ensuring the high-efficiency co-transfection of adenine base editing RNA version 8 (ABE8), plasmid-gRNA (CCCCAATTTGGAAGCAAGAC) used with GFP control plasmid to monitor the transfection efficiency (Figure 4 -1).

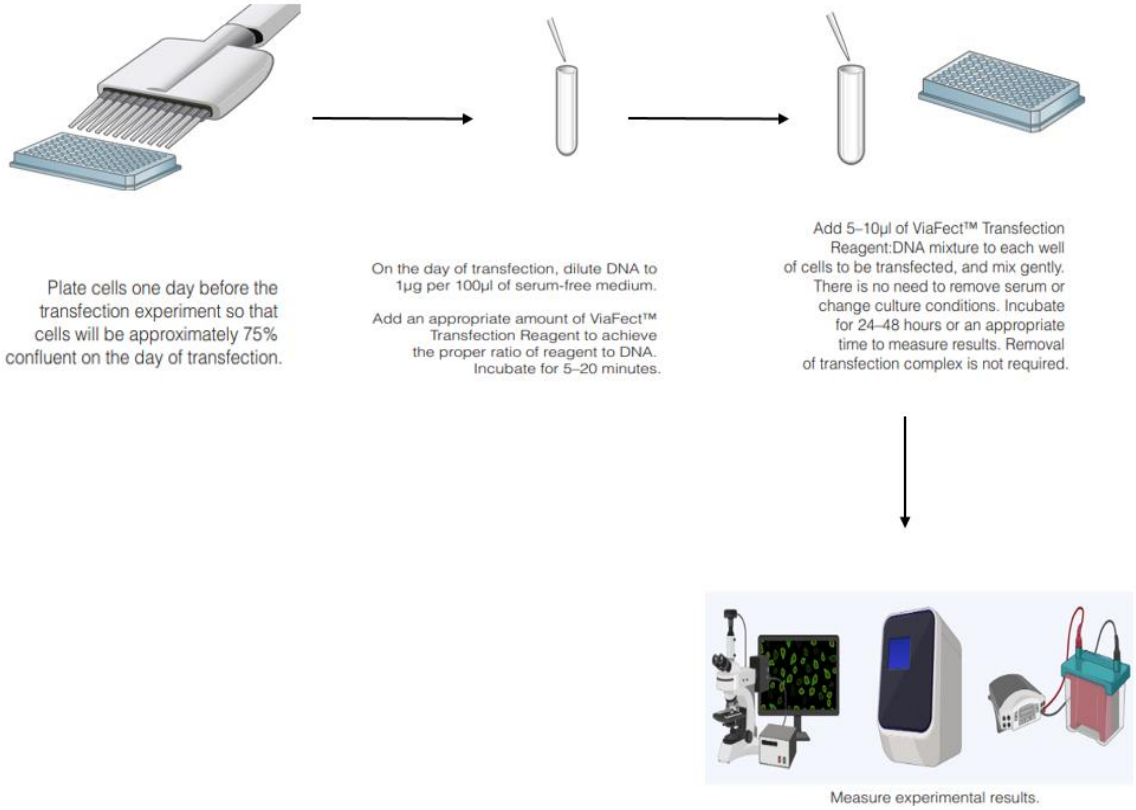


Figure 4-1. General overview of transfection. The basic of transfection by different liposomal based delivery solutions are the same.

4.4.1. Transfection Procedure

The cells were seeded in a 24-well plate at a density of 0.8 × 10^5 cells/ml in 0.5 ml of complete growth medium per well, approximately 18-24 hours before transfection. This

seeding density aimed to achieve a cell confluency of 80% before transfection began. To allow for optimal cell growth and attachment, the cultures were then incubated overnight. Both TransIT-mRNA and mRNA Boost solutions were brought to room temperature to ensure optimal performance. Following this, the solutions were gently shaken to ensure homogeneity. Next, a volume of 100 µl of Opti-MEMI Reduced-Serum Medium was added to a 0.5 ml microtube for the preparation of the transfection complex. One microliter (1 µg) of RNA, obtained from a 1 µg/µl stock solution, was then pipetted into the microtube. Additionally, 500 ng of control GFP plasmid was added to the mixture. The combined solution was then gently mixed using a pipette to achieve homogeneity. To facilitate the formation of transfection complexes, 2 µl of mRNA Boost Reagent was added to the diluted RNA and plasmid mixture. This was followed by the addition of another 2 µl of TransIT-mRNA Solution. The resulting mixture was essential for efficient delivery of the RNA into the cells. Subsequently, the microtube containing the transfection complex was incubated at room temperature for 5 minutes. This incubation period allowed sufficient time for the formation of stable complexes between the RNA, the transfection reagent (TransIT-mRNA), and the mRNA Boost Reagent. Exceeding 5 minutes of incubation is not recommended, as it could potentially decrease the efficacy of the transfection process.

4.4.2. Extracting the DNA

Seventy two hours after transfection first, the growth medium was carefully discarded. The cells were then gently washed two times with PBS solution. For the extraction of genomic DNA, a freshly prepared lysis buffer was utilized. This buffer was composed of 10 mM Tris-HCl (pH 7.5), 0.05% Sodium Dodecyl Sulfate (SDS), and 25 µg/ml proteinase K (New England Biolabs, cat. no. P8107S, USA), was added directly to each well. Afterward, 150 µl of this buffer was added to each well of a 48-well plate. The plates were then incubated at 37°C, allowing for thorough cell lysis over a period of 1-2 hours. Following this, the enzyme inactivation step was applied, where the samples were subjected to 80°C for 30 minutes. This final step ensured the deactivation of proteinase K, preserving the integrity of the extracted genomic DNA.

4.4.3. Assessment of DNA Purity and Concentration

The quality and quantity of extracted DNA were assessed utilizing a Nano Drop spectrophotometer (NanoDrop 1000 Spectrophotometer, PEG Lab GmbH, Germany). The concentration of the isolated DNA recorded 50-100 ng/µl, indicative of an acceptable yield. The absorbance ratios (260/280 nm) mostly ranged from 1.8 to 2.2. This spectral assessment is a reliable metric for purity, confirming the extracted DNA was free from protein or other contaminant interferences. The consistent achievement of these metrics across samples underscores the effectiveness and reliability of the extraction process, affirming the suitability of the acquired DNA for subsequent experimental applications

4.4.4. PCR Amplification

The DNA regions of interest (Col1a1 promoter) were amplified by PCR. The PCR master mix ingredient was as follows in the table 4-1:

Table 4-1. PCR reaction master mix for amplification of Col1a1 promoter.

Component	Amount (µL)	Final concentration	Sequences
OneTaq Hot Start Master Mix with Standard Buffer	12.5	$2x$	–
PCR1 forward primer	0.5	0.5 µM	GTCCCAGAAAGAAAGTACAAGGG
PCR1 reverse primer	0.5	0.5 µM	TGGAGAGCTGGGAGGAACC
Lysis mix with harvested gDNA	3	–	–
Nuclease-free H2O	8.5	–	–
Total reaction volume	25	–	–

PCR reaction condition was performed under the following condition in indicated in table 4-2.

Table 4-2. PCR reaction condition for amplification of Col1a1 promoter.

Cycle No.	Denaturation	Annealing	Extension
1	95 °C, 3 min	–	–
2-35	95 °C , 30 sec	55 °C , 30 sec	72 °C , 6 sec
36	–	–	72 °C, 5 min

4.4.5. Agarose Gel Electrophoresis

To verify the successful amplification of the target DNA, 1.5% agarose gel electrophoresis was performed. 1.5g of agarose was dissolved in 100ml of 1X TAE buffer by heating in microwave. Once the agarose was completely dissolved and the solution was slightly cooled, 7µl of peq GREEN DNA/RNA dye (Avantor, Catalog No. 732-3196, USA,) was added and mixed thoroughly. The liquid agarose was then poured into a gel-casting tray and allowed to solidify at room temperature. Then, 8 µl of the PCR product was mixed with 2 µl of 6X gel loading dye (New England Biolabs, cat. no. B7025S, USA) and loaded into the wells. To determine the size of the amplified DNA amplicons, 5 µl of a 100 bp DNA Ladder (New England Biolabs, cat. no. N3231S, USA) was also loaded alongside the samples. Next, electrophoresis was performed at 130V for approximately 40 minutes or until the dye front was three-quarters down the gel. Finally, the gel was visualized and documented using the FastGene® FAS-BG LED BOX Imaging System (Nippon Genetics, Japan), and the presence of the desired PCR band was confirmed.

4.4.6. Purification and Sanger Sequencing

The PCR products of the expected size were purified using the PureLink™ PCR Purification Kit (Thermo Fisher Scientific, cat. no. K310001, USA). The purified DNA was then subjected to Sanger sequencing (Eurofins Scientific, Germany). For the sequencing reaction, 5 µl of primer was combined with 5 µl of PCR product.

4.5. Amplicon Next Generation Sequencing (NGS)

DNA testing before main experiment included measuring concentration, sample integrity and purity. Sample integrity and purity were detected by agarose gel electrophoresis (Concentration of Agarose Gel:1.5% Voltage:130V, Electrophoresis Time:50 min). Next, 1µg genomic DNA was randomly fragmented by Covaris. The fragmented genomic DNA were selected by magnetic beads to an average size of 200-400bp. Moreover, fragments were end repaired and then 3’ adenylated. Adaptors were ligated to the ends of these 3’ adenylated fragments. This process was to amplify fragments with adaptors from previous step and PCR products were purified by the magnetic beads.

The double stranded PCR products were heat denatured and circularized by the splint oligo sequence. The single strand circle DNA (ssCir DNA) were formatted as the final library then

QC qualified library. Following this the library was amplified with phi29 to make DNA nanoball (DNB) which have more than 300 copies of one molecular. The DNBs were load into the patterned nanoarray and pair end 100/150 bases reads were generated in the way of combinatorial Probe-Anchor Synthesis (cPAS). Finally, the generated FASTQ data were used for analysing of editing efficiency by CRISSPResso2.

4.6. Analysis of Base Editing Sanger Data Using EditR

EditR is a software designed to determine the exact site of base editing (Kluesner et al., 2018). This chapter describes the methodological approach of employing EditR for accurate and efficient base editing analysis-using sanger sequencing data.

4.6.1. Inputing Data to EditR

In the first step, the raw sanger data, contained in an .ab1 file, was submitted. Subsequently, the gRNA sequence was entered to match the target site. It is important to note that for gRNA sequences that are reverse complements to the .ab1 file, the option "Guide sequence is reverse complement" was selected.

4.6.2. Quality Control (QC) of raw data and analysis

To check the quality of the data, the "Data QC" tab was navigated to review any abnormalities appearing in the signal and noise plot. Samples confined to an area of consistent noise were approved to continue the analysis. Then, the system automatically directed to the "Predicted Editing" tab. In this step, the gRNA protospacer chromatogram and the underlying chart were scrutinized. Significant base calls were color-coded, and multiple clues beneath a single base call suggested base editing.

4.6.3. Report Generation

To obtain a comprehensive analytical summary, the "Download Report" option was selected. This action generated a detailed report encompassing all pertinent data and findings from the analysis.

4.6.4. Results and Interpretation

The Data QC tab provided an illustrative understanding of the sanger data's robustness. It began with a display of the entire peak area before any filtering, often showing lesser quality at the start and end sequences. A fundamental filtering technique was employed, extracting the initial 20 bases and any others falling below one-tenth of the mean peak area.

4.6.5. Predicted Editing

This section began with a graphical representation of peak areas throughout the guide sequence. Noteworthy base calls were accentuated with color, adhering to the P-value threshold for base editing identification.

4.7. Quatitative PCR Analysis (qPCR)

Fibroblast cells in culture were the source from which total RNA was extracted. For this goal the RNeasy Mini Kit (Qiagen, Cat. No. 74004, Germany) was employed. From the extracted RNA,1 µg used as the template for synthesizing complementary DNA (cDNA) using the RNeasy Micro Kit (Qiagen, Cat. No. 74004, Germany) and dilute this to a total of 100ul with ddH2O for q PCR tests.

Following synthesis, 1 µl from the cDNA was used for a 10µl PCR amplification. The quantitative assessment of the threshold-cycle value was conducted using the Applied Biosystems™ PowerUp™ SYBR™ Green Master Mix (ABI, PowerUp™ SYBR™ Green, Cat. No. A25742, USA) on an Applied Biosystems Fast 7500 Real-Time PCR System. It is noteworthy to mention that GAPDH gene applied as the gene of reference for normalization purposes. To determine the relative expression metrics of the target genes, the 2^−Δtc method was employed, where Δtc signifies the threshold cycle difference between GAPDH and Col1a1 genes. The primer sequences are outlined in Table 4-3.

Table 4-3. Sequence of forward and reverse primers of Col1a1 and GAPDH genes.

Oligo name	Sequence(′5– 3′)	Final concentration
Forward primer (Col1a1)	AGAGCATGACCGATGGATTC	$1x$
Reverse primer (Col1a1)	AGGCCTCGGTGGACA	$1x$
Forward primer (GAPDH)	CAGCCTCGTCCCGTAGACAA	$1x$
Reverse primer (GAPDH)	CAATCTCCACTTTGCCACTGC	$1x$

The amplification were done using the following reaction condition and set up the parameters (Table 4-4 and Figure 4-2).

Table 4-4. Reaction condition of qPCR using SYBR™ Green Master Mix.

Component	Amount (µL)	Final concentration
PowerUp™ SYBR™ Green Master Mix	5	$2x$
Forward primer (Table 2)	0.67	0.67 µM
Reverse primer (Table 2)	0.67	0.67 µM
cDNA	3	–
Nuclease-free H2O	0.66	–
Total reaction volume	10	–

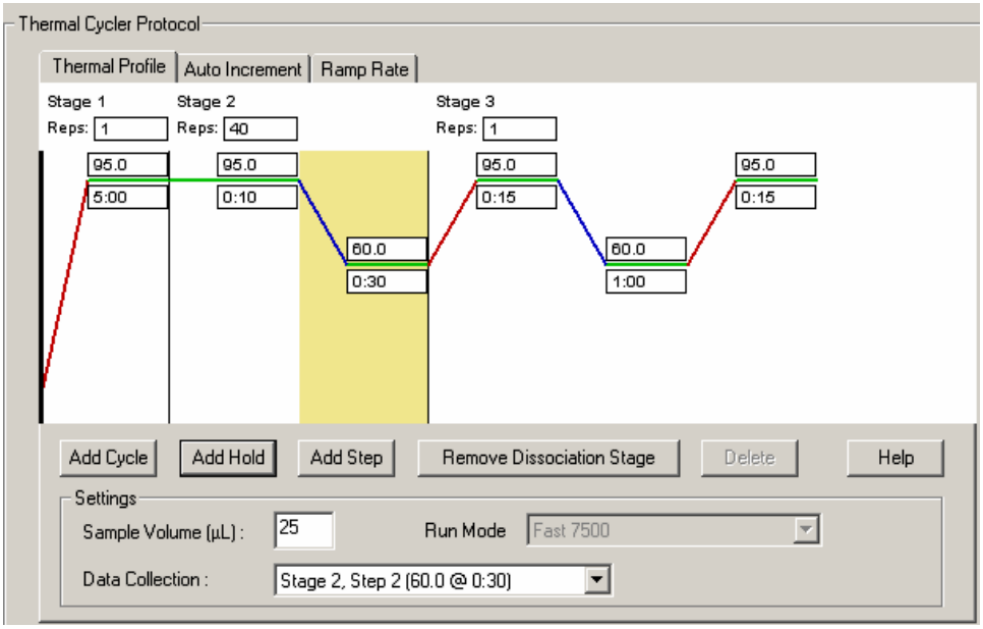


Figure 4-2. Thermal profile and set up the parameters for qPCR.

4.8. Hydroxyproline assay

A colorimetric assay kit (Sigma-Aldrich-Chemie GmbH, Munich, Germany) was employed for the measurement of Hydroxyproline in the supernatant of cell culture plates, following the manufacturer's protocol. Initially, the supernatant of fibroblasts underwent hydrolysis using concentrated hydrochloric acid (HCl, ~12 M) at 120 °C for 3 hours. To facilitate the evaporation of resulting fluid, plates were placed in a 60°C oven for sample drying.

Subsequently, 100 µL of the Chloramine T/Oxidation Buffer Mixture was added to each sample and standard well, followed by incubation at room temperature for 5 minutes. Next, 100 µL of the Diluted DMAB Reagent (4-dimethylamino benzaldehyde) was added to each sample and standard well, with an incubation period of 90 minutes at 60°C. To establish a standard curve, additional microplate wells were filled with known dilutions of hydroxyproline. The optical density of each well was measured using an absorbance microplate reader at 560 nm. The hydroxyproline content in samples was subsequently determined by referencing the optical density values against the standard curve.

4.9. Immunostaining

The process began by seeding $1\times10\text{^}4$ dissociated single cells onto round cover slips, which had been meticulously placed in 12-well plates. With the aim of fostering optimal growth conditions, each well was supplemented with 0.5 mL medium (DMEM, 10% FBS and 1% Penicillin/Streptomycin). Following the initial seeding, the plates were transferred into a humidified incubator (37°C, %5 CO2). After 48 hours cells were fixed with 4% paraformaldehyde (PFA) for 15-30 minutes at room temperature. The cells should be washed three times with ice-cold PBS. Next, non-specific binding sites were blocked by incubating in blocking solution (3% BSA) for 1 hour at room temperature. Following this, the primary antibody (Abcam, anti-collagen 1 a1, Cat. No. Ab21286, UK) was diluted to a 1:200 ratio in the appropriate buffer (3% BSA in PBS). The diluted primary antibody was applied to the samples, followed by overnight incubation at 4°C in a humidified chamber. Then, samples were washed three times with PBS to remove unbound primary antibody, with each wash lasting for 5 minutes.

Subsequently, the secondary antibody (Thermo Fisher Scientific, Alexa Fluor™ 555, Cat. No. A-21430, USA) was diluted to a 1:1000 ratio in the appropriate buffer (3% BSA in PBS). The diluted secondary antibody was applied to the samples, protected from light, and incubated for 1-2 hours at room temperature. Samples were again washed three times with PBS to remove unbound secondary antibody, each wash lasting for 5 minutes.

Finally, slides were mounted using by mounting medium (Thermo Fisher Scientific, ProLong Gold antifade mountant, Cat. No. P10144, USA) and allowed to dry, then were ready for visualization under a fluorescence microscope. It should be metioned that samples were protected from light during and after the application of fluorescently labelled antibodies. Moreover, aappropriate controls, including a no-primary-antibody control, were used to assess non-specific secondary antibody binding.

To analyze the results, post-staining images of both edited and non-edited cells were first captured using fluorescence microscopy (Zeiss, Apotome, Germany). Then, ImageJ, an open-source software, was utilized to compare the intensity of red fluorescence between the edited and non-edited cells (Schindelin et al., 2015). The images were opened in ImageJ to quantify the intensity of the red color, representing the presence of Collagen 1 a1 in the cells.

4.10. Cell Proliferation and Clonogenic Assay

Cell proliferation and clonogenic assays were performed to assess the survival and proliferation potential of MCF7 cells on a new matrix with less collagen. Cell proliferation assays were performed as described previously (J. G. Kim et al., 2010). Briefly, MCF7 cells were seeded at a density of 10,000 cells per six-well plate coated with ECM derived from isogenic fibroblasts. The cells were maintained in DMEM supplemented with 10% FBS. After the indicated time points, the cells in each well were trypsinized and counted to determine cell viability.

Clonogenic assays were used to assess the ability of MCF7 cells to form colonies (Rafehi et al., 2011). Cells were seeded onto 12-well plates at a density of 100 cells per well. Cultures were maintained for 12 days in complete culture media to allow the formation of macroscopic colonies. The colonies were fixed with 4% PFA and stained with 0.5% crystal violet. Finally, the numbers of visible colonies containing at least 50 cells in size were counted.

4.11. Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

All MS/MS experiments were performed in the CECAD proteomics facility (Cologne University, Germany) on a Q Exactive Plus Orbitrap mass spectrometer that was coupled to an EASY nLC (both Thermo Scientific).

4.11.1. Protein Preparation and Digestion

On the first day, the urea lysis buffer was prepared by combining 8M Urea/50 mM TEAB buffer with a 50x protease inhibitor cocktail (20 µL of 50x inhibitor to 1 mL of 8M Urea). To lyse the cells, approximately 1,000,000 cells were used, and the cell pellet was lysed using the prepared urea lysis buffer. Then, a sufficient volume of urea lysis buffer was added to achieve a final concentration of ≥ 6M Urea (Hedrick et al., 2015; Koyuncu et al., 2021).

Next, chromatin shearing was performed using a Bioruptor (or equivalent sonicator) with a cycle of 30/30 seconds. The lysate was then centrifuged at 20,000 × g for 5 minutes. Cell debris was discarded, and the protein concentration of the supernatant was determined using an absorbance reader. An aliquot of 50 µg of protein lysate was transferred to a new 1.5 mL tube.

Then, disulfide bonds were reduced by adding DTT to a final concentration of 5 mM. The mixture was vortexed thoroughly and incubated at 25°C for 1 hour. Subsequently, carbamidomethylation of cysteines was achieved by adding chloroacetamide (CAA) to a final concentration of 40 mM. The sample was vortexed again and incubated in the dark for 30 minutes, followed by dilution with 50 mM TEAB buffer to achieve a final urea concentration of ≤ 2M. Trypsin was then added at a 1:75 enzyme-to-substrate ratio (trypsin:protein), and the mixture was incubated overnight at 25°C (Koyuncu et al., 2021).

4.11.2. Sample Purification Procedure

The samples were purified using stage-tips by following the established protocol. The samples were first acidified with 10% formic acid. Then, they were centrifuged at maximum speed for 5 minutes to pellet any precipitates. The supernatant, containing the desired peptides, was carefully transferred onto the stage-tips. The stage-tips were then centrifuged at 2,600 rpm for 5 minutes to allow the peptides to bind to the stationary phase within the stage-tip.

The stage-tips were washed with 30 µL of Buffer A to remove any unbound contaminants. This washing step was followed by centrifugation at 2,600 rpm for 3 minutes. Next, the stage-tips

were further washed with two consecutive washes of 30 µL each of Buffer B. After each wash, the stage-tips were centrifuged at 2,600 rpm for 3 minutes to remove any remaining contaminants with higher affinity for Buffer B compared to the peptides of interest.

Finally, a syringe was used to thoroughly dry the stage-tips, ensuring complete removal of any residual solvent. Following confirmation of dryness (potentially using Figure 4-3 as a reference), the stage-tips were stored at 4°C for future use (Rappsilber et al., 2007).

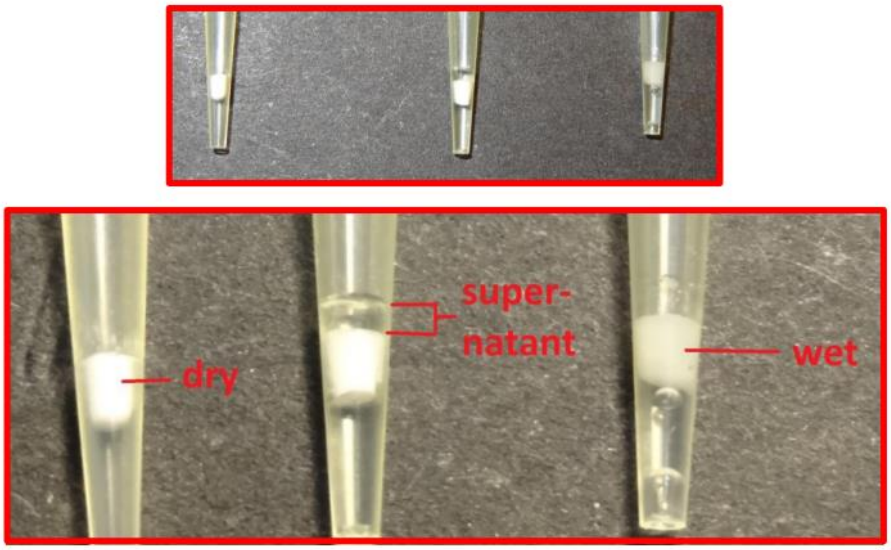


Figure 4-3. This image provides insight into the stagetip's physical appearance, highlighting its dryness crucial for effective sample preparation in mass spectrometry-based proteomics.

4.12. Proteomics Data Analysis

The proteomics data were analysed using Perseus (Version 2). The software enabled a comprehensive examination of the protein data, aiding in the extraction of relevant information by visualization. Results from the Perseus analysis were visualized as various plots, including a volcano plot, scatter plot and heatmap. These formats are commonly used for their clarity and effectiveness in presenting proteomics data (Tyanova & Cox, 2018).

4.13. Transcriptome Experiments Procedures (RNA-sequencing)

Strand-Specific Transcriptome Library Construction Protocol (DNBSEQ)

The library preparation and sequencing were conducted by Beijing Genomics Institute (BGI). Strand specific transcriptome library construction was completed by enriching mRNA from total RNA, sequenced by DNBSEQ high-throughput platform, and followed by bioinformatics analysis (Sudhakaran et al., 2023; K. Wang et al., 2021).

4.13.1. Experimental Protocols

mRNA was isolated from the total RNA using magnetic beads with attached oligo(dT) (Green & Sambrook, 2019). Following the isolation of mRNA from total RNA, mRNA molecules were fragmented into small pieces using a fragmentation approach. To ensure efficient priming, primers were then added to the sample and mixed thoroughly. The mixture was then incubated at 65°C for 5 minutes on a thermal cycler. After incubation, the first-strand synthesis reaction system was added to initiate cDNA synthesis. Subsequently, a second-strand synthesis reaction system was prepared, specifically by using dUTP instead of dTTP in the nucleotide mix. This mixture was incubated on a thermomixer at 16°C for 2 hours to synthesize the second-strand cDNA. As a result, the reaction product was purified with magnetic beads.

In the next step, the end repair and "A-tailing" reaction system was prepared. This reaction mixture contained enzymes and nucleotides necessary for the following steps. The cDNA sample was then incubated on a thermal cycler for a specific duration. Under the action of these enzymes, the incubation facilitated the repair of the "sticky ends" generated during reverse transcription of the double-stranded cDNA. Additionally, an adenine (A) base was added to the 3' end of the cDNA fragments (Sudhakaran et al., 2023).

4.13.2. Library Quality Control

The library was validated using the Agilent Technologies 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). To prepare the final library, double-stranded PCR products were first heat-denatured to form single strands. These single strands were then circularized with the help of a splint oligonucleotide (J. Li et al., 2023).

4.13.3. Sequencing

The starting DNA fragments (library) were enzymatically amplified (with phi29) to create DNA nanoballs (DNBs) containing over 300 copies of a single molecule. These DNBs were loaded onto a patterned chip with tiny holes (nanoarray). A sequencing technique called combinatorial Probe-Anchor Synthesis (cPAS9) was used on the BGISEQ-500 platform (BGI) to generate single-stranded reads of 50 bases (or paired-end reads of 100/150 bases) (J. Li et al., 2023).

4.14. Scanning Electron Microscopy (SEM)

Dr. Frank Nitsche from the General Ecology department at the University of Cologne conducted the SEM analysis. For scanning electron microscopy (SEM), samples were fixed in 4% PFA at room temperature for 60 minutes. Following fixation, the samples underwent post-fixation staining with 1% osmium tetroxide for 10 minutes. They were then dehydrated through a graded ethanol series: 30%, 50%, 70%, 80%, 90%, and 96%, with each step involving two washes in the respective ethanol concentration and a 30-minute soak (Kang et al., 2019; Tanaka & Mitsushima, 1984). Instead of critical point drying, a final dehydration was performed using a 50:50 hexamethyldisilazane (HMDS)-ethanol solution for 30 minutes, followed by two washes with pure HMDS and a 45-minute incubation. Finally, the samples were dried, mounted on sample holders, sputter-coated with a 120 Å layer of gold, and examined using the FEI Quanta 250 FEG SEM (Dwiranti et al., 2019; Nitsche, 2016).

4.15. Preparation of Extracellular Matrix (ECM)

This section details the initial steps involved in isolating the extracellular matrix (ECM) produced by fibroblast cultures (Franco-Barraza et al., 2016).

The experiment began with a fibroblast culture that had reached a semi-confluent state, meaning approximately 80% of the culture dish surface was covered by cells. This density ensured sufficient ECM production while allowing for efficient lysis solution penetration in the next step. Then, the culture medium was carefully aspirated from the dish. The dish was then gently rinsed two times with aliquots of Dulbecco's Phosphate-Buffered Saline (DPBS).

Crucially, during rinsing, the pipette tip was positioned against the dish wall rather than the bottom, avoiding disruption of the cells and the ECM located at the bottom of the dish.

A pre-warmed (37°C) 1 mL of extraction buffer (containing 0.5% (v/v) TRITON™ X-100 (Thermo Fisher Scientific, cat. no. 85111, USA) and 20 mM NH4OH) was then gently added to the cell culture. The extraction buffer is formulated to lyse (break open) the fibroblast cells, releasing their intracellular components and leaving behind the ECM. The process of cell lysis was monitored using light microscopy. This allowed to observe the gradual disappearance of intact cells as the extraction buffer took effect. The incubation continued at 37°C until no intact cells were visualized, typically taking between 5 to 15 minutes.

Following cell lysis, the procedure proceeded with the removal of cellular debris. Without aspirating the extraction buffer, 2 to 3 mL of $DPBS^-$ were slowly added to the culture dish. This dilution step aimed to weaken the concentration of cellular debris. The $DPBS^-$ was gently pipetted onto the side of the dish to minimize disruption of the newly formed matrix. The dishes were then stored overnight at 4°C in $DPBS^-$ to further prevent disturbing the delicate ECM. This dilution process was performed cautiously to prevent turbulence that could detach the freshly exposed ECM layer from the surface. As carefully as possible, the diluted cellular debris was aspirated using a pipette. It was crucial to avoid complete aspiration of the liquid layer to ensure the matrix surface remained hydrated throughout the process. The entire volume of liquid was not aspirated to prevent removing or damaging the ECM layer. Next, 2 ml of fresh $DPBS^-$ were gently added to the dish, followed by gentle aspiration. This washing step was repeated two additional times using only DPBS, with continued focus on avoiding liquid turbulence or disturbing the matrix.

Finally, the matrix-coated plates were covered with at least 3 mL of fresh DPBS containing 100 U/mL penicillin and 100 µg/mL streptomycin. The plates were then sealed with Parafilm® strips and stored for 2 to 6 weeks at 4°C for further use.

4.16. Bioinformatics (in-Silico) Analysis

RNA-seq analysis was performed using the Dr. Tom Multi-Omics Data Mining System (Bai et al., 2022).

4.16.1. Transcriptome (RNA-seq) Analysis Pipeline

The sequencing data was filtered with SOAPnuke by removing reads containing sequencing adapter, reads whose low-quality base ratio (base quality less than or equal to 15) is more than 20%, and reads whose unknown base ('N' base) ratio is more than 5%. Afterwards, clean reads were obtained and stored in FASTQ format (Figure 4-4).

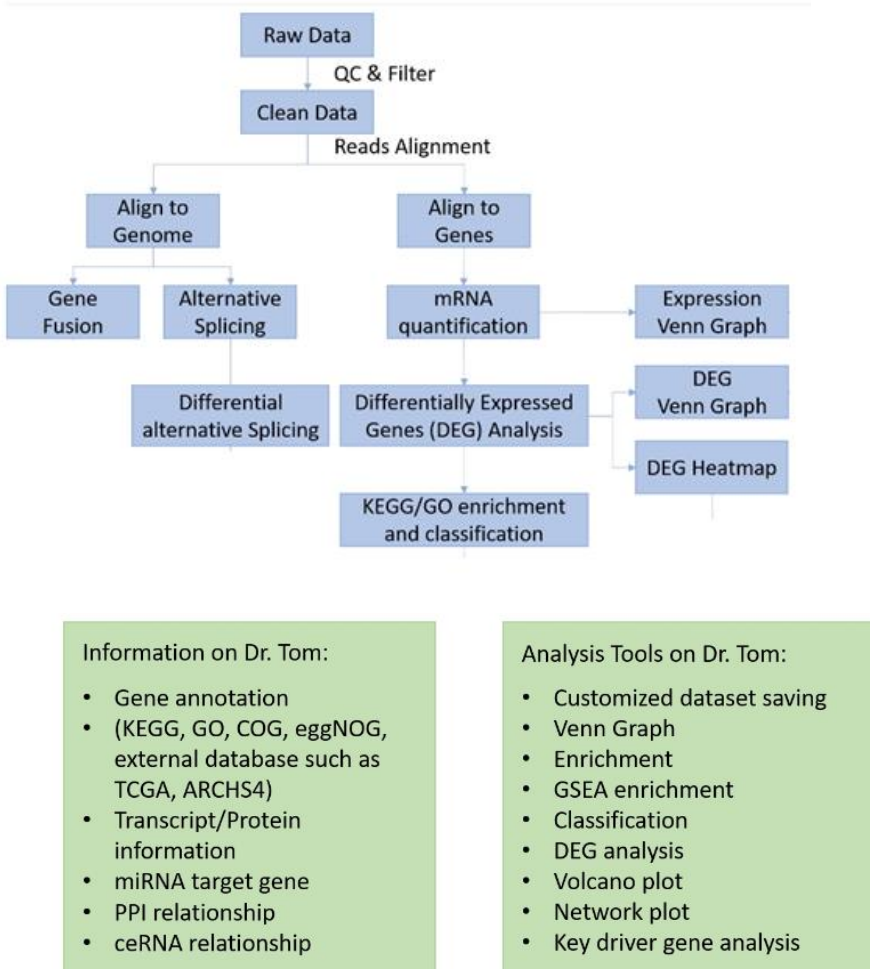


Figure 4-4. This comprehensive figure provides a visual representation of the Transcriptome (RNA-seq) Analysis Pipeline, highlighting each critical step in the process.

The clean reads were aligned to the reference genome utilizing HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts 2). Subsequently, fusion genes and differential splicing

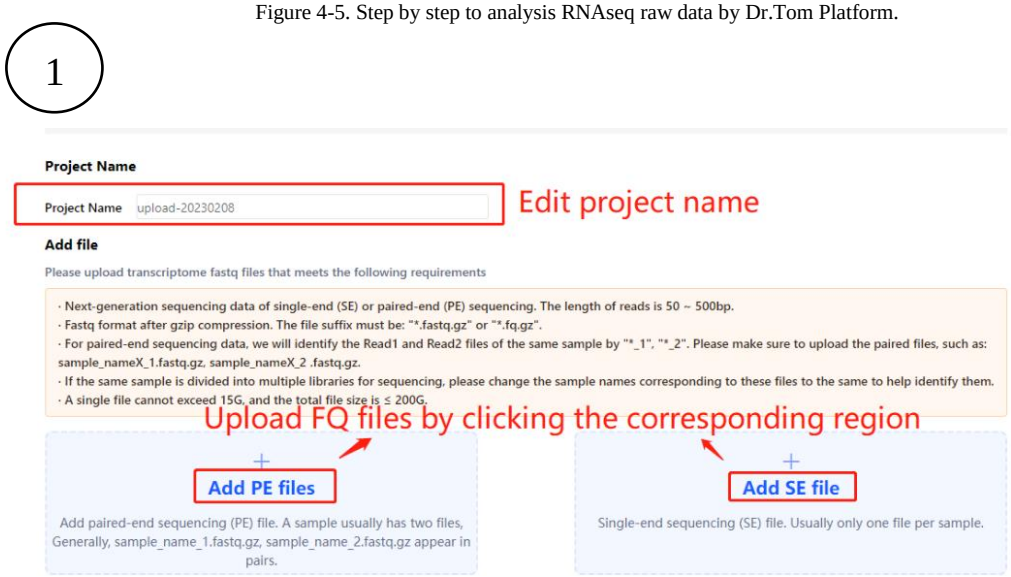
genes (DSGs) were identified using Ericscript (0.5.5-5) and rMATS (v4.1.1, RNA-Seq Splicing Analysis), respectively. The clean reads were aligned to the gene set using Bowtie2 for RNA identification. The gene expression level was determined using RNA-Seq by RSEM (Expectation-Maximization,v1.2.28), which provided read count, FPKM (Fragments Per Kilobase per Million mapped reads), and TPM (Transcripts Per Million). Differential expression gene (DEG) analysis was conducted using DESeq2 (or DEGseq), with a Q value threshold of ≤ 0.05. The DEG heatmap was generated using heatmap based on the DEG analysis outcomes.

For a deeper understanding of phenotype changes, Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genome (KEGG) enrichment analyses of annotated differentially expressed genes were carried out using Phyper based on the Hypergeometric test(Aoki-Kinoshita & Kanehisa, 2007; Young et al., 2010). The significance levels of terms and pathways were adjusted with a stringent Q value threshold (Q value ≤ 0.05).

4.17. Transcriptome Data Analysis by Dr.Tom Platform

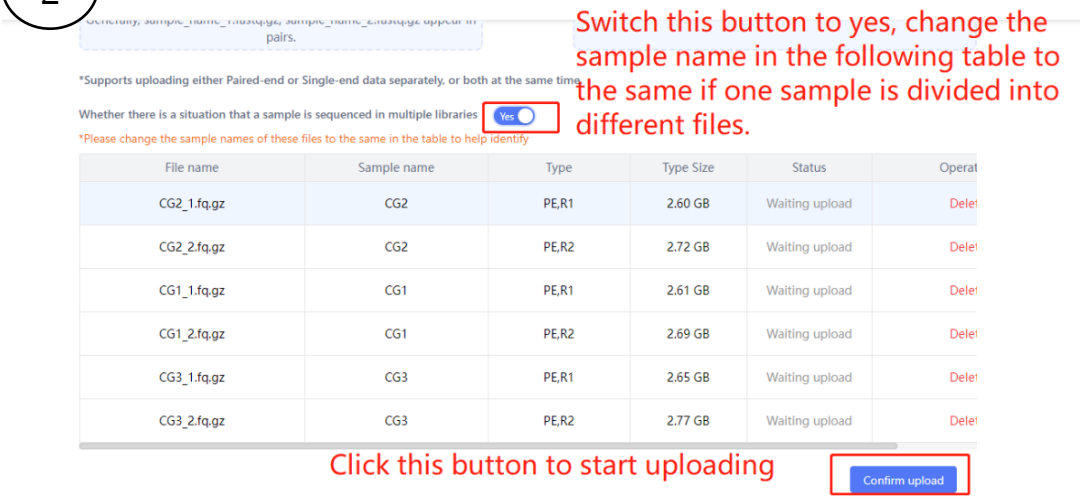
The system was accessed by signing in. Once logged in, the 'Basic data' tab was selected. The then clicked on 'Upload fastq' to initiate the upload process. However, if the 'Upload fastq' button was unavailable, it meant that the account had used up its allotted number of sample uploads (Figure 4-5).

Figure 4-5. Step by step to analysis RNAseq raw data by Dr.Tom Platform.



The FASTQ upload project name was edited as needed. The user then proceeded to choose between uploading paired-end (PE) files or single-end (SE) files by clicking the designated region for the desired file type.

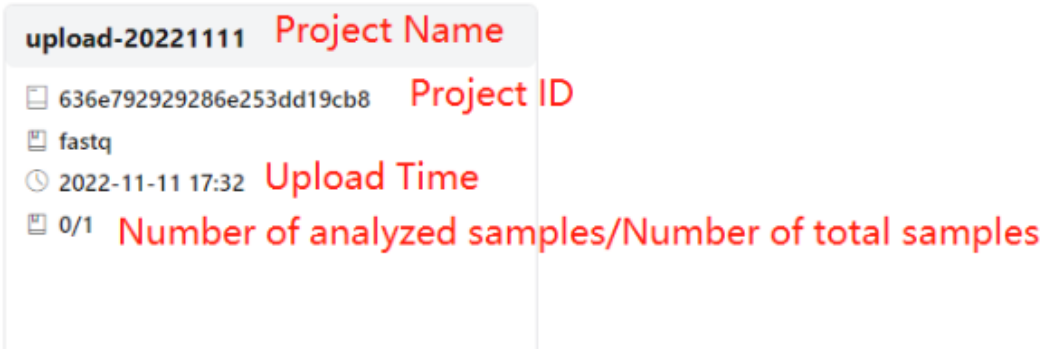
2



The user toggled the designated button to "Yes" if the data for a single sample was split across multiple FASTQ files. This indicated to the system that multiple files belonged to the same sample. In the provided table, a unique sample name was assigned to each individual FASTQ file to facilitate identification during subsequent processing steps.

Once all the files for the batch upload were selected, the information carefully reviewed for accuracy. Following verification, "Confirm upload" was pressed to initiate the transfer of the FASTQ files. It was important to keep this page open throughout the upload process until all files were successfully uploaded. This ensured that the files wouldn't need to be re-selected in case of any interruption.

3



After the user confirmed the upload by clicking "Confirm upload," the system initiated the transfer of the FASTQ files. This process established a new project within the "Basic data" tab. Once all files were successfully uploaded, project details became visible, including the project name, a unique project ID, the time and date of project creation, and the number of samples analyzed compared to the total number of samples included in the project.

4

5

Graph setting and download

4.18. Gene Enrichment Data Analysis

Various databases, such as the KEGG pathway database and GO, were utilized for enrichment analysis (Consortium, 2019; Kanehisa et al., 2017). This analysis was performed on the previously selected genes. By analysing the enrichment within these databases, we were able to gain insights into the dominant metabolic pathways and biological processes that these genes are likely involved in (Figure 4-6). The enrichment analysis employed the hypergeometric test to identify the pathways or functions that were significantly enriched in the selected genes compared to the entire gene background.

It is worth to mentioned that the KEGG pathway analysis, comprised curated pathway maps depicting our understanding of molecular interaction, reaction, and relation networks across various categories including metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases, and drug development, was applied for deep understanding of pathway modifications and related pathophysiologies (L. Chen et al., 2015). Furthermore, the GO resources offered a computational depiction of our current scientific understanding of the functions of genes (or the proteins and non-coding RNA molecules they produce) across diverse organisms, from humans to bacteria. It was used for describing our biological knowledge across three facets: molecular function, cellular component, and biological process (Carbon et al., 2017).

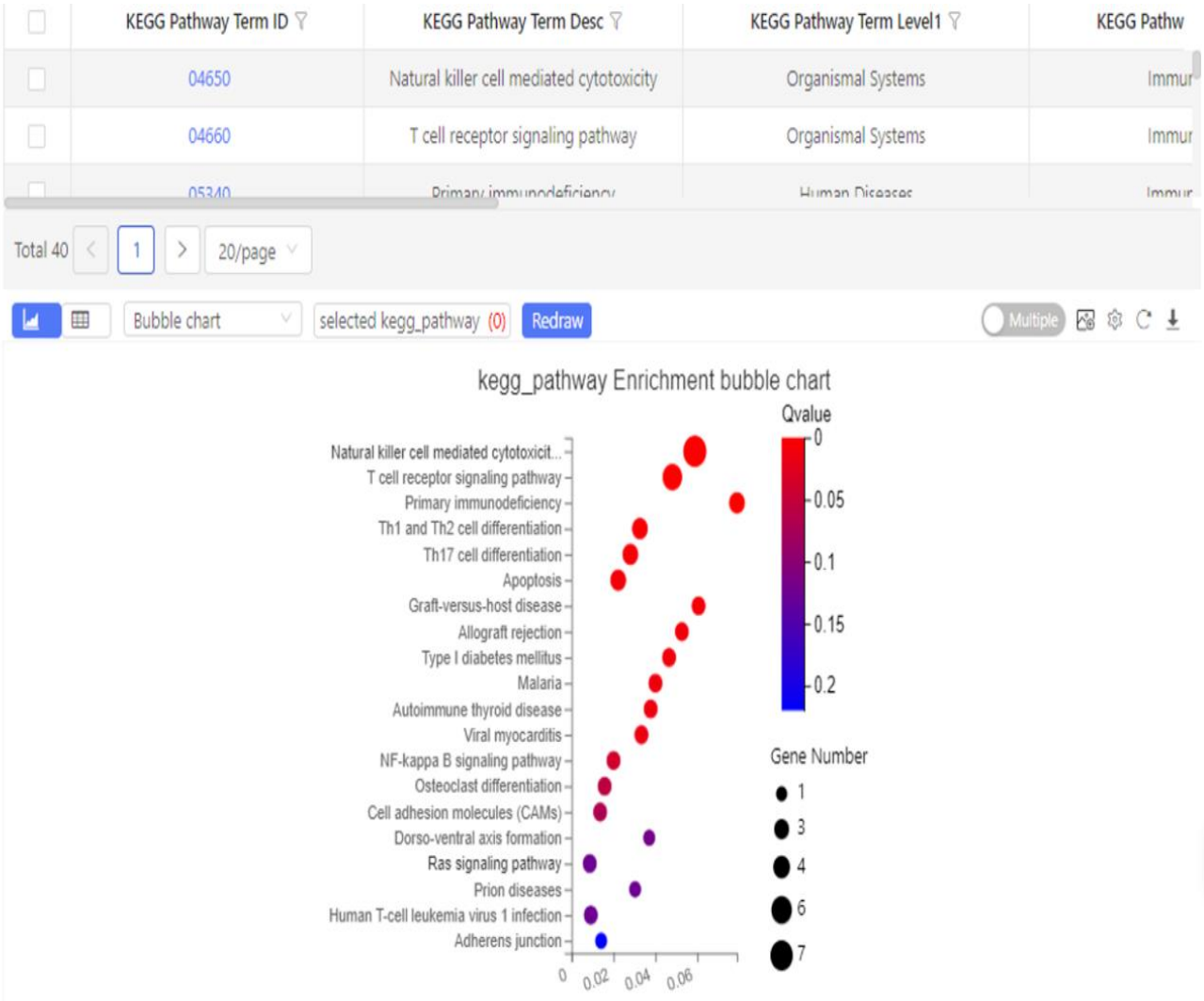


Figure 4-6. (Case Demo: KEGG pathway enrichment) shows the pathways the selected genes are enriched. By default, the graph shows the top 20 pathways sorted by q-value from small to large.

4.19. Protein Network Analysis

4.19.1. Protein-Protein Interaction (PPI) Networks by STRING

Protein interaction network analysis was performed using the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database to investigate potential functional associations. STRING is an online database used to evaluate protein-protein interaction networks inside cells (Szklarczyk et al., 2021). Indeed, STRING was utilized to understand the metabolic pathways in which a particular protein is involved. In other words, we used STRING to comprehend the metabolic pathways in which a particular protein is involved, to discern potential protein functions based on their known interactions, or to identify possible drug targets in a disease-related pathway (Szklarczyk et al., 2023).

The primary focus of STRING investigation was on Protein-Protein Interaction (PPI), encompassing both direct (physical) interactions and indirect (functional) associations. The data in STRING were sourced from various places, including experimental data, computational prediction methods, and public text collections (De Las Rivas & Fontanillo, 2010). This made it a comprehensive resource for obtaining an overview of any given protein's known and predicted interactions. Interactions in STRING were assigned a score based on the evidence supporting each interaction, aiding users in judging the reliability of predicted interactions. Finally, the analysis results were visualized in a graphical interface, indicating an interactive network visualization of the proteins and their interactions (Shaukat et al., 2021).

4.19.2. Starting Point to Work with STRING

The input form on STRING's homepage was used to initiate protein interaction analysis. Several search options were available. The "name" option allowed to enter a single protein name or a list of protein names for analysis. It was also possible to directly input an amino acid sequence, regardless of format, for the protein of interest. Moreover, the "Organism-Based" option facilitated checking if the database contained information for the specific species relevant to this study. Finally, by selecting "Protein Families," it was possible to conduct a broader search by investigating clusters of orthologous groups instead of focusing on individual protein entries within a single organism (Muley, 2023).

The entry process involved entering the protein of interest using its name or identifier. Then the organism type either from a provided dropdown menu or by typing it into the input field

was selected. An autosuggestion feature assisted in this selection process. Notably, general terms encompassing multiple organisms, such as "Mammals," also be used for the search (Figure 4-7).

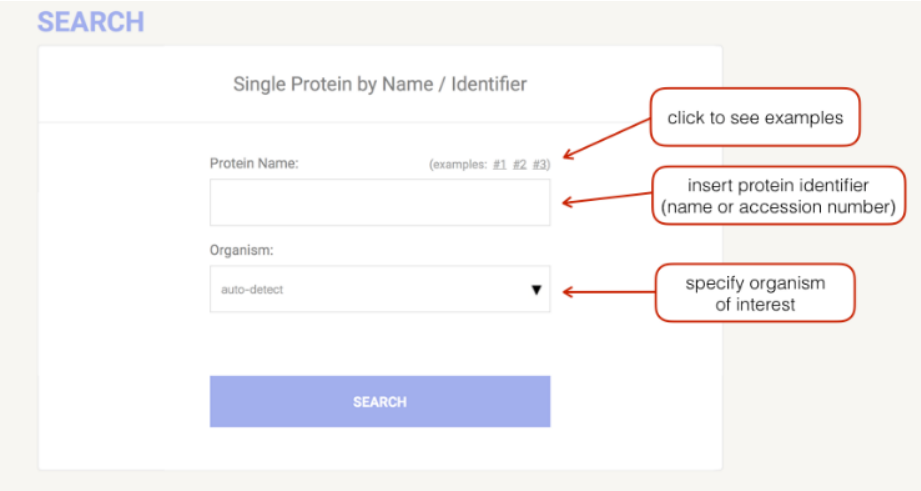


Figure 4-7. Entry process of STRING.

4.20. Statistical analysis

An unpaired student's t-test was used to analyse data between two groups. $P<0.05$ was considered to indicate a statistically significant difference. Data are presented as the mean ± standard error of the mean. All analyses were performed using Graphpad Prism 9 software.

5. Results

In this chapter, I describe the results of experiments arising from CRISPR-Cas9 based gene editing calling adenine base editing (ABE8) in collagen gene expression research. During these experiments, we aimed to edit Col1a1 gene promoter in vitro inside fibroblasts as the major resource of collagen production and evaluate the results.

The specialized gRNA was strategically designed to target both AA inside the CCAAT box located in the promoter region of the Col1a1 gene. The procedures that we adopted led to the identification of base editor combinations associated with the NGG PAM sequence. Notably, our approach demonstrated a remarkable proficiency in accurately converting AA nucleotide pairs into GG nucleotides.

5.1. Genotyping Promoter of Wildtype Fibroblasts

5.1.1. Sanger sequencing

The concentration of the extracted DNA was measured to be around 100 ng/µl using Nanodrop. The 260/280 ratio of [~2.1], which falls within the optimal purity range, indicated minimal protein contaminationn.The purified PCR product was subjected to Sanger sequencing. A clear and interpretable chromatogram was obtained, indicating high-quality sequencing data (Figure 5-1). To confirm the identity and specificity of our sequence and to explore any potential known motifs or homologous regions, we utilized the BLAST (Basic Local Alignment Search Tool) database.

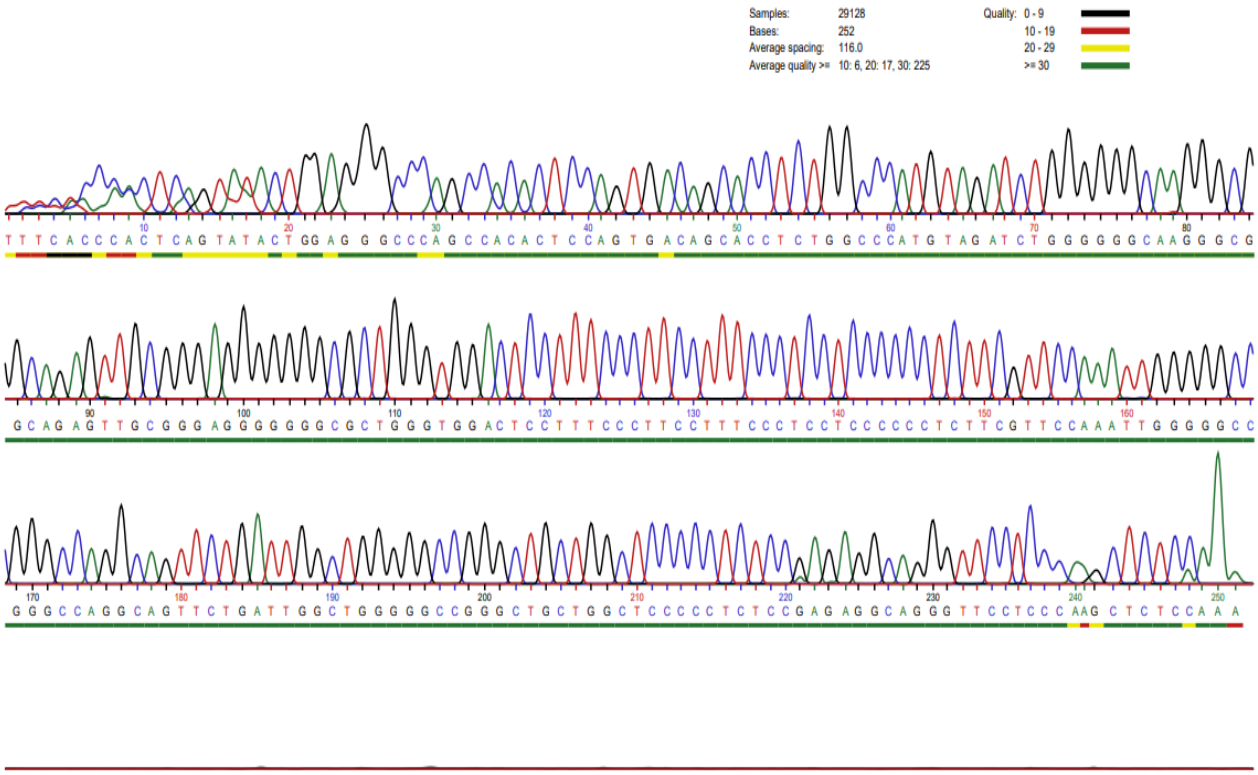


Figure 5-1. A visual representation of the Sanger sequencing chromatogram, displaying the distinct peak patterns characteristic of Col1a1 promoter in wild type fibroblasts. The peaks, in varied colours representing the four DNA bases providing a reference of unedited fibroblasts.

Chromatogram Analysis of the COL1A1 Promoter Upon obtaining the Sanger sequencing data for the COL1A1 promoter region, a detailed examination of the corresponding chromatogram was conducted to ensure data accuracy and quality.

Clarity and Resolution The chromatogram displayed distinct peaks for each nucleotide, indicative of a high-quality sequencing run. The sharpness and separation of the peaks suggest minimal background noise, thereby increasing confidence in the decoded sequence.

Peak Heights and Homogeneity Consistent peak heights were observed throughout the chromatogram, reflecting uniform incorporation of each dideoxynucleotide during the sequencing reaction. This consistency further supports the reliability of the sequencing data.

5.2. NGS Data

Alignment and Overview Using the CRISPResso2, the amplicon NGS reads were aligned to the reference sequence of the CCAAT box in the COL1A1 promoter section.

Utilizing the CRISPResso tool, we quantified the base conversion frequency from the amplicon NGS data. This analysis provided insights into the efficiency of our base editing experiment targeting AA inside the CCAAT box within the COL1A1 promoter region.The similarity of conversion rates between the forward reads (R1) and reverse reads (R2), remaining around the 18%(Figure 5-3), affirms the consistency and reliability of our base editing process. The subtle differences between the R1 and R2 rates fall within the typical range of variability observed in amplicon NGS analyses.

5.3. Genomic Modification Efficiency

In our genome editing experiment, it is vital to differentiate between the genomic regions that underwent successful modification and those that remained untouched. This differentiation provides insights into the efficiency and specificity of our CRISPR system. The following pie chart visually represents the outcomes of our experiment, offering a clear comparison between modified and unmodified reads (Figure 5-2). A conversion efficiency around 18% suggests that a substantial portion of the target cells underwent the desired base edit. This rate offers promising potential for downstream applications.

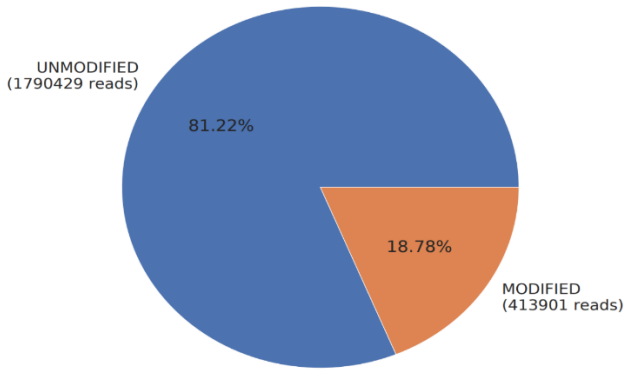


Figure 5-2. Base editing NGS results. Of the total aligned reads, 1,790,429 sequences remained unmodified, whereas 413,901 showed genomic modifications post-editing.

UNMODIFIED (1,790,429 reads) These are reads that match the reference sequence. This means that in these instances, the CRISPR system did not induce any detectable changes to the

DNA sequence. In other words, these are cells where the genome-editing tool did not cause any modification, or if it did, the DNA repair mechanism restored the original sequence without introducing indels or mutations.

MODIFIED (413,901 reads) These are reads that have been altered from the reference sequence. This indicates that the CRISPR system successfully induced a change in these instances. Figure 5-3 is showing more details of editing and non-editing among NGS data.

5.3.1. Nucleotide Percentage Quality Analysis

A crucial aspect of assessing editing NGS data is the nucleotide quality within the whole sequencing data. Reliable nucleotide calls are critical to discern genomic modifications from potential sequencing errors. To this end, we employed CRISPResso to analyze the quality of the nucleotides in our sequencing reads.

Figure 5-3. illustrating the nucleotide percentage quality across our sequencing data which represents a distinct quality percentile, providing insights into the reliability of nucleotide calls after CRISPR editing.

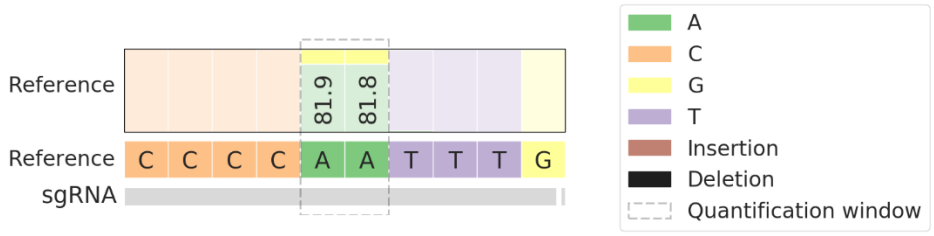


Figure 5-3. Nucleotide percentage quality distribution of sequencing reads after CRISPR editing. This visualization highlights the proportion of nucleotides at various quality scores, emphasizing the editing and non-editing regions.

By analyzing the nucleotide percentage, we aimed to ensure that our genomic modifications interpretations are based on high-quality sequencing data, minimizing the risk of misinterpretations due to potential sequencing anomalies.

5.3.2. Mutation Position Distribution in Base Editing

Base editing, unlike traditional CRISPR-Cas9 genome editing, does not cause double-strand breaks but rather induces point mutations (usually C-to-T or A-to-G changes) directly. The

specificity and distribution of these mutations are crucial in understanding the precision and efficiency of the base editor used. 16.6% (365,226 reads) is the highest proportion. 16.6% of the sequences showed mutations at a specific genomic position, suggesting this might be the most prevalent site of base editing (Figure 5-4).

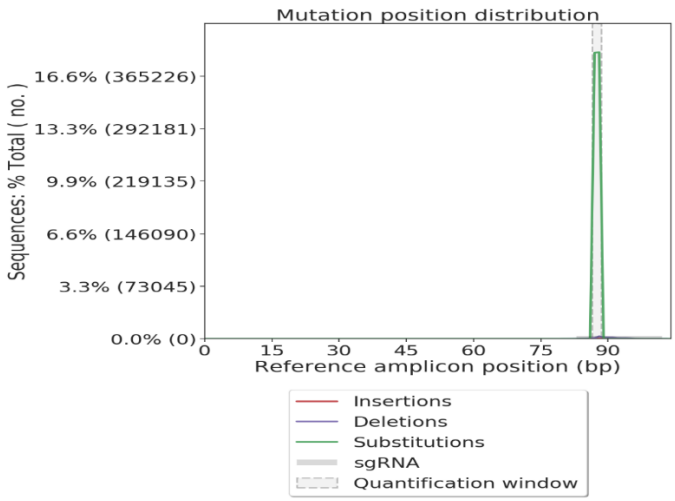


Figure 5-4. Distribution of mutation positions observed in post-base editing NGS data, as analysed by CRISPResso2.

5.4. Sanger Sequencing results via BaseEditr Platform

To gain a meticulous understanding of our base editing outcomes, Sanger sequencing as a versatile tool for single-nucleotide resolution was employed. To enhance the interpretative power of our sequencing data, we utilized the BaseEditr platform, explicitly designed for evaluating Sanger sequencing outcomes in the context of base editing.

5.4.1. Sanger Sequencing Validation

To validate the precision and accuracy of our base editing endeavour, we used Sanger sequencing for obtaining a clear transition from the AA peak to a GG peak, and confirming the successful modification of the target dinucleotide.

5.4.2. Analysis with EditR

To further quantify and analyse the editing efficiency, we utilized EditR – a specialized tool for analysing editing outcomes from Sanger sequencing data. EditR results showed an editing

efficiency that was in concordance with our NGS amplicon sequencing results. This consistency across different techniques underscores the robustness and reliability of our genome editing procedures. A graphical representation from EditR shows the proportion of sequences that underwent the AA to GG conversion in the COL1A1 promoter region compared to unedited sequences (Figure 5-5). This result is confirming the patterns observed in our NGS amplicon sequencing so clearly adding an additional layer of validation to our experimental outcomes.

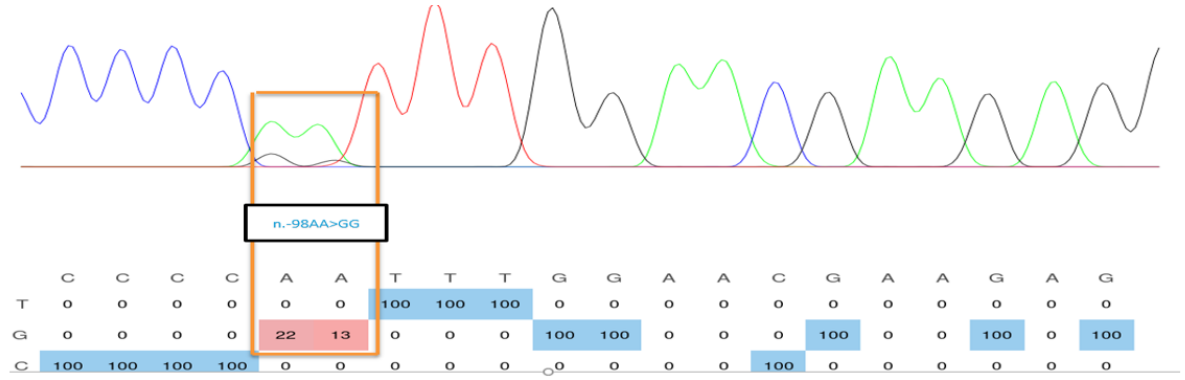


Figure 5-5. EditR analysis of adenine base editing efficiency. This figure illustrates the analysis of adenine base editing efficiency using Sanger sequencing data by EditR tool.

Furthermore, the nucleotide quality scores that are integral for discerning genuine edits from potential sequencing errors, were consistently high across the edited regions, thus supporting our confidence in the identified base modifications.

5.5. Isolation and Verification of Single Base-Edited Fibroblast

Clonal Expansion Approach Following our definitive confirmation of successful base editing via NGS and Sanger sequencing, our research trajectory was centred on the meticulous task of isolating clonal cells that had undergone the anticipated genetic modifications for further applications. To achieve this crucial step, we applied the clonal expansion in the 96-well plates format. After about two weeks by the sequencing of selected colonies from different well of a 96 well plate of our results indicated the successful clonal expansion objectives (Figure 5-6).

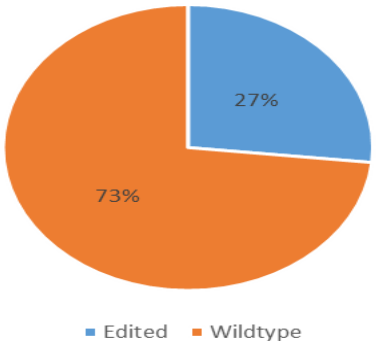


Figure 5-6. The colonies selected from random wells of a 96-well plate. About 27% of then shows genetic editing, and remaining (73 %) were unedited.

5.5.1. Chromatograph Findings

To further substantiate our sequencing results and offer a visual representation of the base edit, we present a detailed chromatograph. This illustrative data distinctly delineates A to G transition in single colonies. Notably, the chromatograph clear and distinct peaks serve as a robust testament to the purity and integrity of our clonal samples, emphasizing that the genetic modifications observed are not also available at wildtype fibroblats (Figure 5-7a) but are the direct consequence of our base editing interventions at edited cells (Figure 5-7b).

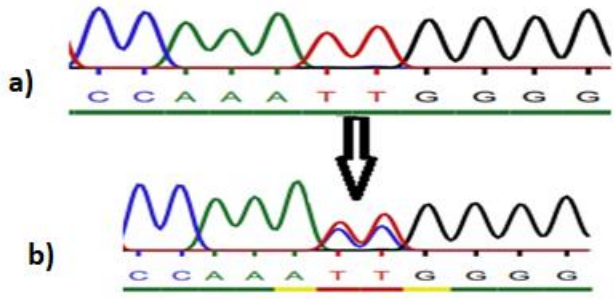


Figure 5-7. Chromatograph showing the promoter of Col1a1 before (a) and after editing (b).

5.6. Collagen1a1 mRNA Expression

Graphical Representation The accompanying bar graph visually represents the relative expression levels of collagen 1a1 mRNA in both groups of edited (EC) and wild type cells (WT). Both WT fibroblasts and EC mean expression values are depicted, and their respective standard deviations (SD) are illustrated to provide a sense of data dispersion around the mean (Figure 5-8).

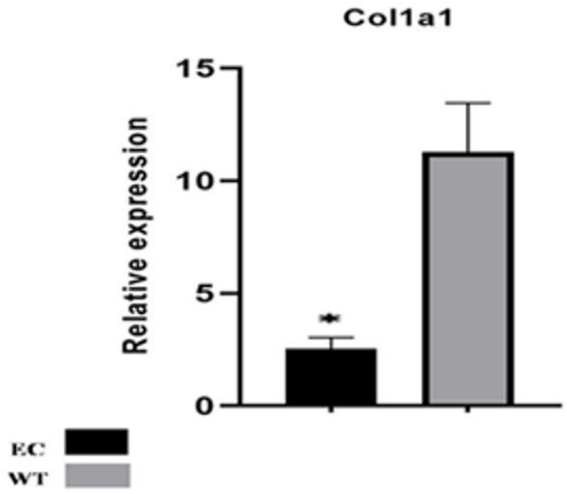


Figure 5-8. The differences in collagen 1a1 mRNA expression levels between base edited cells(EC) and wild-type fibroblasts(WT) are statistically significant (p-value < 0.05).

Note: Given the inherent variability in mRNA and to ensure the robustness of data, GAPDH gene as a standard housekeeping gene was used for normalization. This gene was selected for its known stable expression across various experimental conditions. Replicate measurements were taken to provide a reliable average and to reduce potential technical errors.

Statistical Findings A statistical t-test was applied to evaluate the significance of the observed differences in mRNA expression levels between groups. The results yielded a p-value < 0.05, showing that the differences in collagen 1a1 mRNA expression levels between wild-type fibroblasts and base edited cells are statistically significant. This finding demonstrates a meaningful alteration in collagen 1a1 expression as a result of the base editing process.

5.7. Evaluation of Col1A1 Protein in Base-edited and Control Cells

5.7.1. Semi-Quantitative Analysis

To investigate the protein expression levels of Col1a1 in ECs compared to WTs, an indirect immunostaining approach as a semi-quantitative approach was used.

Fluorescence intensities from the stained cells were quantified using ImageJ software revealed a significant difference in Col1a1 protein between the EC and WT cells (Figure 5-9). This finding confirms a meaningful alteration in collagen 1 expression as a result the base editing process.

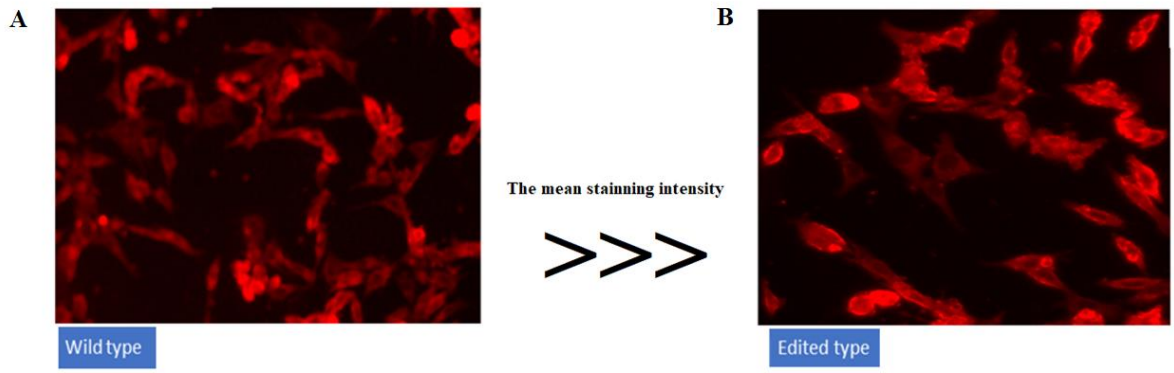


Figure 5-9. Indirect immunostaining of Col1a1 of wildtype (A) and edited cells (B). A difference in fluorescence intensity (using ImageJ software) indicates reduced collagen I in the matrix of edited cells (B) in comparison with the matrix of wildtype (A).

5.7.2. Hydroxyproline measurement

A colorimetric assay kit (Sigma-Aldrich) was employed for the measurement of hydroxyproline in the supernatant of cell culture plates. We also confirmed a reduction in collagen by measuring collagen hydroxyproline content which was consistent with mRNA and MS/MS results (Figure 5-10).

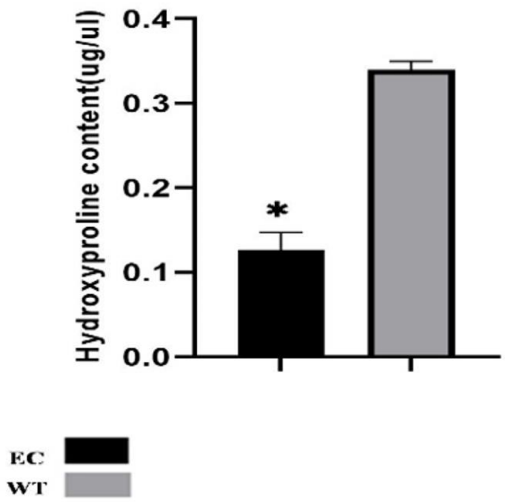


Figure 5-10. Hydroxyproline content in edited cells (EC) and wild-type (WT) cells. The graph shows the hydroxyproline content in EC and WT groups. The WT group has a higher hydroxyproline content in comparison with the EC group (P-value<= 0.05).

5.8. Proteomic results of Base Editing: A Tandem Mass Spectrometry Approach

5.8.1. Differential Proteomic Profiling

To characterize the proteome results of Col1a1 promoter editing, a meticulous proteomic assay was executed by employing tandem mass spectrometry in biological triplicates. Upon analysis of proteomics data, a plethora of proteins demonstrated significant deviation in their abundance between the EC and WT cells illustrating a statistical significance threshold of $p < 0.05$.

5.8.2. Volcano Plot Results

To visualize the landscape of proteomic alterations, a volcano plot was constructed (Figure 5-11). This plot showing the -log10 (p-value) against the log2-transformed fold change, offering a view of the protein changes in terms of statistical significance.

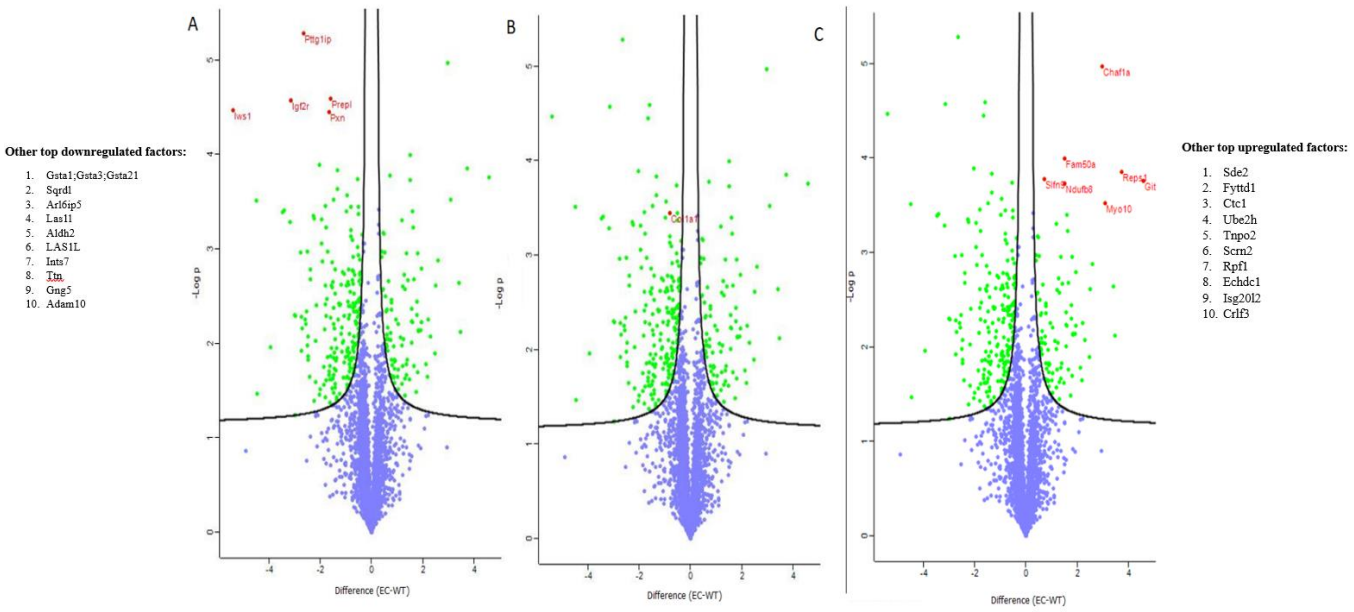


Figure 5-11. Volcano plots illustrating the differential protein expression between base-edited (EC) and wild-type (WT) cells. The x-axis indicated the log2-transformed fold change, while the y-axis displays the -log10 (p-value). Proteins significantly downregulated are shown in upper left side, and those upregulated are shown in in upper right side. The horizontal line marks the significance threshold at $p < 0.05$. Top dowregulted (A) and upregulated (C) protein are labelled. Col1a1 protein is among downregulated proteins (B).

To detect the impact of base editing on specific proteins, we conducted a targeted data analysis focusing on collagen 1a1, a protein of paramount interest in our study. This investigation was promoted by data acquired in biological triplicates (threshold of $p < 0.05$). Indeed, upon analysing the filtered data, a profile plot was constructed to explore the abundance patterns of Col1a1 across the triplicate samples (Figure 5-12). This box plot demonstrateing a consistent and significant reduction trend in the abundance of Col1a1 protein in EC samples when compared with the WT counterparts ($p < 0.05$).

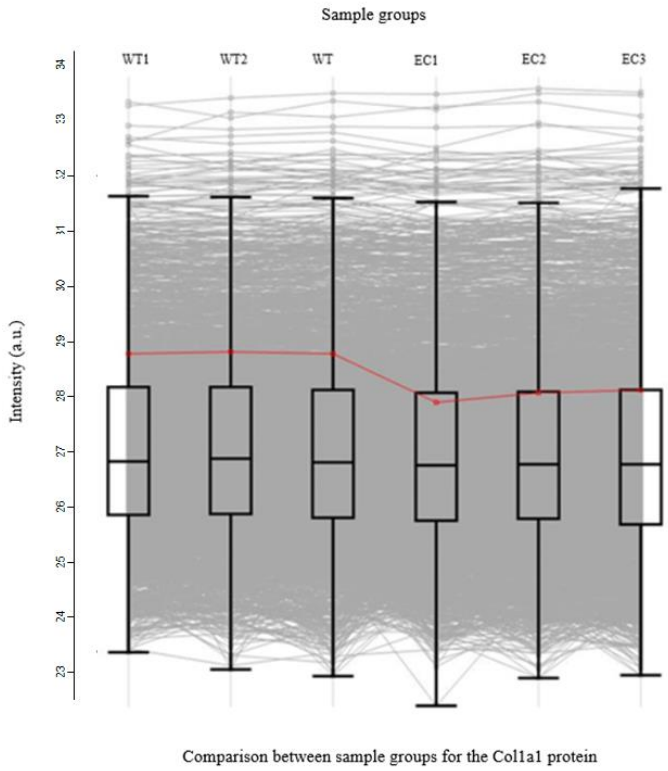


Figure 5-12. Profile plot derived from Perseus software shows significant and consistent reduction in the abundance of Collagen 1a1 protein between and wild-type (WT) and base-edited cells (EC) across triplicate samples. The y-axis represents the normalized protein abundance, while the x-axis shows the individual replicates. Base-edited samples are depicted in [EC1,EC2,EC3], and wild-type samples in [WT1,WT2,WT3]. Error bars represent standard deviation.

5.8.3. Differential Proteomic Profiles: A Heatmap Analysis

To further elucidate the subtle proteomic differences between EC and WT cells, a heatmap was generated based on data derived from biological triplicates. This graphical approach allows for a more detailed observation of the expression patterns across samples.

The heatmap shows distinct clusters of proteins that were either upregulated or downregulated in the EC cells. The colour gradient provides a clear visual contrast of the differential protein abundances between groups (Figure 5-13).

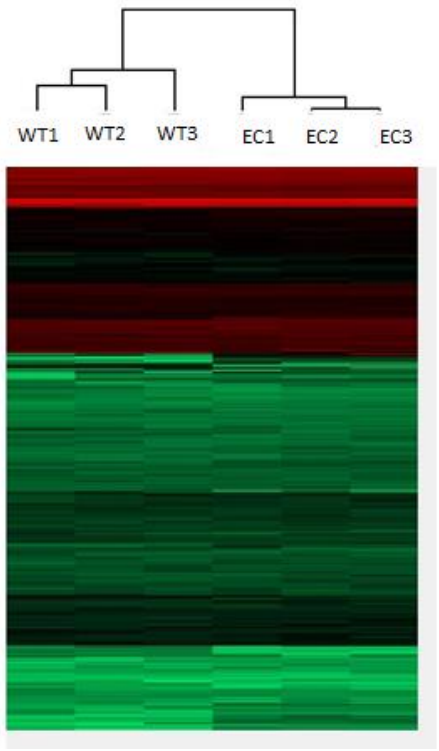


Figure 5-13. Heatmap illustrating the differential protein expression profiles between wildtype (WT) base-edited (EC) groups. Columns correspond to individual replicates. The colour gradient conveys protein abundance levels between WT groups (WT1,WT2,WT3) to EC groups (EC1,EC2, EC3).

5.9. Computational biology analysis results

5.9.1. Protein-Protein Interaction (PPI) Networks Analysis Highlight the Central Role of Chromatin Assembly Factor 1 Subunit A

Protein-protein interaction (PPI) network results highlight the central role of chromatin assembly factor 1 subunit A (CAF-1A) upregulation in WT cells. In other words, in our comparative proteomics study between EC and WT cells, a number of proteins exhibited differential expression. Indeed, our tandem mass spectrometry data showed that CAF-1A is the first rank among significant upregulated list. This protein is known to be a core component of the CAF-1 complex, which is believed to play a critical role in chromatin assembly during DNA replication and repair. The CAF-1 complex is composed of three subunits: CAF-1 p150 (encoded by the CHAF1A gene), CAF-1 p60 (encoded by the CHAF1B gene), and CAF-1 p48 (encoded by the RBBP4 gene). These subunits work together to facilitate the deposition of histones onto newly replicated DNA during DNA synthesis(Volk & Crispino, 2015).

To gain insights into the potential role and interactions of CAF-1A in the cellular environment, we performed a PPI network analysis using the STRING computational platform. The resulting

network, presented in Figure 5-14, places CAF-1A at the core, surrounded by its primary interacting partners. This interconnection suggests that the change CAF-1A may have downstream effects on Myc protein regulation via CHAF1B which is itself a key oncogenic factor (Dang, 2012). This indirect link between CAF-1A and Myc highlights the significance of our findings indicating a possible role of CAF-1A dysregulation after applying our editing also on possible initiation of oncogenic signalling cascades.

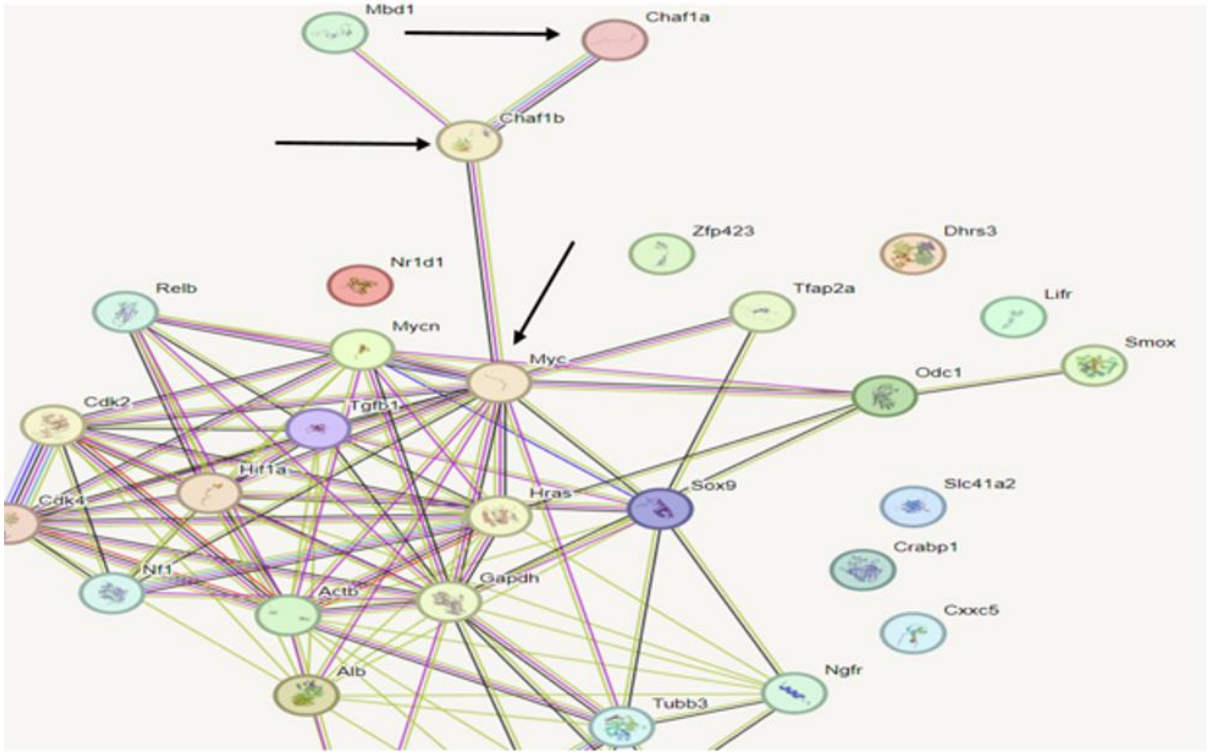


Figure 5-14. Protein–protein interaction (PPI) network focused on chromatin assembly factor 1 subunit A (Chaf1a, the gene encoding the CAF-1A protein). Interactions are represented as lines connecting proteins. The line between two proteins means an interaction score no less than 0.4, and the more interactions with other proteins, the more important this protein is.

Moreover, from the network point of view, CAF-1A does not function alone. Its upregulation might have several effects across the protein network, potentially influencing DNA replication, repair processes, and chromatin assembly dynamics in edited cells.

5.9.2. Possible Role of Pituitary Tumor-Transforming Gene 1 Protein-Interacting Protein (PTTG1IP) after Promoter Editing

Our proteomics analysis distinctly revealed a pronounced downregulation of PTTG1IP. PTTG1IP, known for its potential roles in cellular activities and oncogenic pathways, offers an intricate aspect of our protein alteration landscape in edited cells. PTTG1IP, as its name suggests, is a key interacting protein for the Pituitary tumor-transforming gene 1 (PTTG1). This interaction is vital as PTTG1IP may aid in the nuclear translocation of PTTG1. Given the significance of nuclear translocation events in regulating cellular functions, the downregulation of PTTG1IP may result in cascading effects in cellular signalling pathways, especially those related to cell proliferation, differentiation, and potential oncogenic transformations.

Next, we employed the STRING computational platform centred on PTTG1IP to show the possible consequence of PTTG1IP downregulation and its wide interplay in the cellular proteomic network. This analysis yielded a comprehensive PPI network with Trp53 as the hub (downstream of Pttg1 ip), as depicted in Figure 5-15. The Trp53 gene, also known as the p53 gene, is a critical tumor suppressor gene found in humans and many other organisms. It plays a pivotal role in regulating cell division and preventing the formation of tumors (Brady & Attardi, 2010). The p53 protein, encoded by the Trp53 gene, acts as a transcription factor, meaning it regulates the expression of other genes.

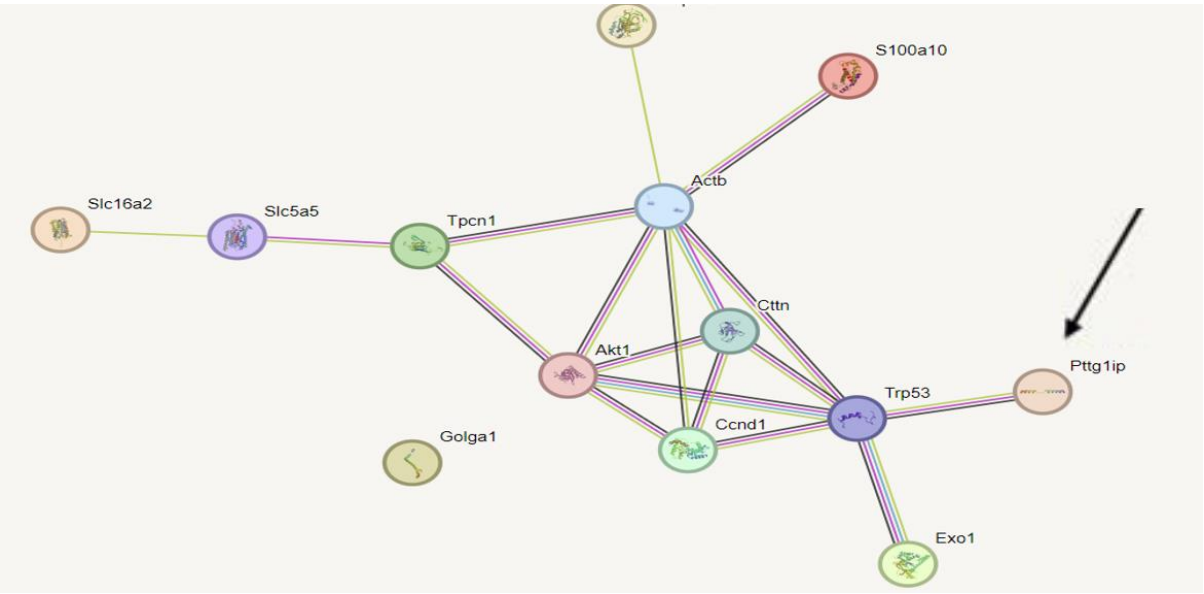


Figure 5-15. Protein–protein interaction (PPI) network emphasizing the hub postion of PTTG1IP. The line between two proteins means an interaction score no less than 0.4, and the more interactions with other proteins, the more important this protein is.

The intricate web of interactions underscores the crucial role of PTTG1IP, in potentially influencing several cellular proteins. This network show the wide perspective in which PTTG1IP operates and offers clues about the larger implications of its downregulation in base-edited cells.

5.10. Transcriptomic Results (High-throughput RNA Sequencing)

To investigate the transcriptional impact of our base editing strategy, RNA sequencing (RNA-seq) was conducted on both EC and WT samples. This section describe the findings of RNA-seq experiments.

5.10.1. Transcriptomic Landscape of Base-edited Cells

A meticulous exploration of the transcriptomic alterations caused by base editing was undertaken to understand the molecular variations after Col1a1 promoter editing in fibroblasts. (Figure 5-16).

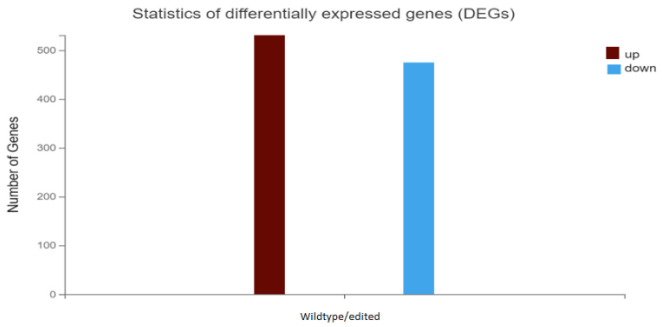


Figure 5-16. The X-axis represents the alignment scheme of differentially expressed proteins (DEGs) for each group, and the Y-axis represents the corresponding number of differentially expressed genes (DEGs). Red represents the number of DEGs up-regulated (531), and blue represents the number of DEGs down-regulated (475).

From this analysis, we identified 1,005 differentially expressed genes (DEGs) in the EC group in comparision with the WT group. This includes 531 genes upregulated, while ans 475 downregulated genes. The extensive list of DEGs underscores the comprehensive impact of

our base editing strategy on cellular transcriptional dynamics, necessitating further exploration to decode the functional implications of these changes.

5.10.2. Heatmap Visualization of Gene Expression Profiles

One of the most insightful graphical representations to emerge from our RNA-seq data was the heatmap, which is featured in the Figure 5-17. This visualization method allowed for a comprehensive understanding of the expression patterns from specific genes across both base-edited and wild type fibroblast samples.

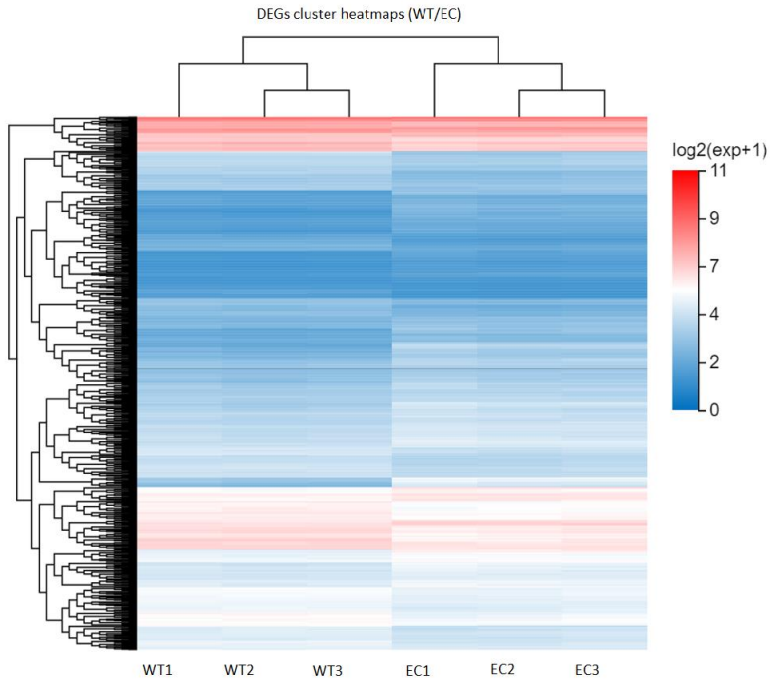


Figure 5-17. The horizontal axis is the log2 of sample (expression value +1), and the vertical axis is the gene. Under the default colour matching, the warmer the colour block has the higher the expression level and the colder the colour block has the lower the expression.

5.10.3. Key Observations from the Heatmap

Differential Expression Patterns The heatmap revealed distinct clusters of genes that were dysregulated in WT and EC samples. This emphasizes the molecular differences induced by our the base editing process.

Sample Consistency Biological replicates of the same type clustered closely together, validating the consistency of experimental procedures and the reliability of our RNA-seq data.

To better understanding the biological implications of these observed patterns, the differentially expressed genes (DEGs) were further subjected to pathway and functional enrichment analyses, which the results are presented in the subsequent sections.

5.11. Gene Ontology Analysis Reveals Distinct Patterns of Dysregulated Transcripts in ECs Compared to WTs

Biological Process (BP) In ECs, the dysregulated transcripts were predominantly enriched in key biological processes such as cellular processes which could encompass a variety of cellular functions including cell division, and cell-to-cell communication. Furthermore, biological regulation processes were prominent, suggesting a role in governing cellular activities and homeostasis. The metabolic processes covering both anabolic and catabolic activities, indicated a significant level of biochemical alteration as well. Additionally, genes implicated in the 'response to stimuli category suggest that ECs may have heightened or altered responsiveness to environmental factors or signaling molecules.

Cellular Component (CC) The data in figure 5-18 highlights the spatial distribution of the dysregulated genes within cellular architecture. A high enrichment was observed in components like 'organelle parts of cells' potentially indicating dysfunction or alteration in subcellular structures such as mitochondria, endoplasmic reticulum, or Golgi apparatus. Similarly, enrichment in the 'membrane' category points to the importance of membrane-bound processes like signaling or transport. Protein-containing complexes were also prominent, which could imply a change in multi-protein assemblies involved in cellular functions. Lastly, the 'extracellular region' enrichment suggests changes in the cell's interaction with its external environment, possibly affecting processes like adhesion, migration, or signaling.

Molecular Function (MF) On the functional level, 'binding' was a major category that came into focus, covering a wide array of molecules like ions, small metabolites, and even macromolecules like proteins and nucleic acids. 'Catalytic activity' indicated a high enrichment in genes coding for enzymes involved in a myriad of biochemical reactions. Furthermore, 'transcription regulatory activity' was notably enriched, pointing towards a possible alteration in the gene expression profiles and regulatory networks within ECs.

In summary, the figure 5-18 presents a comprehensive map of gene enrichment across multiple GO domains, revealing intricate and likely interconnected changes in ECs compared to WTs.

These findings could serve as a crucial foundation for future studies aimed at understanding the underlying mechanisms of EC-specific behaviors or dysfunctions.

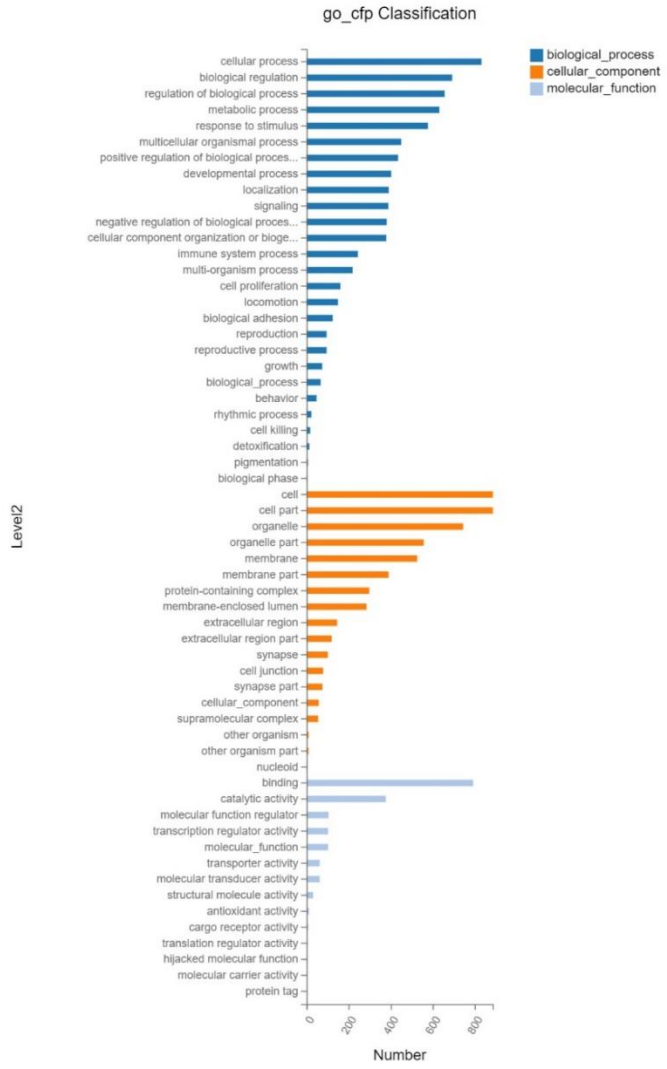


Figure 5-18. The X-axis represents the number of genes annotated to the GO entry, and the Y-axis represents the GO functional classification.

5.12. Network Results of DEGs and Enriched Pathways

5.12.1. KEGG Pathway Analysis

The identified key driver genes were subjected to KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathway analysis to elucidate their functional roles and to identify significantly enriched pathways. It assessed the likelihood that the observed gene set resulted from random chance, and pathways were considered significantly enriched at a p-value threshold of less than 0.05.

Note: In the KEGG pathway, annotation classification (the phyper function in R software) was used to perform the enrichment analysis and to calculate the P-value. In addition, the Q-value was obtained by correction of P-value. Generally, the function of Q-value <= 0.05 is regarded as a significant enrichment.

A detailed interaction network was generated to show the connections between the key driver genes and enriched KEGG pathways. This network is visually represented in figure 5-19.

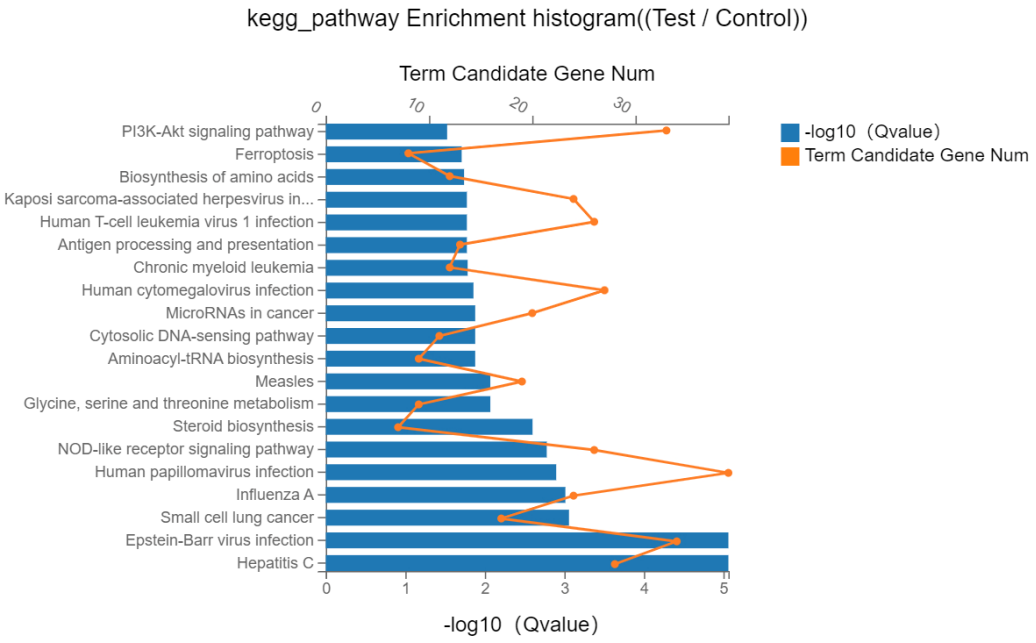


Figure 5-19. A histogram showing the distribution of enriched KEGG diseases associated with the key driver genes in base-edited and wild type fibroblasts. The X-axis represents the different KEGG disease pathways, while the Y-axis displays the count or frequency of key driver genes associated with each disease pathway.

5.13. Key Identified Pathways

5.13.1. PI3K-Akt Signalling Pathway

Number of Genes: 33 key genes were mapped to this pathway (Fig 5-20A).

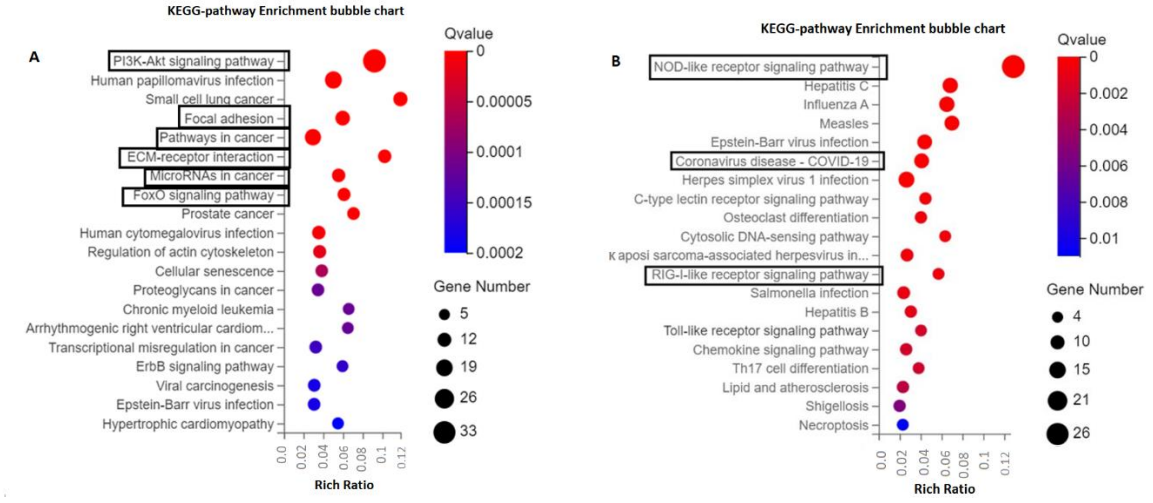


Figure 5-20. Signaling pathway enrichment analysis identified PI3K-Akt (A) and NOD-like receptors (B) as top enriched pathways in the differentially expressed protein.

5.13.2. NOD-like Receptor Signalling Pathway

Number of Genes: 26 key genes were found to be part of this pathway. These 26 genes were identified to contribute in various aspects of inflmmation responses (Fig 5-20B).

Functional Significance: This pathway is central to the immune response and is involved in the detection of pathogens and cellular damage.

Functional Significance: The phosphoinositide 3-kinase- Protein Kinase B (PI3K-Akt) pathway plays a crucial role in cellular processes like growth, proliferation, differentiation, motility, survival, and intracellular trafficking. Its dysregulation has been implicated in various pathologies, including organ fibrosis (Figure 5-21).

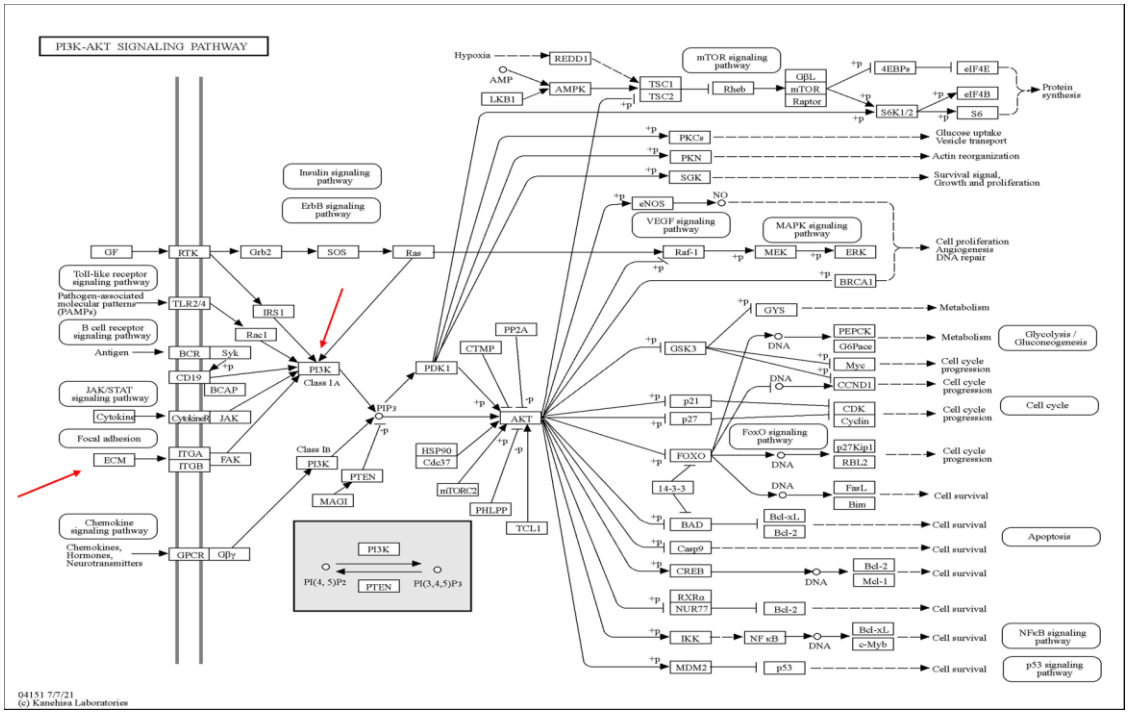


Figure 5-21. The ECM as a key player which addresses PI3K-Akt pathway. (adapted from KEGG database, http://www.genome.jp/kegg/pathway.html).

5.14. Analysis of Potential off-Targets

We employed both computational and experimental approaches to assess thoroughly the off-target effects of ABE8e and sgRNA treatment. The computational tool Cas-OFFinder was used to identify potential off-target sites in the genome with PAM sequences with three or fewer mismatches to the intended protospacer. In parallel, we used proteomics approach to detect if mutations at the DNA level could result in significant dysregulation of the corresponding gene's protein expression between ECs and WTs. Cas-OFFinder predicted five potential off-target sites, detailed in the table 5-1. One off-target sites were located in the intergenic non-coding region and four were in exons, all of which led to silent mutations. Interestingly, we found no significant dysregulation in protein expression within proteomics data related to the off-target edits. Collectively, these findings from genome and proteome wide off-target analyses of ABE8e base-edited fibroblasts did not reveal off-target mutations of anticipated clinical relevance.

Table 5-1. Analysis of potential off-target editing at protein level as predicted by Cas-OFFinder.

	Genes	Target	Chromosome	Position	Direction	Mismatches
1	Dtnb	crRNA: CCCCAATTTGGAACGAAGAGNGG DNA: CCCCAtTTgGGAAtGAAGAGTGG	chr12	3741185	-	3
2	Dlgap1	crRNA: CCCCAATTTGGAACGAAGAGNGG DNA: CCCCAAaTTGGAAaGAAGAcGGG	chr17	70569595	+	3
3	Gm2005	crRNA: CCCCAATTTGGAACGAAGAGNGG DNA: gCCCAATTTGaAACcAAGAGAGG	chr1x	26440891	-	3
4	Tenm2	crRNA: CCCCAATTTGGAACGAAGAGNGG DNA: CCCCAAcTTGtAACGAAaAGAGG	chr11	37480871	-	3
5	Itga3	crRNA: CCCCAATTTGGAACGAAGAGNGG DNA: CCCCAATTTGGAACGAAGAGGGG	chr11	94936108	-	0

5.15. Extracellular matrix (ECM) results

Extracellular matrix (ECM) serves as the architectural foundation of tissues, influencing cellular behaviour, tissue integrity, and overall functionality. The composition, topography, and mechanical properties of the ECM play crucial roles in determining tissue-specific cell functions including adhesion, migration, proliferation, and differentiation (Winkler et al., 2020). In the realm of tissue engineering and regenerative medicine, the ability to modulate and control the ECM characteristics can offer several advantages, especially in designing constructs that closely mimic desired tissue properties. This becomes particularly convincing when considering the use of edited cells (Jo et al., 2021).

This part of the project focused on using SEM microscopy to visualize edited fibroblast morphologies and their interaction with ECM. The SEM analysis of ECs cultures grown on 2D substrates revealed normal morphologies that were uniformly spread throughout the culture It should be mentioned that the normal morphology of fibroblasts has been well-described in scientific texts and literature for several decades. Typically, fibroblasts exhibit a spindle-shaped, elongated, and flattened morphology. Our SEM findings also show these morphologies, indicating that editing has no negative effect on the morphology. In other words, editing the ECM followed by editing Col1 at this level has no negative impact on fibroblast morphology (Figure 5-22).

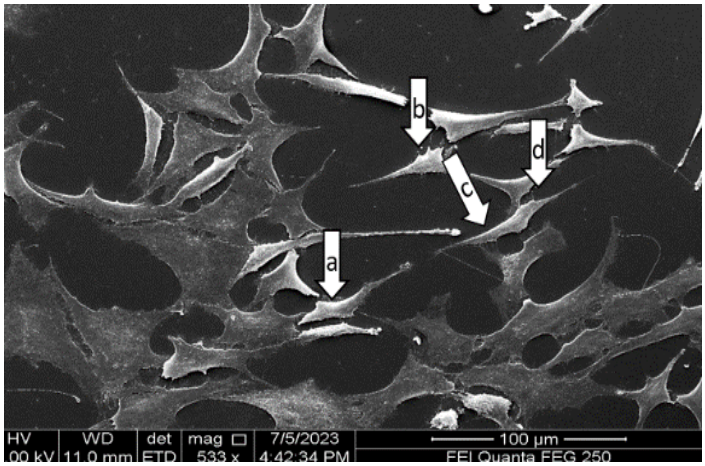


Figure 5-22. Edited fibrobalsts have an elongated (a) or spindle shape (b) and flat form (c and d) Cell-Cell Interactions.

In our results, presented in the figures 5-23-A and 5-23-B, we elucidating the multifaceted interactions between cells and the ECM, a crucial aspect that fundamentally influencing cellular behaviour and functionality.

Figure 5-23-A provides a comprehensive overview of the edited Cell-ECM interactions. It vividly displays various essential aspects including the points of adhesion (a), showing the initial contact and attachment points of edited cells to the ECM. The figure also highlights cell extensions embedding into the matrix, demonstrating the interpenetration of cellular structures within the ECM, thereby ensuring stability and enhanced communication (b). Furthermore, areas (c) where the surrounding ECM organization is apparent reveal the organized and intricate structure of the ECM, which plays a pivotal role in cellular guidance and support.

In the figure 5-23-B, the intricate dynamics of a single fibroblast intimately attached to the ECM are captured. This feature illustrates the uniform and dynamic interplay between cellular structures and the ECM which are crucial for cellular survival, function, and adaptation to environmental cues.

These insights into edited Cell-ECM interactions are paramount for better understanding of the complex cellular behaviours in different physiological and pathological contexts. Understanding these interactions opens avenues for innovative therapeutic strategies, especially in tissue repair and regenerative medicine, by manipulating the cellular environment.

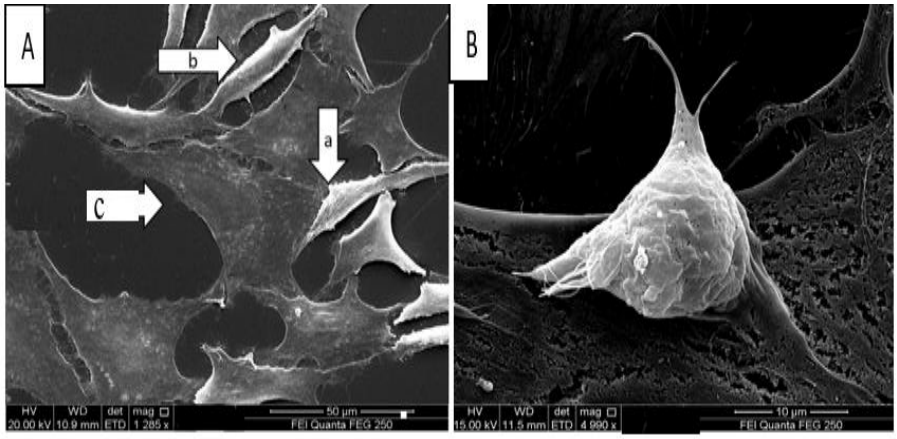


Figure 5-23-A. The edited Cell-ECM Interactions: This includes points of adhesion, cell extensions embedding into the matrix (a and b), and areas where the surrounding ECM organization is evident (c). Figure 5-23-B. A single fibroblast intimately attached to the ECM, illustrating the dynamic interplay between cellular structures and extracellular matrix components.

The scanning electron microscopy (SEM) image presented here depicts the extracellular matrix (ECM) of fibroblasts, with a predominant presence of collagen fibers. This high-magnification view, captured at 12,785x, reveals the intricate and dense network of collagen fibrils, which provide structural support and strength to the ECM. The sample was analyzed using an FEI Quanta FEG 250 SEM, operating at an accelerating voltage of 20.00 kV and a working distance of 11.0 mm. This detailed visualization underscores the crucial role of collagen in maintaining tissue integrity and facilitating cellular communication within the ECM (Figure 5-24).

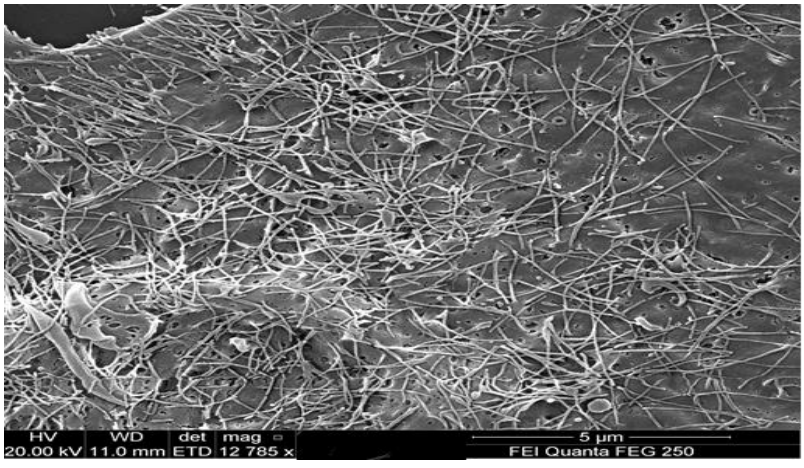


Figure 5-24. The SEM fibrillar structures: These fibers (mostly collagen) appear as thin, elongated forming a mesh-like structure.

Moreover, the comparative phase-contrast microscopy images revealed both the edited and wild type fibroblasts exhibiting typical spindle-shaped and flattened morphologies, characteristic of normal fibroblasts. This indicates that the collagen1a1 gene editing did not alter the fundamental cellular structure. Thus, the edited fibroblasts maintain their normal morphology similar to that of the wild type fibroblasts (Figure 5-25).

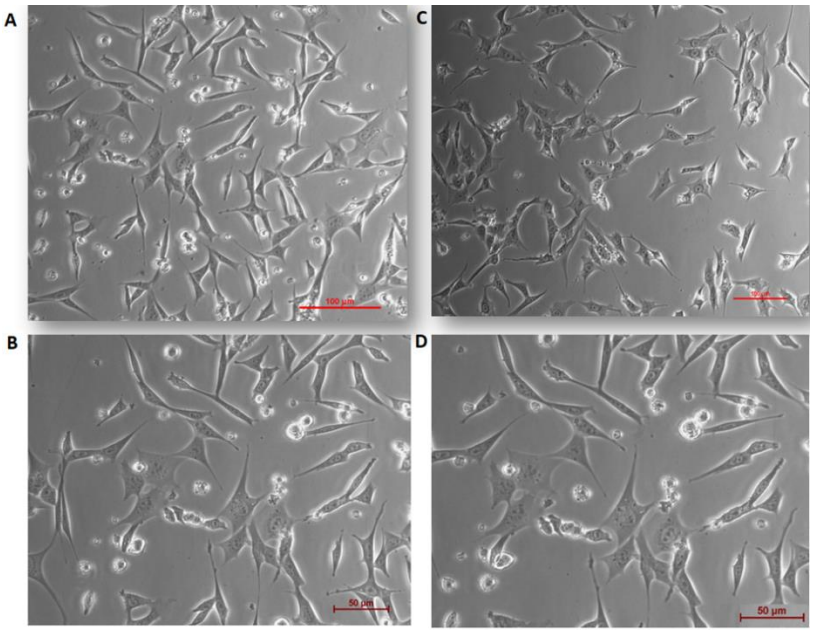


Figure 5-25. Comparative phase-contrast microscopy images of fibroblasts. Panels A (10x magnification) and B (20x magnification) show the morphology of edited fibroblasts, while panels C (10x magnification) and D (20x magnification) display the morphology of wild type fibroblasts. In both edited and wild type fibroblasts, the cells exhibit characteristic spindle-shaped and flattened morphologies typical of normal fibroblasts, indicating that the collagen1a1 editing process did not alter the fundamental cellular structure (morphology). Scale bars represent 100 µm (A, C) and 50 µm (B, D).

5.16. Comparative proteome profiling on fibroblast adaptability to ECM changes

Next, we investigated whether WT cells would exhibit compensatory mechanisms in collagen synthesis when exposed to reduced collagen ECM. Indeed, given the challenges associated with precisely targeting all fibroblasts in vivo, WT cells may compensate for collagen turnover by increasing synthesis. To address this issue, we utilized cultured EC and WT cells in a standard two-dimensional (2-D) cell culture to produce an ECM bioscaffold over a 5-day incubation period. Subsequently, we decellularized the EC and WT fibroblasts were seeded onto the decellularized matrix for proteome profiling (Fig.5-26).

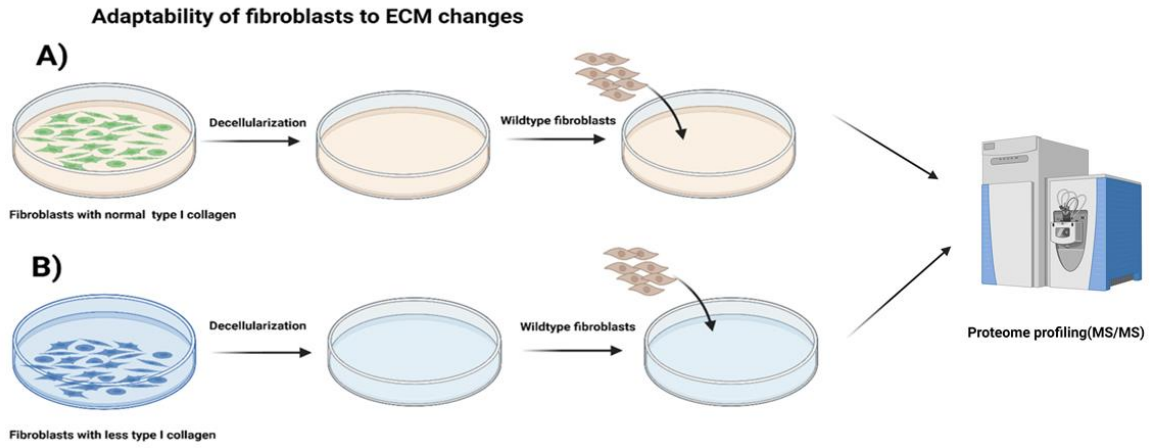


Figure 5-26. Generating two ECM bioscaffold from WTs (A) and ECs (B) and analyzing WT fibroblasts response using proteome profiling.

Our findings revealed that WT fibroblasts do not exhibit increased synthesis of collagen 1A1 when exposed to the decellularized ECM generated by ECs compared to the natural ECM.Interestingly, the expression trend of the procollagen C-protease enhancer protein (PCOLCE) gene is the same as Col1a1 which is compatibale with our expectation due to this fact that PCOLCE has the unique ability to accelerate procollagen maturation and coregulated with COL1A1 (Figure 5-27).

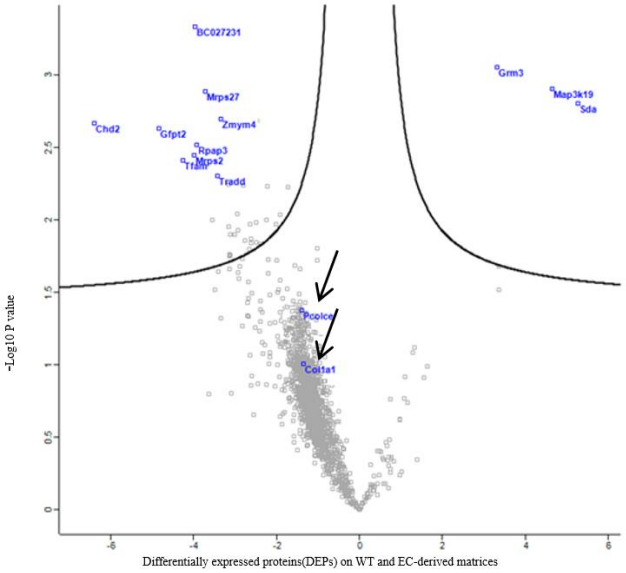


Figure 5-27. Wild-type fibroblasts exposed to decellularized ECM from ECs did not show significantly increased synthesis of Collagen 1A1. The expression trend of the procollagen C-protease enhancer protein (PCOLCE) gene is the same.

Moreover, to obtain a general understanding of changes in the proteome of WT cells after exposing to new microenvironment (derived from ECs) we analysed differences at proteom level. Downregulated proteins that were reduced during adaptation to the new matrix were enriched in proteins that belonged to mitochondrial biogenesis and mitochondrial transcription-translation process (Table 5-2).

Table 5-2. Dysregulated proteins were clustered according to KEGG analysis and as illustrated the pathways listed are primarily related to mitochondrial functions.

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Synthesis Of UDP-N-acetyl-glucosamine R-HSA-446210	0.00006902	0.008731	221.80	2125.08
2	Transcriptional Activation Of Mitochondrial Biogenesis R-HSA-2151201	0.00007795	0.008731	42.05	397.79
3	Mitochondrial Translation Elongation R-HSA-5389840	0.0003024	0.01077	26.04	211.06
4	Mitochondrial Translation Initiation R-HSA-5368286	0.0003024	0.01077	26.04	211.06
5	Mitochondrial Translation Termination R-HSA-5419276	0.0003024	0.01077	26.04	211.06
6	Mitochondrial Translation R-HSA-5368287	0.0003722	0.01077	24.20	191.07
7	RORA Activates Gene Expression R-HSA-1368082	0.0003734	0.01077	83.13	656.15
8	Mitochondrial Biogenesis R-HSA-1592230	0.0003848	0.01077	23.92	188.04

Furthermore, gene ontology analysis using the functional annotation chart tool provided by DAVID (The Database for Annotation, Visualization and Integrated Discovery) showing that dysregulated proteins were belonged to clusters mostly related to mitochondrial metabolism and biogenesis processes as well (Table 5-3).

				Count	P_Value	Benjamini
Annotation Cluster 1	**Enrichment Score: 19.68**					
UP_KW_DOMAIN	Transit peptide	RT		41	4.6E-25	1.0E-23
UP_SEQ_FEATURE	TRANSIT:Mitochondrion	RT		38	5.2E-23	6.1E-20
UP_KW_CELLULAR_COMPONENT	Mitochondrion	RT		53	7.2E-19	2.5E-17
GOTERM_CC_DIRECT	mitochondrion	RT		62	9.6E-19	3.2E-16
GOTERM_CC_DIRECT	mitochondrial inner membrane	RT		32	2.4E-16	4.0E-14
Annotation Cluster 2	**Enrichment Score: 7.66**					
GOTERM_CC_DIRECT	mitochondrial inner membrane	RT		32	2.4E-16	4.0E-14
GOTERM_BP_DIRECT	mitochondrial translation	RT		14	5.3E-12	7.3E-9
GOTERM_CC_DIRECT	mitochondrial large ribosomal subunit	RT		10	3.5E-9	3.9E-7
UP_KW_MOLECULAR_FUNCTION	Ribonucleoprotein	RT		17	4.4E-8	1.8E-6
UP_KW_MOLECULAR_FUNCTION	Ribosomal protein	RT		14	7.4E-8	1.8E-6
GOTERM_CC_DIRECT	ribonucleoprotein complex	RT		18	1.8E-7	1.5E-5
GOTERM_CC_DIRECT	ribosome	RT		14	3.0E-7	2.0E-5
GOTERM_MF_DIRECT	structural constituent of ribosome	RT		12	2.3E-5	1.1E-2
GOTERM_BP_DIRECT	translation	RT		14	2.8E-5	1.3E-2
KEGG_PATHWAY	Ribosome	RT		11	4.6E-5	2.9E-3
Annotation Cluster 3	**Enrichment Score: 3.83**					
GOTERM_MF_DIRECT	oxidoreductase activity	RT		18	1.1E-4	2.7E-2
UP_KW_MOLECULAR_FUNCTION	Oxidoreductase	RT		18	1.4E-4	2.2E-3
UP_KW_LIGAND	NAD	RT		10	2.1E-4	5.9E-3
Annotation Cluster 4	**Enrichment Score: 3.05**					
GOTERM_CC_DIRECT	mitochondrial respiratory chain complex III	RT		5	4.5E-6	2.5E-4
GOTERM_BP_DIRECT	mitochondrial electron transport, ubiquinol to cytochrome c	RT		5	5.1E-6	3.4E-3
KEGG_PATHWAY	Chemical carcinogenesis - reactive oxygen species	RT		13	1.2E-5	2.3E-3
UP_KW_BIOLOGICAL_PROCESS	Electron transport	RT		9	2.3E-5	1.6E-3
GOTERM_CC_DIRECT	respiratory chain	RT		7	2.3E-5	9.6E-4
KEGG_PATHWAY	Oxidative phosphorylation	RT		10	3.1E-5	2.9E-3
UP_KW_BIOLOGICAL_PROCESS	Respiratory chain	RT		7	6.7E-5	2.1E-3

Table 5-3. Functional annotation clustering by DAVID. The data is grouped into four top clusters of annotations, based on related protein functions: Within each cluster, various terms are listed which mostly associated with mitochondrial metabolism.

Considering the crucial role of PPIs in pivotal cellular processes such as signal transduction and transcriptional regulation, as well as the fundamental importance of macromolecular interactions in understanding biological systems, our study extended to include a network analysis of enriched mitochondrial proteins. This approach applied to map out the intricate web of interactions within the mitochondria, providing insights into possible functional dynamics (Fig. 5-28).

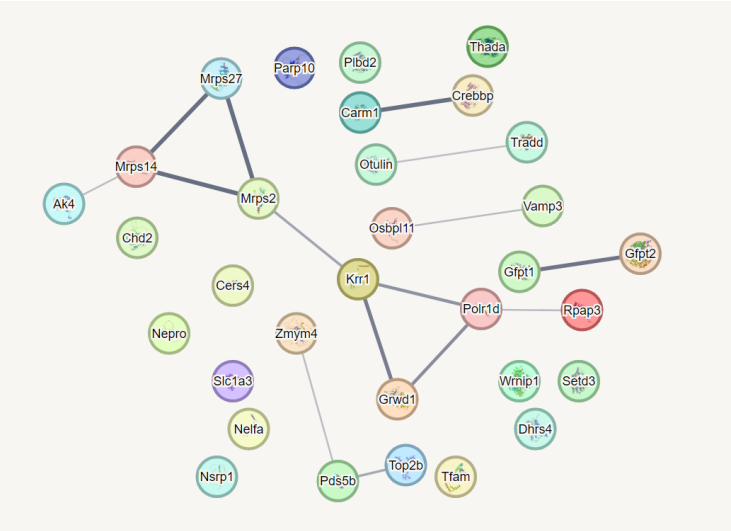


Figure 5-28. A network diagram of protein-protein interactions (PPIs) between the top dysregulated proteins in WT fibroblasts after exposure to EC-derived ECM, using the STRING database. The network nodes represent proteins, and the edges represent predicted or experimentally validated interactions. The thickness of each edge represents the interaction strength.

5.17. Distinctive molecular signatures of breast cancer cells in response to the EC-derived matrix

To investigate the impact of reduced collagen density on cancer cell growth within the tumor microenvironment, mammary epithelial cancer cells (MCF-7) were cultured in ECM derived from ECs and WTs (Fig. 5-29).

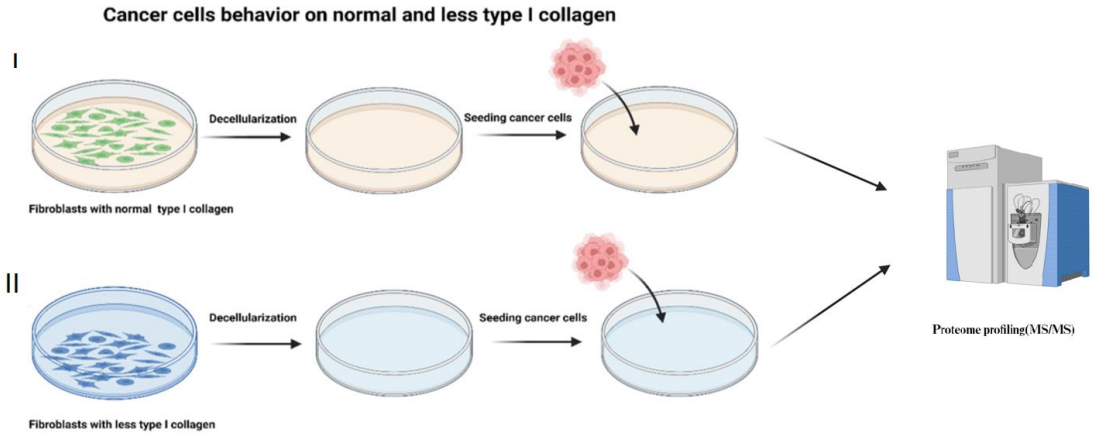


Figure 5-29. Experimental design for investigating the behavior of MCF7 cancer cells on matrices derived from WT and EC cells.

Cell proliferation assays showed that cancer cells in ECM derived from ECs, exhibited reduced proliferation compared to those in ECM derived from WTs (Fig. 5-30).

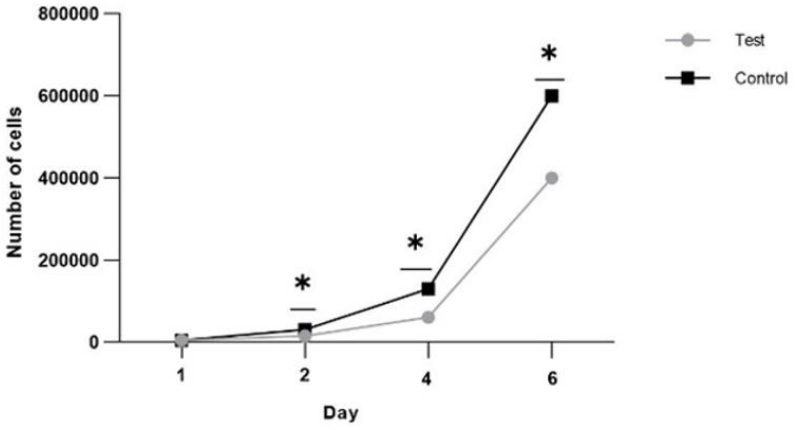


Figure 5-30. Cell proliferation assay: MCF-7 cells were cultured in ECM derived from control (wildtype) and test (eidted cells) at the indicated time points.

These results indicate that promoter editing of Col1a1, which reduce collagen expression possibly induce cell apoptosis, and reduce MCF7 cell proliferation. This result was further supported by a clonogenic assay (colony forming) demonstrating that the number of apoptotic cells is lower in MCF-7/eECM cells compared to MCF-7/ECM (Fig. 5-31).

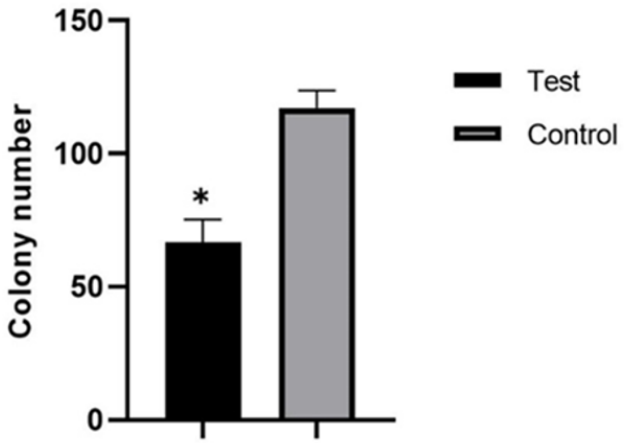


Figure 5-31. Clonogenic assay (Colony forming) results assay results: The number of colonies beween cells whose were cultured in ECM derived from control (wildtype) and test (eidted cells) (P-value<= 0.05).

Next, we compared the global proteomic profiles of matched pairs of MCF-7/eECM and MCF-7/ECM in order to identify pathways that may be responsible for the cell growth inhibition and the general understanding molecular signature associated cancer cells (Fig. 5-32).

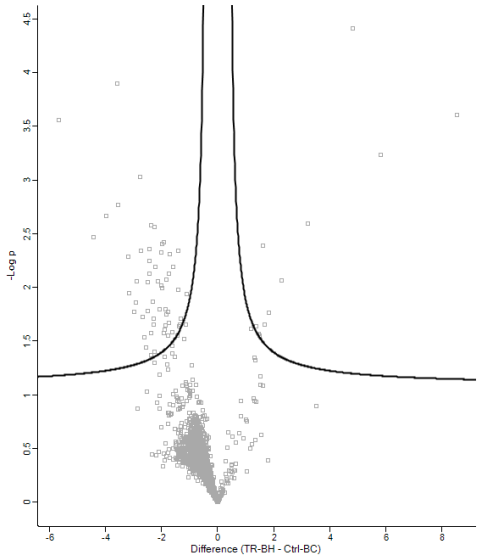


Figure 5-32. Comparative proteomic profiling between MCF7 whose were cultured in ECM derived from control (wildtype) and test (edited cells).

Finally, we have identified that the top cluster of enriched proteins is focal adhesions (FAs) (Fig. 5-33).

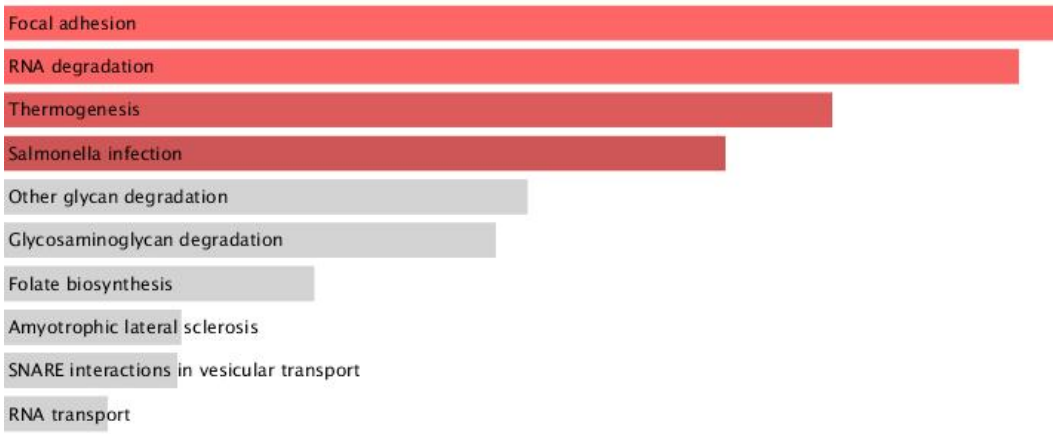


Figure 5-33. The enriched focal adhesion proteins were identified as the top enriched pathway of differentially expressed proteins (DEPs). The results are the averages from three independent experiments.

6. Discussion

To date, published literature has not documented any evidence in terms of editing the promoter section of Col1a1 gene. Our versatile approach, especially when considering various disorders associated with the overexpression of collagen stands as a pioneering effort in this field which can open novel avenues for research and potentially offer an effective therapeutic interventions for different conditions influenced by collagen overexpression.

Our study launched on a pioneering journey to explore the potential of ABE8 gene editing in converting the CCAAT box into the CCGGT version within the promoter of Col1A1gene in fibroblasts. This sequence is located in the promoters of many genes and plays a crucial role in initiating gene transcription (Laporte et al., 2014; Martyn et al., 2017; Traxler et al., 2016).

This alteration was hypothesized to have an impact on the reduction of Col1a1 gene in fibroblasts (the main source of collagen production), implying a breakthrough in developing a new window for anti-fibrotic medications. The conducted research yielded positive outcomes, demonstrating a successful conversion from AA nucleotides into to GG nucleotides by adenine base editor.

Our research findings are promising in the area of gene editing for collagen gene expression modulation. The successful transformation from AA to GG in the CCAAT box of the Coll1a1 inside the basal promoter of fibroblasts indicatting a significant step towards precision in gene editing technologies. This precise modification led to a decrease in collagen 1a1 mRNA and protein, substantiating the initial hypothesis and opening innovative avenue for anti-fibrotic interventions. Our results highlight the versatility of ABE8 gene editing for non-coding DNA editing (promoter), not only in successfully altering specific DNA sequences but also in achieving desired downstream effects with no off-target mutations that are clinically significant.

In the following sections, we will further unfold the multifaceted significance of our research, eexploring into various aspects that underscore its impactful contribution to the field. The exploration of these dimensions will further reinforce the pivotal role of our research in enhancing the development and utilization of physiologically relevant in vitro and in vivo models, paving the way for more comprehensive and accurate investigations into the molecular mechanisms of various diseases.

6.1. Expanding the Horizons of A-to-G Base Editing

In the current research, we have for the first time investigated the capacity of the ABE8 system to perform reliable A-to-G base editing at the promoter of collagen1a1 gene. Our findings confirm the ABE8 strong ability for A-to-G base editing within fibroblasts, underscoring the availability of located NGG (PAM) in this part of fibroblast genome. Indeed, a major limitation in utilizing base editing (BEs) is the necessity for an appropriately located NGG-PAM. This fact has been restricted the target of interest within a tight operational window that significantly limiting the number of accessible target locations (Zuo et al., 2019). To overcome this challenge, multiple research teams have engineered enhanced and sophisticated BE versions with either different PAM prerequisites or broader PAM compatibilities (Molla & Yang, 2019). For instance, Kim et al. presented a series of BEs by replacing the SpCas9 with alternative Cas9 variants, enabling the targeting of specific DNA regions with new positioned PAM such as NGCG sequences (Y. B. Kim et al., 2017). Marion et al also introduced a method to create a zebrafish model for melanoma predisposition based on the simultaneous base editing of multiple genes but using near PAM-less base editor (Rosello et al., 2022). However, despite these advancements, PAM constraints continue to hinder the precise targeting of certain gene loci. Altogether, our results considerably expand the BE applications to introduce new nucleotides in the promoter of collagen 1 gene or possibly other genes. In other words, our versatile approach stands as a precise method for mimicking human pathologies by effectively introducing permanent point mutations in promoters without PAM concern.

Furthermore, the evidence of post-editing indicates a direct correlation between the CCAAT box editing and gene regulation which is compatibale with previous studes that this box is key in the gene expression (Dolfini et al., 2009; Lekstrom-Himes & Xanthopoulos, 1998). This observation is pivotal, providing a molecular insight into the regulatory mechanisms governing gene expression and presenting a novel approach to modulate these processes for further investigation on gene expression studies. For example, how the new box (CCGGT) influenced on RNA polymerase II to initiate of transcription and how related transcription factors such as NF-Y reacte with this new box.

6.2. The Potential and Eficiency of Synthetic RNA Delivery into Fibroblasts

Deep sequencing results revealed the profound ability of ABE8 to invoke site-specific, base substitutions with an efficiency of around 18% in the fibroblast cells. These findings shine a light on the new genotypic mosaicism generated by ABE, opening up a panorama of opportunities for studying fibrotic diseases. Because editing 100% of fibroblasts could lead to a huge reduction in collagen, which is essential for maintaining organ homeostasis. This reduction could cause tissue rupture and render your main approach ineffective. Therefore, a mosaic editing strategy, where only a portion of cells (e.g., 18%) are edited, could help balance collagen turnover in the target organ. This way, the reduction in collagen would be at a level that likely would not disrupt organ homeostasis.

This observed efficiency might be attributed to the ABE RNA. It is postulated that ABE RNA possesses certain inherent advantages over other sstructures such as plasmids and ribonucleoproteins (RNPs) (Jiang et al., 2020; Sürün et al., 2020). This is a significant observation as the use of RNPs, although effective, but RNPs are associated with higher expenses, potentially making it a less accessible option for widespread applications (Anzalone et al., 2020; Porto et al., 2020).

It is worth to mention that the utilization of RNA delivery is a versatile approach for delivering genome-editing tools into embryos (Hashimoto & Takemoto, 2015; Ji Liu et al., 2019; Ohtsuka et al., 2018). This method involves in vitro transcription and subsequent purification of synthetic RNA encoding base editors. When united with a guide RNA, the synthetic RNA can be co-introduced into single-cell zygotes of various species, including mice, humans, rabbits, rats, and zebrafish, through electroporation or direct injection (Gurumurthy et al., 2022; Patton et al., 2018; Q. Yan et al., 2014). This procedure has been shown to produce point mutations with notable efficiency. Such research firmly positions the RNA delivery of base editors into embryos as a versatile and effective technique for creating animals with customized point mutations. Hence, our in vitro methodology in terms of synthetic RNA usage for targeting the CCAAT box has successfully laid the groundwork for in vivo studies across various diseases and making animal models by targeting promoter of different genes.

Furthermore, we used synthetic RNA over plasmid DNA for due to its reported rapid expression, which is known to diminish unintended off-target editing. This choice also avoids the hazard of random plasmid DNA integration into the genome, leading to enhanced overall

editing efficacy (Leonhardt et al., 2014; Tavernier et al., 2011). This could be a possible explanation that why we did not observe any off-target effects.

Additionally, clinical application of gene editing for treatment goals using an in vivo RNA system is also suitable for fibroblasts. The synthetic RNA is gaining attention as a type of therapeutic agent for addressing various diseases. A potential method for delivering these RNAs therapeutically is through lipid nanoparticles (LNPs). Notably, vaccines using lipid nanoparticle-mRNA are broadly utilized in clinical settings for COVID-19, highlighting a significant advancement for RNA therapies (Baden et al., 2021; Dörrie et al., 2020; Phua et al., 2013). Hence, subsequent research, assessing the delivery of ABE8e mRNA alongside LNP technology for targeting fibroblasts in different organs, holds substantial promise. The RNA-centric gene editing treatments can be economically and expansively synthesized, establishing it as an influential framework for the innovation of novel, cost-efficient remedies. After replacing AA nucleotides in the CCAAT, amplicon NGS sequencing confirmed that edited transcripts following treatment with ABE8e had GG nucleotides. This was also evident in Sanger sequencing chromatogram. Moreover, reduction of mRNA and Col1a1 protein expressed in edited cells were significant. These observations give promise that this level of genomic DNA edition achieved using ABE8e in our study will be possibly sufficient to generate new CCGGT box inside promoters in vivo as well.

6.3. Striking the Balance: The Role of Mosaicism in Gene-Editing Trials and its Implications for Collagen I Modulation

Since the dawn of gene-editing technologies, enhancing the efficiency of targeted gene editing has consistently emerged as a pivotal focus within the scientific efforts. A myriad of research endeavours have been dedicated to increase the efficacy of gene-editing approaches i.e. targeting the maximum number of cells as much as possible. For instance, Ryan et al. constructed a modified CRISPR/Cas system specifically designed for simultaneous editing of multiple genes. They reported the ability to precisely delete targeted chromosomal segments with an efficiency spanning from 70 to 100% (Cobb et al., 2015).

In the area of base editing also, numerous trials have been conducted to boost the editing efficiency. The laboratory led by David Liu –as stands at the forefront of this innovative research- has been introduced various types of Adenine Base Editors (ABE), Cytosine Base Editors (CBE), and Prime Editors (PE) to enhance editing precision and efficiency (Gaudelli

et al., 2017). A comparative analysis reveals that ABE8, in comparision with for instance ABE2, exhibits a marked improvement in editing efficiency, indicating the continual progress in the domain of gene editing technologies (P. J. Chen & Liu, 2023) to have higher efficiency.

These trials are entirely understandable that maximum editing, holds significant implications for numerous single-gene disorders such as thalassemia (the most abundant single gene disorder) , hemophilia (Frangoul et al., 2021; Malech, 2021), and sickle cell anemia (Park & Bao, 2021) as well as multifactorial diseases like cancer. The objective is clear: to efficiently and effectively eliminate all disease-causing pathogenic mutations in all cells. These endeavours are not just in vitro; some of these sophisticated editing approaches are already being applied in clinical settings or are part of ongoing official clinical trials (Chiesa et al., 2023; H. Zhang et al., 2021).

While most investigations emphasize to increase editing efficiency to target maximum number of cells, a notable innovation in our work diverges from this trend. Indeed, our research uniquely focuses on the Col1a1 gene expression, the most abundant protein in the human body, holding vital functional significance in the whole normal body physiology (Mienaltowski & Birk, 2014; Prockop & Kivirikko, 1984). Reducing collagen expression, if excessive, can pose risks to the target organ. In other words, reducing expression should not be excessive, and efforts should not be made to target maximum cells in order to prevent potential risks associated with excessive expression reduction. A delicate balance must be maintained, as excessive editing could inadvertently compromise tissue integrity, leading to potential tissue rupture due to low level of collagen protein (Deshmukh et al., 2016; Tsamis et al., 2013).

This aspect underscores our work highlighting the criticality of cautious in the editing of crucial genes, to avoid inadvertently triggering adverse physiological consequences (Ricard-Blum, 2011). Indeed, generating a mosaism genotype (i.e. not to target all cells) is a new offer of our project to have a balance of the favourite phenotype for targeting collagen I or other ECM proteins because if all cells be edited possibly organ haemostasis will be disrupted. Our project introduces the idea of a mosaic genotype which can be considered for a balanced phenotype when targeting collagen I or other ECM proteins.

6.4. Heterochromatine Is not a Barrier in Col1a1 Promoter Editing Approach

Cas9 aiming at a DNA site within a highly condensed locus of the genome (heterochromatin) results in reduced binding (Wu et al., 2013). This diminished binding might be attributed to limited accessibility to the PAM sequence (Hotta & Ellis, 2008). The target sequence's heterochromatin state can limit the Cas9 movement, making it essential to consider targeting areas with less condensed chromatin for more effective results (Knight et al., 2015). Given the successful editing at the CCAAT binding site, we can draw an additional inference: this region of the genome likely contains less heterochromatin or dense DNA, which aligns with prior research indicating that promoter regions of highly transcribed genes are typically not condensed (Leidescher et al., 2022; Van Steensel & Belmont, 2017).

6.5. Developing Novel Approach to Address the Therapeutic Gaps in Fibrotic Disorders

Fibrosis is characterized by an excessive of fibrous connective tissue, like collagen and fibronectin parts of the ECM. This accumulation within and around inflamed or injured tissue may result in permanent scars, organ dysfunction, and potentially death (Hao et al., 2022; Piek et al., 2016). These features are common in advanced stages of different diseases including liver and kidney fibrosis, idiopathic pulmonary fibrosis (IPF), heart failure and so on (Schuppan & Kim, 2013; Wiśniewska et al., 2021). Additionally, fibrosis significantly contributes to various chronic autoimmune diseases such as scleroderma (Gabrielli et al., 2009), rheumatoid arthritis (Roschmann & Rothenberg, 1987), Crohn’s disease (Chao Li & Kuemmerle, 2014), ulcerative colitis(I. O. Gordon et al., 2014), myelofibrosis (Abou Zahr et al., 2016), and systemic lupus erythematosus (Chalayer et al., 2014). Beyond this, fibrosis plays a crucial role in tumor invasion (J. Y.-F. Chung et al., 2021), metastasis (Cox & Erler, 2014), chronic graft rejection (Lundvig et al., 2012), and the emergence of numerous progressive myopathies (Corallo et al., 2017). Despite the growing evidence for fibrogenesis as a primary factor for morbidity and mortality in numerous conditions, limited treatment strategies are available that specifically address the underlying factors in pathogenesis of fibrosis (Wynn, 2007; Wynn & Ramalingam, 2012).

Interestingly, the proposal to hinder the self-assembly of collagen I has been suggested as an approach for antifibrotic treatment. Nevertheless, this concept has thus far attracted limited attention in the domain of fibrosis due to low efficacy (Knüppel et al., 2017). Alternatively, initiatives have been executed to assess the obstruction of collagen crosslinking by the enzyme

lysyl oxidase-like 2, particularly a subsequent step to spontaneous fibril formation that fortifies existing fibrils. Of late, a phase II study employing a monoclonal anti–lysyl oxidase-like 2 antibody was terminated due to an absence of efficacy (Knüppel et al., 2017).

Pirfenidone, as a crucial example, approved by the US FDA, serves as a treatment for idiopathic pulmonary fibrosis (IPF), showing notable effectiveness in halting the progression of fibrosis in animal models of IPF, and heart failure patients. It acts by inhibiting the over-expression of collagen type I and heat shock protein 47, a collagen-specific molecular chaperone stimulated with transforming growth factor-β1 in vitro (Lopez-de la Mora et al., 2015). Despite its efficacy, the administration of pirfenidone presents several challenges. It has a brief half-life of 2.4 hours in the body, demonstrating slow absorption post-oral administration and is primarily expelled through urine. Furthermore, clinical studies highlight its initial hepatic metabolism, leading to certain adverse effects post-oral administration, underscoring the need for careful management and consideration of these factors in treatment planning (Ghazipura et al., 2022; Lewis et al., 2021).

It is also worth to mention that, collagen augmentation is not confined to the area of organ fibrosis. An analysis conducted using the UALCAN datasets (http://ualcan.path.uab.edu/) reveals that the mRNAs COL1A1 is significantly upregulated in a variety of human tumor tissues. These include kidney renal clear cell carcinoma (KIRC), liver hepatocellular carcinoma (LIHC), lung adenocarcinomas (LUAD), lung squamous cell carcinoma (LUSC), and stomach adenocarcinoma (STAD), all showing increased levels compared to their corresponding normal tissues. This insight highlights the extensive and diverse involvement of collagen in various pathological conditions, emphasizing the critical need for targeted therapeutic interventions (Chandrashekar et al., 2022).

An abundance of experimental data demonstrating that pharmaceutical inhibitors aimed at targeting type I collagen exhibit anti-tumor properties. Several of these compounds are currently under evaluation in different phases of clinical trials, displaying their potential role in cancer therapeutics. But also there has been several disappointing clinical trial results(Shi et al., 2022a). For example, synthetic inhibitors of Matrix Metalloproteinase (MMPIs) underwent examination for application across various cancer types and contrary to the highly encouraging preclinical data, all trials failed to diminish tumor progress or enhance overall survival. Additionally, MMPIs manifested significant side effects (Winer et al., 2018). So, it is clear that, while numerous anti-fibrotic treatments appear hopeful within experimental frameworks,

clinical evidence remains disappointing and inconclusive (Fang et al., 2017). Therefore, making an effective anti-fibrotic medication is an important unmet need in clinics and remains the need for quick in vitro screening tools to check main anti-scar compounds.

Here, we show, to the best of our knowledge for the first time, that adenine base editing technology (RNA-based) downregulate collagen I, an critical player for the initiation and pursuing of fibrosis. This suggests possible therapeutics perspective not in the area of fibrotic organs but also anti-tumor agents. Indeed, our method provides new avenue for the concept of inhibition of collagen production as a promising antifibrotic or anti-tumor strategy. Notably, this is a particularly interesting finding in fibroblasts, in which the fibrotic process starts by high expression of Col1a1.

In fact, ABE editing has the potential to synergize with emerging fibrosis treatments including siRNA (Ishiwatari et al., 2013; J.-C. Wang et al., 2010), other small-molecule drugs (Ito et al., 2017; Nanthakumar et al., 2015) or antisense oligonucleotides that target collagen or related factors (Hagiwara et al., 2007; Nishino et al., 2003). Several approaches for collagen reduction to remedy fibrotic conditions are being tested in clinical trials (C. Z. C. Chen & Raghunath, 2009). It is not yet known which strategy is safest or most effective (Collins & Raghu, 2019; McPherson et al., 2017). However, our base editing approach offers several potential advantages. First, generating a permanent change at DNA level by precise ABE8 editing reduce the concentration of collagen 1a1 (the primary determinant of pathogenic collagen concentration in most organs) more effectively than siRNA or small molecules because these traditional approaches can decrease the fraction, of collagen but these reductions are temporary and after stopping treatment, the collagen will increase as those interventions are not at DNA level and so due to repeated administration potential toxicities and costs increase. Hence, there is an unmet need for additional treatments to manage fibrosis permanently and reduce the long-term complications in fibrotic diseases or possibly reverse fibrosis.

When it comes to the clinical setting it should be mention that base editing primarily overcome the double-strand breaks (DSBs) produced by conventional nucleases gene editing resulting in random combinations of insertions and deletions at the targeted site, alongside extensive rearrangements, loss of chromosomes, chromothripsis, and the triggering of the p53 DNA damage response (Haapaniemi et al., 2018; Song et al., 2020; Zuccaro et al., 2020). On the other hand, base editing does not require necessarily DNA delivery, which is a requirement for gene therapy or homology-directed repair. By contrast, base editing using RNA directly

converts CCAAT into a CCGGT allele with no requirement for exogenous DNA. ABE8 variants has the potential to increase editing efficiency if needed and phenotypic rescue, or might reduce the required dosage in comparison with other version of ABE such as ABE7 (Jeong et al., 2021).

6.6. Deciphering the Molecular Landscape of Edited Fibroblasts: Insights from Multi-Omics Analysis

Another crucial element of our research involves employing multi-omics technologies (transcriptomics and proteomics) to characterize the molecular signature of post-editing fibroblasts. These high-throughput omics technologies have substantially revolutionized medical research. Therefore, when it comes to collagen I modification- which plays a role in several signalling pathways in the cell- a significant concern belongs to the potential impact on related signalling pathways involving collagen, and may inadvertently induce side effects in cell phenotype and directing cells into novel pathological conditions. To address this, we employed high-throughput omics technologies, including transcriptomics (RNA-seq) and proteomics (tandem mass spectrometry), to investigate the molecular signature of edited cells and execute a signature comparative analysis with wild-type cells.

This analysis focusing on the key pathways such as the Integrin 69, Notch70, and BMP71 signalling pathways - which collagen involved in- revealed no significant abnormal results when compared with wild type or non-edited cells. This finding shows the safety and stability of COL1A1 reduction in these key cellular pathways, affirming no unintended alterations or disruptions in relavent cellular signalling. In other words, these pathways are crucial for collagen-related functions and play important roles in the development of various diseases. However, since we observed no changes in these pathways, it suggests that reducing collagen at this level is safe and does not significantly affect pathological processes. This strategic of omics approach significantly enhances the ability to uncover detailed biological processes, promoting a more focused and effective drug development landscape. Actually, it accelerates the transition from laboratory research to patient care, infusing the path with increased understanding and innovative approaches.

It should be mentioned that with the advent of omics technologies, biology has become increasingly dependent on data generated at these levels, which together is called as “multi-omics” data (Subramanian et al., 2020;Zheng et al., 2023). Availability of multi-omics data has

revolutionized the field of medicine and biology by creating avenues for integrated system-level approaches to generate multiple datasets from a single sample and enhancing hypothesis creation. These approaches aid in uncovering a spectrum of biological, molecular functions and mechanisms, alongside revealing various associations and correlations (Santiago-Rodriguez & Hollister, 2021;Walejko et al., 2018). Despite their remarkable capabilities and potential, numerous challenges must be meticulously assessed for the prosperous design and accomplishment of a multi-omics study such a sample collection, storage, and data analysis (Y.-Y. Wang et al., 2022;Wörheide et al., 2021).

6.7. Elucidating the Impact of the Phosphoinositide 3-Kinase (PI3K)/Akt Signaling Pathway and inflammatory responses in Fibrotic Diseases: Insights from Base Editing Strategies

Our findings, which emphasize the repressing of the PI3K/Akt pathway (as a key fibrotic signalling pathway), align with other studies demonstrating that targeting the CCAAT promoter of Col1a1 gene exerts an antifibrotic effect. This suggests a synergistic potential in therapeutic approaches focusing on these molecular targets to combat fibrosis.

Extensive investigations have shown that phosphoinositol-3 kinase (PI3K)/Akt signaling pathway has a critical role in regulating the occurrence, progression and pathological formation of cardiac fibrosis through regulating cell survival, apoptosis, growth, cardiac contractility and the transcription of related genes through a series of factors such as mammalian target of rapamycin (mTOR), glycogen synthase kinase 3 (GSK-3), forkhead box proteins O1/3 (FoxO1/3), and nitric oxide synthase (NOS) (Dou et al., 2019). Moreover, it is well demonstrated that PI3K and AKT have important implications in the process of cell proliferation, and collagen synthesis, promoting the progression of hepatic fibrosis (Peng et al., 2017).

Furthermore, the contribution of the PI3K/AKT in lung fibrotic processes is increasingly significant, and several PI3K/AKT inhibitors currently under clinical evaluation in idiopathic pulmonary fibrosis (IPF) (J. Wang et al., 2022). Additionally, PI3K/Akt/mTOR plays a crucial role during the process of kidney fibrosis. For example, Kim et al. reported that autophagy is activated by Akt/mTOR signaling pathway which plays a protective role in renal tubular injury and renal fibrosis (KIM et al., 2012).

Inflammation has developed into a highly intricate process over millions of years, demonstrating the host's necessity to effectively handle numerous harmful agents while keeping proper checks and balances. While an uncontrolled inflammatory response is uncommon and can result in acute or chronic diseases, inflammation is usually a crucial and beneficial reaction for survival. It comprises a wide network of cellular interactions supported by an extensive range of molecules (Flavell et al., 2008).

One of the significant signaling pathways that was notably affected in edited cells compared to unedited cells involves inflammatory responses such as the NOD-like receptor pathway and pathways related to inflammatory responses to viral infections like COVID-19. The activation of these pathways in edited cells indicates that after the reduction of collagen, the cells attempt to respond to this new situation, primarily through an inflammatory response. In other words, cells perceive this collagen reduction, especially in the extracellular matrix, as a type of damage and respond by activating inflammatory pathways.

Stromal cells, including fibroblasts, hold considerable promise as therapeutic targets because they play a crucial role in organizing and sustaining inflammatory infiltrates (Jordana et al., 1994). So, in the future, potential clinical applications of our approaches should take inflammatory responses into consideration.

6.8. Tailoring Delivery Modalities: The Promise of Liposomal Platform

Another crucial aspect of our research project was the use of a liposomal-based delivery method to target gene of interest. This technique with its unique benefits and limitations, demonstrated effectiveness in our gene editing experiments (Hejabi et al., 2022). So, as we move toward in vivo studies it is essential to consider alternative strategies such as functionalized lipid-like nanoparticles for in vivo RNA delivery of base editing to specifically target fibroblasts in different organs (Yin et al., 2016; X. Zhang et al., 2020). While our research has predominantly concentrated on fibroblasts recognizing the potential to target other cell types is essential. In other words, the continued progress in our understanding of delivery technology will play a crucial role in investigating different delivery methods in different targeted cells which can contribute significantly in the effectiveness of possible therapeutic interventions.

6.9. Engineering Promoter Locus: A Gateway to Deciphering Transcriptional Dynamics and in Vitro Models

In our novel in vitro model, a deliberate mutation was generated inside the promoter of Col1a1 within fibroblast genome. This intentional alteration paves the way for further functional investigations by other researchers. For instance, studies on how changing adenine to guanine can influence gene regulation mechanisms, such as the attachment of the transcription factors to the CCAAT box, can be pursued. These explorations could potentially unlock new avenues in transcription factor (TF) and gene regulation studies which may contribute significantly to our understanding of cellular and molecular processes and offering insights for innovative research strategies.

Further application of our research could be impactful in functional studies. For example, in clinical genetics or paediatrics contexts, the discovery of new mutations in patients consistently poses challenges for clinicians due to the unpredictability of their effects on patient phenotypes (Kurian et al., 2017; Nicolosi et al., 2022). While in silico and computational approaches provide some predictive insight, the necessity for wet lab functional studies remains a substantial hurdle in the way of patient management. Our methodology transcends this barrier, not only presenting a new model of transgenic fibroblasts carrying novel mutations in CCAAT box but also offering a versatile protocol for generating in vitro models by base editing. In other words, this strategy can be applied to generate new in vitro models targeting the CCAAT box in various cell types, thereby broadening the spectrum of investigational opportunities.

6.10. Transgenic Fibroblasts: Insights into Cellular Morphology and Adaptability

In this research, we are introducing not only a novel in vitro model but also a novel ECM scaffold that are produced by transgenic fibroblast that holds potential for various applications.

Analysis by SEM of seeded cells on extracted ECM from transgenic cells uncovered normal cell morphologies consistently distributed throughout the culture (Dalby et al., 2003). For instance, the prevalence of a spindle-shaped body, commonly observed in normal fibroblasts, reinforces the acknowledged heterogeneity of these cells, as observed in our SEM trials (Murata et al., 2019). Specifically, observations revealed three unique cell phenotypes, aligning

with previous reports of morphological variation in fibroblasts. This highlights the presence of distinct cellular subsets, indicative of the extensive heterogeneity and adaptability of fibroblasts to the new ECM. This brief communication shed light on thesis hypothesis that possibly reducing collagen I in vivo will not affect the morphology of fibroblasts (Murata et al., 2019).

6.11. Moving beyond CRISPRi Limitations through Promoter-Targeting Base Editing

By shifting focus to another aspect of our finding, it should be mentioned that manipulating gene expression with precision is pivotal for both deciphering gene functionalities and designing genetic regulatory framework studies. CRISPR interference (CRISPRi), which facilitates the targeted suppression of transcription in different systems such as bacterial and human cellular systems, has been introduced for these goals (Larson et al., 2013). Briefly, when the sgRNA is directed towards the promoter region, it can physically obstruct the interaction between essential cis-acting DNA elements and their corresponding trans-acting transcription factors, resulting in the inhibition of transcription initiation (S. J. Liu et al., 2017).

CRISPRi offers precision in gene regulation but faces drawbacks. It can sometimes cause many off-targets inside the genome and may not effectively reduced the expression genes of interest as well as its gene silencing is entirely reversible (Larson et al., 2013). Additionally, introducing CRISPRi components into cells poses complexities, especially concerning efficient delivery and expression (R. Zhang et al., 2021). The expression investigations by CRISPRi for lasting effects might not always be ideal beacuase does not induce permanent genetic changes (Ghavami & Pandi, 2021; Morelli et al., 2021). Overexpressing its components might stress cells, and there are potential immune response concerns in long-term therapeutic applications.

Our innovative protocol to target promoters by base editing has been meticulously designed to address the aforementioned limitations of CRISPRi. By targeting specific promoter regions, our method may ensure comprehensive modulating of target genes. The strategy is useful for both bacterial and mammalian systems so can eliminate the obstacles often seen in CRISPRi methods. In essence, our tailored promoter-targeting strategy stands as a superior alternative, effectively mitigating the main challenges associated with CRISPRi.

In conclusion, we provide evidence in fibroblasts for the application of ABE to target the promoter of Col1A1 to introduce a novel antifibrotic landscape based of gene editing. We demonstrate therapeutically beneficial editing using transient and non-viral delivery vectors.

The remaining risks of such clinically viable base editor therapies must be carefully weighed against their benefits. This makes fibrotic organs and cancers, which at present can be cured only by organ transplantation and suboptimal approaches, ideal candidates for further investigations.

Our study into the area of gene-editing technologies, particularly focusing on ABE and its impact on the Col1A1 basal promoter is offering an interesting universal therapeutic landscape for universal disorders. By considering the potential of our novel in vitro model and the derived ECM scaffold produced by transgenic fibroblasts, we have delineated an innovative framework for developing novel antifibrotic treatment. Furthermore, our versatile approach of targeting promoters through base editing effectively mitigates the limitations associated with the CRISPRi technique, addressing the modulation of target genes and suggesting the applications adaptability across various cellular systems.

6.12. Innovative Cancer Research: Reducing Collagen I to Inhibit Tumor Growth

In the next part of our project, we utilized interdisciplinary ideas to develop a novel approach to cancer investigation. Specifically, we examined the intricate relationship between collagen I reduction and the subsequent oncogenic processes. Actually, a hallmark of many cancers is characterized by the excessive deposition of extracellular matrix (ECM) components, particularly collagens (Pickup et al., 2014; Winslow et al., 2016) In certain cancers, such as breast cancer, the desmoplastic stroma can comprise up to 90% of the tumor mass, which is correlated with poor prognosis. The ECM plays a pivotal role in modulating the hallmarks of cancer; for instance, an increased density of collagen I is known to enhance mammary tumor initiation, growth, and invasion. Consequently, targeting ECM synthesis in stromal fibroblasts presents itself as a promising strategy to mitigate cancer progression.

Our research focused on the impact of a novel microenvironment with reduced collagen I on breast cancer cells. We found that exposing these cells to such an environment led to the downregulation of focal adhesion (FA) related pathways, which in turn resulted in a reduced proliferation rate of the cancer cells. Notably, focal adhesions are critical regulators of the transcriptional response in both cancer and fibrosis conditions (Lagares & Kapoor, 2013; Tilghman & Parsons, 2008). This suggests that the reduction in collagen I density could significantly influence the signaling responses observed in cancer cells, thereby highlighting

the potential of ECM modulation as a therapeutic avenue in cancer treatment which is compatible with previous reports (Murphy et al., 2020; Venning et al., 2015).

6.13. Limitations and Challenges

Ttargeting fibroblasts, the main collagen producers, within a living organism remains a challenge. Functionalized nanoparticles offer a promising approach for smarter delivery. However, collagen is crucial for tissue stability, and reducing it excessively could disrupt this balance. A potential solution is editing a small number of fibroblasts, creating a mosaic of edited and unedited cells to maintain tissue integrity.

Furthermore, more research is needed to assess the long-term effects and safety of permanently editing the Col1a1 promoter region. Unforeseen genetic instability or secondary health problems could arise.

Additionally, high-throughput multi-omics technologies, while powerful, generate complex data sets that can be challenging to interpret accurately. Ensuring consistency across studies and conditions is crucial but difficult.

Finally, the application of CRISPR-Cas9 and other gene editing technologies in humans faces strict ethical and regulatory scrutiny. Navigating these regulations and gaining approval for clinical use can be time-consuming and complex. Developing and delivering gene editing therapies on a large scale also presents logistical challenges. Ensuring that these therapies are accessible to a wide range of patients, particularly those in low-resource settings, remains a significant hurdle.

6.14. Future perspective

While our research has provided evidence and insights at the cellular and molecular level within the boundaries of an in vitro environment, the real-world efficacy of these findings will be determined by their performance in comprehensive in vivo trials. Diseases that currently suffer from suboptimal treatment modalities or those, which require drastic interventions like organ fibrosis, stand to benefit significantly if our strategies prove efficacious in the complex milieu of in vivo. The specific modulation of collagen within these derived ECMs could enable the study of various Cell-ECM interactions under various pathological conditions, including fibrosis and tumor pathogenesis. Furthermore, the application of the ECM in vitro setting open

up avenues for an optimal disease modelling, especially in fields like tissue engineering and regenerative medicine. This could lead to the development of improved in vitro disease models, offering a more physiologically relevant context for drug screening and therapeutic interventions.

Targeting the CCAAT promoter in the Col1a1 gene within the context of the PI3K)/Akt signaling pathway has shown promising results for the future application of antifibrotic medications. This approach could potentially be effective in treating fibrosis across various organs, indicating a broad-spectrum therapeutic potential.

One of the most challenges as we explore more deeply is the delivery approaches. Our current work used the liposomal-based delivery systems. However, the need for employing more efficient, targeted, and biocompatible delivery systems is arising. These systems should aim to ensure the precision of delivery while minimizing side effects, an aspect even more critical when considering in vivo applications. Lastly, the relationship between reduced collagen dynamics and the tumor microenvironment warrants deeper investigation. Given collagen pivotal role in the tumor microenvironment, a comprehensive understanding of how tumor cells interact, behave, and proliferate within a collagen-modulated matrix stands as a research priority. If addressed correctly, this could bridge the fields of fibrosis research and oncology and offering innovative therapeutic strategies.

To make a long story short, while the horizon seems promising, rigorous scientific inquiry, technological adaptation, and a commitment to translational research will be vital in ensuring that our findings translate to substantial clinical benefits.

7. Table of Figures

8. References

Abou Zahr, A., Salama, M. E., Carreau, N., Tremblay, D., Verstovsek, S., Mesa, R., Hoffman, R., & Mascarenhas, J. (2016). Bone marrow fibrosis in myelofibrosis: pathogenesis, prognosis and targeted strategies. *Haematologica*, *101*(6), 660.

Alwin, S., Gere, M. B., Guhl, E., Effertz, K., Barbas, C. F., Segal, D. J., Weitzman, M. D., & Cathomen, T. (2005). Custom zinc-finger nucleases for use in human cells. *Molecular Therapy*, *12*(4), 610–617.

Anzalone, A. V, Koblan, L. W., & Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. In *Nature biotechnology* (Vol. 38, Issue 7, pp. 824–844). Nature Publishing Group US New York.

Anzalone, A. V, Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., & Raguram, A. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 1.

Aoki-Kinoshita, K. F., & Kanehisa, M. (2007). Gene annotation and pathway mapping in KEGG. *Comparative Genomics*, 71–91.

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., & Creech, C. B. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England Journal of Medicine*, *384*(5), 403–416.

Bai, B., Liu, Y., Fu, X.-M., Qin, H.-Y., Li, G.-K., Wang, H.-C., & Sun, S.-L. (2022). Dysregulation of EZH2/miR-138-5p Axis contributes to radiosensitivity in hepatocellular carcinoma cell by downregulating hypoxia-inducible factor 1 alpha (HIF-1α). *Oxidative Medicine and Cellular Longevity*, *2022*.

Baldari, S., Di Modugno, F., Nisticò, P., & Toietta, G. (2022). Strategies for efficient targeting of tumor collagen for cancer therapy. *Cancers*, *14*(19), 4706.

Becker, S., & Boch, J. (2021). TALE and TALEN genome editing technologies. *Gene and Genome Editing*, *2*, 100007.

Bhogal, R. K., Stoica, C. M., McGaha, T. L., & Bona, C. A. (2005). Molecular aspects of regulation of collagen gene expression in fibrosis. *Journal of Clinical Immunology*, *25*, 592–603.

Birk, D. E., & Brückner, P. (2010). Collagens, suprastructures, and collagen fibril assembly. In *The extracellular matrix: an overview* (pp. 77–115). Springer.

Brady, C. A., & Attardi, L. D. (2010). p53 at a glance. *Journal of Cell Science*, *123*(15), 2527–2532.

Capecchi, M. R. (2022). The origin and evolution of gene targeting. *Developmental Biology*, *481*, 179–187.

Carbon, S. ea, Dietze, H., Lewis, S. E., Mungall, C. J., Munoz-Torres, M. C., Basu, S., Chisholm, R. L., Dodson, R. J., Fey, P., & Thomas, P. D. (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Research*, *45*(D1).

Chalayer, É., Ffrench, M., & Cathébras, P. (2014). Bone marrow fibrosis as a feature of systemic lupus erythematosus: a case report and literature review. *Springerplus*, *3*, 1–10.

Chandrasegaran, S., & Carroll, D. (2016). Origins of programmable nucleases for genome engineering. *Journal of Molecular Biology*, *428*(5), 963–989.

Chandrashekar, D. S., Karthikeyan, S. K., Korla, P. K., Patel, H., Shovon, A. R., Athar, M., Netto, G.

J., Qin, Z. S., Kumar, S., & Manne, U. (2022). UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia*, *25*, 18–27.

Chen, C. Z. C., & Raghunath, M. (2009). Focus on collagen: in vitro systems to study fibrogenesis and antifibrosis _ state of the art. *Fibrogenesis & Tissue Repair*, *2*, 1–10.

Chen, L., Chu, C., Lu, J., Kong, X., Huang, T., & Cai, Y.-D. (2015). Gene ontology and KEGG pathway enrichment analysis of a drug target-based classification system. *PloS One*, *10*(5), e0126492.

Chen, P. J., & Liu, D. R. (2023). Prime editing for precise and highly versatile genome manipulation. *Nature Reviews Genetics*, *24*(3), 161–177.

Chiesa, R., Georgiadis, C., Syed, F., Zhan, H., Etuk, A., Gkazi, S. A., Preece, R., Ottaviano, G., Braybrook, T., & Chu, J. (2023). Base-Edited CAR7 T Cells for Relapsed T-Cell Acute Lymphoblastic Leukemia. *New England Journal of Medicine*.

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J.-S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research*, *24*(1), 132–141.

Chu, S. H., Packer, M., Rees, H., Lam, D., Yu, Y., Marshall, J., Cheng, L.-I., Lam, D., Olins, J., & Ran, F. A. (2021). Rationally designed base editors for precise editing of the sickle cell disease mutation. *The CRISPR Journal*, *4*(2), 169–177.

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., & Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, *33*(5), 543–548.

Chung, H. J., Steplewski, A., Chung, K. Y., Uitto, J., & Fertala, A. (2008). Collagen fibril formation: a new target to limit fibrosis. *Journal of Biological Chemistry*, *283*(38), 25879–25886.

Chung, J. Y.-F., Chan, M. K.-K., Li, J. S.-F., Chan, A. S.-W., Tang, P. C.-T., Leung, K.-T., To, K.-F., Lan, H.-Y., & Tang, P. M.-K. (2021). TGF-β signaling: from tissue fibrosis to tumor microenvironment. *International Journal of Molecular Sciences*, *22*(14), 7575.

Chylinski, K., Le Rhun, A., & Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biology*, *10*(5), 726–737.

Cobb, R. E., Wang, Y., & Zhao, H. (2015). High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synthetic Biology*, *4*(6), 723–728.

Collins, B. F., & Raghu, G. (2019). Antifibrotic therapy for fibrotic lung disease beyond idiopathic pulmonary fibrosis. *European Respiratory Review*, *28*(153).

Consortium, G. O. (2019). The KEGG Pathway analysis, which comprised curated pathway maps depicting our understanding of molecular interaction, reaction, and relation networks across various categories including metabolism, genetic information processing, environmental information. *Nucleic Acids Research*, *47*(D1), D330–D338.

Corallo, C., Cutolo, M., Volpi, N., Franci, D., Aglianò, M., Montella, A., Chirico, C., Gonnelli, S., Nuti, R., & Giordano, N. (2017). Histopathological findings in systemic sclerosis-related myopathy: fibrosis and microangiopathy with lack of cellular inflammation. *Therapeutic Advances in Musculoskeletal Disease*, *9*(1), 3–10.

Cox, T. R., & Erler, J. T. (2014). Molecular pathways: connecting fibrosis and solid tumor metastasis. *Clinical Cancer Research*, *20*(14), 3637–3643.

Cradick, T. J., Ambrosini, G., Iseli, C., Bucher, P., & McCaffrey, A. P. (2011). ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites. *BMC Bioinformatics*, *12*(1), 1–10.

Crotti, S., Piccoli, M., Rizzolio, F., Giordano, A., Nitti, D., & Agostini, M. (2017). Extracellular matrix and colorectal cancer: how surrounding microenvironment affects cancer cell behavior? *Journal of Cellular Physiology*, *232*(5), 967–975.

Dalby, M. J., Childs, S., Riehle, M. O., Johnstone, H. J. H., Affrossman, S., & Curtis, A. S. G. (2003). Fibroblast reaction to island topography: changes in cytoskeleton and morphology with time. *Biomaterials*, *24*(6), 927–935.

Daliri, K., Hescheler, J., & Pfannkuche, K. P. (2024). Prime Editing and DNA Repair System: Balancing Efficiency with Safety. *Cells*, *13*(10), 858.

Dang, C. V. (2012). MYC on the path to cancer. *Cell*, *149*(1), 22–35.

De Las Rivas, J., & Fontanillo, C. (2010). Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Computational Biology*, *6*(6), e1000807.

Deshmukh, S. N., Dive, A. M., Moharil, R., & Munde, P. (2016). Enigmatic insight into collagen. *Journal of Oral and Maxillofacial Pathology: JOMFP*, *20*(2), 276.

Devkota, S. (2018). The road less traveled: strategies to enhance the frequency of homology-directed repair (HDR) for increased efficiency of CRISPR/Cas-mediated transgenesis. *BMB Reports*, *51*(9), 437.

Dolfini, D., Zambelli, F., Pavesi, G., & Mantovani, R. (2009). A perspective of promoter architecture from the CCAAT box. *Cell Cycle*, *8*(24), 4127–4137.

Dörrie, J., Schaft, N., Schuler, G., & Schuler-Thurner, B. (2020). Therapeutic cancer vaccination with ex vivo RNA-transfected dendritic cells—an update. *Pharmaceutics*, *12*(2), 92.

Dou, F., Liu, Y., Liu, L., Wang, J., Sun, T., Mu, F., Guo, Q., Guo, C., Jia, N., & Liu, W. (2019). Aloe-emodin ameliorates renal fibrosis via inhibiting PI3K/Akt/mTOR signaling pathway in vivo and in vitro. *Rejuvenation Research*, *22*(3), 218–229.

Duarte, B. D. P., & Bonatto, D. (2018). The heat shock protein 47 as a potential biomarker and a therapeutic agent in cancer research. *Journal of Cancer Research and Clinical Oncology*, *144*, 2319–2328.

Dwiranti, A., Masri, F., Rahmayenti, D. A., & Putrika, A. (2019). The effects of osmium tetroxide post-fixation and drying steps on leafy liverwort ultrastructure study by scanning electron microscopy. *Microscopy Research and Technique*, *82*(7), 1041–1046.

Eisinger-Mathason, T. S. K., Zhang, M., Qiu, Q., Skuli, N., Nakazawa, M. S., Karakasheva, T., Mucaj, V., Shay, J. E. S., Stangenberg, L., & Sadri, N. (2013). Hypoxia-dependent modification of collagen networks promotes sarcoma metastasis. *Cancer Discovery*, *3*(10), 1190–1205.

Engel, J., & Bächinger, H. P. (2005). Structure, stability and folding of the collagen triple helix. *Collagen: Primer in Structure, Processing and Assembly*, 7–33.

Exposito, J.-Y., Valcourt, U., Cluzel, C., & Lethias, C. (2010). The fibrillar collagen family. *International Journal of Molecular Sciences*, *11*(2), 407–426.

Fan, D., Takawale, A., Lee, J., & Kassiri, Z. (2012). Cardiac fibroblasts, fibrosis and extracellular matrix remodeling in heart disease. *Fibrogenesis & Tissue Repair*, *5*(1), 15.

Fane, M., & Weeraratna, A. T. (2020). How the ageing microenvironment influences tumour progression. *Nature Reviews Cancer*, *20*(2), 89–106.

Fang, L., Murphy, A. J., & Dart, A. M. (2017). A clinical perspective of anti-fibrotic therapies for cardiovascular disease. *Frontiers in Pharmacology*, *8*, 186.

Filippova, J., Matveeva, A., Zhuravlev, E., & Stepanov, G. (2019). Guide RNA modification as a way

to improve CRISPR/Cas9-based genome-editing systems. *Biochimie*, *167*, 49–60.

Flavell, S. J., Hou, T. Z., Lax, S., Filer, A. D., Salmon, M., & Buckley, C. D. (2008). Fibroblasts as novel therapeutic targets in chronic inflammation. *British Journal of Pharmacology*, *153*(S1), S241–S246.

Franco-Barraza, J., Beacham, D. A., Amatangelo, M. D., & Cukierman, E. (2016). Preparation of extracellular matrices produced by cultured and primary fibroblasts. *Current Protocols in Cell Biology*, *71*(1), 10–19.

Frangogiannis, N. G. (2021). Cardiac fibrosis. *Cardiovascular Research*, *117*(6), 1450–1488.

Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., Foell, J., de la Fuente, J., Grupp, S., & Handgretinger, R. (2021). CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. *New England Journal of Medicine*, *384*(3), 252–260.

Gabrielli, A., Avvedimento, E. V, & Krieg, T. (2009). Scleroderma. *New England Journal of Medicine*, *360*(19), 1989–2003.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A• T to G• C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471.

Gaudelli, N. M., Lam, D. K., Rees, H. A., Solá-Esteves, N. M., Barrera, L. A., Born, D. A., Edwards, A., Gehrke, J. M., Lee, S.-J., & Liquori, A. J. (2020). Directed evolution of adenine base editors with increased activity and therapeutic application. *Nature Biotechnology*, *38*(7), 892–900.

Gay, S., & Miller, E. J. (1983). Overview: What is collagen, what is not. *Ultrastructural Pathology*, *4*(4), 365–377.

Geng, Y., Deng, Z., & Sun, Y. (2016). An insight into the protospacer adjacent motif of Streptococcus pyogenes Cas9 with artificially stimulated RNA-guided-Cas9 DNA cleavage flexibility. *RSC Advances*, *6*(40), 33514–33522.

Ghavami, S., & Pandi, A. (2021). CRISPR interference and its applications. *Progress in Molecular Biology and Translational Science*, *180*, 123–140.

Ghazipura, M., Mammen, M. J., Bissell, B. D., Macrea, M., Herman, D. D., Hon, S. M., Kheir, F., Khor, Y. H., Knight, S. L., & Raghu, G. (2022). Pirfenidone in progressive pulmonary fibrosis: a systematic review and meta-analysis. *Annals of the American Thoracic Society*, *19*(6), 1030–1039.

Gistelinck, C., Gioia, R., Gagliardi, A., Tonelli, F., Marchese, L., Bianchi, L., Landi, C., Bini, L., Huysseune, A., & Witten, P. E. (2016). Zebrafish collagen type I: molecular and biochemical characterization of the major structural protein in bone and skin. *Scientific Reports*, *6*(1), 21540.

Goldberg, M., & Smith, A. J. (2004). CELLS AND EXTRACELLULAR MATRICES OF DENTIN AND PULP: A BIOLOGICAL BASIS FOR REPAIR AND TISSUE ENGINEERING. *Critical Reviews in Oral Biology & Medicine*, *15*(1), 13–27. https://doi.org/10.1177/154411130401500103

Gordon, I. O., Agrawal, N., Goldblum, J. R., Fiocchi, C., & Rieder, F. (2014). Fibrosis in ulcerative colitis: mechanisms, features, and consequences of a neglected problem. *Inflammatory Bowel Diseases*, *20*(11), 2198–2206.

Gordon, M. K., & Hahn, R. A. (2010). Collagens. *Cell and Tissue Research*, *339*(1), 247–257.

Green, M. R., & Sambrook, J. (2019). Isolation of poly (A)+ messenger RNA using magnetic oligo (dT) beads. *Cold Spring Harbor Protocols*, *2019*(10), pdb-prot101733.

Gurumurthy, C. B., Quadros, R. M., & Ohtsuka, M. (2022). Prototype mouse models for researching SEND-based mRNA delivery and gene therapy. *Nature Protocols*, *17*(10), 2129–2138.

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., & Taipale, J. (2018). CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*, *24*(7), 927.

Hagiwara, S., Iwasaka, H., Matsumoto, S., & Noguchi, T. (2007). Antisense oligonucleotide inhibition of heat shock protein (HSP) 47 improves bleomycin-induced pulmonary fibrosis in rats. *Respiratory Research*, *8*(1), 1–11.

Han, B., Li, W., Sun, Y., Zhou, L., Xu, Y., & Zhao, X. (2014). A prolyl-hydroxylase inhibitor, ethyl-3, 4-dihydroxybenzoate, induces cell autophagy and apoptosis in esophageal squamous cell carcinoma cells via up-regulation of BNIP3 and N-myc downstream-regulated gene-1. *PLoS One*, *9*(9), e107204.

Hao, M., Han, X., Yao, Z., Zhang, H., Zhao, M., Peng, M., Wang, K., Shan, Q., Sang, X., & Wu, X. (2022). The pathogenesis of organ fibrosis: Focus on necroptosis. *British Journal of Pharmacology*.

Hashimoto, M., & Takemoto, T. (2015). Electroporation enables the efficient mRNA delivery into the mouse zygotes and facilitates CRISPR/Cas9-based genome editing. *Scientific Reports*, *5*(1), 11315.

Hedrick, V. E., LaLand, M. N., Nakayasu, E. S., & Paul, L. N. (2015). Digestion, purification, and enrichment of protein samples for mass spectrometry. *Current Protocols in Chemical Biology*, *7*(3), 201–222.

Hejabi, F., Abbaszadeh, M. S., Taji, S., O'Neill, A., Farjadian, F., & Doroudian, M. (2022). Nanocarriers: A novel strategy for the delivery of CRISPR/Cas systems. *Frontiers in Chemistry*, *10*, 957572.

Helm, R. F., & Potts, M. (2012). Extracellular matrix (ECM). In *Ecology of cyanobacteria II: Their diversity in space and time* (pp. 461–480). Springer.

Henderson, N. C., Rieder, F., & Wynn, T. A. (2020). Fibrosis: from mechanisms to medicines. *Nature*, *587*(7835), 555–566.

Hotta, A., & Ellis, J. (2008). Retroviral vector silencing during iPS cell induction: an epigenetic beacon that signals distinct pluripotent states. *Journal of Cellular Biochemistry*, *105*(4), 940–948.

Hu, Q., Gao, L., Peng, B., & Liu, X. (2017). Baicalin and baicalein attenuate renal fibrosis in vitro via inhibition of the TGF-β1 signaling pathway. *Experimental and Therapeutic Medicine*, *14*(4), 3074–3080.

Huang, T. P., Newby, G. A., & Liu, D. R. (2021). Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nature Protocols*, *16*(2), 1089–1128.

Ishikawa, Y., & Bächinger, H. P. (2013). A molecular ensemble in the rER for procollagen maturation. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1833*(11), 2479–2491.

Ishino, Y., Krupovic, M., & Forterre, P. (2018). History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *Journal of Bacteriology*, *200*(7), 10–1128.

Ishiwatari, H., Sato, Y., Murase, K., Yoneda, A., Fujita, R., Nishita, H., Birukawa, N. K., Hayashi, T., Sato, T., & Miyanishi, K. (2013). Treatment of pancreatic fibrosis with siRNA against a collagen-specific chaperone in vitamin A-coupled liposomes. *Gut*, *62*(9), 1328–1339.

Ito, S., Ogawa, K., Takeuchi, K., Takagi, M., Yoshida, M., Hirokawa, T., Hirayama, S., Shin-Ya, K., Shimada, I., & Doi, T. (2017). A small-molecule compound inhibits a collagen-specific molecular chaperone and could represent a potential remedy for fibrosis. *Journal of Biological Chemistry*, *292*(49), 20076–20085.

Jang, H., Jo, D. H., Cho, C. S., Shin, J. H., Seo, J. H., Yu, G., Gopalappa, R., Kim, D., Cho, S.-R., &

Kim, J. H. (2022). Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. *Nature Biomedical Engineering*, *6*(2), 181–194.

Jeong, Y. K., Lee, S., Hwang, G.-H., Hong, S.-A., Park, S., Kim, J.-S., Woo, J.-S., & Bae, S. (2021). Adenine base editor engineering reduces editing of bystander cytosines. *Nature Biotechnology*, *39*(11), 1426–1433.

Jeong, Y. K., Song, B., & Bae, S. (2020). Current status and challenges of DNA base editing tools. *Molecular Therapy*, *28*(9), 1938–1952.

Jiang, T., Henderson, J. M., Coote, K., Cheng, Y., Valley, H. C., Zhang, X.-O., Wang, Q., Rhym, L. H., Cao, Y., & Newby, G. A. (2020). Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. *Nature Communications*, *11*(1), 1979.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821.

Jo, Y., Hwang, S. H., & Jang, J. (2021). Employing extracellular matrix-based tissue engineering strategies for age-dependent tissue degenerations. *International Journal of Molecular Sciences*, *22*(17), 9367.

Jordana, M., Sarnstrand, B., Sime, P. J., & Ramis, I. (1994). Immune-inflammatory functions of fibroblasts. *European Respiratory Journal*, *7*(12), 2212–2222.

Kan, Y., Ruis, B., Lin, S., & Hendrickson, E. A. (2014). The mechanism of gene targeting in human somatic cells. *PLoS Genetics*, *10*(4), e1004251.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361.

Kang, D. J., Lee, S. J., Na, J. E., Seong, M., Yoon, S. Y., Jeong, Y. W., Ahn, J. P., & Rhyu, I. J. (2019). Atmospheric scanning electron microscopy and its applications for biological specimens. *Microscopy Research and Technique*, *82*(1), 53–60.

Karsdal, M. A., Daniels, S. J., Holm Nielsen, S., Bager, C., Rasmussen, D. G. K., Loomba, R., Surabattula, R., Villesen, I. F., Luo, Y., & Shevell, D. (2020). Collagen biology and non-invasive biomarkers of liver fibrosis. *Liver International*, *40*(4), 736–750.

Kauppila, S., Stenbäck, F., Risteli, J., Jukkola, A., & Risteli, L. (1998). Aberrant type I and type III collagen gene expression in human breast cancer in vivo. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, *186*(3), 262–268.

Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B. W., Beal, J., & Vora, S. (2015). Cas9 gRNA engineering for genome editing, activation and repression. *Nature Methods*, *12*(11), 1051–1054.

Kim, J. G., Kim, E. O., Jeong, B. R., Min, Y. J., Park, J. W., Kim, E. S., Namgoong, I. S., Kim, Y. Il, & Lee, B. J. (2010). Visfatin stimulates proliferation of MCF-7 human breast cancer cells. *Molecules and Cells*, *30*, 341–345.

KIM, W., Nam, S. A., Song, H. C., Ko, J. S., Park, S. H., Kim, H. L., Choi, E. J., KIM, Y., Kim, J., & Kim, Y. K. (2012). The role of autophagy in unilateral ureteral obstruction rat model. *Nephrology*, *17*(2), 148–159.

Kim, Y. B., Komor, A. C., Levy, J. M., Packer, M. S., Zhao, K. T., & Liu, D. R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology*, *35*(4), 371–376.

King, T. E., Pardo, A., & Selman, M. (2011). Idiopathic pulmonary fibrosis. *The Lancet*, *378*(9807), 1949–1961.

Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., & Moriarity, B. S. (2018). EditR: a method to quantify base editing from Sanger sequencing. *The CRISPR Journal*, *1*(3), 239–250.

Knight, S. C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L. B., Bosanac, L., Zhang, E. T., El Beheiry, M., Masson, J.-B., & Dahan, M. (2015). Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*, *350*(6262), 823–826.

Knüppel, L., Ishikawa, Y., Aichler, M., Heinzelmann, K., Hatz, R., Behr, J., Walch, A., Bächinger, H. P., Eickelberg, O., & Staab-Weijnitz, C. A. (2017). A novel antifibrotic mechanism of nintedanib and pirfenidone. Inhibition of collagen fibril assembly. *American Journal of Respiratory Cell and Molecular Biology*, *57*(1), 77–90.

Koblan, L. W., Doman, J. L., Wilson, C., Levy, J. M., Tay, T., Newby, G. A., Maianti, J. P., Raguram, A., & Liu, D. R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Biotechnology*, *36*(9), 843–846.

Koblan, L. W., Erdos, M. R., Wilson, C., Cabral, W. A., Levy, J. M., Xiong, Z.-M., Tavarez, U. L., Davison, L. M., Gete, Y. G., & Mao, X. (2021). In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. *Nature*, *589*(7843), 608–614.

Koide, T., & Nagata, K. (2005). Collagen biosynthesis. *Collagen: Primer in Structure, Processing and Assembly*, 85–114.

Komor, A. C., Badran, A. H., & Liu, D. R. (2018). Editing the genome without double-stranded DNA breaks. *ACS Chemical Biology*, *13*(2), 383–388.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. In *Nature* (Vol. 533, Issue 7603, pp. 420–424). Nature Publishing Group.

Koo, T., Lee, J., & Kim, J.-S. (2015). Measuring and reducing off-target activities of programmable nucleases including CRISPR-Cas9. *Molecules and Cells*, *38*(6), 475.

Koyuncu, S., Loureiro, R., Lee, H. J., Wagle, P., Krueger, M., & Vilchez, D. (2021). Rewiring of the ubiquitinated proteome determines ageing in C. elegans. *Nature*, *596*(7871), 285–290.

Krenning, G., Zeisberg, E. M., & Kalluri, R. (2010). The origin of fibroblasts and mechanism of cardiac fibrosis. *Journal of Cellular Physiology*, *225*(3), 631–637.

Kurian, A. W., Li, Y., Hamilton, A. S., Ward, K. C., Hawley, S. T., Morrow, M., McLeod, M. C., Jagsi, R., & Katz, S. J. (2017). Gaps in incorporating germline genetic testing into treatment decision-making for early-stage breast cancer. *Journal of Clinical Oncology*, *35*(20), 2232.

Kurose, H. (2021). Cardiac fibrosis and fibroblasts. *Cells*, *10*(7), 1716.

Kurose, H., & Mangmool, S. (2016). Myofibroblasts and inflammatory cells as players of cardiac fibrosis. *Archives of Pharmacal Research*, *39*, 1100–1113.

Lagares, D., & Kapoor, M. (2013). Targeting focal adhesion kinase in fibrotic diseases. *BioDrugs*, *27*, 15–23.

Laporte, P., Lepage, A., Fournier, J., Catrice, O., Moreau, S., Jardinaud, M.-F., Mun, J.-H., Larrainzar, E., Cook, D. R., & Gamas, P. (2014). The CCAAT box-binding transcription factor NF-YA1 controls rhizobial infection. *Journal of Experimental Botany*, *65*(2), 481–494.

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols*, *8*(11), 2180–2196.

LeBleu, V. S., Taduri, G., O'connell, J., Teng, Y., Cooke, V. G., Woda, C., Sugimoto, H., & Kalluri,

R. (2013). Origin and function of myofibroblasts in kidney fibrosis. *Nature Medicine*, *19*(8), 1047–1053.

Leidescher, S., Ribisel, J., Ullrich, S., Feodorova, Y., Hildebrand, E., Galitsyna, A., Bultmann, S., Link, S., Thanisch, K., & Mulholland, C. (2022). Spatial organization of transcribed eukaryotic genes. *Nature Cell Biology*, *24*(3), 327–339.

Leitinger, B. (2011). Transmembrane collagen receptors. *Annual Review of Cell and Developmental Biology*, *27*, 265–290.

Lekstrom-Himes, J., & Xanthopoulos, K. G. (1998). Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *Journal of Biological Chemistry*, *273*(44), 28545–28548.

Leonhardt, C., Schwake, G., Stögbauer, T. R., Rappl, S., Kuhr, J.-T., Ligon, T. S., & Rädler, J. O. (2014). Single-cell mRNA transfection studies: delivery, kinetics and statistics by numbers. *Nanomedicine: Nanotechnology, Biology and Medicine*, *10*(4), 679–688.

Levine, M., & Levine, M. (2011). Collagen Synthesis, Genetic Diseases, and Scurvy. *Topics in Dental Biochemistry*, 101–112.

Lewis, G. A., Dodd, S., Clayton, D., Bedson, E., Eccleson, H., Schelbert, E. B., Naish, J. H., Jimenez, B. D., Williams, S. G., & Cunnington, C. (2021). Pirfenidone in heart failure with preserved ejection fraction: a randomized phase 2 trial. *Nature Medicine*, *27*(8), 1477–1482.

Li, Chao, & Kuemmerle, J. F. (2014). Mechanisms that mediate the development of fibrosis in patients with Crohn's disease. *Inflammatory Bowel Diseases*, *20*(7), 1250–1258.

Li, Cong, Qiu, S., Liu, X., Guo, F., Zhai, J., Li, Z., Deng, L., Ge, L., Qian, H., & Yang, L. (2023). Extracellular matrix-derived mechanical force governs breast cancer cell stemness and quiescence transition through integrin-DDR signaling. *Signal Transduction and Targeted Therapy*, *8*(1), 247.

Li, J., Chen, Z., Xiao, W., Liang, H., Liu, Y., Hao, W., Zhang, Y., & Wei, F. (2023). Chromosome instability region analysis and identification of the driver genes of the epithelial ovarian cancer cell lines A2780 and SKOV3. *Journal of Cellular and Molecular Medicine*, *27*(21), 3259–3270.

Li, L., Wu, L. P., & Chandrasegaran, S. (1992). Functional domains in Fok I restriction endonuclease. *Proceedings of the National Academy of Sciences*, *89*(10), 4275–4279.

Liu, Ji, Chang, J., Jiang, Y., Meng, X., Sun, T., Mao, L., Xu, Q., & Wang, M. (2019). Fast and efficient CRISPR/Cas9 genome editing in vivo enabled by bioreducible lipid and messenger RNA nanoparticles. *Advanced Materials*, *31*(33), 1902575.

Liu, Jun, Jin, T., Chang, S., Ritchie, H. H., Smith, A. J., & Clarkson, B. H. (2007). Matrix and TGF-β-related gene expression during human dental pulp stem cell (DPSC) mineralization. *In Vitro Cellular & Developmental Biology - Animal*, *43*(3–4), 120–128. https://doi.org/10.1007/s11626-007-9022-8

Liu, S. J., Horlbeck, M. A., Cho, S. W., Birk, H. S., Malatesta, M., He, D., Attenello, F. J., Villalta, J. E., Cho, M. Y., & Chen, Y. (2017). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science*, *355*(6320), eaah7111.

Liu, Y., Yang, G., Huang, S., Li, X., Wang, X., Li, G., Chi, T., Chen, Y., Huang, X., & Wang, X. (2021). Enhancing prime editing by Csy4-mediated processing of pegRNA. *Cell Research*, *31*(10), 1134–1136.

Lopez-de la Mora, D. A., Sanchez-Roque, C., Montoya-Buelna, M., Sanchez-Enriquez, S., Lucano-Landeros, S., Macias-Barragan, J., & Armendariz-Borunda, J. (2015). Role and new insights of pirfenidone in fibrotic diseases. *International Journal of Medical Sciences*, *12*(11), 840.

Lundvig, D. M. S., Immenschuh, S., & Wagener, F. A. (2012). Heme oxygenase, inflammation, and fibrosis: the good, the bad, and the ugly? *Frontiers in Pharmacology*, *3*, 81.

Mack, M., & Yanagita, M. (2015). Origin of myofibroblasts and cellular events triggering fibrosis. *Kidney International*, *87*(2), 297–307.

Maeder, M. L., Thibodeau-Beganny, S., Osiak, A., Wright, D. A., Anthony, R. M., Eichtinger, M., Jiang, T., Foley, J. E., Winfrey, R. J., & Townsend, J. A. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular Cell*, *31*(2), 294–301.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., & Yakunin, A. F. (2011). E. coli. *Nature Reviews Microbiology*, *9*(6), 467–477.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., & Haft, D. H. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736.

Malech, H. L. (2021). Treatment by CRISPR-Cas9 gene editing—a proof of principle. In *New England Journal of Medicine* (Vol. 384, Issue 3, pp. 286–287). Mass Medical Soc.

Malzahn, A., Lowder, L., & Qi, Y. (2017). Plant genome editing with TALEN and CRISPR. *Cell & Bioscience*, *7*(1), 1–18.

Mani, M., Kandavelou, K., Dy, F. J., Durai, S., & Chandrasegaran, S. (2005). Design, engineering, and characterization of zinc finger nucleases. *Biochemical and Biophysical Research Communications*, *335*(2), 447–457.

Martel-Pelletier, J., Boileau, C., Pelletier, J.-P., & Roughley, P. J. (2008). Cartilage in normal and osteoarthritis conditions. *Best Practice & Research Clinical Rheumatology*, *22*(2), 351–384.

Martyn, G. E., Quinlan, K. G. R., & Crossley, M. (2017). The regulation of human globin promoters by CCAAT box elements and the recruitment of NF-Y. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1860*(5), 525–536.

McPherson, S., Wilkinson, N., Tiniakos, D., Wilkinson, J., Burt, A. D., McColl, E., Stocken, D. D., Steen, N., Barnes, J., & Goudie, N. (2017). A randomised controlled trial of losartan as an anti-fibrotic agent in non-alcoholic steatohepatitis. *PloS One*, *12*(4), e0175717.

McVicker, B. L., & Bennett, R. G. (2017). Novel anti-fibrotic therapies. *Frontiers in Pharmacology*, *8*, 318.

Mienaltowski, M. J., & Birk, D. E. (2014). Structure, physiology, and biochemistry of collagens. *Progress in Heritable Soft Connective Tissue Diseases*, 5–29.

Miller, E. J., & Matukas, V. J. (1969). Chick cartilage collagen: a new type of α1 chain not present in bone or skin of the species. *Proceedings of the National Academy of Sciences*, *64*(4), 1264–1268.

Mino, T., Aoyama, Y., & Sera, T. (2009). Efficient double-stranded DNA cleavage by artificial zinc-finger nucleases composed of one zinc-finger protein and a single-chain FokI dimer. *Journal of Biotechnology*, *140*(3–4), 156–161.

Molla, K. A., & Yang, Y. (2019). CRISPR/Cas-mediated base editing: technical considerations and practical applications. *Trends in Biotechnology*, *37*(10), 1121–1142.

Moore-Morris, T., Guimarães-Camboa, N., Yutzey, K. E., Pucéat, M., & Evans, S. M. (2015). Cardiac fibroblasts: from development to heart failure. *Journal of Molecular Medicine*, *93*, 823–830.

Morelli, E., Gulla', A., Amodio, N., Taiana, E., Neri, A., Fulciniti, M., & Munshi, N. C. (2021). CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) to explore the oncogenic lncRNA network. In *Long Non-Coding RNAs in Cancer* (pp. 189–204). Springer.

Muley, V. Y. (2023). Search, Retrieve, Visualize, and Analyze Protein–Protein Interactions from

Multiple Databases: A Guide for Experimental Biologists. In *Protein-Protein Interactions: Methods and Protocols* (pp. 429–443). Springer.

Murata, K., Hirata, A., Ohta, K., Enaida, H., & Nakamura, K. (2019). Morphometric analysis in mouse scleral fibroblasts using focused ion beam/scanning electron microscopy. *Scientific Reports*, *9*(1), 6329.

Murphy, J. M., Rodriguez, Y. A. R., Jeong, K., Ahn, E.-Y. E., & Lim, S.-T. S. (2020). Targeting focal adhesion kinase in cancer cells and the tumor microenvironment. *Experimental & Molecular Medicine*, *52*(6), 877–886.

Najafi, M., Farhood, B., & Mortezaee, K. (2019). Extracellular matrix (ECM) stiffness and degradation as cancer drivers. *Journal of Cellular Biochemistry*, *120*(3), 2782–2790.

Nami, F., Ramezankhani, R., Vandenabeele, M., Vervliet, T., Vogels, K., Urano, F., & Verfaillie, C. (2021). Fast and efficient generation of isogenic induced pluripotent stem cell lines using adenine base editing. *The CRISPR Journal*, *4*(4), 502–518.

Nanthakumar, C. B., Hatley, R. J. D., Lemma, S., Gauldie, J., Marshall, R. P., & Macdonald, S. J. F. (2015). Dissecting fibrosis: therapeutic insights from the small-molecule toolbox. *Nature Reviews Drug Discovery*, *14*(10), 693–720.

Naomi, R., Ridzuan, P. M., & Bahari, H. (2021). Current insights into collagen type I. *Polymers*, *13*(16), 2642.

Nicolosi, P., Heald, B., & Esplin, E. D. (2022). What is a variant of uncertain significance in genetic testing? *European Urology Focus*, *8*(3), 654–656.

Nishino, T., Miyazaki, M., Abe, K., Furusu, A., Mishima, Y., Harada, T., Ozono, Y., Koji, T., & Kohno, S. (2003). Antisense oligonucleotides against collagen-binding stress protein HSP47 suppress peritoneal fibrosis in rats. *Kidney International*, *64*(3), 887–896.

Nitsche, F. (2016). A phylogenetic and morphological re-investigation of Diaphanoeca spiralifurca, Didymoeca elongata and Polyoeca dichotoma (Acanthoecida/Choanomonadida) from the Caribbean Sea. *European Journal of Protistology*, *52*, 58–64.

Ohtsuka, M., Sato, M., Miura, H., Takabayashi, S., Matsuyama, M., Koyano, T., Arifin, N., Nakamura, S., Wada, K., & Gurumurthy, C. B. (2018). i-GONAD: a robust method for in situ germline genome engineering using CRISPR nucleases. *Genome Biology*, *19*(1), 1–15.

Panizo, S., Martínez-Arias, L., Alonso-Montes, C., Cannata, P., Martín-Carro, B., Fernández-Martín, J. L., Naves-Díaz, M., Carrillo-López, N., & Cannata-Andía, J. B. (2021). Fibrosis in chronic kidney disease: pathogenesis and consequences. *International Journal of Molecular Sciences*, *22*(1), 408.

Park, S. H., & Bao, G. (2021). CRISPR/Cas9 gene editing for curing sickle cell disease. *Transfusion and Apheresis Science*, *60*(1), 103060.

Patton, C., Farr III, G. H., An, D., Martini, P. G. V, & Maves, L. (2018). Lipid nanoparticle packaging is an effective and nontoxic mRNA delivery platform in embryonic zebrafish. *Zebrafish*, *15*(3), 217–227.

Peng, R., Wang, S., Wang, R., Wang, Y., Wu, Y., & Yuan, Y. (2017). Antifibrotic effects of tanshinol in experimental hepatic fibrosis by targeting PI3K/AKT/mTOR/p70S6K1 signaling pathways. *Discovery Medicine*, *23*(125), 81–94.

Petersen, B., & Niemann, H. (2015). Molecular scissors and their application in genetically modified farm animals. *Transgenic Research*, *24*, 381–396.

Phan, S. H. (2002). The myofibroblast in pulmonary fibrosis. *Chest*, *122*(6), 286S-289S.

Phua, K. K. L., Leong, K. W., & Nair, S. K. (2013). Transfection efficiency and transgene expression

kinetics of mRNA delivered in naked and nanoparticle format. *Journal of Controlled Release*, *166*(3), 227–233.

Pickup, M. W., Mouw, J. K., & Weaver, V. M. (2014). The extracellular matrix modulates the hallmarks of cancer. *EMBO Reports*, *15*(12), 1243–1253.

Piek, A., De Boer, R. A., & Silljé, H. H. W. (2016). The fibrosis-cell death axis in heart failure. *Heart Failure Reviews*, *21*, 199–211.

Pinzani, M. (2008). Welcome to fibrogenesis & tissue repair. *Fibrogenesis & Tissue Repair*, *1*(1), 1.

Porto, E. M., Komor, A. C., Slaymaker, I. M., & Yeo, G. W. (2020). Base editing: advances and therapeutic opportunities. *Nature Reviews Drug Discovery*, *19*(12), 839–859.

Prockop, D. J., & Kivirikko, K. I. (1984). Heritable diseases of collagen. *New England Journal of Medicine*, *311*(6), 376–386.

Qi, Y., & Xu, R. (2018). synthesis and cancer progressionRoles of PLODs in collagen. *Frontiers in Cell and Developmental Biology*, *6*, 66.

Rafehi, H., Orlowski, C., Georgiadis, G. T., Ververis, K., El-Osta, A., & Karagiannis, T. C. (2011). Clonogenic assay: adherent cells. *JoVE (Journal of Visualized Experiments)*, *49*, e2573.

Rajewsky, K., Gu, H., Kühn, R., Betz, U. A., Müller, W., Roes, J., & Schwenk, F. (1996). Conditional gene targeting. *The Journal of Clinical Investigation*, *98*(3), 600–603.

Rappsilber, J., Mann, M., & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols*, *2*(8), 1896–1906.

Ravassa, S., López, B., Treibel, T. A., San José, G., Losada-Fuentenebro, B., Tapia, L., Bayés-Genís, A., Díez, J., & González, A. (2023). Cardiac Fibrosis in heart failure: Focus on non-invasive diagnosis and emerging therapeutic strategies. *Molecular Aspects of Medicine*, *93*, 101194.

Ricard-Blum, S. (2011). The collagen family. *Cold Spring Harbor Perspectives in Biology*, *3*(1), a004978.

Ricard-Blum, S., Dublet, B., & Van der Rest, M. (2000). *Unconventional collagens*. Oxford University Press, USA.

Rishikof, D. C., Ricupero, D. A., Liu, H., & Goldstein, R. H. (2004). Phenylbutyrate decreases type I collagen production in human lung fibroblasts. *Journal of Cellular Biochemistry*, *91*(4), 740–748.

Rong, Y. S., & Golic, K. G. (2000). Gene targeting by homologous recombination in Drosophila. *Science*, *288*(5473), 2013–2018.

Roschmann, R. A., & Rothenberg, R. J. (1987). Pulmonary fibrosis in rheumatoid arthritis: a review of clinical features and therapy. *Seminars in Arthritis and Rheumatism*, *16*(3), 174–185.

Rosello, M., Serafini, M., Mignani, L., Finazzi, D., Giovannangeli, C., Mione, M. C., Concordet, J.-P., & Del Bene, F. (2022). Disease modeling by efficient genome editing using a near PAM-less base editor in vivo. *Nature Communications*, *13*(1), 3435.

Rosso, F., Giordano, A., Barbarisi, M., & Barbarisi, A. (2004). From cell–ECM interactions to tissue engineering. *Journal of Cellular Physiology*, *199*(2), 174–180.

Santiago-Rodriguez, T. M., & Hollister, E. B. (2021). Multi 'omic data integration: A review of concepts, considerations, and approaches. *Seminars in Perinatology*, *45*(6), 151456.

Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, *82*(7–8), 518–529.

Scholefield, J., & Harrison, P. T. (2021). Prime editing–an update on the field. *Gene Therapy*, *28*(7–8), 396–401.

Schultz, G. S., & Wysocki, A. (2009). Interactions between extracellular matrix and growth factors in wound healing. *Wound Repair and Regeneration*, *17*(2), 153–162.

Schuppan, D., & Kim, Y. O. (2013). Evolving therapies for liver fibrosis. *The Journal of Clinical Investigation*, *123*(5), 1887–1901.

Shaukat, Z., Aiman, S., & Li, C.-H. (2021). Protein-protein interactions: Methods, databases, and applications in virus-host study. *World Journal of Virology*, *10*(6), 288.

Shi, R., Zhang, Z., Zhu, A., Xiong, X., Zhang, J., Xu, J., Sy, M., & Li, C. (2022a). Aberrant type I and type III collagen gene expression in human breast cancer in vivo. *International Journal of Cancer*, *151*(5), 665–683.

Shi, R., Zhang, Z., Zhu, A., Xiong, X., Zhang, J., Xu, J., Sy, M., & Li, C. (2022b). Targeting type I collagen for cancer treatment. *International Journal of Cancer*, *151*(5), 665–683.

Sila-Asna, M., Bunyaratvej, A., Maeda, S., Kitaguchi, H., & Bunyaratavej, N. (2007). Osteoblast Differentiation and Bone Formation Gene Expression in Strontium-inducing Bone Marrow Mesenchymal Stem Cell. *Kobe J. Med. Sci*, *53*(1), 25–35.

Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., & Pâques, F. (2011). Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Current Gene Therapy*, *11*(1), 11–27.

Song, Y., Liu, Z., Zhang, Y., Chen, M., Sui, T., Lai, L., & Li, Z. (2020). Large-fragment deletions induced by Cas9 cleavage while not in the BEs system. *Molecular Therapy-Nucleic Acids*, *21*, 523–526.

Sorushanova, A., Delgado, L. M., Wu, Z., Shologu, N., Kshirsagar, A., Raghunath, R., Mullen, A. M., Bayon, Y., Pandit, A., & Raghunath, M. (2019). The collagen suprafamily: from biosynthesis to advanced biomaterial development. *Advanced Materials*, *31*(1), 1801651.

Staal, J., Alci, K., Schamphelaire, W. De, Vanhoucke, M., & Beyaert, R. (2019). Engineering a minimal cloning vector from a pUC18 plasmid backbone with an extended multiple cloning site. *BioTechniques*, *66*(6), 254–259.

Stefanovic, B., Michaels, H. A., & Nefzi, A. (2021). Discovery of a lead compound for specific inhibition of type I collagen production in fibrosis. *ACS Medicinal Chemistry Letters*, *12*(3), 477–484.

Stefanovic, L., & Stefanovic, B. (2019). Technology for discovery of antifibrotic drugs: phenotypic screening for LARP6 inhibitors using inverted yeast three hybrid system. *ASSAY and Drug Development Technologies*, *17*(3), 116–127.

Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinformatics and Biology Insights*, *14*, 1177932219899051.

Sudhakaran, M., Navarrete, T. G., Mejía-Guerra, K., Mukundi, E., Eubank, T. D., Grotewold, E., Arango, D., & Doseff, A. I. (2023). Transcriptome reprogramming through alternative splicing triggered by apigenin drives cell death in triple-negative breast cancer. *Cell Death & Disease*, *14*(12), 824.

Sun, X., Cui, X., Chen, X., & Jiang, X. (2020). Baicalein alleviated TGF β1-induced type I collagen production in lung fibroblasts via downregulation of connective tissue growth factor. *Biomedicine & Pharmacotherapy*, *131*, 110744.

Sürün, D., Schneider, A., Mircetic, J., Neumann, K., Lansing, F., Paszkowski-Rogacz, M., Hänchen, V., Lee-Kirsch, M. A., & Buchholz, F. (2020). Efficient generation and correction of mutations in

human iPS cells utilizing mRNAs of CRISPR base editors and prime editors. *Genes*, *11*(5), 511.

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., & Bork, P. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605–D612.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., & Pyysalo, S. (2023). The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, *51*(D1), D638–D646.

Tan, J., Zhang, F., Karcher, D., & Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nature Communications*, *10*(1), 439.

Tanaka, K., & Mitsushima, A. (1984). A preparation method for observing intracellular structures by scanning electron microscopy. *Journal of Microscopy*, *133*(2), 213–222.

Tang, L., Fares, H., Zhao, X., Du, W., & Liu, B.-F. (2012). Different endocytic functions of AGEF-1 in C. elegans coelomocytes. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1820*(7), 829–840.

Tavernier, G., Andries, O., Demeester, J., Sanders, N. N., De Smedt, S. C., & Rejman, J. (2011). mRNA as gene therapeutic: how to control protein expression. *Journal of Controlled Release*, *150*(3), 238–247.

Thomson, C. A., Atkinson, H. M., & Ananthanarayanan, V. S. (2005). Identification of small molecule chemical inhibitors of the collagen-specific chaperone Hsp47. *Journal of Medicinal Chemistry*, *48*(5), 1680–1684.

Tilghman, R. W., & Parsons, J. T. (2008). Focal adhesion kinase as a regulator of cell tension in the progression of cancer. *Seminars in Cancer Biology*, *18*(1), 45–52.

Traxler, E. A., Yao, Y., Wang, Y.-D., Woodard, K. J., Kurita, R., Nakamura, Y., Hughes, J. R., Hardison, R. C., Blobel, G. A., & Li, C. (2016). A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nature Medicine*, *22*(9), 987–990.

Tsamis, A., Krawiec, J. T., & Vorp, D. A. (2013). Elastin and collagen fibre microstructure of the human aorta in ageing and disease: a review. *Journal of the Royal Society Interface*, *10*(83), 20121004.

Tyanova, S., & Cox, J. (2018). Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research. *Cancer Systems Biology: Methods and Protocols*, 133–148.

Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., & Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, *435*(7042), 646–651.

van Nieuwenhoven, F. A., & Turner, N. A. (2013). The role of cardiac fibroblasts in the transition from inflammation to fibrosis following myocardial infarction. *Vascular Pharmacology*, *58*(3), 182–188.

Van Steensel, B., & Belmont, A. S. (2017). Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell*, *169*(5), 780–791.

Venning, F. A., Wullkopf, L., & Erler, J. T. (2015). Targeting ECM disrupts cancer progression. *Frontiers in Oncology*, *5*, 224.

Volk, A., & Crispino, J. D. (2015). The role of the chromatin assembly complex (CAF-1) and its p60 subunit (CHAF1b) in homeostasis and disease. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1849*(8), 979–986.

Walejko, J. M., Koelmel, J. P., Garrett, T. J., Edison, A. S., & Keller-Wood, M. (2018). Multiomics approach reveals metabolic changes in the heart at birth. *American Journal of Physiology-Endocrinology and Metabolism*, *315*(6), E1212–E1223.

Wang, J.-C., Lai, S., Guo, X., Zhang, X., de Crombrugghe, B., Sonnylal, S., Arnett, F. C., & Zhou, X. (2010). Attenuation of fibrosis in vitro and in vivo with SPARC siRNA. *Arthritis Research & Therapy*, *12*, 1–9.

Wang, J., Hu, K., Cai, X., Yang, B., He, Q., Wang, J., & Weng, Q. (2022). Targeting PI3K/AKT signaling for treatment of idiopathic pulmonary fibrosis. *Acta Pharmaceutica Sinica B*, *12*(1), 18–32.

Wang, K., Shen, X., & Williams, R. (2021). Sequencing BGI: the evolution of expertise and research organisation in the world's leading gene sequencing facility. *New Genetics and Society*, *40*(3), 305–330.

Wang, Y.-Y., Hsu, S.-H., Tsai, H.-Y., Cheng, F.-Y., & Cheng, M.-C. (2022). Transcriptomic and Proteomic Analysis of CRISPR/Cas9-Mediated ARC-Knockout HEK293 Cells. *International Journal of Molecular Sciences*, *23*(9), 4498.

Watson, C. J. E., & Dark, J. H. (2012). Organ transplantation: historical perspective and current practice. *British Journal of Anaesthesia*, *108*(suppl_1), i29–i42.

Webber, B. R., Lonetree, C., Kluesner, M. G., Johnson, M. J., Pomeroy, E. J., Diers, M. D., Lahr, W. S., Draper, G. M., Slipek, N. J., & Smeester, B. A. (2019). Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nature Communications*, *10*(1), 5222.

Winer, A., Adams, S., & Mignatti, P. (2018). Matrix metalloproteinase inhibitors in cancer therapy: turning past failures into future successes. *Molecular Cancer Therapeutics*, *17*(6), 1147–1155.

Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., & Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nature Communications*, *11*(1), 5120.

Winslow, S., Lindquist, K. E., Edsjö, A., & Larsson, C. (2016). The expression pattern of matrix-producing tumor stroma is of prognostic importance in breast cancer. *BMC Cancer*, *16*, 1–13.

Wiśniewska, J., Sadowska, A., Wójtowicz, A., Słyszewska, M., & Szóstek-Mioduchowska, A. (2021). Perspective on stem cell therapy in organ fibrosis: animal models and human studies. *Life*, *11*(10), 1068.

Wörheide, M. A., Krumsiek, J., Kastenmüller, G., & Arnold, M. (2021). Multi-omics integration in biomedical research–A metabolomics-centric review. *Analytica Chimica Acta*, *1141*, 144–162.

Wu, Y., Liang, D., Wang, Y., Bai, M., Tang, W., Bao, S., Yan, Z., Li, D., & Li, J. (2013). Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell*, *13*(6), 659–662.

Wynn, T. A. (2007). Common and unique mechanisms regulate fibrosis in various fibroproliferative diseases. *The Journal of Clinical Investigation*, *117*(3), 524–529.

Wynn, T. A., & Ramalingam, T. R. (2012). Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nature Medicine*, *18*(7), 1028–1040.

Yan, J., Cirincione, A., & Adamson, B. (2020). Prime editing: precision genome editing by reverse transcription. *Molecular Cell*, *77*(2), 210–212.

Yan, Q., Zhang, Q., Yang, H., Zou, Q., Tang, C., Fan, N., & Lai, L. (2014). Generation of multi-gene knockout rabbits using the Cas9/gRNA system. *Cell Regeneration*, *3*(1), 3–12.

Yin, H., Song, C.-Q., Dorkin, J. R., Zhu, L. J., Li, Y., Wu, Q., Park, A., Yang, J., Suresh, S., & Bizhanova, A. (2016). Therapeutic genome editing by combined viral and non-viral delivery of

CRISPR system components in vivo. *Nature Biotechnology*, *34*(3), 328.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, *11*, 1–12.

Yue, B. (2014). Biology of the extracellular matrix: an overview. *Journal of Glaucoma*, S20.

Zaboikin, M., Zaboikina, T., Freter, C., & Srinivasakumar, N. (2017). Non-homologous end joining and homology directed DNA repair frequency of double-stranded breaks introduced by genome editing reagents. *PloS One*, *12*(1), e0169931.

Zeng, J., Wu, Y., Ren, C., Bonanno, J., Shen, A. H., Shea, D., Gehrke, J. M., Clement, K., Luk, K., & Yao, Q. (2020). Therapeutic base editing of human hematopoietic stem cells. *Nature Medicine*, *26*(4), 535–541.

Zeng, J., Wu, Y., Ren, C., Bonanno, J., Shen, A. H., Shea, D., Gehrke, J. M., Manis, J. P., Gehrke, J. K., & Bauer, D. E. (2019). Therapeutic Base Editing of Human Hematopoietic Stem Cells. *Blood*, *134*, 612.

Zhang, H., Qin, C., An, C., Zheng, X., Wen, S., Chen, W., Liu, X., Lv, Z., Yang, P., & Xu, W. (2021). Application of the CRISPR/Cas9-based gene editing technique in basic research, diagnosis, and therapy of cancer. *Molecular Cancer*, *20*, 1–22.

Zhang, J.-D., Li, A.-T., Yu, H.-L., Imanaka, T., & Xu, J.-H. (2011). Synthesis of optically pure S-sulfoxide by Escherichia coli transformant cells coexpressing the P450 monooxygenase and glucose dehydrogenase genes. *Journal of Industrial Microbiology and Biotechnology*, *38*(5), 633–641.

Zhang, R., Xu, W., Shao, S., & Wang, Q. (2021). Gene silencing through CRISPR interference in bacteria: current advances and future prospects. *Frontiers in Microbiology*, *12*, 635227.

Zhang, X., Zhao, W., Nguyen, G. N., Zhang, C., Zeng, C., Yan, J., Du, S., Hou, X., Li, W., & Jiang, J. (2020). Functionalized lipid-like nanoparticles for in vivo mRNA delivery and base editing. *Science Advances*, *6*(34), eabc2315.

Zheng, Y., Liu, Y., Yang, J., Dong, L., Zhang, R., Tian, S., Yu, Y., Ren, L., Hou, W., & Zhu, F. (2023). Multi-omics data integration using ratio-based quantitative profiling with Quartet reference materials. *Nature Biotechnology*, 1–17.

Zuccaro, M. V, Xu, J., Mitchell, C., Marin, D., Zimmerman, R., Rana, B., Weinstein, E., King, R. T., Palmerola, K. L., & Smith, M. E. (2020). Allele-specific chromosome removal after Cas9 cleavage in human embryos. *Cell*, *183*(6), 1650–1664.

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L. M., Li, Y., & Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*, *364*(6437), 289–292.

Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen - einschließlich Tabellen, Karten und Abbildungen , die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind in jedem Einzelfall als Entlehnung kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertationsschrift noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss der Promotion nicht ohne Genehmigung der / des Vorsitzenden des IPHS-Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von PD.Dr.Kurt Pfannkuche betreut worden.Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem IPHS-Promotionsausschuss unverzüglich mitzuteilen.

Datum/Unterschrift 09.07.2024