Designing a prime editing guide to target the rs1190870 SNP associated with the development of adolescent idiopathic scoliosis in East Asian populations

Evana Jang

Abstract

Several studies have reported the association of the Ladybird Homeobox 1 (*LBX1*) gene with the development of adolescent idiopathic scoliosis (AIS) in East Asian populations. AIS is a type of scoliosis that develops in adolescents with no definite cause and leads to spinal discomfort and complications. Within the regulatory regions of *LBX1*, there are single-nucleotide polymorphisms (SNPs) that are linked to susceptibility to AIS, specifically if an individual inherits what is deemed the risk allele. This paper discusses three SNPs that are linked to AIS within the East Asian population: rs11190870, rs678741, and rs625039. We then dive into the most highly associated SNP – rs11190870 – and design a prime editing guide utilizing the PegFinder website to edit the risk allele T to the non-risk allele A. This paper then discusses the optimal way to deliver the CRISPR complex *in vivo*, through an adeno-associated virus (AAV) directly into skeletal muscle cells. Our study proposes a CRISPR guide to replace a well-studied risk allele, potentially reducing the risk of developing scoliosis, utilizing cutting-edge gene technology that may benefit families with a history of scoliosis.

Introduction

Globally, around 2-3% of the population is affected by scoliosis, a condition where the spine has abnormal curvature [1]. The severity of the curve determines the symptoms a patient experiences. In moderate or even mild cases, this spinal curvature can impact the quality of life of the patient [2]. Namely, it can cause back pain, leg pain, numbness, discomfort, core muscle weakness, uneven shoulders, and even difficulty breathing. Severe cases may further result in additional complications such as organ damage, nerve damage, arthritis, and spinal fluid leakage, if left untreated [2].

Patients with scoliosis are typically diagnosed during early childhood or adolescence. The spinal curvature can take on a C or S shape, and the severity of the condition is measured by the degree of the curve [2]. The degree is measured using a Cobb angle, which is the angle between the top and bottom vertebrae of the spinal curve. A curvature of 0° - 10° indicates a person does not have scoliosis, 10° - 24° indicates a mild case, 25° - 39° indicates a moderate case, and 40° or above indicates a severe case [2].

Approximately 85% of scoliosis cases are mild and do not require heavy treatment such as surgery and bracing. Bracing is commonly used for moderate cases of scoliosis, especially in patients who are still growing. However, there are side effects associated with bracing, as the use of external compressive forces has been found to decrease lung function by reducing lung volume and increasing the effort required for breathing [3]. Surgery is reserved for severe cases of scoliosis and involves the surgeon placing bone grafts to hold the spine in the correct position. Risks of scoliosis surgery include an irreversible loss of normal spinal range of motion, neurological damage, stress fractures, and even death [4]. Strategies to treat scoliosis are continually evolving, seeking to maximize effectiveness while minimizing side effects.

There are three types of scoliosis: congenital scoliosis, syndromic scoliosis, and idiopathic scoliosis [5]. Congenital scoliosis is a deformity from birth due to abnormally formed

vertebrae. On the other hand, syndromic scoliosis develops from medical conditions such as neuromuscular, skeletal, or connective tissue diseases. Scoliosis is considered idiopathic when it has no known cause of disease.

This research paper focuses on adolescent idiopathic scoliosis (AIS) – a type of scoliosis that occurs during late childhood or adolescence, usually when children encounter a growth spurt. AIS is the most common spinal deformity, and about 1-3% of children aged 10-16 years old will have some degree of spinal curvature [5]. While most of the time, the spinal curvature remains unchanged as a person grows, some cases can be progressive – meaning the curvature worsens over time. Around 80% of AIS cases are considered mild, with a curvature degree of around 10-25.

Common signs of AIS include having shoulders at uneven levels and waistline asymmetry, with one hip or rib sticking up more than the other. Treatment beyond regular monitoring is not needed, as symptoms are often mild. However, mild AIS can cause discomfort and mobility issues that may be hard to resolve, as surgery tends not to be an option due to how mild the case is. Back pain is the most pressing complaint of the disease [5]—approximately one out of every four patients with AIS experiences back pain [6].

Females tend to be more susceptible to severe AIS, although the reasoning behind this is not clearly defined. Konieczny et al. found that girls were reported to possess higher Cobb angles than boys. The female-to-male ratio possessing a Cobb angle greater than 40° was 7.2:1. These statistics show that scoliosis in girls tends to progress to a higher severity [7].

Ethnicity may also influence a person's susceptibility to developing AIS. Multiple studies compared the prevalence of AIS in different ethnic groups within similar regions. For instance, Guo et al. examined the prevalence of scoliosis among a group of 2,445 adolescents attending school in Tibet [8]. 47.2% of the population was Han Chinese, and the rest were ethnic minorities. After measuring trunk rotation as an indicator of scoliosis, females of Han ethnicity were found to have larger trunk rotational values than ethnic minorities, indicating a higher prevalence of scoliosis.

While predisposing factors, such as gender and ethnicity, can play a role in the multifactorial etiology of scoliosis, lifestyle factors also significantly contribute to its development. Engaging in sports involving asymmetric loading of the spine can disturb the balancing mechanism of the spine and trigger or progress the development of scoliosis [9]. Modi et al. reported that ballet dancers and rhythmic gymnasts are developing scoliosis at an increased rate because of the asymmetric load that their activities involve, which causes muscle imbalances and thus uneven pressure on the spine [9]. Asymmetric load due to repeated rotational motions or one-hand dominance can be found in most sports, each containing a varying degree of severity. For instance, Modi et al. also examined the prevalence of scoliosis within volleyball players versus a control group of non-volleyball players. They found that the study group had a higher prevalence of scoliosis, which could be attributed to the repeated asymmetrical rotational and bending strains that volleyball players face, producing instability in their muscles.

Athletes who are affected by scoliosis often have to spend more energy performing at the same level as their non-affected peers because muscle imbalance and unevenness can decrease overall strength and agility. For example, scoliosis heavily impacted Zhang Yufei's, a Chinese Olympic gold medalist in swimming, career through making upper body stability more challenging, resulting in additional energy compensation to balance it out [10]. She had to work

even harder, often supplementing swimming training with scoliosis-focused exercises to achieve top performance.

As found in its name, its idiopathic nature signifies that there is no definitive known cause for this disease. Previous studies have concluded that the aetiology of AIS is multifaceted and includes a complex combination of both environmental and genetic factors, and tends to run in families. Grauers et al. analyzed scoliosis data within a population of 64,578 twins in Sweden, estimating that 38% of scoliosis development is due to genetic effects, while 62% is due to environmental factors [11]. Within the genetic aspect of AIS, specific genes are associated with an individual's susceptibility to AIS.

The LBX1 gene and its association with AIS

The ladybird homeobox 1 gene (*LBX1*) is one of the most prominent genes linked to the development of AIS [12]. *LBX1* plays a crucial role in the development of skeletal muscle [12]. The Genotype-Tissue Expression (GTEx) Project, which sequenced the RNA expression from 54 non-diseased tissue sites in approximately 1000 individuals, measured LBX1 expression across multiple tissues for *LBX1*. *LBX1* is highly upregulated in skeletal muscle tissue, with the brain and spinal cord tissue following as the next most highly expressed. *LBX1* is active in muscle precursor cells known as myoblasts, which are undifferentiated muscle cells that eventually mature into muscle fibers [13]. Watanabe et al. reported that, in particular, *LBX1* is expressed in myoblasts that undergo long-distance migration to target locations such as the limbs, diaphragm, and tongue to form muscles there. To aid in guiding myoblasts to their target location, *LBX1* controls the genes that recognize and interpret cues determining the route of migrating muscle precursors [14].

In *LBX1* mutants, Brohmann et al. discovered that the cells migrate less efficiently, and thus fewer precursors reach their target destination. Less efficient migration can cause certain muscle groups to become smaller in size. Besides AIS, mutant *LBX1* has been linked to various diseases, including recessive congenital central hypoventilation syndrome, which is an uncommon genetic disorder that affects breathing and the autonomic nervous system [15]. It is caused by a homozygous frameshift mutation, which is characterised by one or more base pairs being added or deleted in both copies of a gene.

Regarding AIS, previous studies have found that the disease can be caused by single-nucleotide polymorphisms (SNPs) within and surrounding the *LBX1* gene, where certain alleles increase the susceptibility of developing AIS. Specific genetic variants were found to lower or upregulate the expression of *LBX1* in paraspinal muscles, which causes an imbalance in muscle. For instance, for rs1322330, Xu et al. found that patients with a genotype of AA instead of GG had significantly decreased *LBX1* mRNA expression and protein expression in paraspinal muscles [16]. Within the study's sample, the frequency of allele A was considerably higher than in the control group, causing AA to be labeled as the risk allele while GG to be labeled as the non-risk allele.

This paper aims to explore other SNPs that have not been studied to the depth of rs1322330, including rs11190870, rs678741, and rs625039. These variants were chosen because of their prominence in East Asian and/or female groups. Jiang et al. conducted research surrounding the rs11190870 SNP in a Han Chinese population within the Yangtze River region. The SNP was previously found to have an association with AIS in both a Japanese and a Hong Kong population, and Jiang et al. later found that it also had an association with the

Yangtze River region population [17]. Cao et al.'s study examined the rs625039 SNP and found statistical significance that the rs625039 polymorphism was associated with AIS in Asian populations [18]. Cao et al.'s study also examined the rs678741 SNP in a solely female population and found that the G allele could decrease a woman's susceptibility to AIS (2016). Each SNP has its own risk allele and non-risk allele, which can be obtained from the data within the UCSC Genome Browser.

The rs11190870 SNP is located on the 3' flanking region of the *LBX1* gene on chromosome 10q24.31 [19]. The 3' flanking region is the DNA sequence located immediately downstream of the 3' end of the gene, beyond the coding sequence. This region is usually transcribed but not translated into RNA. The 3' flanking region plays a role in determining the pattern and expression level of the gene by interacting with transcription factor proteins and different hormones [20]. Guo et al. investigated the interaction of the SNP rs11190870 and the gene *LBX1* in zebrafish, finding that the risk allele was T and the non-risk allele was C [21]. HEK 293T cells containing the risk allele of rs11190870 had higher transcriptional activity in the promoter region of *LBX1*. The overexpression of the *LBX1* gene caused body curvature in zebrafish embryos, suggesting that the T risk allele was associated with scoliosis development. This may occur because the upregulation of a gene can disrupt normal development of the musculoskeletal and nervous systems, increasing an individual's susceptibility to musculoskeletal disorders, including scoliosis.

The rs625039 SNP is located in the 5' flanking region of the gene *LBX1*, also on chromosome 10q24.31. This region is a sequence of DNA that lies upstream of the transcriptional start site of *LBX1*, unlike the previous SNP located in the downstream area. The 5' flanking region is crucial for regulating gene expression as it contains elements that influence the extent to which a gene is transcribed. The region contains enhancers, which are sequences of DNA that enhance transcription [22]. Luo et al. identified a change to a G at the rs625039 SNP as a risk allele for AIS. Like the risk allele of rs11190870, which overexpresses the *LBX1* gene, the enhancers that the risk allele rs625039 SNP interacts with also elevate gene expression, which can contribute to the disruption of normal musculoskeletal development. Guo et al. found that enhancer-driven overexpression of *LBX1* leads to defective convergent extension movement and body curvature in zebrafish, mirroring the symptoms of scoliosis in humans [21].

Finally, the rs678741 SNP is located in the intron of the *LBX1AS1* gene, which is near the *LBX1* gene also on chromosome 10q24.31. The *LBX1AS1* gene is a DNA sequence that closely resembles the functional gene but is non-coding, meaning it does not directly code for proteins. It does, however, have the potential to regulate the expression of the *LBX1* gene [23]. Previous studies have found that the A allele is the risk allele. A meta-analysis of 34,626 subjects conducted by Cao et al. found that the G allele of rs678741 decreased AIS susceptibility in females [18].

CRISPR treatment approach to target the risk allele

Research has shown that the development of AIS can be attributed partly to genetics. To target the genetic component of the disease's etiology, specific gene editing tools can be utilized to modify an individual's DNA, reducing the chances of developing AIS. In particular, if AIS is known to run in a family's history, these tools can be utilized in offspring to minimize the chances of developing the disease or reverse its progression. Gene editing systems such as

CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9) offer a promising solution, potentially allowing patients to not only avoid the negative symptoms of scoliosis but also invasive treatment procedures, such as surgery. CRISPR/Cas9 is a widely used genetic engineering tool to make insertions, removals, or modifications to DNA [24]. Initially found in prokaryotes as a key defense mechanism against viruses, CRISPR/Cas9 was revolutionized in 2012 by Jennifer Doudna and Emmanuelle Charpentier, who discovered that it could be used to edit DNA [25].

CRISPR/Cas9 uses two essential components – a protein called Cas9 and a guide RNA (gRNA) to make desired genetic changes [26]. The gRNA guides the Cas9 to the target DNA sequence by finding its complementary base pair location. The Cas9 protein then makes double-stranded breaks at the site three base pairs upstream of the Protospacer Adjacent Motif (PAM) sequence, which is a short 2-5 base pair sequence located right next to the target cleavage site that indicates where the guide RNA should bind and make a cut. The cell is then able to repair the double-stranded break through non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ repairs the double-stranded breaks without a DNA template by joining the two DNA fragments together. This process is more prone to error, resulting in small, random insertions or deletions, which can lead to frameshift mutations or premature stop codons. The HDR cell mechanism, on the other hand, is more precise as it utilizes a homologous DNA template to make repairs. CRISPR gene editing utilizes HDR by inserting a donor DNA template to effect the desired genetic changes. The sgRNA carries the donor DNA template with the sequence containing the edits, which is inserted after Cas9 makes the double-stranded breaks.

While CRISPR/Cas9 is a revolutionary technology, it does carry multiple risks. To start, it has been shown to induce a high number of double-stranded breaks at the incorrect site [27]. The double-strand break at the correct site can also cause off-target mutations, including indels, translocations, and duplications. Thus, improvements have been made to the technology to create higher precision and expand its functions. One advancement is Prime Editing (PE), a technology developed in 2019 [27]. Its complex includes a Cas9 nickase, which was created through mutating either the HNH or RuvC catalytic residues of the Cas9 enzyme [24]. The result was a nickase that only generates single-stranded breaks instead of cleaving the DNA. It also contains a Prime Editing guide RNA (pegRNA) that includes the new sequence and leads the complex to the correct site for editing, and a reverse transcriptase domain, which reads the sequence. At the site, the Cas9 nickase utilizes a spacer sequence to hybridize with the DNA strand, and then creates a single-stranded break in the strand containing the target sequence. The reverse transcriptase then reads the new sequence from the pegRNA to begin reverse transcription. The created sequence is then copied into the target DNA sequence, resulting in an altered DNA sequence.

Because PE does not rely on double-stranded breaks, unlike conventional CRISPR/Cas9, it drastically limits the amount of off-target effects when genetically editing [27]. Ochoa-Sanchez et al. also suggest that PE has a broader range of editing capabilities, as CRISPR is only capable of editing small target sequences and not well-suited to editing at the base pair level, while PE can introduce all 12 types of point mutations into the DNA sequence. Additionally, PE provides higher security in ensuring the pegRNA is at the correct site because it hybridizes with different target DNA sequences not once but three times.

Since we aimed to genetically edit a single base pair, we chose prime editing due to its high specificity and precision at a single base pair resolution.

Methods

Investigating LBX1 gene

We used the Genotype-Tissue Expression (GTEx) Portal to explore the RNA expression levels of *LBX1* across different tissue types, hypothesizing that it is most actively translated in skeletal muscle tissues, given its association with AIS. We then visualized *LBX1's* predicted protein structure from The Human Protein Atlas [28].

Known SNPs affecting LBX1 associated with LBX1

After analyzing the *LBX1* gene, we selected three SNPs in the gene's regulatory regions with RefSeq accessions rs11190870, rs678741, and rs625039 that were previously identified to have an association with the development of AIS, especially in the East Asian population [18, 28]. As a source of confirmation of their relation to AIS, we checked the GWAS Catalog to confirm whether each SNP had a "reported trait" of AIS within the website's public datasets and ensured the *P-values* rejected the null hypothesis that there is no association with AIS [28]. The UCSC genome browser was then utilized to gather further data on each SNP [30]. The human GRCh38/hg38 version was selected, and the data gathered included the location of the SNP and twenty sequences flanking each SNP. The Ensembl genome database was then used to determine the risk and non-risk alleles of each SNP by viewing the associations listed from the available data. The population genetics report was then reviewed to determine the percentage of individuals within the East Asian population who carry the risk alleles. The regulatory regions of the *LBX1* containing the SNPs within Ensembl were also analyzed. We chose the rs11190870 because it had the highest association with AIS out of the three SNPs reported to continue additional downstream analysis on Benchling and to design a CRISPR guide [18].

Designing a prime editing guide

Benchling was utilized to visualize the rs1190870 SNP in relation to the *LBX1* gene by first copying the gene and 8,000 base pairs upstream and downstream of the gene from Ensembl into the software. The 20 base pairs flanking the rs11190870 SNPs previously obtained from the UCSC genome browser were copied into the search bar of the Benchling site containing the *LBX1* gene to locate each SNP in the sequence. The SNPs were then labeled utilizing Benchling's annotation tool.

Prime editing consists of five essential components: the spacer, scaffold, RTT, PBS, and tevopreQ1 motif [31]. The scaffold and tevopreQ1 remain constant, and the rest were designed utilizing the pegFinder website [32]. First, 100 base pairs upstream and downstream of the rs11190870 SNP containing the risk allele, T, were obtained from the USCS browser and copied into the wildtype/reference sequence section of the pegFinder website. Then, the identical 100 base pairs upstream and downstream of the rs11190870 SNP containing the non-risk allele, C, was set as the edited/desired sequence. After pasting the sequences, the option for finding PE3/PE3b secondary nicking sgRNAs was set as "yes." The minimum nick distance was set as 0, the maximum nick distance was set as 150 base pairs, and the CRISPR Enzyme was set as the Cas9-NGG enzyme. The pegFinder website then generated possible gRNAs for the target site and ranked them based on precision and efficiency. The top recommendations for

the sgRNA, RT template, PBS, PE3 nicking sgRNA, and PE3b nicking sgRNA were selected to form the pegRNA complex and copied into Benchling to visualize.

Results

LBX1 and SNP Analysis

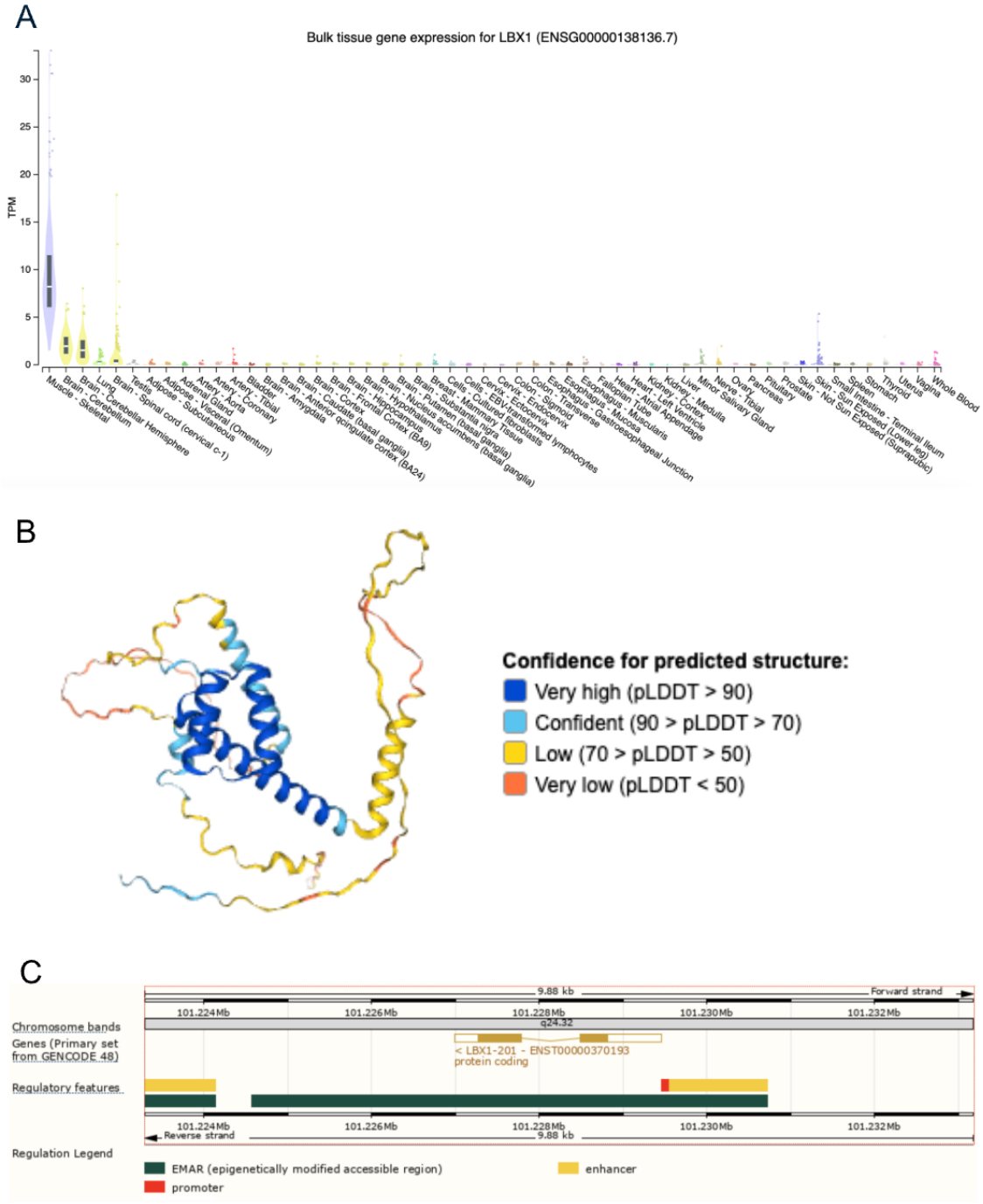

Figure 1: *LBX1* gene results. A) Gene expression of *LBX1* across tissue types in GWAS. B) Predicted protein structure of the wild-type *LBX1* gene from Alphafold v2.3.2 obtained within

The Human Protein Atlas. C) Gene architecture and regulatory region of the *LBX1* gene in Ensembl.

To ensure that the *LBX1* gene is expressed in the tissues we are planning on editing, the spinal muscles, we analyzed the bulk tissue gene expression for *LBX1* using the GTEx Portal. *LBX1* is indeed highly expressed in skeletal muscle tissue as compared to any other tissue **(Figure 1A)**. In a sample of 818 muscle skeletal tissues, the median transcript per million (TPM) was reported to be 8.239 **(Figure 1A)**. The next highest tissue type was the cerebellum, with 1.5 TPM, which is much lower. We then examined the predicted protein structure of the wild-type *LBX1* gene from Alphafold v2.3.2. **(Figure 1B)**. The predicted Local Distance Difference Test (pLDDT) refers to the level of confidence in AlphaFold's prediction. The color coordination indicates that the predicted protein structure has the highest confidence at the core of the protein in the alpha-helices (pLDDT > 90; **Figure B**) and less towards the disordered regions. Generally, AlphaFold was able to predict the whole structure. To visualize the gene architecture of *LBX1* and identify known regulatory elements surrounding the gene, we used Ensembl. We found that it is flanked by an enhancer and promoter region, which regulates gene expression **(Figure 1C)**. The flanking region is where the majority of the SNPs relating to AIS lie.

Table 1: Reported genomic and population data for each SNP significantly associated with AIS

SNP	Chromosome Location	Risk Allele	Non-risk Allele	Sample Size	P-value confirming significance with AIS	Referenced Study	Risk Allele frequency in East Asian population.
rs11190870	10:101219450	T	C	34,626	2×10^{-82}	Kou et al., 2019	54%
rs678741	10:101237824	A	G	10,333	1×10^{-36}	Zhu et al., 2015	44%
rs625039	10:101233892	G	A	34,626	< 0.001	Cao et al., 2016	65%

The three SNPs listed in Table 1 were identified from the reference studies to have a high association with the development of AIS. All three SNPs are located on chromosome 10, which is where the *LBX1* gene resides, and their exact location is listed in column 1 **(Table 1)**. The risk allele, listed in column 2, has a higher association with the disease than the non-risk allele, listed in column 3. To confirm the risk allele's association with AIS, the *P-values* listed in the referenced study of each SNP were included. The minuscule *P-values* for all three studies demonstrate that there is a firm rejection of the null that the SNPs do not have an association with AIS. We used Ensembl to gather the population percentage of the risk allele for all three SNPs. The risk allele, G, of rs626039 had the highest frequency in the East Asian population,

while the risk allele, A, of rs678741 had the lowest frequency in the East Asian population **(Table 1)**.

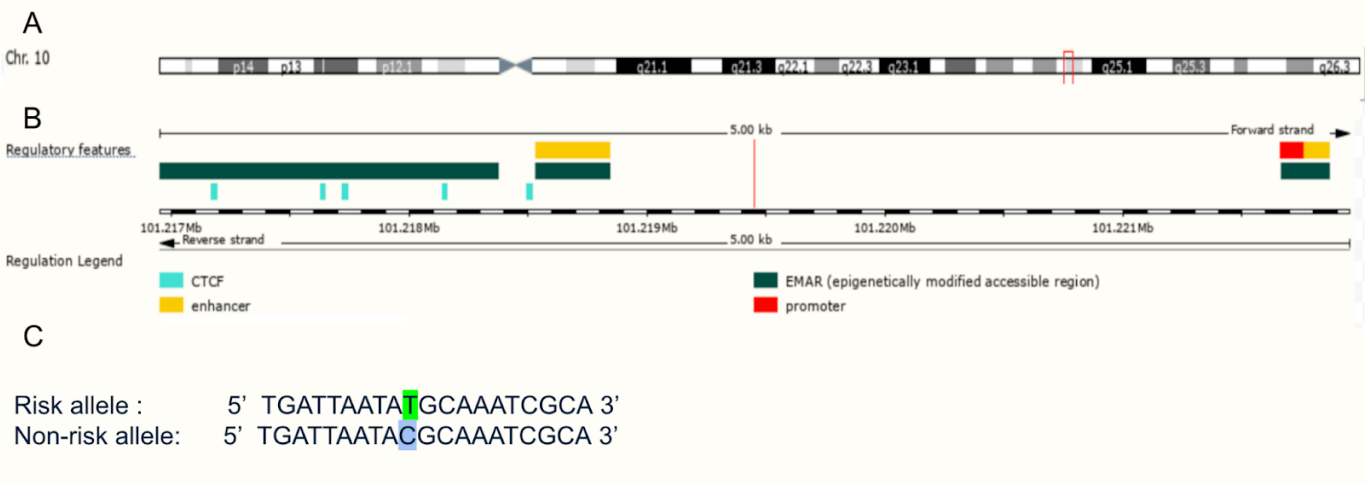

Figure 2: Gene context and flanking allele of rs11190870. A) Rs11190870's Location on Chromosome 10. B) Genomic context of rs11190870 (depicted by the red vertical line). This shows it is located in between an enhancer (yellow) and promoter (red). C) Non-risk allele (blue) and risk allele (green) for the rs111qqqq90870 SNP.

The genetic variant containing the highest association with AIS, rs11190870, was chosen to investigate further and to design a CRISPR guide for. Rs11190870 is located on chromosome 10 at position 10:101219450 **(Figure 1A)**. To visualize the genomic context of rs11190870, we used Ensembl and found that it is situated between an enhancer and promoter **(Figure 1B)**. We then used UCSC to gather both the risk allele and non-risk allele of the SNP and nine base pairs flanking it on both sides **(Figure 1C)**. The top risk allele, highlighted in green, was reported to have a higher association with AIS. The bottom sequence shows the non-risk allele, highlighted in blue, which has a lesser association with AIS.

Prime editing guide results

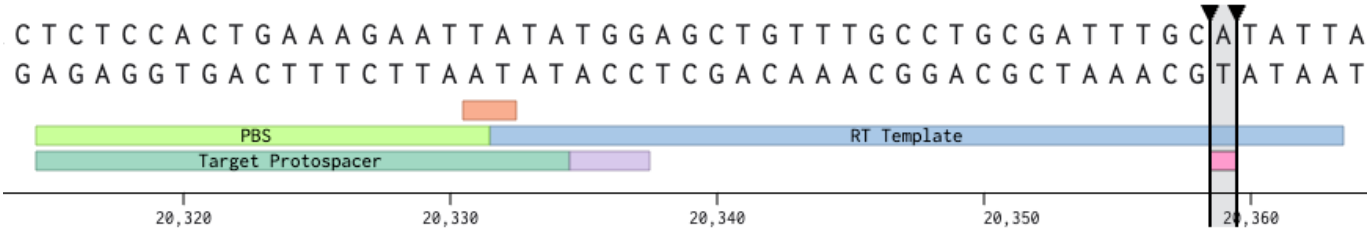

Figure 3. Prime editing complex in Benchling to target the rs1119870 site. The highlighted pink region is the SNP containing the risk allele T. The purple section is the PAM sequence (NGG), and it is located upstream of the edit. Directly upstream of the PAM sequence is the target protospacer, a 20-nucleotide sequence that recognizes and binds to the complementary DNA sequence. The orange region represents the nick site.

The five components of a prime editing complex – the spacer, scaffold, RTT, PBS, and tevopreQ1 motif – were visualized in Benchling **(Figure 3)**. We identified the PAM region for this SNP as TGG on the 5' strand. The nick site is between the T and the A **(Figure 3)** and is three bases upstream of the PAM site, and this is also where the Cas9 nickase makes a single-stranded cut. The PBS site of 5' TCTCCACTGAAAGAAT 3' is directly upstream of the nick, while the RTT site is directly downstream and contains the desired edit. The reverse transcriptase synthesizes the DNA from the nick towards the end and four nucleotides past the SNP. An extra single guide RNA can be utilized in the complex to increase editing efficiency because adding it triggers a DNA response that better incorporates the desired edit [35].

Table 2: Ranked sgRNA candidates in pegFinder

sgRNA sequence	Orientation	Distance to Edit (bp)	Rank
TCTCCACTGAAAGAATTATA	sense	27	1
CCCCGGCCCTCCAACACCAG	antisense	32	2
TTGGACCCCACACCCCGCCC	antisense	49	3
ATGTATTAATTTGCTTTAAC	sense	61	4
CAAACACTCCTTCACACCTT	antisense	68	5

The PegFinder website generated five potential sgRNAs to target the rs1119870 SNP and ranked them based on editing efficiency **(Table 2)**. The closer the sgRNA is to the edit site, the higher its ranking because previous studies found that the closer the distance, the better the gene targeting success [36]. The top-ranked sgRNA sequence is 27 base pairs away from the edit site in the sense direction.

Discussion

The association of the Ladybird Homeobox 1 (*LBX1*) gene to the development of adolescent idiopathic scoliosis (AIS) has been well established [18, 32, 33]. Therefore, SNP variants affecting this gene provide great candidates for pursuing CRISPR gene editing therapies. Given the association of *LBX1* and AIS, we aimed to 1) investigate the *LBX1* gene under a wild-type setting and determine its expression across human tissues, 2) characterize the genetics of 3 SNPs shown to be particularly important for the susceptibility, especially in East Asian populations, and 3) design a prime editing complex for the SNP with the highest association. Understanding the expression profile of the *LBX1* gene and identifying variants that are prominent within East Asian populations can help shed light on potential treatments for this particularly susceptible population.

Rs11190870 was chosen among our three variants, given its highest statistical significance with AIS in a meta-analysis of more than 30k individuals [18], for designing a prime editing CRISPR guide. Because rs11190870 is located in the regulatory region upstream of

LBX1 and is surrounded by an enhancer and promoter, having a disease variant in the regulatory region could have severe consequences on the gene expression and regulation [37].

Previous research has also confirmed the deleterious effects of this SNP. For example, Guo et al. found that HEK 293T cells containing the risk allele T produced higher than normal transcriptional activity in the region [21]. The overexpression of the *LBX1* gene caused body curvature in zebrafish embryos, suggesting that the T risk allele was associated with scoliosis development. This may occur because the upregulation of a gene can disrupt normal development of the musculoskeletal and nervous systems, increasing an individual's susceptibility to musculoskeletal disorders, including scoliosis. Therefore, by editing the SNP to the non-risk allele C, it decreases the chances of irregular spinal growth. This involves targeting the SNPs surrounding the *LBX1* gene, particularly the risk alleles, and genetically editing them to non-risk alleles. It is important to note that inheriting a risk allele for scoliosis is just one of many contributing risk factors combined with other genetic and environmental ones, so correcting the SNP to a non-risk allele will lessen the risk of developing AIS, not completely eradicate the disease [38].

CRISPR/Cas9 gene therapy has been an emerging tool to treat diseases that arise from SNPs. It was employed to treat sickle cell disease (SCD), which results from a single mutation in the sixth codon of the β-globin chain, leading to abnormal hemoglobin polymerization [39]. The treatment was a CRISPR-Cas9 therapy called CASGEVY, which precisely modifies hematopoietic stem cells in patients 12 and above, leading to a drastic decrease in side effects associated with SCD, such as severe pain, tissue damage, and organ damage [39]. The same principle used for CASGEVY can be utilized to create a gene therapy to treat AIS by targeting the associated SNPs.

Given the single-base substitution associated with AIS in our study, prime editing was chosen as the delivery mechanism as opposed to traditional CRISPR/Cas9 editing because of its ability to make precise edits at a base pair level while being less prone to off-target effects. Because CRISPR/Cas9 is better suited for editing small target sequences rather than at the base pair level, prime editing was chosen for our variant of interest, rs11190870, because of its broader range of editing capabilities, particularly point mutations. CRISPR/Cas9's reliance on double-stranded breaks also has a higher risk of off-target mutations, indels, translocations, and duplications compared to prime editing's usage of a single-stranded nick [27].

Besides prime editing, other well-suited CRISPR mechanisms were considered, such as base editing (BE). Base editing works similarly to prime editing as both use the Cas9 nickase to make single-stranded nicks instead of double-stranded breaks [40]. A limitation with base editing is that, within or surrounding some target sites, there can be multiple editable sites that the complex can mistakenly edit, producing off-target effects [41]. PE reduces the risk of off-target effects because it hybridizes with the DNA site three times instead of once to ensure the guideRNA is at the correct site [27]. Thus, PE remains the best mechanism.

Implementing the CRISPR-based prime editing therapy to edit the rs11190870 SNP requires considering multiple factors, including the stage of intervention, the site of delivery, the delivery mechanism, and ethical considerations. To start, the stage of intervention would have to be when the patient's spinal muscles are still growing because *LBX1* is the most transcriptionally active during the development of the musculoskeletal system **(Figure 1A)**, especially the embryonic stage [42]. Therefore, if scoliosis is known to run in a family's history and the patient carries the risk allele, T, the prime editing complex can be delivered either in an

embryonic stage or during childhood before skeletal maturity. Because embryonic cell editing is very restricted in many countries due to its ethical and safety concerns, including causing genetic mosaicism and high risk of off-target effects, it is unlikely to be an option for therapeutic use [43]. Thus, somatic editing during childhood before skeletal maturity remains the best option.

There are multiple delivery mechanisms for inserting a CRISPR system into *in vivo* human gene therapy. The delivery is composed of two key components [26]. The first is the cargo, which includes the Cas 9 and guide RNA components. The three most common cargo approaches are DNA plasmids, mRNA for translating the Cas 9 protein with a separate gRNA, and a ribonucleoprotein complex. The second is the delivery vehicle, which determines whether the complex can be delivered as DNA, mRNA, or a ribonucleoprotein complex [44]. Vehicles that are best suited for *in vivo* therapies are viral systems because of their ability to be used for a wide variety of cell types, high transfection efficiency, and lower rates of cytotoxicity compared to physical methods [45]. The most popular viral vectors are adeno-associated viruses (AAV), and other common ones include lentivirus and adenoviruses.

To select the best delivery mechanism for this scoliosis therapy, the usage of prime editing and the site of delivery, the erector spinae muscles, have to be considered. Although there are several non-viral delivery mechanisms for PE, like liposomes, exosomes, and polymer-based systems [46], many of these have drawbacks such as instability and off-target effects. The viral mechanism with the highest editing efficiency in muscle cells and that has success in treating muscle-related disease in animal models is AAV delivery [47]. Although the PE complex's full length is too large to fit within a single adeno-associated virus, there are methods to split and reassemble the full-length protein complex within dual AAVs for *in vivo* use [35].

While these strategies exist, it is essential to note the risks and challenges with somatic cell CRISPR prime editing with viral delivery. To start, off-target effects remain one of the key challenges to keep in mind. A gene-therapy case in 2003 engineered a retrovirus to administer the gene editing system into cells to treat severe combined immunodeficiency disease [48]. While it succeeded in curing nine out of eleven of the patients, the other two developed cancer because the retrovirus was inserted in or near a cancer-causing gene, causing uncontrolled cell growth. It is challenging to predict the downstream effects of a case where the CRISPR system is inserted in the incorrect location, highlighting the benefit of using PE, which triple-checks the location at which the pegRNA binds. When utilizing viruses as vectors, another risk to consider is the chances of an inflammatory response occurring, which has caused death before [26]. With genetic editing therapies, ethical considerations such as distributive justice, societal equity, and safety concerns should also be considered [39].

This paper presents a potential pegRNA complex that can be utilized to reduce the risk of developing AIS by editing the risk allele of the rs11190870 SNP, once genetic editing technology is more thoroughly understood and the safety of AAV and prime editing is ensured in somatic tissue cell editing. Before making this therapy a reality, the etiology of scoliosis still needs to be fully understood, as risk alleles are just one part of the intricate puzzle contributing to the disease, and just targeting the SNP alone is not enough to treat scoliosis at large. For instance, environmental factors play a significant role in the development of AIS and are not fully grasped yet. Additional clinical studies and ethical considerations surrounding the usage of

CRISPR-Cas/9 editing within humans also have to be undertaken before the guide can be developed.

Conclusion

The purpose of this research project was to identify SNPs in the regulatory regions of the *LBX1* that were highly associated with the etiology of AIS. Numerous studies reported rs11190870 as having one of the highest associations with the disease, making it the most logical SNP to target. A prime editing complex was designed to target the rs1190870 SNP to mitigate the overexpression of the *LBX1* gene, leading to irregular musculoskeletal development. Once the safety and validity of the therapy are verified through further research and experimentation, the complex can expand the treatment options for AIS beyond surgery.

References

[1] American Association of Neurological Surgeons, “Scoliosis.” Apr. 30, 2024. [Online]. Available: https://www.aans.org/patients/conditions-treatments/scoliosis/
[2] Cleveland Clinic, “Scoliosis.” The Cleveland Clinic Foundation, Jan. 26, 2024.
[3] F. Di Maria *et al.*, “Immediate Effects of Sforzesco® Bracing on Respiratory Function in Adolescents with Idiopathic Scoliosis,” *Healthcare*, vol. 9, no. 10, p. 1372, Oct. 2021, doi: 10.3390/healthcare9101372.
[4] H.-R. Weiss and D. Goodall, “Rate of complications in scoliosis surgery – a systematic review of the Pub Med literature,” *Scoliosis*, vol. 3, no. 1, p. 9, Dec. 2008, doi: 10.1186/1748-7161-3-9.
[5] F. Altaf, A. Gibson, Z. Dannawi, and H. Noordeen, “Adolescent idiopathic scoliosis,” *BMJ*, vol. 346, no. apr30 1, pp. f2508–f2508, Apr. 2013, doi: 10.1136/bmj.f2508.
[6] J. A. Janicki and B. Alman, “Scoliosis: Review of diagnosis and treatment,” *Paediatr. Child Health*, vol. 12, no. 9, pp. 771–776, Nov. 2007, doi: 10.1093/pch/12.9.771.
[7] M. R. Konieczny, H. Senyurt, and R. Krauspe, “Epidemiology of adolescent idiopathic scoliosis,” *J. Child. Orthop.*, vol. 7, no. 1, pp. 3–9, Feb. 2013, doi: 10.1007/s11832-012-0457-4.
[8] H. Guo *et al.*, “Ethnic Disparity in the Incidence of Scoliosis Among Adolescents in Tianzhu Tibetan Autonomous County, China,” *Front. Public Health*, vol. 10, p. 791550, Apr. 2022, doi: 10.3389/fpubh.2022.791550.
[9] H. Modi, S. Srinivasalu, S. Smehta, J.-H. Yang, H.-R. Song, and S. W. Suh, “Muscle imbalance in volleyball players initiates scoliosis in immature spines: a screening analysis,” *Asian Spine J.*, vol. 2, no. 1, pp. 38–43, June 2008, doi: 10.4184/asj.2008.2.1.38.
[10] N. S. Yan, “‘Butterfly Queen’ Zhang Yufei: Olympic Champion Battling Scoliosis Source URL : ‘Butterfly Queen’ Zhang Yufei: Olympic Champion Battling Scoliosis | The Standard https://www.thestandard.com.hk/health-and-wellness/article/66415/,” *The Standard*, Sept. 24, 2024. [Online]. Available: https://www.thestandard.com.hk/health-and-wellness/article/66415/Butterfly-Queen-Zhang-Yufei-Olympic-Champion-Battling-Scoliosis
[11] A. Grauers, I. Rahman, and P. Gerdhem, “Heritability of scoliosis,” *Eur. Spine J.*, vol. 21, no. 6, pp. 1069–1074, June 2012, doi: 10.1007/s00586-011-2074-1.
[12] Y. Wang *et al.*, “Role of differentially expressed LBX1 in Adolescent Idiopathic Scoliosis (AIS) paraspinal muscle phenotypes and muscle-bone crosstalk through modulating myoblasts,” in *Studies in Health Technology and Informatics*, X.-C. Liu and J. G. Thometz, Eds., IOS Press, 2021. doi: 10.3233/SHTI210425.
[13] S. Watanabe, S. Kondo, M. Hayasaka, and K. Hanaoka, “Functional analysis of homeodomain-containing transcription factor Lbx1 in satellite cells of mouse skeletal muscle,” *J. Cell Sci.*, vol. 120, no. 23, pp. 4178–4187, Dec. 2007, doi: 10.1242/jcs.011668.
[14] H. Brohmann, K. Jagla, and C. Birchmeier, “The role of *Lbx1* in migration of muscle precursor cells,” *Development*, vol. 127, no. 2, pp. 437–445, Jan. 2000, doi: 10.1242/dev.127.2.437.
[15] K. Cui *et al.*, “Genetic identification of medullary neurons underlying congenital hypoventilation,” *Sci. Adv.*, vol. 10, no. 25, p. eadj0720, June 2024, doi: 10.1126/sciadv.adj0720.
[16] L. Xu *et al.*, “A Functional SNP in the Promoter of LBX1 Is Associated With the Development of Adolescent Idiopathic Scoliosis Through Involvement in the Myogenesis of

Paraspinal Muscles," *Front. Cell Dev. Biol.*, vol. 9, p. 777890, Nov. 2021, doi: 10.3389/fcell.2021.777890.
[17] H. Jiang *et al.*, "Association of rs11190870 near LBX1 with adolescent idiopathic scoliosis susceptibility in a Han Chinese population," *Eur. Spine J.*, vol. 22, no. 2, pp. 282–286, Feb. 2013, doi: 10.1007/s00586-012-2532-4.
[18] Y. Cao, J. Min, Q. Zhang, H. Li, and H. Li, "Associations of LBX1 gene and adolescent idiopathic scoliosis susceptibility: a meta-analysis based on 34,626 subjects," *BMC Musculoskelet. Disord.*, vol. 17, p. 309, July 2016, doi: 10.1186/s12891-016-1139-z.
[19] L. Dai, Y. Deng, N. Li, L. Xie, M. Mao, and J. Zhu, "Discontinuous microduplications at chromosome 10q24.31 identified in a Chinese family with split hand and foot malformation," *BMC Med. Genet.*, vol. 14, no. 1, p. 45, Dec. 2013, doi: 10.1186/1471-2350-14-45.
[20] L. P. Bignold, "The cell-type-specificity of inherited predispositions to tumours: review and hypothesis," *Cancer Lett.*, vol. 216, no. 2, pp. 127–146, Dec. 2004, doi: 10.1016/j.canlet.2004.07.037.
[21] L. Guo *et al.*, "Functional Investigation of a Non-coding Variant Associated with Adolescent Idiopathic Scoliosis in Zebrafish: Elevated Expression of the Ladybird Homeobox Gene Causes Body Axis Deformation," *PLOS Genet.*, vol. 12, no. 1, p. e1005802, Jan. 2016, doi: 10.1371/journal.pgen.1005802.
[22] D. Mercatelli, L. Scalambra, L. Triboli, F. Ray, and F. M. Giorgi, "Gene regulatory network inference resources: A practical overview," *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, vol. 1863, no. 6, p. 194430, June 2020, doi: 10.1016/j.bbagrm.2019.194430.
[23] M. Luo, Y. Zhang, S. Huang, and Y. Song, "The Susceptibility and Potential Functions of the LBX1 Gene in Adolescent Idiopathic Scoliosis," *Front. Genet.*, vol. 11, p. 614984, Jan. 2021, doi: 10.3389/fgene.2020.614984.
[24] M. Asmamaw and B. Zawdie, "Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing," *Biol. Targets Ther.*, vol. Volume 15, pp. 353–361, Aug. 2021, doi: 10.2147/BTT.S326422.
[25] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity," *Science*, vol. 337, no. 6096, pp. 816–821, Aug. 2012, doi: 10.1126/science.1225829.
[26] C. A. Lino, J. C. Harper, J. P. Carney, and J. A. Timlin, "Delivering CRISPR: a review of the challenges and approaches," *Drug Deliv.*, vol. 25, no. 1, pp. 1234–1257, Jan. 2018, doi: 10.1080/10717544.2018.1474964.
[27] A. Ochoa-Sanchez *et al.*, "Prime Editing, a Novel Genome-Editing Tool That May Surpass Conventional CRISPR-Cas9," *ReGEN Open*, vol. 1, no. 1, pp. 75–82, June 2021, doi: 10.1089/regen.2021.0016.
[28] M. Uhlén *et al.*, "Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, p. 1260419, Jan. 2015, doi: 10.1126/science.1260419.
[29] M. Cerezo *et al.*, "The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity," *Nucleic Acids Res.*, vol. 53, no. D1, pp. D998–D1005, Jan. 2025, doi: 10.1093/nar/gkae1070.
[30] G. Perez *et al.*, "The UCSC Genome Browser database: 2025 update," *Nucleic Acids Res.*, vol. 53, no. D1, pp. D1243–D1249, Jan. 2025, doi: 10.1093/nar/gkae974.

[31] J. L. Doman, A. A. Sousa, P. B. Randolph, P. J. Chen, and D. R. Liu, “Designing and executing prime editing experiments in mammalian cells,” *Nat. Protoc.*, vol. 17, no. 11, pp. 2431–2468, Nov. 2022, doi: 10.1038/s41596-022-00724-4.
[32] R. D. Chow, J. S. Chen, J. Shen, and S. Chen, “A web tool for the design of prime-editing guide RNAs,” *Nat. Biomed. Eng.*, vol. 5, no. 2, pp. 190–194, Sept. 2020, doi: 10.1038/s41551-020-00622-8.
[33] I. Kou *et al.*, “Genome-wide association study identifies 14 previously unreported susceptibility loci for adolescent idiopathic scoliosis in Japanese,” *Nat. Commun.*, vol. 10, no. 1, p. 3685, Aug. 2019, doi: 10.1038/s41467-019-11596-w.
[34] Z. Zhu *et al.*, “Genome-wide association study identifies new susceptibility loci for adolescent idiopathic scoliosis in Chinese girls,” *Nat. Commun.*, vol. 6, no. 1, p. 8355, Sept. 2015, doi: 10.1038/ncomms9355.
[35] Z. Zhao, P. Shang, P. Mohanraju, and N. Geijsen, “Prime editing: advances and therapeutic applications,” *Trends Biotechnol.*, vol. 41, no. 8, pp. 1000–1012, Aug. 2023, doi: 10.1016/j.tibtech.2023.03.004.
[36] D. P. Santos, E. Kiskinis, K. Eggan, and F. T. Merkle, “Comprehensive Protocols for CRISPR/Cas9-based Gene Editing in Human Pluripotent Stem Cells,” *Curr. Protoc. Stem Cell Biol.*, vol. 38, p. 5B.6.1-5B.6.60, Aug. 2016, doi: 10.1002/cpsc.15.
[37] S. G. Coetzee, S. Pierce, P. Brundin, L. Brundin, D. J. Hazelett, and G. A. Coetzee, “Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson’s disease etiology,” *Sci. Rep.*, vol. 6, no. 1, p. 30509, July 2016, doi: 10.1038/srep30509.
[38] T. Cheng, E. Einarsdottir, J. Kere, and P. Gerdhem, “Idiopathic scoliosis: a systematic review and meta-analysis of heritability,” *EFORT Open Rev.*, vol. 7, no. 6, pp. 414–421, June 2022, doi: 10.1530/EOR-22-0026.
[39] A. Singh, H. Irfan, E. Fatima, Z. Nazir, A. Verma, and A. Akilimali, “Revolutionary breakthrough: FDA approves CASGEVY, the first CRISPR/Cas9 gene therapy for sickle cell disease,” *Ann. Med. Surg.*, vol. 86, no. 8, pp. 4555–4559, Aug. 2024, doi: 10.1097/MS9.0000000000002146.
[40] E. M. Porto, A. C. Komor, I. M. Slaymaker, and G. W. Yeo, “Base editing: advances and therapeutic opportunities,” *Nat. Rev. Drug Discov.*, vol. 19, no. 12, pp. 839–859, Dec. 2020, doi: 10.1038/s41573-020-0084-6.
[41] H. A. Rees and D. R. Liu, “Base editing: precision chemistry on the genome and transcriptome of living cells,” *Nat. Rev. Genet.*, vol. 19, no. 12, pp. 770–788, Dec. 2018, doi: 10.1038/s41576-018-0059-1.
[42] W. Jennings *et al.*, “Paraspinal muscle ladybird homeobox 1 (LBX1) in adolescent idiopathic scoliosis: a cross-sectional study,” *Spine J.*, vol. 19, no. 12, pp. 1911–1916, Dec. 2019, doi: 10.1016/j.spinee.2019.06.014.
[43] A. Plaza Reyes and F. Lanner, “Towards a CRISPR view of early human development: applications, limitations and ethical concerns of genome editing in human embryos,” *Development*, vol. 144, no. 1, pp. 3–7, Jan. 2017, doi: 10.1242/dev.139683.
[44] R. Roberts, “CRISPR Delivery Methods: Cargo, Vehicles, and Challenges,” *Synthego*, July 2025, [Online]. Available: https://www.synthego.com/blog/delivery-crispr-cas9
[45] S. Demirci, K. Essawi, P. Germino-Watnick, X. Liu, W. Hakami, and J. F. Tisdale, “Advances in CRISPR Delivery Methods: Perspectives and Challenges,” *CRISPR J.*, vol. 5, no. 5, pp. 660–676, Oct. 2022, doi: 10.1089/crispr.2022.0051.

[46] E. Ivanova, “How Various Drug Delivery Methods Could Aid in the Translation of Genome Prime Editing Technologies,” *Genet. Res.*, vol. 2022, p. 7301825, 2022, doi: 10.1155/2022/7301825.
[47] M. Tabebordbar *et al.*, “In vivo gene editing in dystrophic mouse muscle and muscle stem cells,” *Science*, vol. 351, no. 6271, pp. 407–411, Jan. 2016, doi: 10.1126/science.aad5177.
[48] J. Kaiser, “Seeking the Cause of Induced Leukemias in X-SCID Trial,” *Science*, vol. 299, no. 5606, pp. 495–495, Jan. 2003, doi: 10.1126/science.299.5606.495.