Molecular Therapy
Nucleic Acids
Original Article

CRISPR/Cas9-Directed Reassignment of the GATA1 Initiation Codon in K562 Cells to Recapitulate AML in Down Syndrome

Kevin M. Bloh,[1,2] Pawel A. Bialk,[1] Anilkumar Gopalakrishnapillai,[2] E. Anders Kolb,[2] and Eric B. Kmiec[1]

[1]Gene Editing Institute, Helen F. Graham Cancer Center & Research Institute, Christiana Care Health Services, Inc., Newark, DE 19713, USA; [2]Nemours Center for Childhood Cancer Research, Alfred I. duPont Hospital for Children, Wilmington, DE 19803, USA

Using a CRISPR/Cas9 system, we have reengineered a translational start site in the *GATA1* gene in K562 cells. This mutation accounts largely for the onset of myeloid leukemia in Down syndrome (ML-DS). For this reengineering, we utilized CRISPR/Cas9 to generate mammalian cell lines that express truncated versions of the Gata1s protein similar to that seen in ML-DS, as determined by analyzing specific genetic alterations resulting from CRISPR/Cas9 cleavage. During this work, 73 cell lines were clonally expanded, with allelic variance analyzed. Using Tracking of Indels by DEcomposition (TIDE) and Sanger sequencing, we defined the DNA sequence and variations within each allele. We found significant heterogeneity between alleles in the same clonally expanded cell, as well as among alleles from other clonal expansions. Our data demonstrate and highlight the importance of the randomness of resection promoted by non-homologous end joining after CRISPR/Cas9 cleavage in cells undergoing genetic reengineering. Such heterogeneity must be fully characterized to predict altered functionality inside target tissues and to accurately interpret the associated phenotype. Our data suggest that in cases where the objective is to rearrange specific nucleotides to redirect gene expression in human cells, it is imperative to analyze genetic composition at the individual allelic level.

INTRODUCTION

Down syndrome (DS), trisomy 21, is the most common chromosomal abnormality in childhood and occurs in 1 in 1,000 births. Among several physical and cognitive abnormalities, children with DS also have a 10- to 20-fold increased risk of leukemia relative to other children.[1] Additionally, children with DS who are younger than 4 years of age have a 500-fold increased risk for acute myeloid leukemia (ML-DS).[2] Nearly half of all diagnosed ML-DS cases are of the acute megakaryoblastic (AMKL) phenotype. The driver events in ML-DS have long appeared to be unique to patients with DS.[3] Wechsler et al.[4] reported that nearly all patients with ML-DS share mutations in the essential hematopoietic transcription factor, *GATA1*, located on the X chromosome. Although the specific mutations may vary, all occur in exon 2, resulting in utilization of an alternate exon 3 initiation site. The functional but truncated protein, Gata1s, has a deficient or absent amino-terminal transactivation domain.[5,6]

Nearly all children with DS-AML diagnosed before age 4 years have an exon 2 *GATA1* mutation. Although Gata1s is sufficient to cause myelodysplasia, progression to DS-AML requires additional mutations, commonly in the cohesin components, *CTCF*, epigenetic regulators, JAK family kinases, *MPL*, *SH2B3*, and RAS pathway components. Trisomy 21, Gata1s cells are dysplastic and mimic leukemic phenotypes, but are not necessarily leukemic.[7,8] Acquisition of a subsequent leukemogenic mutation is still needed for the malignant transformation.

CRISPR/Cas9 causes a blunt-end double-strand break (DSB) 3 bp upstream of a PAM sequence (NGG in *Streptococcus pyogenes* Cas9, used in this paper).[9,10] This DSB is repaired via the natural pathway of non-homologous end joining (NHEJ), involving resection of individual free ends in a 3′ to 5′ direction and patching of the break site via alignment and polymerization of microhomology sequences.[11] The effects of CRISPR/Cas9-mediated resection is stochastic, with resection patterns ranging from deletions and insertions (indels) 1 bp to hundreds in length.[12–14] This variety of gene-editing products is useful in the ablation of whole-gene function, as is the case in creating genetic knockouts.[15] The clinically relevant molecular restructuring of ML-DS presents a unique challenge for gene-editing technologies. Instead of the widely used, broad-based strategies of gene knockout or forced insertion, this genetic engineering problem requires a more surgical approach. Here, we use the precise gene-targeting capabilities of CRISPR/Cas9 to surgically modify the *GATA1* gene and alter gene function.[16–20] Herein, we describe the successful creation of Gata1s-expressing cells by utilizing a CRISPR/Cas9 system that causes a DSB 2 bp upstream of the Met1 initiation site. We also provide detailed analyses for tracking

Received 16 March 2017; accepted 9 April 2017;
http://dx.doi.org/10.1016/j.omtn.2017.04.009.

Correspondence: Eric B. Kmiec, Gene Editing Institute, Helen F. Graham Cancer Center & Research Institute, Christiana Care Health Services, Inc., Newark, DE 19713, USA.
E-mail: eric.b.kmiec@christianacare.org

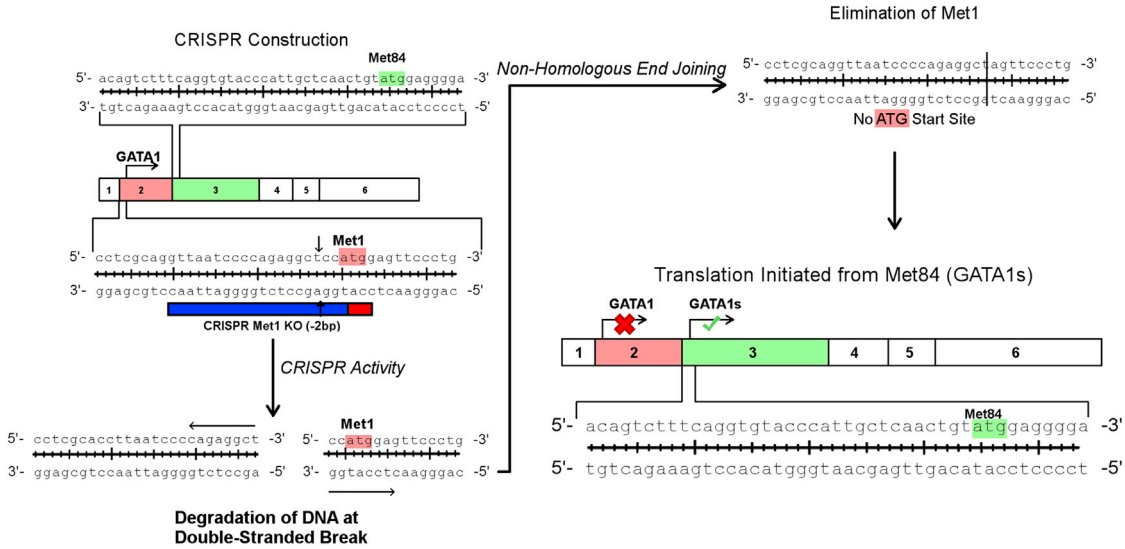


Figure 1. Diagram of Experimental Workflow

A CRISPR/Cas9 system was designed using the MIT CRISPR Design Tool (http://crispr.mit.edu) to cleave the *GATA1* sequence 2 bp upstream of the Met1 initiation codon. These blunt ends will then be resected and reannealed via the non-homologous end joining pathway (NHEJ). If resection occurs across the Met1 initiation site in both alleles, this will cause exclusive production of Gata1s.

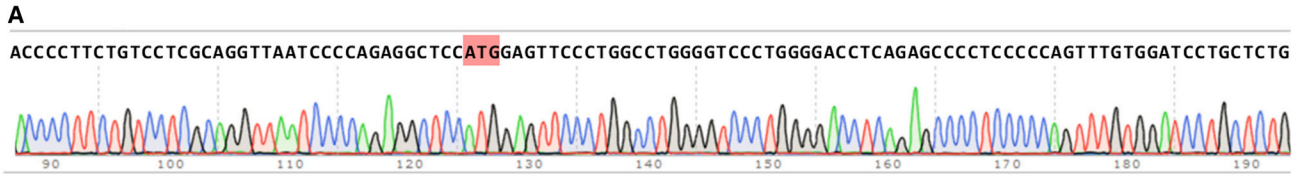
allelic variance after CRISPR/Cas9 editing of exon 2 of *GATA1* in human cells.

RESULTS

The goal of this work is to alter the normal translation start site of the GATA1 gene, thereby creating a truncated Gata1 variant known as Gata1s. The strategy is to use a CRISPR/Cas9 system to disable an ATG sequence at the normal Met1 translation start site of the gene. This action will enable a second initiation codon to be utilized, one located 84 bases downstream in exon 3 (Met 84). Initiation of translation from a transcript starting at Met 84 will produce the truncated protein variant known as Gata1s. CRISPR/Cas9 activity induces DNA cleavage near the Met1 site and DNA resection through the NHEJ pathway, eliminating the Met1 sequence. This was accomplished through the analysis of two individual transfections performed in K562 cells, producing 74 total clonally expanded populations.

Figure 1 outlines our experimental strategy and rationale for the mutagenesis of the first translation initiation codon. As seen in the diagram, DNA breakage was targeted at a position two bases upstream (−2) from the targeted Met 1 site. Because NHEJ occurs on free ends of DSB sites, the adjacency of the CRISPR/Cas9 cleavage to the Met1 site is vital to appropriate indel formation and low off-target effects. Because the underlying DNA sequence is critical for designing active CRISPR/Cas9, we sequenced 462 bases across the *GATA1* gene within the region of interest. The DNA sequence is presented in Figure 2A, and the target ATG sequence is shaded. By integrating this information into the analytical CRISPR Design Tool website (http://crispr.mit.edu), we generated several possibilities for the CRISPR/Cas9 system (Figures 2B and 2C). In addition, information was also gathered that defined a ranking of potential off-target sites for CRISPR/Cas9. This ranking is based on the equation published by Hsu et al.,[21] which utilizes as factors: (1) weighted combination of mismatches, (2) location of these sequences within genome (sequences located within exonic regions of genes are weighted more heavily than those located within intronic sequences), and (3) the steric hindrance created by individual and consecutive mismatches at their specific locations within the overall 20 bp sequence. The off-site target score, out of 100, and the specific locus are also highlighted in Figure 2C. A CRISPR/Cas9 complex that will induce DNA cleavage at the −2 bp position relative to the Met1 site was chosen because it had the lowest potential for off-site activity, yet would induce a DNA break closest to the target site.

The CRISPR/Cas9 system, expressed from plasmid pX458, which also carries a single copy of a wild-type EGFP gene, was introduced into K562 cells by lipofection.[22] The experimental design is illustrated in Figure 3A, with the ultimate analysis of sequence alteration confirmed at both the genotypic and the phenotypic levels. After 72 hr, the cells were sorted for EGFP expression using a FACSAria, and single cells scoring $EGFP^+$ were placed into individual wells (Figures 3B and 3C). Individual cells were allowed to expand for a period of 3–6 weeks, after which time the clonal expansions were isolated for DNA and functional protein analyses. Figure 3D exhibits a representative Sanger sequence trace file, demonstrating the breakdown of the peaks directly downstream of the CRISPR/Cas9 complex cut site. Successful DNA cleavage of the *GATA1* gene would be seen if the DNA sequence upstream of the cut site appears contiguous without sequence breakdown. In contrast, the DNA sequence downstream would be reflected as multiple peaks at the same position, although the spacing of the peaks should remain constant. These predictions are fulfilled and



B

C

Top 5 Exonic Off-target Sequences	Mismatch	Off-target Score	Locus
5'-GTTAATCCCCAGAGGCTCCATGG-3'			
5'-GTTAGTCCCCAGAGCCTCCATGG-3'	2	2.3	chr12:+19440459
5'-GCTGAGCCCCAGAGGCTCCATGG-3'	3	1.5	chr1:+144300931
5'-GCTGAGCCCCAGAGGCTCCATGG-3'	3	1.5	chr1:-149616341
5'-GTCAGTCCCCAGAGGCTCCATGG-3'	3	1.4	chr11:-66612265
5'-GCCGATCCTCAGAGGCTCCATGG-3'	4	0.8	chr7:-138520564

Figure 2. Verification of K562 *GATA1* Sequence and CRISPR/Cas9 Design for Elimination of Met1

(A) The Met1 initiation site in K562 cells was verified by PCR prior to CRISPR design. (B) Once the sequence was verified, a CRISPR seed sequence was designed, which allowed for a DSB to be caused 2 bp upstream of the Met1 initiation codon. (C) This seed sequence was examined for potential off-target sites using the standard equation for off-target efficiency of CRISPR seed sequences.[21] The top five exonic off-targets are displayed with mismatches from the seed sequences highlighted in orange. Percentage off-target efficiency scores are also shown.

displayed in Figure 3D, indicating that specific cleavage by CRISPR/Cas9 at the designated Met1 site had occurred. Trace files of Sanger sequencing obtained from individual clones can be analyzed by the software program Tracking of Indels by Decomposition (TIDE), to determine the individual sequences within the multi-peaked breakdown product after CRISPR/Cas9 activity.[23] TIDE utilizes an algorithm that isolates individual sub-sequences determined by similar sub-peak intensity within the breakdown sequence. By aligning each of the sub-sequences with a control parental band, the indel pattern of individual alleles within the clonal population can be determined and exhibited in a simple format. This software allows for the qualitative analysis of individual clonal populations and is a useful intermediate step in overall analysis of the cleavage profile in each clonal population.

Figure 4 displays the results of Sanger sequencing of several individual clonally expanded populations. Figures 4A–4G exhibit the variety of altered genomes obtained from clonal populations expanded from individual cells as a result of the CRISPR/Cas9 cleavage activity. Letter designations A–G in Figures 4, 5, 6, and 7 correspond to various analyses of the same clonal population of CRISPR/Cas9-edited cells (i.e., Figures 4A, 5A, 6A, and 7A all correspond to analysis from the same expanded population, clone A). Each clonal population is presented along with its corresponding TIDE analysis, showing a different indel profile with varying lengths on either or both alleles. The percentage of the mixed-peak sequence falling into each individual sub-sequence is listed above each bar. In our hands and others, CRISPR/Cas9 systems are often efficient enough to execute site-specific DNA cleavage in a biallelic fashion, rarely showing activity on only one allele.[13,24,25] Figure 4A illustrates the activity of the CRISPR/Cas9 system on a clonal population where one *GATA1* allele was resected 6 bp, with the other allele resected 4 bp. Figure 4B displays another distinct CRISPR cleavage event, exhibiting one allele apparently uncut, whereas the other allele displays a 22 bp deletion. Positioned adjacent to each TIDE readout is the analysis of sequence decomposition, showing the decomposition of the single-peak sequence in relation to the CRISPR/Cas9 cut site (indicated by a blue dotted line). The increase in sequence breakdown prior to the cut site indicates that resection after CRISPR/Cas9-mediated cleavage occurred both upstream and downstream of the cut site.

Figure 4C displays another example of biallelic CRISPR activity, with one allele being resected 4 bp and the other allele 38 bp. This example also reveals signal decomposition prior to the selected cut site, indicating that resection of the 38 bp deletion occurred partially upstream of the cleavage site (the decomposition extends too far upstream of the cut site to be caused solely by the 4 bp deletion). A similar indel pattern can be observed in Figure 4D, with both alleles incurring a 5 bp deletion and a 12 bp deletion. Figure 4E displays a clone containing only one deletion that could accurately be detected with the TIDE software. The maximum indel size that can be distinguished by the software is 50 bp, indicating that the second allele contains a deletion larger than 50 bp, with the higher background levels and lower R^2 value being attributed to the TIDE analysis's attempts to align the second allele to the parental sequence. Finally, Figures 4F and 4G both display heterozygosity, with the resection of the affected allele occurring partially upstream of the cut site; Figure 4F contains one 12 bp deletion, and Figure 4G contains a 22 bp deletion.

The data given by TIDE analysis are useful to intermediate data analysis, especially when evaluating clonally expanded populations. However, as seen in Figure 5, utilizing TIDE as a terminal analytical tool in

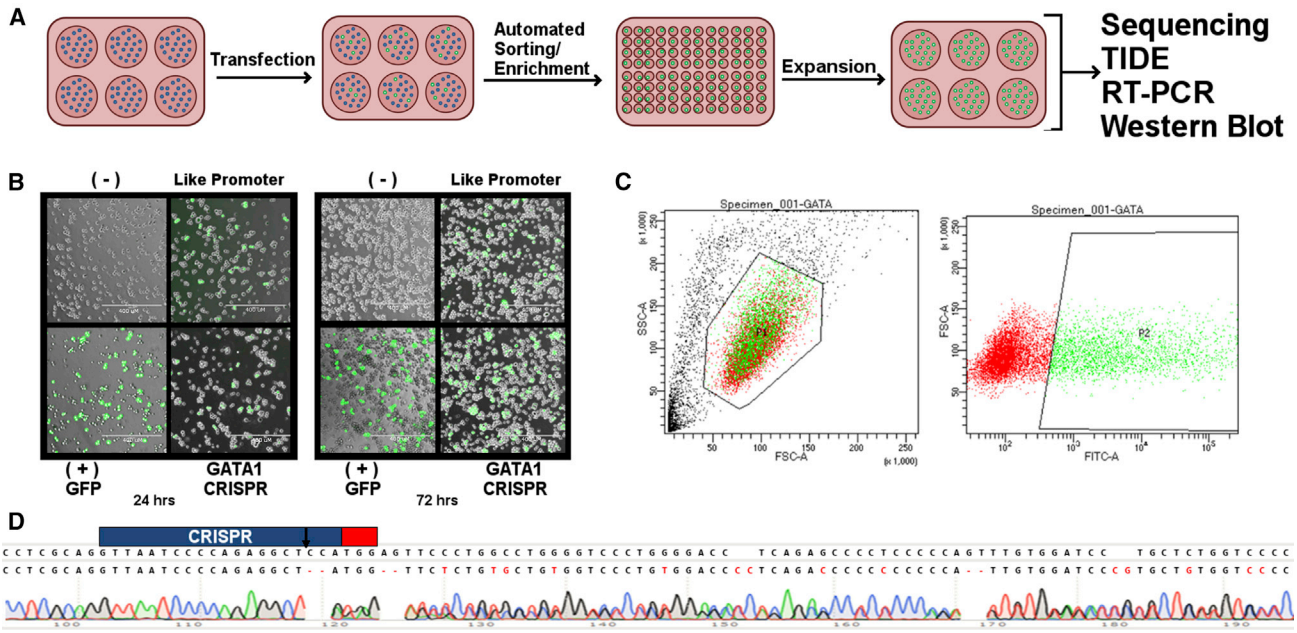


Figure 3. Effective Transfection and Enrichment of CRISPR/Cas9-Affected K562 Cells via Single-Cell Sorting

(A) The workflow of the experiment is displayed. Seventy-two hours post-lipofection, cells were sorted into individual wells of a 96-well dish using the FACSAriaII. Individual clones were then expanded from individual wells until able to be analyzed and cultured under normal conditions. (B) Overall representative images exhibiting the relative transfection efficiencies of the experimental cells in relation to the positive GFP control, under a CMV promoter, and the Like Promoter control, which expressed GFP under a CAG enhancer. (C) FACSAriaII readout of experimental K562s. (D) Representative example of signal decomposition downstream of CRISPR/Cas9 cut site.

data analysis can lead to an incomplete analysis of the obtained data. Figure 5 demonstrates the further allelic analysis performed using TIDE data. Because resection can occur both upstream and downstream of the cut site, an overall deletion of three or more bases does not implicitly demonstrate that the Met1 site was ablated in that allele; only resection of three or more bases downstream of the cut site is necessary for Met1 elimination. This situation is in fact observed in Figure 5A (the individual populations examined in Figures 4A–4G correspond with the samples presented in Figures 5A–5G). TIDE analysis for Figure 5A shows that a 4 bp and a 6 bp deletion occurred in both alleles of the population. This is verified when using the TIDE data to manually analyze the Sanger trace file. By determining which two bases exist at each position in the sequence, the actual, exact sequence of each allele can be accurately extrapolated. This is an example where, even though the TIDE analysis alone shows that both alleles showed a deletion greater than 3 bp, one allele exhibited resection entirely upstream of the cut site, leaving the Met1 site untouched.

This disparity between TIDE readout and primary data is shown more vividly in Figure 5B. The TIDE analysis determined that one allele has a 22 bp deletion, whereas the other allele is untouched. Examining the trace file, however, reveals that the allele determined as unbroken in fact contains a single thymine insertion directly at the cut site, leaving the Met1 site intact. The second allele of clone B also displays a different breakage pattern; the allele contains both a 26 bp deletion combined with a 5 bp insertion (labeled in orange) that eliminates the Met1 site. Clone C, upon initial allelic analysis, exhibited a potential example of a biallelic Met1 knockout. Allele 1 showed a 3 bp deletion, appearing to eliminate a portion of the Met1 site. Allele 2, in comparison, contains a 36 bp deletion, removing a large section of exon 2 including the Met1 site. This genetic restructuring, however, was later determined to be incorrect, with allele 1 containing a 2 bp deletion that retained the Met1 site. Figure 5D shows an example where both alleles contain an insertion and a deletion event. Allele 1 contains a 13 bp deletion and a single thymine insertion. However, resection occurred entirely upstream of the cut site, leaving the Met1 site intact. The second allele displays a deleted Met1 site, with another 13 bp deletion, this time paired with a 6 bp insertion. Examining the trace file in Figure 5E allows for an accurate determination of the length of the deletion that could not be analyzed with TIDE. This allele contains a 220 bp deletion, eliminating both Met1 and a majority of exon 2. The other allele contains an intact Met1, with a 5 bp deletion upstream of the cut site, combined with a single thymine insertion. Figures 5F and 5G both show similar sequence decomposition, with one allele with a single thymine insertion and the other allele containing a deletion. Figure 5F contains an 11 bp deletion on the second allele, but leaves the Met1 site intact. Figure 5G, on the other hand, shows ablation of the Met1 site on its second allele, with a 21 bp deletion both upstream and downstream of the cut site. In both of these examples, the TIDE analysis of the sequence determined that one allele remained wild-type. However, further analysis shows that both alleles actually

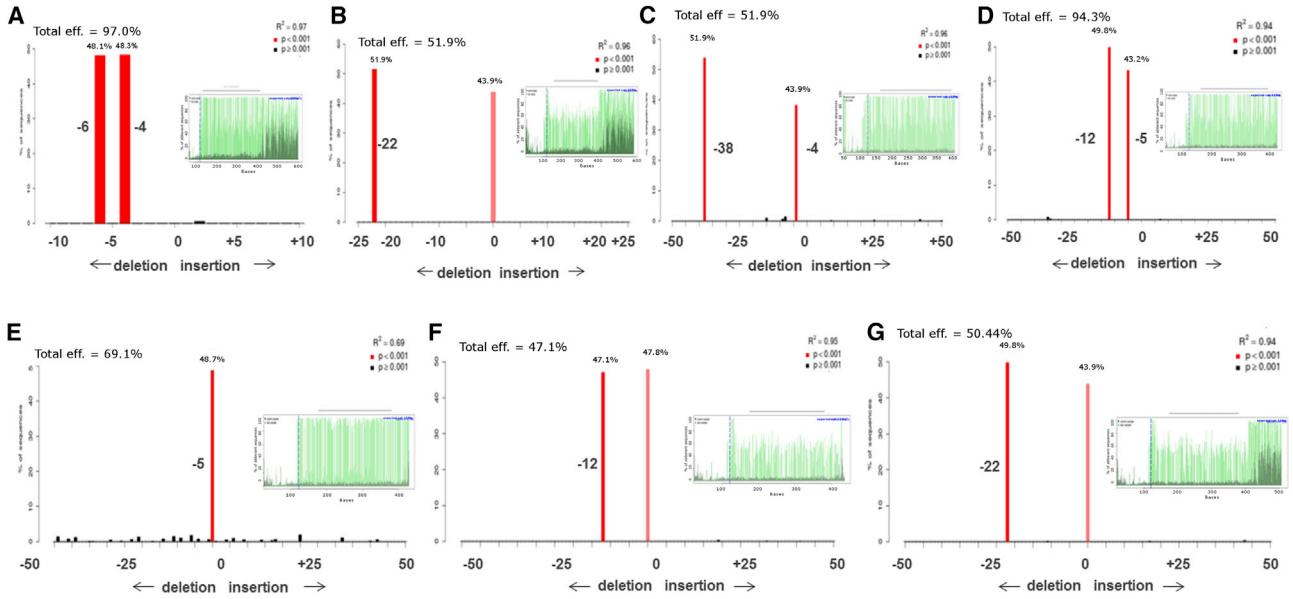


Figure 4. CRISPR/Cas9 Treatment Causes Heterogeneity of Products through Random Resection Lengths via Non-homologous End Joining

Each module depicts the Tracking of Indels by DEcomposition (TIDE) analysis of an individual clonal population, containing two *GATA1* alleles showing independent CRISPR/Cas9-mediated resection. Each bar graph depicts the identity of the indel event on each allele, with an estimation of the percentage of the population exhibiting each decomposition sequence above. Each graph also displays the increase in sequence aberrations in relation to the CRISPR/Cas9 cut site, as shown to be an increase in signal decomposition (in green) in relation to the DSB site (blue dotted line). Gray bars above the aberrant sequence analysis depict the portion of the aberrant sequence used to generate the identity of each indel profile.

contain single thymine insertions. Thymine insertion is seen in every +1 insertion event analyzed across all 73 isolated clones. The individual allelic analysis of each of the 26 clonally expanded populations is shown in Figure S1.

The K562 cell line karyotype demonstrates the presence of two X chromosomes.[26] The wild-type *GATA1* gene transcript is modified into two splice variants in wild-type K562 cells: one splice variant, including exon 2, encodes full-length Gata1; and the other, with exon 2 excluded, encodes Gata1s.[27,28] Because of this, wild-type K562 cells show expression of both long-form and short-form *GATA1* mRNA (Figure 6), with levels of expression weighted toward the short-form mRNA. Clones A and E exhibit severe knockdown, although not complete elimination, of full-length *GATA1* mRNA expression, exhibiting highly weighted expression of the Gata1s-encoding mRNA splice variant. The CRISPR/Cas9-mediated resection shows less effect on the expression profile of clones B, C, and G, which exhibit transcription profiles similar to wild-type K562 cells. Clones D and F display similar expression profiles, with relatively equal levels of long-form and short-form transcripts. These variants were sequenced in order to determine which allele is being expressed in each clonally expanded population (Figure S2). Sequences obtained from each sample determined that each short-form *GATA1* mRNA sequence encoded an identical Gata1s splice variant, lacking exon 2.

The protein expression levels of Gata1 and Gata1s, analyzed via western blot, are displayed in Figure 7. In wild-type K562 cells, the mRNA expression profile is weighted toward expressing short-form mRNA (Figure 6). The protein expression, in comparison, is skewed heavily toward full-length Gata1. This demonstrates that in K562 cells, the full-length *GATA1* mRNA is translated at a higher efficiency than the short-form splice variant.[27] Clone F shows a similar characteristic; although both variants exhibit at comparable amounts at the mRNA level, the protein expression profile is heavily weighted toward full-length Gata1 expression, with very little Gata1s formed. Another example is observed in data from clones A and E; although there is virtually no long-form mRNA expressed in these populations, the protein levels of Gata1 and Gata1s are equal. Clone B, in comparison, maintains its mRNA expression profile at the protein level; both profiles are weighted toward the short-form transcript and Gata1s, respectively. This is also seen in clone D, where equal levels of Gata1 and Gata1s are observed. Finally, clones C and G exhibit high knockdown of full-length Gata1, although their transcription profiles display the presence of low levels of *GATA1* mRNA.

Figure 8 displays the allelic analysis and protein expression of two important clones, generated from a second transfection under the same protocol, with the TIDE and allelic analysis of clones 67 and 72 (Figure 8A). The DNA sequence of each clone reveals an elimination of the *GATA1* Met1 initiation site. Clone 67 displays a 36 bp deletion in allele 1 and a 34 bp deletion in allele 2. Again, the discrepancy in deletion length with the TIDE results is expected, as the spacing of residue outputs of the Sanger sequencing was not constant. Clone 72 shows a 20 bp deletion in allele 1 and an 18 bp deletion in allele 2.

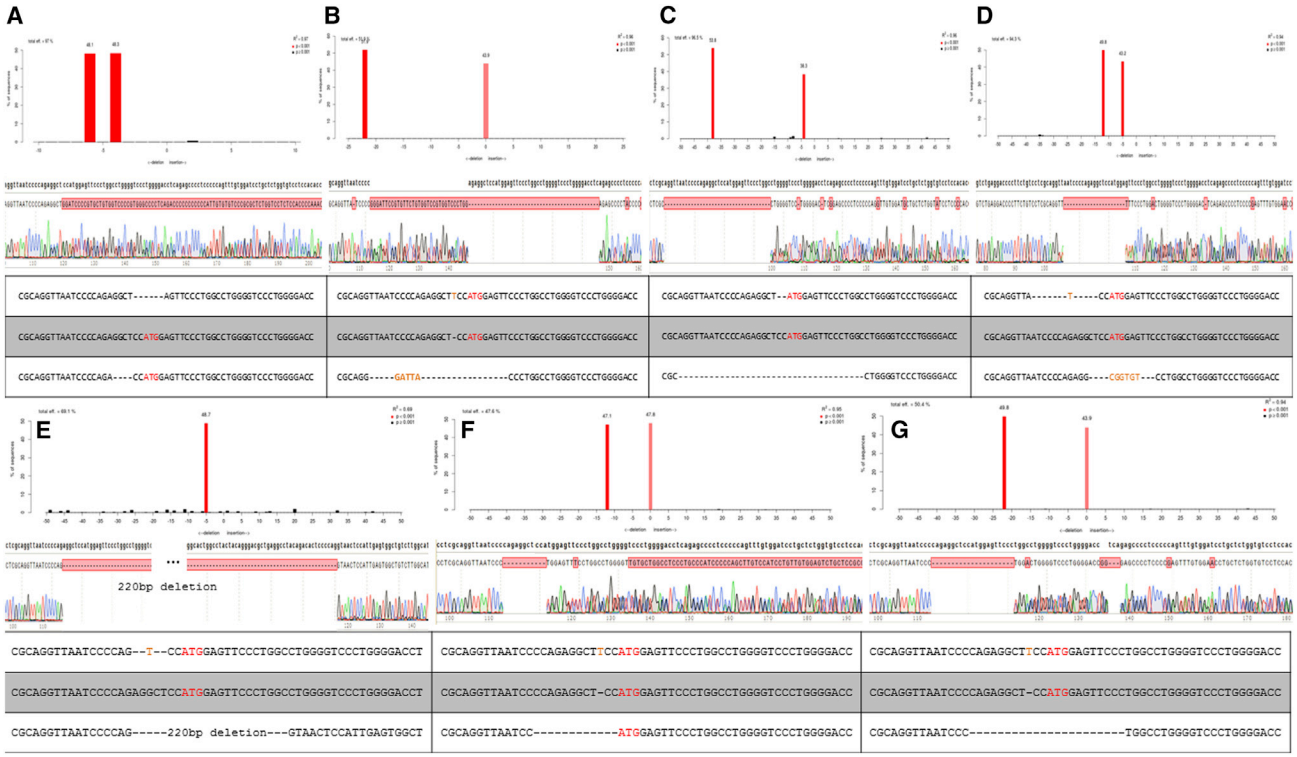


Figure 5. Allelic Analysis Using TIDE Readout Allows for Examination of NHEJ Resection Patterns at the Level of Individual Alleles

Examining the sequence trace file using the TIDE analysis as a guideline allows for the accurate determination of individual allelic resection patterns in samples with two alleles. Each clone from Figure 4 underwent allelic analysis, with the results shown above. Each panel displays TIDE analysis readout, chromatogram sequence aligned with the WT *GATA1* sequence, and allelic analysis readout. Wild-type sequence is displayed highlighted in gray, with the two allelic sequences above and below. The Met1 site is highlighted in red, and insertions are highlighted in orange.

Because spacing of the Sanger sequence was constant, the TIDE results aligned with the actual results of allelic analysis. The top panel in Figure 8B displays the complete ablation of long-form Gata1 in a patient sample from a child diagnosed with ML-DS, whereas the bottom panel shows complete ablation of long-form Gata1 and shift of expression to Gata1s in clones 67 and 72 (normalized values of full-length Gata1: 0.0055 and 0.051, respectively). This result is verified in Figure 8C. Expression of full-length Gata1 and Gata1s was measured by densitometry and normalized to the actin levels of their loading control. The full-length *GATA1* expression of both clones fell within background levels, indicating complete abrogation of the full-length gene product. Further analysis of these Gata1s clones revealed an increase in expression of Gata1s protein. Clone 67 displayed a nearly 10-fold increase in Gata1s expression compared with wild-type, and clone 72 displayed a 4-fold increase.

DISCUSSION

A mutation in the *GATA1* gene is observed in almost all cases of ML-DS diagnosed before 4 years of age.[29,30] A particular class of mutations that account for a majority of these cases results in the exclusive or near-exclusive expression of the Gata1s splice variant of the protein, lacking the first 83 aa of the full-length protein.[5] The main objective of this work was to develop a model cell line for the analysis of downstream genetic instability triggered by the ablation of full-length *GATA1* expression. The intrinsic pathway that leads to the knockout ability of CRISPR/Cas9 is NHEJ, a stochastic process. As such, using a specifically designed CRISPR/Cas9 system can lead to a multitude of unique products of DNA resection at the DSB site.[12,13,25,31] Because of this, analysis of genetically modified clones on a cell-by-cell basis is necessary in order to verify products expressing the appropriate Met1 deletion. Through this process, we were successful in the creation of this genetically altered cell line.

The efficiency of CRISPR/Cas9 comes from the sequence of the sgRNA component of the complex. As such, sgRNA sequences must have a perfect match to the sequence at the target site, with the DSB as close to the site of change as possible. To minimize the potential of the CRISPR/Cas9 to cleave off-target sites, we utilized the CRISPR Design Tool (http://crispr.mit.edu/). The protospacer used in this experiment (5′-GTTAATCCCCAGAGGCTCCATGG-3′) was used with an off-target score of 63. Perfect on-target homology results in a score of 100, with the introduction of mismatches and their location in the seed sequence relative to each other, and the sequence as a whole returning a score lower than 100. These

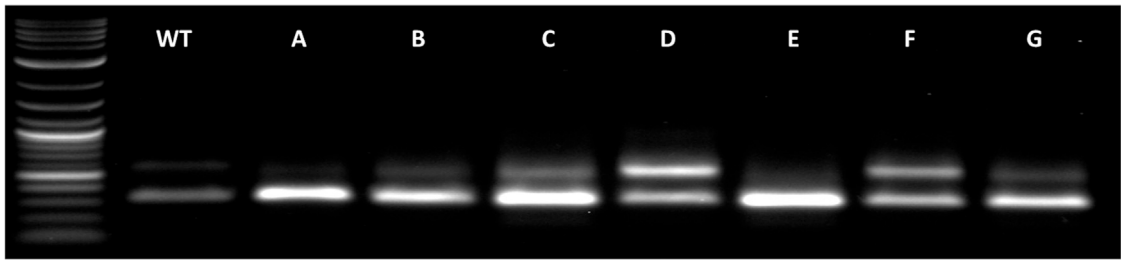


Figure 6. CRISPR/Cas9-Mediated Resection at the *GATA1* Met1 Initiation Site Allows for the Expression of Short-Form mRNA via Disruption of the Intron 1 to Exon 2 Splice Site

RT-PCR was performed on whole-cell cDNA in order to obtain an accurate determination of the effect each CRISPR/Cas9-mediated editing event had on *GATA1* transcription. Short-form mRNA, indicated by the mRNA amplicons of length 325 bp, will encode only Gata1s, as the entirety of exon 2 is excised. Long-form mRNA encodes either Gata1 or Gata1s, contingent on the presence or absence of an unresected Met1 site, respectively.

individual scores are then aggregated to form the total off-target score. The highest single off-target score obtained was 2.5, with the highest exonic off-target site scoring a 2.3. Because of its lack of significant off-target sites and proximity to the Met1 site, this protospacer was chosen to be used in the experiment.

The collection of 25 clones exhibited a variety of different insertion/deletion (indel) profiles, as seen in Figures 4, 5, 8, and S1, respectively. Although most clones display vastly different resection patterns, even between alleles, there was one recurring indel motif in many clonal populations. Of the 25 isolated clones, 19 populations contained at least one allele where a single thymine insertion occurred at the DSB site, with 9 of them being homozygous, containing the single thymine insert on both alleles. Four more clones contained a single thymine insertion along with a deletion of various lengths. In every case with a single base insertion, in fact, the base inserted is a thymine.

A distinct theme is seen in clones B, F, and G regarding the disparity between TIDE analysis in Figure 4 and allelic analysis in Figure 5. During TIDE analysis, the mixed-peak sample is broken down into each of the constituent decomposition products, with each of these decomposition products being compared to determine what the indel profile of each clone contains. However, the products are determined based on sequences of homology on either side of the cut site within each of the decomposition products; any mismatches that occur outside these homology sequences are not utilized in the determination of indel length. Because of this, misaligned Sanger sequencing has the potential to return TIDE analysis results that show indels 1 bp shorter or longer than the actual indels. The allelic analysis reveals that clones B, F, and G in fact contain a single thymine insertion in each of the alleles that the TIDE analysis displayed as wild-type. A related issue can be seen analyzing CRISPR activity that leads to a combination of insertions and deletions, as seen in clones B and D. Because sequence length between each of the alignment windows is the determining factor in indel length, only the net size of the indel is returned in the TIDE analysis. In clone B, a 22 bp deletion is observed in the TIDE analysis. Allelic analysis reveals that this deletion is actually 27 bp long, compounded with a 5 bp insertion; the net change in the sequence length is 22 bp. Both issues are compounded in clone D, where a combination indel combined with atypically spaced Sanger sequencing. This situation leads the TIDE software to return a deletion of 5 bp when the allele in question obtained a deletion of 13 bp compounded with an insertion of 6 bp. In this clone, the combination of misspaced Sanger sequencing, deletion, and insertion leads to a predicted deletion 2 bp shorter than the net change in sequence length and 8 bp shorter than the actual deletion.

These examples highlight an important limitation in the operation of TIDE software. When utilizing the software to analyze CRISPR activity in populations of cells, or when looking for overall gene knockouts, the TIDE software may be able to be used as a terminal stage of data analysis. Both of these levels of analysis allow for a degree of latitude in the exact size of indels formed. TIDE is still incredibly useful when employing CRISPR/Cas9 to create a more exact change, such as in this experiment. However, the TIDE readout must then be further investigated to obtain a more exact allelic profile and cannot be used as the terminal step in data analysis.

Figure 6 demonstrates the effects of the CRISPR/Cas9-mediated resection on the expression levels Gata1 and Gata1s in each of clones A–G at the mRNA level. *GATA1* in wild-type K562 cells is expressed in two splice variants, both with and without exon 2. Because exon 2 contains *GATA1*'s Met1 site, the smaller splice variant codes for Gata1s naturally, whereas the longer variant encodes full-length Gata1. Expression of both splice variants of *GATA1* is seen in cell lines and patient samples.[27,28,32–34] We observe this same expression profile in our samples, with the various CRISPR/Cas9-mediated resections affecting the balance between long-form and short-form *GATA1* expression consistent with mRNA and protein expression profiles seen in various ML-DS patients with a variety of different mutations at the Met1 site.[27,28,33] In each clone, the short-form mRNA sequence was identical to the wild-type (the exon 1 to exon 3 splice site is shown in Figure S2).

The expression of genes on the X chromosomes is tempered through the process of lyonization. One X chromosome is randomly inhibited and bound in heterochromatin during early blastocyst development. This chromosome is replicated normally during S-phase; furthermore, the same X chromosome is inhibited in each generation of

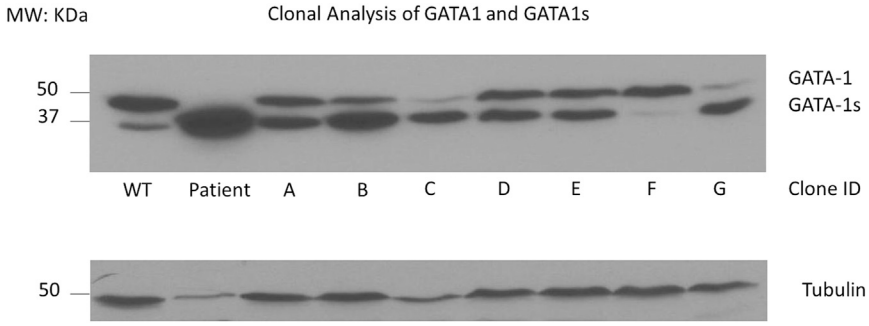


Figure 7. CRISPR/Cas9-Mediated Resection at the *GATA1* Met1 Initiation Site Leads to Altered Expression of Gata1 and Gata1s Proteins

Relative total levels of Gata1 and Gata1s of wild-type K562s, each clonally expanded population, and blood samples from a ML-DS patient were examined via western blot. Full-length Gata1 is shown above with a protein band at approximately 50 kDa, with Gata1s represented by a band at 37 kDa. Tubulin protein level controls are also shown.

daughter cells. However, this is not the case for all genes; depending on the cell, up to 15% of X-linked genes can escape inactivation. *GATA1* in particular shows slight escape ability.[35] This effect can be seen in several of the sequences of the long-form mRNA amplicons (data not shown). Figure 7 shows the protein levels of Gata1 and Gata1s in each of the clones. When compared with the mRNA expression profiles, it becomes apparent that the mRNA for full-length *GATA1* is preferentially expressed as compared with Gata1s mRNA.

Clone C highlights an important issue that is often overlooked in the generation of cell lines using this more surgical form of gene editing. When the overall goal is to modify precisely, rather than totally abrogate, gene expression, a detailed allelic profile must be generated at both the genotypic and the phenotypic levels; a population-based genotyping without clonal expansion could lead to a level of insufficient data with which to make valid conclusions. For example, our own initial allelic analysis of clone C indicated that a 3 bp deletion had occurred, and that the initial adenine residue had been resected. However, mRNA and protein expression analysis determined that the clone in fact was affected with only a 2 bp deletion in allele 1, allowing the Met1 site to remain intact. Because of this lack of a biallelic knockout in all 25 of the initial clones, a second transfection was performed, with a further 48 clones being successfully isolated. It was from this transfection that clones 67 and 72, both biallelic knockouts, were generated.

Clone 67 displayed the most patient-like protein expression profile. Expression of full-length Gata1 was completely ablated, with the corrected Gata1 protein levels falling well within background levels. Furthermore, expression of Gata1s was augmented greatly in the clone, with a 9.9-fold increase in Gata1s compared with wild-type. This profile matches the *GATA1* expression profile of the ML-DS patient sample, indicating the success of the genetic surgery strategy.

Clone 72 also displayed complete abrogation of full-length Gata1 and a 4.0-fold increase in Gata1s expression. The differences in the degree of resection of the *GATA1* gene between both clones open the possibility of disruption of the 3D orientation of the *GATA1* promoter sequence. This, combined with the magnification of an individual cell's minor genetic or epigenetic variations as a product of clonal expansion, confound the determination of an exact mechanism for the differing expression profiles. Both clones show complete elimination of full-length Gata1, and so both clone 67 and clone 72 are evidence of the success of this genetic surgery. However, the higher levels of Gata1s protein expression in clone 67 elevate the clone to the level of replicating the *GATA1* expression profile of the patient sample.

The protein expression profiles of the clones demonstrate that the long-form *GATA1* mRNA is preferentially translated. The exact mechanism behind this is unclear, but may be related to the stronger Kozak sequence located proximal to the Met1 site, in comparison with the Met84 site. At the Met1 site, the −4, −2, −1, and +4 sites of the start sequence match that of the consensus Kozak sequence, whereas the Met84 site match only at the −4 and +4 sites.[36,37] This increased homology to the consensus Kozak sequence may be a contributing factor to the preferential translation of full-length Gata1. Furthermore, altering of these Kozak sequences through CRISPR/Cas9-mediated resection may have altered the translation efficiency of the resulting mRNA transcripts, even if the Met1 site itself remains intact. In clones A and E, the proportional levels of long-form mRNA are much lower than the amount of full-length Gata1 protein being expressed. Clones C and G, in contrast, showed lower levels of full-length Gata1 protein than would be expected, given the amount of long-form mRNA seen in Figure 6. Interestingly, the mRNA expression profiles and the protein expression profiles are identical in clones B and D. The CRISPR/Cas9-mediated resections that occurred led to a variety of combinations of exonic splice-site variations and Kozak sequence permutations, leading to such vastly different proportions of full-length Gata1 and Gata1s.

Although the variable effects of CRISPR/Cas9-mediated knockouts have been explored extensively, the levels and identity of indel formation on a larger scale require further investigation. These findings may hold value in examining the identity and efficacy of oligonucleotide-mediated gene manipulation or insertion, or in the utilization of CRISPR/Cas9 alone as a surgical tool for gene editing. These cells can now be used to determine downstream mutations as a result of an altered start site in the process of leukemogenesis. Furthermore, our experimental approach is of significant value in studies of DS-induced pluripotent stem cells (DS-iPSCs) to elucidate a pathway that reveals where genetic "hits" occur to enable ML-DS. Our long-term goal is to examine how these hits can occur in non-transformed

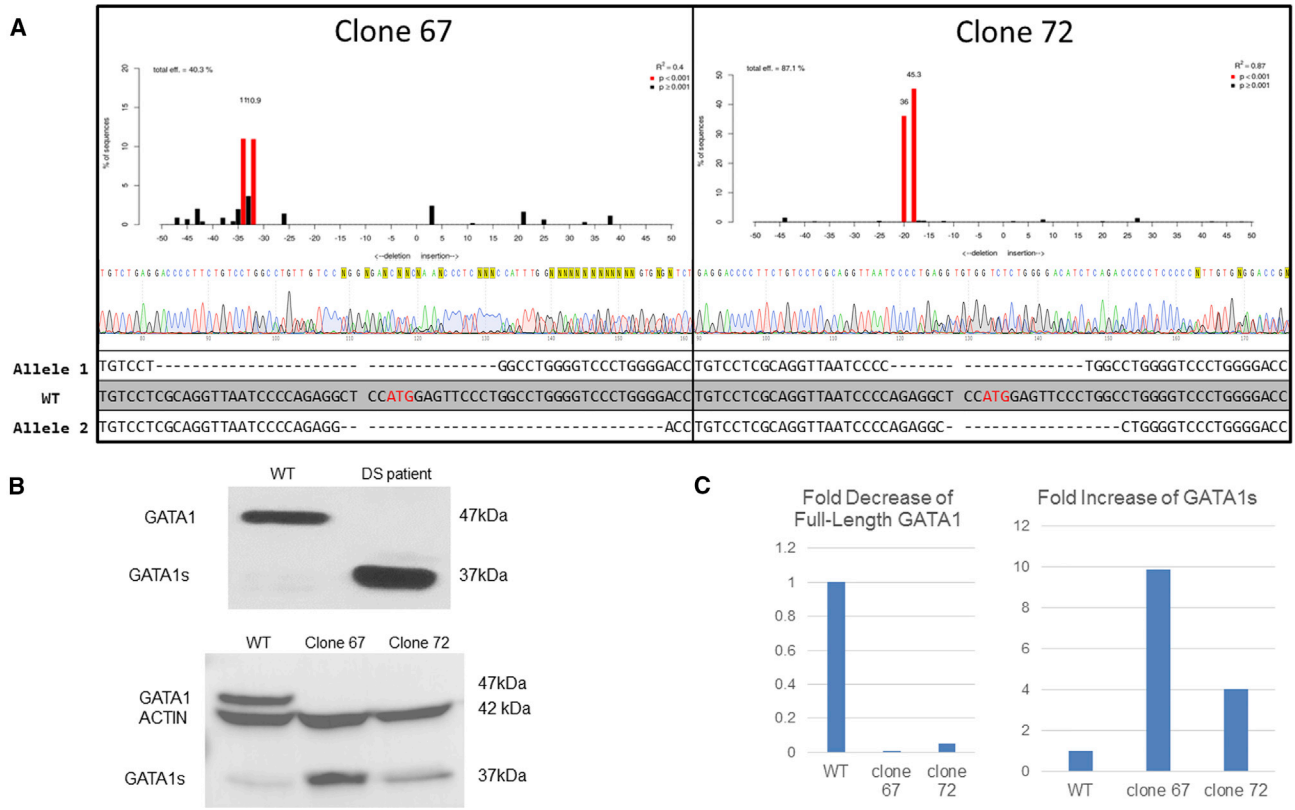


Figure 8. Biallelic CRISPR/Cas9-Mediated Elimination of the *GATA1* Met1 Initiation Site Leads to Complete Ablation of Full-Length Gata1

(A) TIDE and allelic analysis of clones 67 and 72 are shown, displaying biallelic elimination of the *GATA1* Met1 initiation site. (B) Relative total levels of Gata1 and Gata1s of wild-type K562s and blood samples from a ML-DS patient were examined via western blot in the left panel. The right panel shows relative total levels of Gata1 and Gata1s in CRISPR/Cas9-treated clonally expanded colonies 67 and 72. Full-length Gata1 is shown above with a protein band at approximately 47 kDa, with Gata1s represented by a band at 37 kDa. Actin protein level controls are also shown. (C) Fold decrease of full-length Gata1 was determined in clones 67 and 72 via densitometry and normalized to WT actin control. Fold increase of Gata1s expression was determined in the same manner.

cells and to develop a more clinically relevant mechanism for identifying druggable pathway targets.

MATERIALS AND METHODS

Experimental Design

K562 cells were treated with a CRISPR/Cas9 system designed to cleave 2 bp upstream of the *GATA1* Met1 initiation codon site. The population was enriched for cells that successfully took up the CRISPR plasmid, and the cells were clonally expanded after single-cell isolation. Upon expansion, genomic DNA was extracted from each clonal population and analyzed via Sanger sequencing and TIDE analysis. Individual allelic analysis was performed using the data obtained via TIDE. Clones exhibiting varied degrees of CRISPR/Cas9-mediated resection were analyzed for mRNA and protein expression via RT-PCR and western blot, respectively.

Cell Culture, Transfection, and Single-Cell Cloning

K562 cells were cultured in Iscove's Modified Dulbecco's Medium (ATCC), supplemented with 10% fetal calf serum and 1% penicillin/streptomycin. Cells were maintained at 37°C and 5% CO_2.

For transfections, K562 cells were seeded in six-well plates at a density of 5×10^5 cells/well in 1.5 mL of complete media. Transfection complexes were formed at room temperature in 500 µL of Opti-MEM according to the optimized LTX protocol (Thermo Fisher Scientific) and added drop-wise to K562 cells, followed by 72 hr incubation. Cells were sorted into a 96-well plate with a FACSAriaII flow cytometer (BD Biosciences), with an individual positively transfected cell sorted to each well. Clones were expanded into larger plates as the individual clones reached confluence, with DNA isolation occurring when cells reached confluence in a six-well plate (1×10^6 cells/mL).

CRISPR/Cas9 Plasmid Design and Construction

The *GATA1* gene sequence was entered into the Zhang lab's[25] online generator (http://crispr.mit.edu/), and the appropriate CRISPR guide sequence that binds in close proximity to target (−2 bp upstream of Met1) was chosen. The CRISPR was constructed using standard cloning methods following the latest oligo annealing and backbone cloning protocol with single-step digestion-ligation. The CRISPR guide sequence was cloned into the pX458 backbone vector (plasmid

48138; Addgene), a human codon optimized pSpCas9 and chimeric guide RNA expression plasmid with a 2AEGFP, purchased through Addgene (https://www.addgene.org). Following construction, clones were verified by DNA sequencing by Eton Bio Incorporated.

PCR, RT-PCR, and Western Blot

Cellular gDNA was isolated from pellets of 2×10^6 K562 cells using the QIAGEN DNeasy Blood and Tissue Kit (cat. ID 69506). PCR was performed using AmpliTaq Gold Fast PCR Master Mix (Thermo Fisher Scientific) on isolated gDNA, with amplification parameters optimized for an amplicon size of 462 bp (forward [FWD] primer: 5′-gggaggtgggaaggagaaatatggag-3′, reverse [REV] primer: 5′- cctcacagtggtattctgacctagcc-3′; obtained from Integrated DNA Technologies). Amplicon size was verified on 1% agarose gel, and PCR products were verified by DNA sequencing by Eton Bio Incorporated.

For RT-PCR, total cellular RNA was isolated using TRIzol reagent from Thermo Fisher Scientific (cat. ID 15596026). cDNA was synthesized from total RNA using iScript cDNA Synthesis Kit from Bio-Rad Laboratories (cat. ID 170-8891) following the manufacturer's protocol. Purified cDNA was then amplified using AmpliTaq Gold Fast PCR Master Mix (Thermo Fisher Scientific), optimized for differentially sized amplification of long-form and short-form *GATA1* cDNA conversions (FWD RT primer: 5′- gatcacactgagcttgccacatcc-3′, REV RT primer: 5′- gcactattggggacagggagtg-3′; obtained from Integrated DNA Technologies). Mixed amplicons were separated on a 1% agarose gel, and individual bands were excised and purified using a QIAquick Gel Extraction Kit (cat. ID 28704). Products were then verified by DNA sequencing by Eton Bio Incorporated.

Protein isolation from individual clones was performed using SDS lysis buffer, with protein concentrations determined and normalized via the Bio-Rad DC Protein Assay (cat. ID 5000111). Western blot was performed using standard procedures and transferred to nitrocellulose membrane using the Bio-Rad Mini Transblot Cell (cat. ID 1703930). Gata1 and Gata1s were stained with rabbit anti-Gata1 unconjugated antibody (cat. no. 4591; Cell Signaling Technology), with tubulin being stained with rabbit anti-tubulin unconjugated antibody (cat. no. 2146; Cell Signaling Technology). Secondary antibody used to visualize protein was anti-rabbit IgG, HRP-linked antibody (cat. no. 7074; Cell Signaling Technology). Stained blots were visualized using Amersham ECL Prime Reagent (cat. ID RPN2232; GE Life Sciences) on X-ray film, with an exposure time of 10 s.

Statistical Analysis

TIDE analysis was performed on all Sanger sequences obtained from all clones as part of allelic analysis.[23] The p value used to determine significance of potential sequence decompositions was $p = 0.001$.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and can be found with this article online at http://dx.doi.org/10.1016/j.omtn.2017.04.009.

AUTHOR CONTRIBUTIONS

K.M.B. and P.A.B. designed the experiments. K.M.B. performed the transfection and culturing of the clonal cell populations before and after single-cell sorting. Isolation, PCR, and sequencing prep were performed by K.M.B. and P.A.B. Allelic analysis was designed and completed by K.M.B. RT-PCR and western blots were performed by A.G. and K.M.B. K.M.B., E.A.K., and E.B.K. wrote the manuscript. A.G., E.A.K., and E.B.K. conceived the ideas and directed the work.

ACKNOWLEDGMENTS

We thank Dr. Lynn Opdenaker for outstanding technical assistance with the FACS analysis, and Drs. Nicholas Petrelli (CCHS) and Sonali Barwe (Nemours) for important discussions at various stages during this project. We also acknowledge the valuable advice and input from all the members of the Kmiec laboratory. This project was supported by grants 1P20GM109021 (to E.A.K. and E.B.K.) and P20 GM103446 (to A.G.) from the NIH and by the Leukemia Research Foundation of Delaware (to E.A.K.).

REFERENCES

1. Mitelman, F., Heim, S., and Mandahl, N. (1990). Trisomy 21 in neoplastic cells. Am. J. Med. Genet. Suppl. *7*, 262–266.
2. Zipursky, A., Poon, A., and Doyle, J. (1992). Leukemia in Down syndrome: a review. Pediatr. Hematol. Oncol. *9*, 139–149.
3. Hasle, H., Niemeyer, C.M., Chessells, J.M., Baumann, I., Bennett, J.M., Kerndrup, G., and Head, D.R. (2003). A pediatric approach to the WHO classification of myelodysplastic and myeloproliferative diseases. Leukemia *17*, 277–282.
4. Wechsler, J., Greene, M., McDevitt, M.A., Anastasi, J., Karp, J.E., Le Beau, M.M., and Crispino, J.D. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nat. Genet. *32*, 148–152.
5. Greene, M.E., Mundschau, G., Wechsler, J., McDevitt, M., Gamis, A., Karp, J., Gurbuxani, S., Arceci, R., and Crispino, J.D. (2003). Mutations in GATA1 in both transient myeloproliferative disorder and acute megakaryoblastic leukemia of Down syndrome. Blood Cells Mol. Dis. *31*, 351–356.
6. Kaneko, H., Kobayashi, E., Yamamoto, M., and Shimizu, R. (2012). N- and C-terminal transactivation domains of GATA1 protein coordinate hematopoietic program. J. Biol. Chem. *287*, 21439–21449.
7. Gruber, T.A., and Downing, J.R. (2015). The biology of pediatric acute megakaryoblastic leukemia. Blood *126*, 943–949.
8. Yoshida, K., Toki, T., Okuno, Y., Kanezaki, R., Shiraishi, Y., Sato-Otsubo, A., Sanada, M., Park, M.J., Terui, K., Suzuki, H., et al. (2013). The landscape of somatic mutations in Down syndrome-related myeloid disorders. Nat. Genet. *45*, 1293–1299.
9. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.
10. Sánchez-Rivera, F.J., and Jacks, T. (2015). Applications of the CRISPR-Cas9 system in cancer biology. Nat. Rev. Cancer *15*, 387–395.
11. Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell Res. *18*, 134–147.
12. Xu, P., Tong, Y., Liu, X.Z., Wang, T.T., Cheng, L., Wang, B.Y., Lv, X., Huang, Y., and Liu, D.P. (2015). Both TALENs and CRISPR/Cas9 directly target the HBB IVS2-654 (C > T) mutation in β-thalassemia-derived iPSCs. Sci. Rep. *5*, 12065.
13. Schumann, K., Lin, S., Boyer, E., Simeonov, D.R., Subramaniam, M., Gate, R.E., Haliburton, G.E., Ye, C.J., Bluestone, J.A., Doudna, J.A., and Marson, A. (2015). Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. Proc. Natl. Acad. Sci. USA *112*, 10437–10442.
14. Cho, S.W., Kim, S., Kim, J.M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. *31*, 230–232.

15. Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Cai, W., Yang, G., Bronson, R., Crowley, D.G., et al. (2014). CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature *514*, 380–384.

16. Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature *527*, 192–197.

17. DeWitt, M., Magis, W., Bray, N.L., Wang, T., Berman, J.R., Urbinati, F., Muñoz, D.P., Kohn, D.B., Walters, M.C., Carroll, D., et al. (2016). Efficient correction of the sickle mutation in human hematopoietic stem cells using a Cas9 ribonucleoprotein complex. BioRxiv, Published online January 15, 2016. http://dx.doi.org/10.1101/036236.

18. Li, H., Shi, J., Huang, N.-J., Pishesha, N., Natarajan, A., Eng, J.C., and Lodish, H.F. (2016). Efficient CRISPR-Cas9 mediated gene disruption in primary erythroid progenitor cells. Haematologica *101*, e216–e219.

19. Xie, F., Ye, L., Chang, J.C., Beyer, A.I., Wang, J., Muench, M.O., and Kan, Y.W. (2014). Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. Genome Res. *24*, 1526–1533.

20. Song, B., Fan, Y., He, W., Zhu, D., Niu, X., Wang, D., Ou, Z., Luo, M., and Sun, X. (2015). Improved hematopoietic differentiation efficiency of gene-corrected beta-thalassemia induced pluripotent stem cells by CRISPR/Cas9 system. Stem Cells Dev. *24*, 1053–1065.

21. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

22. Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., et al. (2013). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. *31*, 1137–1142.

23. Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168.

24. Bialk, P., Rivera-Torres, N., Strouse, B., and Kmiec, E.B. (2015). Regulation of gene editing activity directed by single-stranded oligonucleotides and CRISPR/Cas9 systems. PLoS ONE *10*, e0129308.

25. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

26. Naumann, S., Reutzel, D., Speicher, M., and Decker, H.J. (2001). Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ hybridization, and comparative genomic hybridization. Leuk. Res. *25*, 313–322.

27. Rainis, L., Bercovich, D., Strehl, S., Teigler-Schlegel, A., Stark, B., Trka, J., Amariglio, N., Biondi, A., Muler, I., Rechavi, G., et al. (2003). Mutations in exon 2 of GATA1 are early events in megakaryocytic malignancies associated with trisomy 21. Blood *102*, 981–986.

28. Calligaris, R., Bottardi, S., Cogoi, S., Apezteguia, I., and Santoro, C. (1995). Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor. Proc. Natl. Acad. Sci. USA *92*, 11598–11602.

29. Hitzler, J.K. (2007). Acute megakaryoblastic leukemia in Down syndrome. Pediatr. Blood Cancer *49* (*Suppl 7*), 1066–1069.

30. Malinge, S., Izraeli, S., and Crispino, J.D. (2009). Insights into the manifestations, outcomes, and mechanisms of leukemogenesis in Down syndrome. Blood *113*, 2619–2628.

31. Yin, H., Xue, W., Chen, S., Bogorad, R.L., Benedetti, E., Grompe, M., Koteliansky, V., Sharp, P.A., Jacks, T., and Anderson, D.G. (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. Nat. Biotechnol. *32*, 551–553.

32. Halsey, C., Tunstall, O., Gibson, B., Roberts, I., and Graham, G. (2010). Role of GATA-1s in early hematopoiesis and differences between alternative splicing in human and murine GATA-1. Blood *115*, 3415–3416.

33. Kanezaki, R., Toki, T., Terui, K., Xu, G., Wang, R., Shimada, A., Hama, A., Kanegane, H., Kawakami, K., Endo, M., et al. (2010). Down syndrome and GATA1 mutations in transient abnormal myeloproliferative disorder: mutation classes correlate with progression to myeloid leukemia. Blood *116*, 4631–4638.

34. Reeder, C.B., Reece, D.E., Kukreti, V., Chen, C., Trudel, S., Laumann, K., Hentz, J., Pirooz, N.A., Piza, J.G., Tiedemann, R., et al. (2010). Once- versus twice-weekly bortezomib induction therapy with CyBorD in newly diagnosed multiple myeloma. Blood *115*, 3416–3417.

35. Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404.

36. Kozak, M. (1984). Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. Nature *308*, 241–246.

37. De Angioletti, M., Lacerra, G., Sabato, V., and Carestia, C. (2004). Beta+45 G → C: a novel silent beta-thalassaemia mutation, the first in the Kozak sequence. Br. J. Haematol. *124*, 224–231.