

# Background Information for Fangoul et al (V3) (Propel Literature Review Oct 2022)

## The Disease and Treatment Strategies

### 1) What do we know about hemoglobin, the molecule affected by **Thalassemia and Sickle Cell Anemia**?

Both diseases have a defect in the production of hemoglobin (Hb). Hb is a 4 protein tetramer in red blood cells (RBCs) that is responsible for transporting one oxygen molecule per monomer to cells and one carbon dioxide molecule per monomer from cells.

Most adult Hg tetramers consists of two  $\alpha$  (alpha) Hb subunits and two  $\beta$  (beta) Hb subunits; this  $\alpha_2\beta_2$  tetramer is referred to as adult hemoglobin (**HbA**). However, there are two  $\beta$ -like Hb subunits called  $\delta$  (delta) and  $\gamma$  (gamma) that can take the place of  $\beta$  Hb subunit in the tetramer. In adults, a small portion of Hb tetramers is of the form ( $\alpha_2\delta_2$ ), also known as **HbA2**. During fetal development, however, most of the Hg tetramer is of the form ( $\alpha_2\gamma_2$ ). This is what is referred to as fetal hemoglobin (**HbF**). Once a child is born, expression of  $\gamma$  Hb subunit is repressed and expression of  $\beta$  Hb subunit induced causing replacement of  $\alpha_2\gamma_2$  HbF with primarily  $\alpha_2\beta_2$  HbA.

In thalassemia, there is insufficient or no expression of  $\beta$  Hb subunits, so there is a serious deficiency in hemoglobin that develops after birth. Thalassemia is fundamentally a defect in protein quantity.

In sickle cell anemia, a point mutation in the  $\beta$  Hb subunit causes a conformational change in the presence of low oxygen concentrations that causes the mutant  $\alpha_2\beta^{SC}2$  HbS tetramers to polymerize, effectively forming an insoluble mass in RBCs. This massive polymerization of hemoglobin within them causes the RBCs to acquire a deformed sickle shape that causes them to get tangled up with each other, blocking small blood vessels and causing painful oxygen deprivation (ischemia) downstream of the blocks. That pain and damage is experienced as a sickle cell crisis. Sickle cell anemia is fundamentally a defect in protein quality.

### 2) How do you reconcile the recessive behavior of **Thalassemia mutations with their molecular defect**?

**Recessive** mutations are those that must affect both copies of a gene in a diploid organism in order for a mutant phenotype to evince itself. **Dominant** mutations are those where only one copy of a gene needs to be affected in a diploid organism for the mutant phenotype to emerge.

**Most recessive mutations are loss of function mutations** (this is only a rule of thumb, because there are clear exceptions to this rule that I don't have time to discuss). Like most diploids, humans have two copies of every autosomal chromosome (i.e. not X and Y sex chromosomes)

and thus two copies of every gene on these chromosomes. Hence, if one of these genes simply loses its function, often the remaining wild-type intact gene can provide sufficient function for people to enjoy reasonably normal health. Only if an individual inherits two mutant genes, one from each parent, will they experience a severe problem.

The thalassemia gene mutation is an example of such a loss of function mutation. The mutant  $\beta$  Hb gene (HBB) is simply not expressed. Hence, carriers with one mutant and one wild-type HBB gene will still be able to express enough wild-type  $\beta$  Hb to make sufficient HbA tetramer protein for decent oxygenation. Loss of both genes is required to experience the disease, making the mutation recessive

3) **How do you reconcile the recessive behavior of Sickle Cell mutations with their molecular defect?**

**Most dominant mutations are gain of function mutations** (this is also only a rule of thumb, because there are also clear exceptions to this rule that I don't have time to discuss). They often produce defective proteins that act in a new toxic manner to poison a biological process, even in the presence of a wild-type copy of the gene that produces the wild-type protein.

At first glance, the sickle cell mutation is a gain of function mutation that creates a new toxic property: the polymerization of conformationally altered sickle cell hemoglobin HbS for the brief window of time (~1-2 sec) after it has released its oxygen in capillaries. But sickle cell anemia behaves as a recessive mutation, as if the presence of wild-type  $\beta$  Hb subunits can interfere with polymerization induced by the  $\beta^{SC}$  Hb subunit. How can this be?

The primary reason is that the presence of roughly equal amounts of wild-type  $\beta$  Hb and mutant  $\beta^{SC}$  Hb means that roughly 50% of the tetramers will be hybrid  $\alpha_2\beta\beta^{SC}$ , 25% will be  $\alpha_2\beta_2$  (HbA), and only 25% will be  $\alpha_2\beta^{SC}_2$  (HbS), effectively reducing the concentration of polymerization competent HbS by 4-fold. This lower concentration of HbS leads to a significantly lower polymerization rate, too slow to result in polymerization of transiently deoxygenated HbS.

Hence, replacement of both wild-type  $\beta$  Hb subunit genes with mutant  $\beta^{SC}$  subunit genes is needed to see the disease phenotype, making the mutation recessive.

4) **What type of strategies can you use to fix the molecular defect in  $\beta$  Hb subunit production?**

Simple Answer: Correct or Compensate

One way is to correct the primary defect. That would involve correcting the defective  $\beta$  Hb genes so that it can be expressed at normal levels (for thalassemia) or so that a normal unmutated version of the protein can be expressed (for sickle cell anemia).

However, current efforts, as described by Fangoul et al., are instead aimed toward compensating for the defect by inducing synthesis of  $\alpha_2\gamma_2$  HbF simply by preventing the  $\gamma$  Hb subunit from being repressed postnatally (i.e. after birth) so that its expression persists into adulthood.

Remember, as discussed above, you don't have to completely replace the sickle cell HbS. It is sufficient to make enough of the wild-type HbF to presumably disrupt the polymerization of HbS molecules and convert the doubly mutant sickle cell patient to a phenotypic status equivalent to that of a mutant/wt carrier.

In order to understand how to disrupt the postnatal repression of  $\gamma$  Hb subunit, we first need to review how gene expression is regulated.

**5) A molecular biology primer: how is gene expression regulated?**

Genes are expressed by recruitment and activation of RNA Polymerase to the DNA at the beginning of genes. From there RNA Polymerase proceeds down the DNA **transcribing** an RNA copy of the DNA strand that is oriented 5' to 3' from the beginning of the gene to the end. (Remember the complementary DNA strand is oriented in an "antiparallel" manner in a 3' to 5' direction).

In most eukaryotic cells, only short (a few hundred nucleotide) segments of the initial RNA synthesized encode amino acids that will make up the final protein product. These segments, which are called **exons** are interspersed among much longer segments of noncoding RNA called introns. By a process called **splicing**, the **introns** are excised leaving a shorter final messenger RNA (mRNA) comprised of the exons stitched together in their original order. This spliced mRNA is transported out of the nucleus where it is recognized by ribosomes that **translate** the mRNA into proteins.

**6) How do the sequences at a gene control its regulation?**

Simple Answer: Promoter and Enhancers through recruitment of protein transcription factors and regulators

DNA sequences within a few thousand nucleotides of the beginning of the gene form the **promoter** of the gene. These promoter sequences direct the binding of transcriptional regulatory proteins that either facilitate or prevent the recruitment and initiation of RNA polymerase at the promoter.

However, what happens at promoters is also influenced by additional sequences that can be located tens of thousands of nucleotides away from the promoter and can work from upstream, downstream or even within the gene. When these sequences direct the binding of transcriptional regulatory proteins that enhance the activation of RNA polymerase at the promoter, they are called **enhancers**.

**7) How do genes get regulated differently in different cell types if each gene has the same fixed set of sequence elements associated with it?**

Although the DNA sequence elements that can control a gene's expression are fixed and identical among all cell types, whether any specific sequence element participates in that gene

regulation in a particular cell type depends on whether that cell type expresses the transcriptional regulators recruited by that element. Thus, the particular subset of transcriptional regulators that are expressed in a cell type will determine which subset of sequence elements are used to regulate a gene in that cell type.

This means that if you mutate a sequence element that can regulate a gene, it will only affect the expression of the gene in cell types that use that element.

**8) How is expression of the  $\gamma$  Hb subunit regulated in RBC precursor cells.**

Human RBCs have no nuclei and are effectively stores of hemoglobin protein that were stockpiled during the differentiation of RBC precursor cells

The primary  $\beta$ -like Hb gene transcriptionally expressed in fetal RBC precursor cells is the  $\gamma$  Hb gene. After the birth of a baby, its RBC precursor cells repress the transcription of the  $\gamma$  Hb gene in part by inducing expression of a transcriptional repressor called BCL11A. Hence, the final differentiated RBC cells contain little if any HbF.

**9) How is BCL11A repressor protein induced in RBC precursor cells?**

This protein induction arises from the transcriptional induction of the BCL11A gene in RBC precursor cells. We don't know all the transcription factors and sequence elements that together cause this transcriptional induction. However, the authors previously identified at least one **enhancer element** in one of the introns that is necessary for full induction of BCL11A and therefore full repression of HbF. In a previous paper (Cancer et al. 2015 Nature) the authors had performed a systematic mutagenesis analysis of this BCL11A intron to narrow down the region in the intron where mutations would impair BCL11A transcription in RBC precursor cells and thereby cause increased expression of HbF in differentiated RBCs.

Thus, in principle we have a potential strategy to treat both thalassemia and sickle cell anemia. Mutate a specific enhancer sequence in the BCL11A gene to prevent the postnatal induction of the BCL11A repressor protein in RBC precursors. One of the repression targets of BCL11A, the  $\gamma$  Hb subunit gene, will not be repressed, allowing the subunit to persist in adult RBCs.

The next sections address the practical aspects of carrying out this strategy at the molecular biological level.

## **The Molecular Biology of CRISPR-Induced Mutagenesis**

**10) How can one target the mutation of a specific DNA sequence?**

This question raises two fundamental and distinct concepts that will need to be addressed separately: (1) how do you "**find**" a specific DNA sequence; and (2) how do you **mutate** the

found sequence? We discuss the finding, or targeting, here, and discuss the mutating in a later section.

How can we search through billions of nucleotides in the human genome and find a unique sequence of nucleotides? We take advantage of Nature, which has evolved some amazing protein machines to do this. The basic principle Nature uses is **sequence complementarity** to a **single-stranded probe sequence**. These protein machines bind the probe sequence and use to look through huge amounts of double-stranded sequence to find a segment with one strand that can base pair with the single stranded probe because the two are complementary. Not only must the protein be able to rapidly scan through huge stretches of DNA, transiently and locally separating the two strands to test for base pairing, it must be able to detect when there is sufficient base pairing to call it a match.

### 11) What molecular tools are available to find/target specific DNA sequences?

Simple Answer: HR and CRISPR

We know of two molecular biological processes that can carry out a sequence complementarity search through double-stranded DNA

The first to be discovered was **homologous recombination (HR)**. For its probe, the HR machinery uses single-stranded DNA exposed at the ends of linear double-stranded DNA molecules. This exposed single strand is generated by a another protein machinery that is recruited to double-stranded ends and specifically resects the 5' strand at each end. In this manner the HR machinery, which was evolved to precisely repair double-stranded DNA breaks, can use the two ends on either side of a break to find a DNA segment that is identical (like the copy that was made during DNA replication) or nearly identical (like the homologous segment on the homologous chromosome) to the broken segment and use the unbroken segment as a DNA template to repair the broken segment.

The second process that can use sequence complementarity to search through double-stranded DNA is the **CRISPR**-based adaptive immunity that evolved to protect bacteria from reinfection by bacteriophages (bacterial viruses) it had previously survived. A wide variety of CRISPR systems have evolved in different bacteria. One of the simplest ones has been harnessed as a two-component system that can perform the fundamental tasks characteristic of all CRISPR mechanisms (1) **search** for a specific DNA sequence based on complementarity to a single stranded RNA probe sequence; (2) double stranded **cleavage** of the DNA sequence exhibiting complementarity. The two components consist of the protein **Cas9** and a **guide RNA** that provides a 20 nucleotide probe for the complementarity search. This CRISPR-Cas9 system has revolutionized our ability to target DNA sequences for mutation.

### 12) After finding a specific sequence how does CRISPR-Cas9 induce mutations there?

Simple Answer: It doesn't

CRISPR systems evolved **to search and destroy**, not search and mutate. CRISPR systems provide a memory of recent bacteriophage infections by storing short segments from the genomes of these bacteriophages in a special array called CRISPR in the bacterial genome. In the CRISPR array these short segments are called “**spacers**”. RNA transcripts of these spacers provide probes to search for any reappearance of these bacteriophage genomes by looking for these segments in those genomes. Because the spacer sequences originated from those genomes, the same sequence in the genome is called a **protospacer**. Should the CRISPR protein machinery detect reinfecting bacteriophages, it will cleave both DNA strands of the protospacer to initiate nuclease digestion and destruction of the reinfecting genome. By these means, the CRISPR system provides a surveillance mechanism to protect against reinfection

**13) Why doesn't the CRISPR system cause bacteria to commit suicide by cleaving its own CRISPR array?**

Simple Answer: PAM sequences, which adds an additional element of specificity in CRISPR targeting

The guide RNA transcribed from a spacer in the CRISPR array can neither find nor cleave that spacer because both events require additional short sequences just 3' of the protospacer called “protospacer adjacent motifs” of PAM sequences. For the CRISPR-Cas9 system the sequence motif is 5'-NGG-3, where N is any of the four nucleotides.

In short, not any 20 nt sequence can serve as a protospacer CRISPR target. It has to be a 20 nt sequence just 5' of a PAM sequence. By ensuring that only spacer sequences without the PAM are incorporated within its CRISPR array, bacteria avoid targeting their own genomes for destruction.

**13) How then does CRISPR-Cas9 cleavage lead to mutations?**

Short Answer: It is the cells that do the mutating in their effort to repair the CRISPR cleavage.

A double strand chromosomal break is highly lethal event if not repaired. It will lead to detached chromosomal arms containing hundreds of genes but no centromere getting “lost” during chromosome segregation. Hence, these breaks are alarming disasters, and cells have evolved multiple mechanisms to correct DNA breaks, including breaks generated by CRISPR induced cleavage. The two major mechanisms are **homologous recombination (HR)** and **nonhomologous end-joining (NHEJ)**.

As discussed in section 11 above, HR provides an **error free** way to repair DNA breaks by finding within the genome and using as repair template an identical or nearly identical copy of the segment that was cleaved. Thus, HR using endogenous homologous repair templates can interfere with CRISPR stimulated mutagenesis by restoring the wild-type sequence.

NHEJ, on the other hand provides an **error prone** mechanism to repair DNA breaks by forcing ligation of the broken ends. This can lead to small nucleotide insertions or deletion (**indels**) at

the cleavage site, especially if nucleases have an opportunity to “nibble” at the broken ends before they are ligated together.

Importantly, HR repair leaves an unmutated DNA segment that can once again be recognized by the CRISPR machinery and be cleaved. On the other hand, NHEJ repair results in an indel mutation of the protospacer, making it resistant to further CRISPR cleavage. Thus, there is a strong selective advantage for cleavage resistant indel mutations to accumulate after CRISPR-induced cleavage.

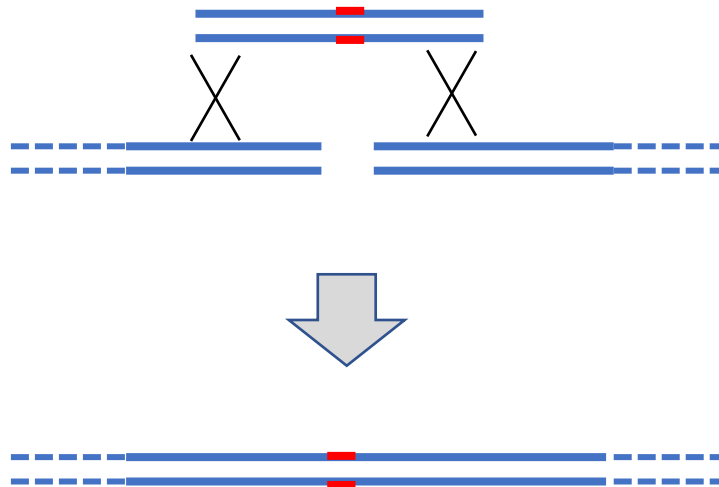
This NHEJ mechanism is how the indel mutations in the BCL11A enhancer were generated in Fangoul et al.

**14) Is there a way to use CRISPR-Cas9 to precisely edit DNA sequences?**

The generation of indels is a relatively crude and random way of generating mutations. It is effective if you just want to destroy the function of a sequence by randomly introducing short insertions and deletions.

But what if you wanted to make a precise nucleotide change? For example, what if you wanted to convert the sickle cell mutation to wild-type? How would you do that?

The trick is to subvert the error-free HR repair pathway by introducing an exogenous homologous but altered DNA segment to serve as the template for break repair. Note, this system is more complicated in having three components, not just two.



**15) How are the CRISPR components introduced into cells**

The easiest way to introduce the two-component CRISPR-Cas9 system is by introducing it through transfection of DNA that can express both Cas9 and the guide RNA. Ideally the DNA is transiently maintained in the cells so that the cells are eventually purged of the system after the desired genetic alterations have been introduced. However, unless you can select against the DNA, there is always the possibility that some of this DNA will get integrated into the genome, making some cells retain the CRISPR-Cas9 system.

An alternative is to biochemically produce the Cas9-RNA complex then introduce it into cells by permeabilizing the membrane through electroporation. This is the method used in Fangoul et al. This method ensures that DNA sequences encoding the CRISPR-Cas9 system do NOT inadvertently persist in the cells.

### **Concern over CRISPR Target Specificity**

**16) Doesn't the 20 nt guide RNA and PAM sequence provide high target specificity for the human genome?**

In principle if all 20 nucleotides of the guide RNA must be matched to a protospacer just 5' of a PAM sequence with two Gs to complete a search, that sequence is unlikely to appear randomly somewhere else as an OFF-target site. With each position having 4 possible nucleotides, the chance of a specific 22 nucleotide sequence randomly appearing elsewhere would be 4 to the power of 22 which is roughly one out of 16 trillion.

With a human genome of 6 billion nucleotides this back of the envelope calculation suggests that 22 nucleotides is sufficient to provide a specific search with little or no chance of an OFF-target match. Even if one or two nucleotide mismatches were tolerated we would still find such randomly generated imperfect matches occur at frequencies of roughly one out of a trillion.

**17) How many of the 20 nucleotides in the guide RNA are critical for targeting?**

In reality, however, only the 3' most 10 nucleotides must be stringently matched. A lot of mismatches are tolerated in the 5' most 10 nucleotides. Hence the minimum search specificity is closer to 4 to the power of 12 (when you include the two PAM Gs) or on the order of finding a randomly generated OFF target site at about one in 16 million. Even if there are some partial restrictions on the identity of the 5' most 10 nucleotides and the frequency of off target sequences rises to one in a billion, one should expect the possibility of multiple off target hits.

**18) Can homologous sequences within a genome further disrupt target specificity?**

The calculations performed above were based on the oversimplifying assumption that the human genome is equivalent to a random string of nucleotides. In reality, functionally important sequences are often reutilized throughout the genome. They can range in size from protein binding sites to whole genes that underwent variable rounds of gene duplication to



create a gene family. This proliferation of functionally important homologous sequences provides an additional potential source of OFF-target matches or near matches.

This is why the excitement of having a powerful genetic search tool like CRISPR for the human genome has to be tempered by the realization that in its current form this tool poses a real risk of OFF-target effects.

## The Organismal and Cell Biology of CRISPR Genetic Editing

### 19) What are the different levels at which one can, in principle, perform CRISPR gene therapy?

Simple answer: (1) postnatally by bone marrow transplants (2) prenatally in primordial somatic tissues and organs; (3) in the zygote, which will affect all somatic tissue plus the germline.

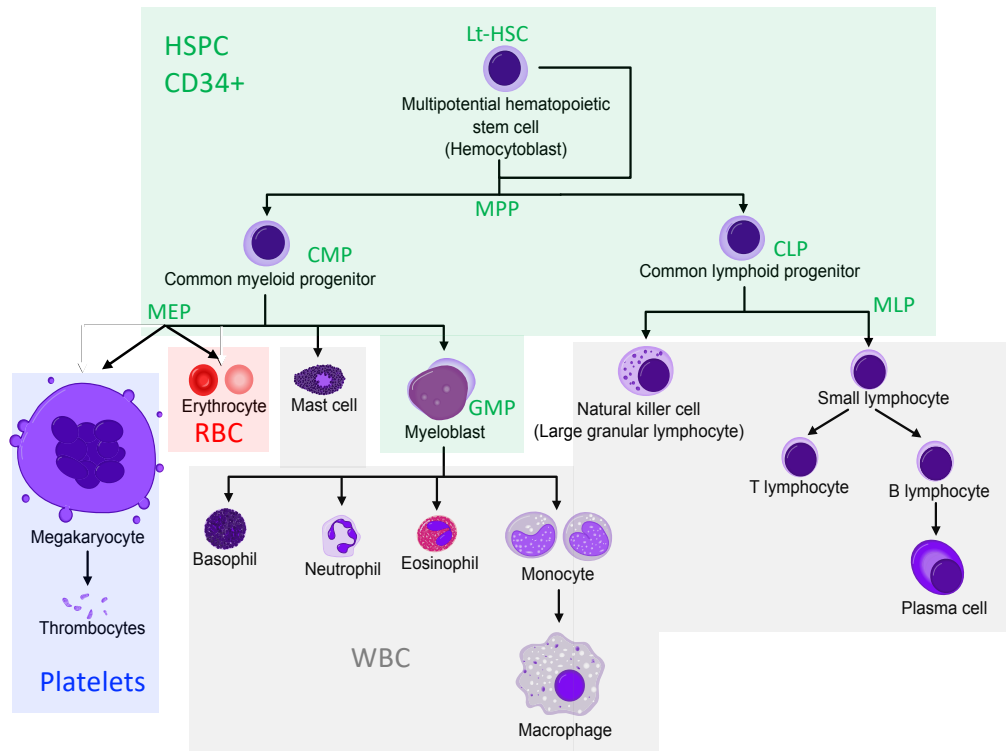
The fixed cellular architecture of tissues and organs established during prenatal development places significant restrictions on the ability to perform gene therapy on most tissues and organs postnatally. The majority of tissues and organs, e.g. the heart or kidney, undergo little cell turnover. In this setting, gene therapy would have to be performed on every terminally differentiated cell to genetically alter the tissue or organ, and this would be precluded by the tight three-dimensional packing of cells. Some organs like skin and gut epithelium are continually self-renewing through proliferation and differentiation of epithelial stem cells. In principle, one would only need to target all these stem cells for gene therapy. But even for the gut epithelium, where stem cells are accessible from the gut lumen, the stem cells are so widely distributed over the entire epithelial surface that successfully targeting all of them is a daunting task. Only one tissue is readily amenable to gene replacement postnatally, namely blood. This is because you can remove the hematopoietic (blood producing) stem cells from their normal environment in the bone marrow, perform CRISPR gene therapy on these cells in the test tube, then reinject the genetically altered stem cells back into the patient where they will home in on the bone marrow to reconstitute the hematopoietic tissue. An example of gene therapy on the hematopoietic system through bone marrow transplantation is the topic of Frangoul et al.

What if you want to perform CRISPR gene therapy on a solid tissue and organ? Theoretically, this is best attempted early in development when the tissue or organ first takes shape, has fewest cells, and has most of its cells accessible to injected CRISPR machinery. This would involve **in utero** injection of the CRISPR machinery to those primordial tissues or organs in the **early fetus**. Early experimentation with this type of strategy is being tried in animal model systems, e.g. mice.

Finally, if you need to alter the genetic makeup of all the cells in the body, then you need to perform the gene therapy on **zygotes**, which will not just affect all the somatic cells but the germline as well. Hence, the repercussions of inadvertent genetic consequences and unpredictable OFF-target perturbations are much greater because you will affect all the descendants of the treated individual. This zygotic CRISPR gene therapy is being actively studied in animal models, but deep ethical concerns and insufficient target specificity have held up its implementation in humans.

20) What is the hematopoietic lineage and which of the cells in this lineage would you treat with CRISPR?

Hematopoiesis from Hematopoietic Stem and Progenitor Cells (HSPC CD34+)



Hematopoietic Stem and Progenitor Cells (HSPC) are a mixed population of cells (highlighted in green above) that give rise to all the differentiated white blood cells (WBC), red blood cells (RBC) and platelets. At the very top of the lineage is the Longterm Hematopoietic Stem Cell (Lt-HSC) which can proliferate to generate more Li-HSC. Some of these cells can also start down the differentiation path by generating Multipotent Progenitors (MPP), which in turn will generate other HSPCs with more restricted lineages. These include the Common Lymphoid Progenitor (CLP), Common Myeloid Progenitor (GMP), Multilymphoid Progenitor (MLP), Granulocyte-Monocyte Progenitor (GMP), and the Megakaryocyte-Erythrocyte Progenitor (MEP).

Only the Lt-HSC has clear stem cell function, i.e. the ability to divide indefinitely, so in one sense it is the best target for genetic alterations as it allows these alterations to persist indefinitely in the body. However, it is easiest to isolate the entire mixed population of HSPC using antibodies against the CD34 surface marker, so this entire population is what Fangoul et al. genetically alter with CRISPR. And altering the downstream progenitor cells will allow faster production of the altered final differentiated cells in patients. So in practice, the entire HSPC population of cells highlighted in green in diagram are treated with CRISPR.

**21) How do you obtain HSPCs from the patient?**

Simple Answer: by blood draw after special preparation of the patient

HSPCs mainly reside in the bone marrow, where hematopoiesis occurs. It is invasive and painful to extract those HSPCs from the bone marrow. Now special cell growth factors (like filgrastim and plerixafor) are injected into the patient for 4-6 days to stimulate these HSPC to overproliferate causing them to spill out into the peripheral blood system where they can easily be collected by drawing blood. This increase in HSPC concentration in the peripheral blood is called **mobilization**.

The collected peripheral blood is centrifuged to separate the blood into 4 basic fractions based on their different densities: red blood cells, leukocytes (white blood cells), platelets, and plasma. This process is called **apheresis**. The leukocyte fraction is collected because the HSPC reside in this fraction. These HSPC can be separated out with antibodies specific for the CD34 surface antigen found on HSPC but not on differentiated blood cells.

**22) Are the modified HSPCs we inject back into the patient (termed CX001 in paper) a uniform or mixed population of cells**

Simple Answer: mixed population containing both cell type heterogeneity and genetic heterogeneity.

It is important to realize that we are not going to be dealing with a well-defined uniform population of cells. If you draw a schematic for how a cell is altered by CRISPR before you return it to the patient, it is easy to make the implicit and mistaken assumption that all the cells are going to look identical to the one you drew out.

In reality, we are dealing with a mixed population of cells containing several layers of heterogeneity. First, the starting population contains multiple different cell types that only share the expression of the CD34 protein on their surface. Second, there is some randomness in whether a gene will be successfully altered by CRISPR or not. You can try to skew the probability in favor of alteration by trying to introduce more of the CRISPR machinery into cells, but it's ever going to be 100%, the more machinery the more chances for OFF-target issues. Third, there are 2 copies of the BCL11A gene in each cell. Presumably altering both copies will generate a greater increase in HbF than just altering one copy. But there will be some randomness in whether any particular cell has one or both copies of its genes altered. Fourth, if you depend on NHEJ to generate an indel alteration, there will be some randomness in the exact sequence of the insertion and/or deletion that is generated on any DNA copy of the gene.

**23) Why not obtain a well characterized uniform population of CTX001 cells by clonal expansion from a single cell?**

Simple Answer: there are multiple disadvantages and risks of clonal expansion

Clonal expansion would have to be done from a single Lt-HSC stem cell, causing you to lose the advantage of having multiple altered progenitor cells capable of more rapidly regenerating the final differentiated blood cell populations. You are also committing to a specific sequence alteration of the two genes and possibly OF-target alteration. Given our ignorance of what is best or at least least toxic, better to provide a mix and provide some opportunity for the best and healthiest cells to come out on top.

Finally, and most importantly, clonal expansion will require culturing cells for weeks, if not months. Not only is this impractical clinically, but having cells growing many generations in highly abnormal in vitro conditions, will select for unexpected spontaneous genetic alterations.

So we are stuck with both lineage and genetic heterogeneity in our CTX001 cell population, whose heterogeneity will change after injection into the patient due to different selective pressures and requirements for survival and proliferation. The best one can do is try to characterize a few key aspects of that heterogeneity both before and after injection in the patient.

**24) How can we characterize the genetic heterogeneity generated at the BCL11A locus of the CTX001 cells?**

Simple Answer: Random sequencing of that locus from the heterogenous population of genomic DNA.

How the BCL11A locus was sequenced is not well described in Fangoul et al. However, in the end they were able to compare and collate many sequences reads randomly generated from the genomic DNA isolated from the mixed population of CTX001 cells. Several sequences occurred repeatedly (see Suppl Fig. 1). However, they also discovered some wild-type sequence reads, indicating that not every copy of the BCL11A was mutated. The percent of sequence reads that show an alteration from wild type is termed **Percent of Allelic Editing**.

We do not know what percent of cells have both of their BCL11A genes mutated, only one mutated, or neither mutated. But if we assume the mutation of each gene is completely independent of each other, and we observed that 70% of the genes were mutated (i.e underwent Allelic Editing), then we would expect 49% of the cells would have both of their gene mutated, 9% of the cells would have both genes wild-type, and 42% of cells having one mutant and one wild-type gene. The only way to know exactly what those percentages are is to sequence hundreds of single cell genomes

**25) How do the injected mixed population of CTX001 cells take over hematopoiesis in the bone marrow?**

Simple Answer: Eliminate the competition from endogenous HSPCs

You can't transplant a new heart into a patient until you first remove his old one. Similarly, you can't transplant new HSPCs into a patient's bone marrow until you first remove the original

residents. The cellular niches that support HSPCs and their hematopoietic function need to be vacated so that the injected HSPCs can find and occupy those niches. This is done by killing off the original HSPC residents with chemicals and/or radiation.

This extermination of the original HSPC population is called **myeloablative conditioning** in Fangoul et al. it is done by the addition of Busulfan, which crosslinks the two strands of DNA to each other, blocking active DNA replication, which requires the two parental template strands to separate. This severe disruption of DNA replication pushes actively dividing cells like HSPCs into apoptosis but spares the vast majority of cells in your body, which are not dividing.

With the bone marrow purged of most of its original HSPCs, there is now a place for the injected HSPCs to occupy and start producing differentiated blood cells. This process of finding the proper home and functioning productively is called **engraftment**. Engraftment can take several weeks to become apparent and full productivity may take several months.

**26) How does the patient survive without robust hematopoiesis between myeloablation and CTX001 engraftment?**

Simple Answer: With a lot of caution and transfusion support.

With myeloablation you have no ability to produce RBCs, crippled your defense against infection (reduced lymphocytes, neutrophils, and macrophages), and can't clot your blood (reduced platelets). So you are extremely vulnerable, and many of the early adverse events at the start of the transplantation will be related to these deficiencies. Until your genetically modified HSPCs can start replenishing these missing cells, you will need blood cell replacement from periodic blood transfusions from other people (i.e. **allogeneic transfusions**). Those are the transfusions labeled in Fig. 2 E and F as "**post transfusion support**).

For sickle cell disease there is an additional vulnerability. Sickle cell patient go into crisis when their bodies are stressed. Being treated with what is effectively poison for myeloablative conditioning, becoming severely immunosuppressed, and being unable to produce RBCs for several weeks is serious stress. Their RBCs are going to be sickling all over the place. What to do?

Solution, before the myeloablation and CTX001 infusion, replace the patients sickle susceptible RBCs with healthy normal RBCs from another person. This **replacement transfusion** is done gradually by bleeding the patient a bit (ironically, essentially what was done in 17<sup>th</sup> century medicine) and replacing the lost blood with a transfusion of normal RBCs. In Fangoul et al., the SCD patient underwent 8 weeks of RBC replacement transfusion before his HSPCs were mobilized.