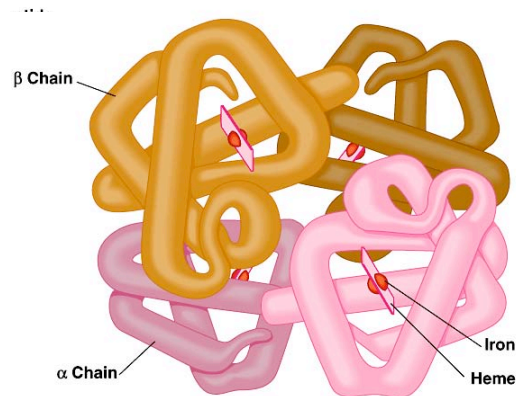


**AP LAB \_\_\_\_: SICKLE CELL ANEMIA & THE HEMOGLOBIN GENE****Using Bioinformatics in Medicine****Student Lab Guide**

Sickle cell anemia is one of the most common genetic diseases in the United States with its highest incidence in African Americans. The disease affects red blood cells and is potentially lethal. The cause of the disease has been well-documented. Hemoglobin is a protein made of four subunits: 2 alpha polypeptide chains and 2 beta polypeptide chains. A mutation in the gene for the beta hemoglobin subunit changes the 6<sup>th</sup> amino acid (glutamic acid → valine) in the polypeptide chain. This change causes the hemoglobin molecules to stick together and to form fibers under low blood oxygen conditions.



This causes red blood cells to become distorted from their normal round shape to a sickle (crescent) shape. Consequently, these sickle-shaped red blood cells clump together and clog small blood vessels causing fever, great pain, and damage to organs (including brain damage). Breakdown of damaged red blood cells also causes anemia, physical weakness, and heart failure.

The mutant allele (Hb S) is surprisingly common in the population. It is found in 1 out of 14 African Americans, which means approximately 2 million Americans are carrying the allele. This allele frequency results in 72,000 Americans having the disease.



The mutant allele (Hb S) is recessive to the normal allele (Hb A). Homozygous (Hb SS) recessive individuals have sickle cell disease. Heterozygotes (Hb AS) are carriers. If two carriers have children, each child has a 25% chance of having the full sickle cell disease. However, many carriers are not aware they have the allele. The prevalence of the mutant allele and the consequent high incidence of sickle cell disease creates a significant individual and public health burden. It would be a great benefit to the population if the medical community developed a genetic screening test to identify carriers so they could be offered genetic counseling and pre-natal testing.

**YOUR ASSIGNMENT:**

The National Institutes of Health have allocated funds and awarded your team the grant to develop a simple, inexpensive DNA test for the sickle cell allele that we can use to efficiently screen large groups of people.

**THE PLAN OF ACTION**

- Identify the DNA sequence for the normal hemoglobin allele.
- Identify the DNA sequence for the sickle cell mutant hemoglobin allele.
- Design DNA probes to detect normal and sickle cell hemoglobin alleles in blood samples.

**IMPLEMENTING THE PLAN****1. Identify the DNA sequence for the normal hemoglobin allele**

Thanks to the Human Genome Project, the DNA code for all 23 human chromosomes has been sequenced. This information is organized and stored in databases that are freely available to both the public and the scientific community and accessible online at a number of Web sites. You just need to know how to use the analysis tools that enable you to sort through all that DNA code. One of the goals of this exercise is to teach you how to use these tools that are freely available.

- a. We will be using the genome database available through the University of California at Santa Cruz (UCSC).
- b. Go to the UCSC Genome Browser: <http://genome.ucsc.edu>
- c. Click on the “Genome Browser” link in the side navigation bar (See Figure 1).

This will bring you to the “Human Genome Browser Gateway”—a tool that allows you to search the human genome and retrieve gene information in an organized way.

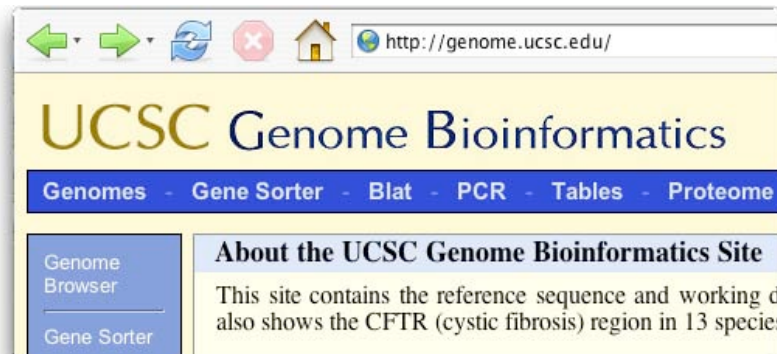


Figure 1. The Genome Browser homepage at University of California at Santa Cruz (UCSC).

- d. The Genome Browser asks you to:
  - (1) choose which genome you want to search in the genome pulldown menu. As you can see there are DNA sequence data stored here for a number of other organisms besides humans.

- (2) You need to choose which version of the database to use.
- (3) Identify which gene you are searching for.

From experience, we know the best way to query the Genome Browser database is with the following settings (see Figure 2):

- human genome
  - July 2003
  - HBB (this is the standard notation for the Hemoglobin beta chain)
  - image width can be left at the default 620
- e. Click the “submit” button

The screenshot shows the 'Human Genome Browser Gateway' search interface. At the top, it says 'The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#). Software Copyright (c) The Regents of the University of California. All rights reserved.' Below this is a search form with four fields: 'genome' (dropdown menu set to 'Human'), 'assembly' (dropdown menu set to 'July 2003'), 'position' (text input field containing 'HBB'), and 'image width' (text input field containing '620'). A 'Submit' button is to the right of the 'image width' field. Below the form, there is a link: '[Click here to reset](#) the browser user interface settings to their defaults.' and a button: 'Add Your Own Custom Tracks'.

Figure 2. The settings for the Human Genome Browser to search for the DNA sequence of the human hemoglobin beta chain.

- f. The search results are returned to you on the next page. You are specifically interested in the “RefSeq Genes”. This is the link to the “reference sequence”: the sequence that is considered the authoritative version. (see Figure 3):

The screenshot shows the 'RefSeq Genes' section of the search results. It lists three genes with their RefSeq IDs and descriptions:

- [HBB at chr11:5203302-5204770](#) - (AF181832) Hemoglobin beta subunit variant (Fragment).
- [HBB at chr11:5203270-5204877](#) - (BC007075) hemoglobin, beta
- [HBBP at chr3:45240961-45242758](#) - (AF438313) Ras-induced senescence 1 (P40BBp).

Below this, there is a section titled 'RefSeq Genes' with one entry:

- [HBB at chr11:5203271-5204877](#) - (NM\_000518) beta globin

Figure 3. The link to the Reference Sequence for the human hemoglobin beta chain.

Built into the link to the HBB reference gene sequence is a lot of information:

**Question 1.**

Copy the RefSeq link below:

\_\_\_\_\_

Look at the text of the RefSeq link. Hypothesize as to what each part means:

HBB \_\_\_\_\_

chr11 \_\_\_\_\_

5211004 \_\_\_\_\_

5212610 \_\_\_\_\_

- g. Click on the HBB reference sequence link to view the HBB gene on chromosome 11 (see Figure 4).

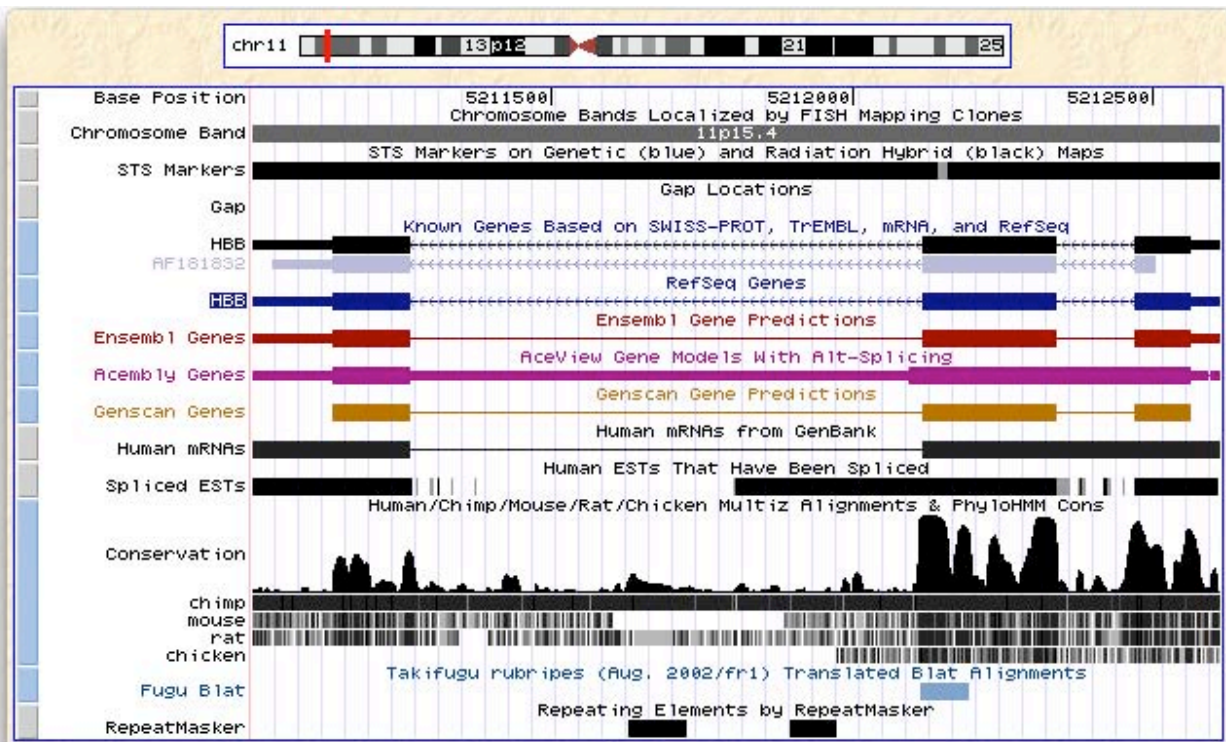


Figure 4. The Genome Browser view containing a portion of chromosome 11, showing the human hemoglobin beta chain (HBB) gene.

This is a lot of information!

Near the top of the page is an illustration of the entire human chromosome 11 (see Figure 5).



Figure 5. An illustration of the entire human chromosome 11 from the UCSC Genome Browser.

### **Question 2.**

What do you think the red line on the illustration of human chromosome 11 is marking?

---



---

### **Question 3.**

If the HBB gene is at base 5.2 million, how long do you think human chromosome 11 is (in bases)?

---



---

- h. The position of the HBB gene is shown about 4 lines down in the main panel (labeled HBB). The reference sequence (RefSeq Genes) for HBB is shown in blue. The thicker parts of the line are exons and the arrows are introns. The direction of the arrow shows the direction of transcription (See Figure 6).

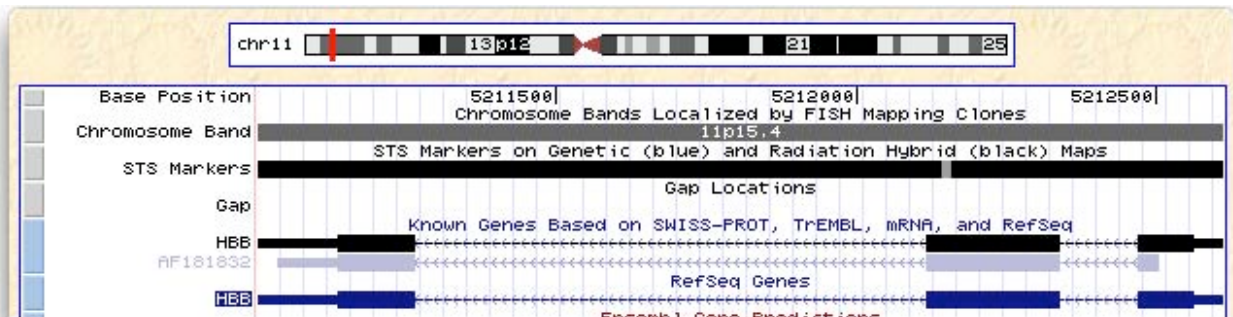


Figure 6. The HBB gene on human chromosome 11.

- i. Above the chromosome illustration are a series of navigational tools, we'll call the "move and zoom" controls (see Figure 7). Practice using the left and right "move" arrows and the "zoom in" and "zoom out" controls. If you get lost, just type "HBB" in the "position" textbox and hit the "jump" button to get back to this original view.

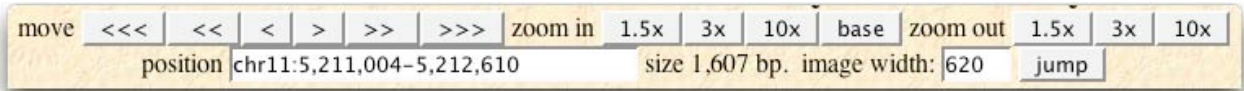


Figure 7. The "move and zoom" controls in the UCSC Genome Browser.

- j. Now zoom out 30x. How are going to do that?

Also move a little to the right with one click of the right arrow. This brings up a cluster of interesting HB genes (HBD, HBG) in the main panel (See Figure 8).

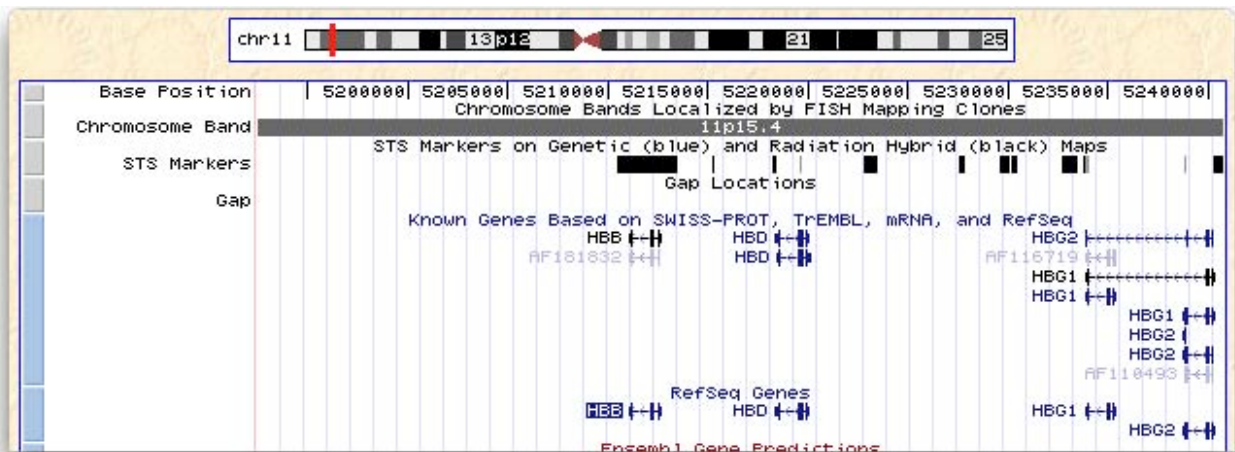


Figure 8. The cluster of HB genes on human chromosome 11.

**Question 4.**

What are the HBD and HBG genes on human chromosome 11?

---



---



---

Hypothesize why these HB genes are found in a cluster?

---



---



---

- k. Click on the HBB reference sequence (RefSeq Genes) shown in blue. This brings you to the RefSeq page for the HBB gene. Scroll down to the link “Genomic Sequence from assembly” (See Figure 9).

**Position:** [chr11:5211005-5212610](#)  
**Band:** 11p15.4  
**Genomic Size:** 1606  
**Strand:** -

**Links to sequence:**

- [Predicted Protein](#)
- [mRNA Sequence](#) may be different from the genomic sequence.
- [Genomic Sequence](#) from assembly

Figure 9. The link to the Genomic Sequence for the HBB gene.

- l. Click on “Genomic Sequence from assembly” link. This brings you to a formatting page that allows you to structure the way your DNA sequence will be displayed. Most of the default settings are fine. Just make sure that the radio button is selected for “Exons in upper case, everything else in lower case.” Hit the “submit” button.
- m. Now you finally have your HBB gene sequence (See Figure 10). Copy the sequence and paste it into a word processing program.

**Question 5.**

How many introns? \_\_\_\_\_

How many exons? \_\_\_\_\_

Are there more DNA bases in the introns or the exons of the HBB gene?

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Do you think this is a large or small gene?

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

```
>hg16_refGene_NM_000518 range=chr11:5211005-5212610
ACATTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAACAGACACC
ATGGTGCATCTGACTCCTGAGGAGAAGTCTGCCCTTACTGCCCTGTGGGG
CAAGGTGAACCTGGATGAAGTTGGTGGTGGAGCCCTGGGCAGgtggtat
caaggttacaagacaggtttaaggagaccaatagaaactgggcatgtgga
gacagagaagactcttgggttctgatagggactgaactctctgcctat
tggctctatttcccacccttagGCTGCTGGTGGTCTACCCTGGACCCAG
AGGTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGG
CAACCCTAAGGTGAAGGCTCATGGCAAGAAGTGTCTGGTGCCTTAGTG
ATGGCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGT
GAGTGCACCTGACAAGCTGCAGTGGATCCTGAGAACTCAGGgtgag
tctatgggacgcttgatgtttcttcccctctcttctatgggttaagtt
catgtcataggaaggggataagtaaacagggtagcagtttagaatgggaac
agacgaatgatgcatcagtggaagctcagcagatcgttttagttctct
ttatttgcgtgtcctaacaattgtttcttcttggtttaattcttgcctct
ttttttctctccgcaatttttactattatacttaatgccttaacatt
gtgataacaaaagaaaatactctgagatacattaaagtaacttaaaaaa
aaactttacacagctgcctagtacattactatttggaaatatagtgtgct
ttatttgcattatcataatctcctactttatttcttttatttttaatt
gatacataatcattatacatatttatgggttaagtgtaattgttttaata
tgtgtacacataatgaccaaatacagggtaatttgcatttgaatttttaa
aaaaagctttcttcttttaataatactttttgtttatcttatttcttaata
ctttcccatactcttcttctcagggcaataatgatacaaatgatcatcgc
ctctttgcaccattctaaagaataaacagtgataatttctgggttaaggca
atagcaataatctctgcatataaaatatttctgcatataaaatgttaactgat
gtaagaggtttcatattgctaataagcagctacaatccagctaccattctg
cttttatttatggttgggataaaggctggatattctgagctccaagctag
gcccttttgcataateatgttcatacctcttctctcccccagCTCCT
GGGCAACGTGCTGGTCTGTGTGCTGGCCATCAGTCTTGGCAAGAATTC
CCCCACAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGGCTAAT
GCCCCGGCCCAAGTATCACTAAGCTCGCTTTCTTGTCTGCAATTCT
ATTAAGGTTCTTTGTTCCCTAAGTCCAACCTACTAAACTGGGGGATATT
ATGAAGGCCTTGACATCTGGATTCTGCCTAATAAAAAACATTTATTT
CATTGC
```

Figure 10. The genomic DNA sequence for the human HBB gene.

## 2. Identify the DNA sequence for the sickle cell mutant hemoglobin allele.

Now its time to get the sickle cell mutant hemoglobin allele.

- Go back to the Genome Browser main panel that showed the HBB gene on the map of chromosome 11 (as seen in Figure 4).
- To get the sequence of the mutant allele, we need to add information about genetic variations to the map. To do this, scroll all the way to the bottom of this Web page to the section entitled "Variation and Repeats". Set the SNPs pulldown to "pack". SNPs are single nucleotide polymorphisms — single base mutations (See Figure 11). Hit the "refresh" button.



Figure 11. The "Variation and Repeats" section of the Genome Browser.

- Now the main panel has more information in it. At the bottom of the main panel there is a section entitled "Simple Nucleotide Polymorphisms". Each of these is a known mutation of the human HBB gene. We are interested in the "rs334" polymorphism. This is the sickle cell mutation (See Figure 12).

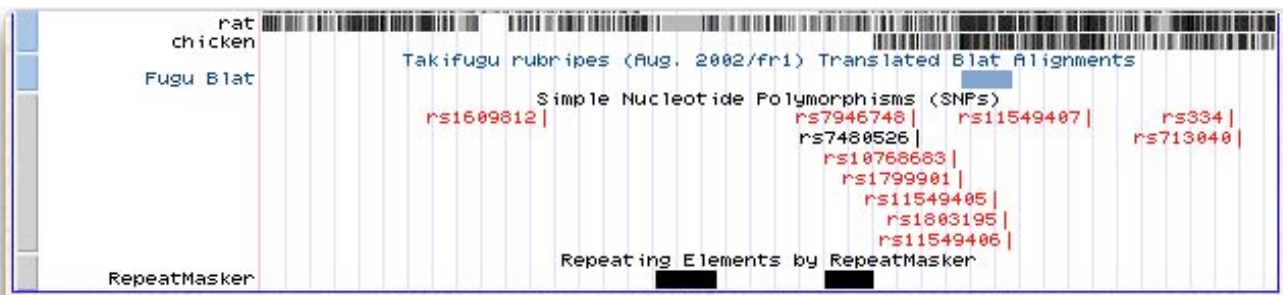


Figure 12. Single nucleotide polymorphisms of the human HBB gene.



**Question 6.**

Is the sickle cell mutation in an intron or an exon? \_\_\_\_\_

Does this make sense? \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Is the mutation at the beginning or end of the gene?

\_\_\_\_\_

\_\_\_\_\_

- d. Click on the “rs334” polymorphism and you will be forwarded to a page summarizing the information about this SNP (See Figure 13).
- The “Sequence in Assembly” is a part of the sequence of the normal allele.
  - The “Alternate Sequence” is a part of the mutant allele showing the single base mutation.

### Simple Nucleotide Polymorphism (SNP) rs334

**Position:** [chr11:5212541-5212541](#)

**Band:** 11p15.4

**Genomic Size:** 1

[View DNA for this feature](#)

**Average Heterozygosity:** Not Known

**Standard Error of Avg. Het.:** Not Known

**Functional Status:** coding-nonsynon, reference

**Validation Status:** no-information

**Allele1:** A

**Allele2:** T

**Sequence in Assembly:** catggtgcacctgactcctgAggagaagtctgccgttactg

**Alternate Sequence:** catggtgcacctgactcctgTggagaagtctgccgttactg

**Variant Source:** OTHER

**Variant Type:** SNP

Figure 13. Summary information for SNP rs334, the sickle cell mutation of the HBB gene.

- e. Copy these two lines of DNA sequence into your word processing file. You will need them in a minute.

**Question 7.**

What is the single nucleotide change that creates the sickle cell mutation?

---



---

### 3. Design DNA probes to detect normal and sickle cell hemoglobin alleles in blood samples.

These two short sequences are enough to design the DNA probes that we will use to detect both the normal and sickle cell alleles in human blood samples. This will allow us to screen people, so we can alert them if they are carriers and can also be used as a prenatal test to detect sickle cell disease.

- a. We need to use a probe design tool. One of the best probe design programs, Primer3, is available for free on the Web from MIT. Either Google “Primer3” or go to:

[http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)

Primer3 is very complex and powerful, but it can be used without too much hassle by leaving most of the advanced settings at their default values.

- b. Paste the sickle cell DNA sequence fragment into the Primer3 text box. Place “curly brackets ({} )” 10 bases on either side of the mutation. This will force Primer3 to include that mutation base within the probe. (See Figure 14).

Also use the following settings:

- check “Pick hybridization probe”
- uncheck “Pick left primer” and “Pick right primer”

Primer3		<a href="#">disclaimer</a>	<a href="#">source code</a>
pick primers from a DNA sequence		<a href="#">cautions</a>	<a href="#">FAQ</a>
Paste source sequence below (5'→3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FA format ok. Please N-out undesirable sequence (vector, ALUs, LINES, etc.) or use a <a href="#">Mispriming Library (repeat library)</a> :			
<input type="text" value="NONE"/>			
<pre>catggtgcacc{tgactcctgTggagaagtc}tgccgttactg</pre>			
<input type="checkbox"/> Pick left primer or use left primer below.	<input checked="" type="checkbox"/> Pick hybridization probe (internal oligo) or use oligo below.	<input type="checkbox"/> Pick right primer or use right primer below (5'→3' on opposite strand).	

Figure 14. Primer3 probe design tool.

c. You also need to adjust the settings for the size and hybridization temperature of the primer. Scroll far down on the page to the “Hyb Oligo (Internal Oligo) General Conditions” section (See Figure 15). Use the following settings:

- Hyb Oligo Size— Min: 12, Opt: 14, Max: 16
- Hyb Oligo Tm— Min: 40, Opt: 50, Max: 60

Hyb Oligo (Internal Oligo) General Conditions						
<a href="#">Hyb Oligo Size:</a>	Min	<input type="text" value="12"/>	Opt	<input type="text" value="14"/>	Max	<input type="text" value="16"/>
<a href="#">Hyb Oligo Tm:</a>	Min	<input type="text" value="40.0"/>	Opt	<input type="text" value="50.0"/>	Max	<input type="text" value="60.0"/>
<a href="#">Hyb Oligo GC%:</a>	Min:	<input type="text" value="20.0"/>	Opt:	<input type="text"/>	Max:	<input type="text" value="80.0"/>

Figure 15. Probe design settings in the Primer3 probe design tool.

d. Click the “Pick Primers” button. You will now get a probe design to optimally hybridize with the sickle cell gene in a human blood sample (see Figure 16). Record your probe sequence for Question 8.

## Primer3 Output

```
Using 1-based sequence positions
OLIGO      start  len  tm  gc%  any  3' seg
INTERNAL_OLIGO  15  20  59.99  60.00  5.00  4.00 ctctctgTggagaagtctgcc
SEQUENCE SIZE: 41
INCLUDED REGION SIZE: 41

1 catggtgcacctgactcctgTggagaagtctgccgttactg
   ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
```

Figure 16. Probe sequence obtained from Primer3 probe design tool.

e. Repeat this probe design process with the normal HBB gene sequence fragment. Record your probe sequence for Question 8.

### **Question 8.**

Write the sequence of both probes below:

normal \_\_\_\_\_

sickle cell \_\_\_\_\_

Name \_\_\_\_\_

Period \_\_\_\_\_

- f. You now can send your probe sequences to your local DNA synthesis lab and you will be ready to start testing for sickle cell next week.

**Question 9.**

Now that you have two probes for the normal and mutant alleles, how can you use them to distinguish between normal, sickle cell disease and carrier genotypes?

---

---

---

---

---

---

---

---

---

---

Why do we need two probes for the genetic test: one probe for the normal and one for the mutant allele?

---

---

---

---

---

---

---

---

---

---

**EXTRA CREDIT:** Find the closest commercial DNA synthesis lab and estimate the cost of two 15 base probes.

---

---