

PCR

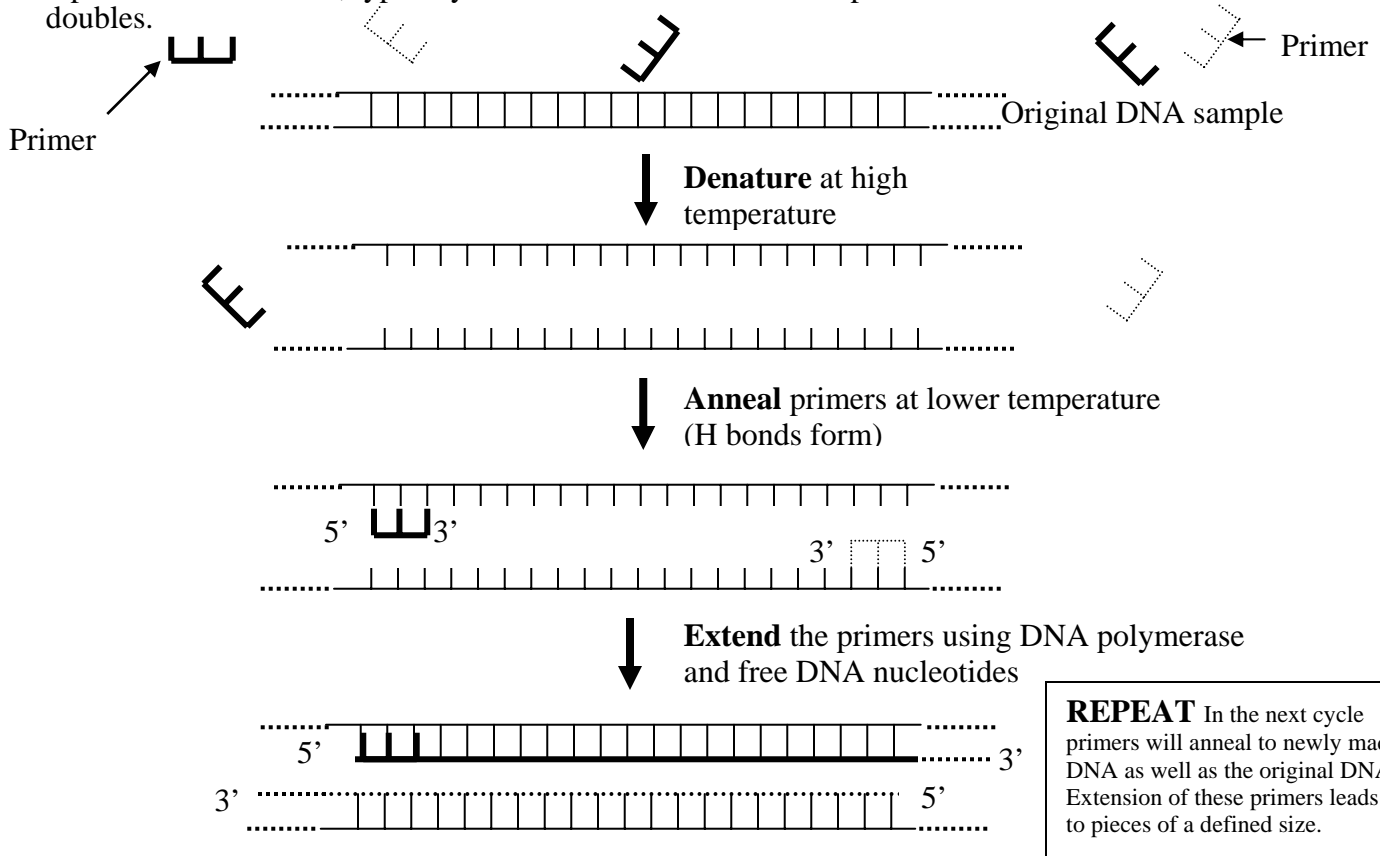
What is PCR?

The **polymerase chain reaction** (PCR) is a relatively recent invention (1983) but is widely used in biology and forensic analysis. It enables researchers to **amplify** (make more of) a selected piece of DNA present in a sample. This sounds mundane, but is actually incredibly useful. PCR enables investigators to study DNA present in bloodstains, or in the cells clinging to a hair at the scene of a crime. It has allowed the isolation of DNA from preserved extinct animals (but does not enable us to recreate dinosaurs!) or the isolation of a desired gene out of all of the DNA present in an organism's genome. It has been used to show that HIV is present in the bodies of infected individuals years after initial infection (usually the HIV virus is present at such low levels that, while it is evident from antibodies that the body has been exposed to virus, the virus itself cannot be found. Now, using PCR, doctors can assess the amount of virus a patient carries). The Nobel Prize was awarded to Kary Mullis for the invention of PCR (a true off-beat Berkeleyite, he went surfing to celebrate).

Overview of the reaction

In most types of PCR, short single-stranded pieces of DNA called **oligonucleotides** or "**oligos**" (that can be made to order) must **hybridize or anneal** (bind via their complementary nucleotides) to the DNA flanking the desired sequence. Thus, for most applications, you need to know something about the sequence flanking the DNA you want to amplify. These oligos are also referred to as **primers** since they will serve to mark the starting line of DNA synthesis (they "prime" the synthesis to get it started).

The steps of a basic reaction are as follows: The crude DNA sample is added to a tube along with a solution of salts at the correct pH, the two primers, and all four deoxyribonucleotides. The sample is heated to denature the crude DNA into single strands. The tube is then cooled somewhat so that the primers can hybridize to their complementary or "**target**" sequences (there is so much primer in the tube that the target sequence is more likely to bind to an primer than to find its true partner again). Then **DNA polymerase** is used. This enzyme (protein) adds nucleotides to the 3' ends of the annealed primers, using the crude DNA as a template. This three step cycle-- **denature, anneal, extend**-- is repeated over and over, typically about 30 times. With each repetition the amount of the desired DNA doubles.



The discovery of microorganisms that can live at very high temperature (around boiling temperature), and the isolation of their DNA polymerases made PCR much easier. As you may know, most proteins are not stable when heated to high temperature, as must be done to denature the DNA during PCR (think of what happens to egg protein when you heat it). So DNA polymerase isolated from a typical source like *E. coli* (a type of bacteria found in the gut) would only last for one cycle, as it would be "cooked" during the next denaturation step and have to be added again (30 times! very tedious!). However, heat-stable DNA polymerases from organisms like *Thermus aquaticus* (e.g. **Taq polymerase**) are stable through at least thirty cycles, so are only added once. Yeah! Another nifty development was the PCR machine, which heats and cools according to a program set by the user. Otherwise, the researcher has to have three waterbaths or hot blocks at different temperatures (one for denaturing, one for annealing, and one for extending) and move tubes from one to the other every minute or so for several hours. Ugh! Aren't you glad you don't have to do that!

Our PCR project

To do PCR you must have a source of trace DNA, two primers that will stick to that DNA with their 3' ends pointing toward each other, the correct salt and pH conditions, a bit of magnesium, all four nucleotides, and of course the enzyme that makes the DNA, DNA polymerase. We will amplify DNA from your cheek cells. We will use primers that bind on either side of a section of human DNA with the catchy name **vesicular monoamine transporter 2 (VMAT2)**. This is the gene examined by Dean Hamer and colleagues and described in his book [The God Gene](#).

Within the VMAT2 gene there are many bases that can vary from person to person. Information on these differences, called SNPs (single nucleotide polymorphisms), is collected by researchers because they may explain difference in how the gene performs or they can serve as useful variations to follow in DNA-fingerprinting-type studies. Many VMAT2 SNPs are listed in the government National Center for Biotechnology Information (NCBI) database of SNPs:

http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=6571&chooseRs=all.

Another website lists VMAT2 SNPs as part of a study of how differences in membrane transporter proteins may affect how patients respond to drugs (<http://pharmacogenetics.ucsf.edu/index.html> - follow "available data" link). We will be looking at three SNPs drawn from the latter list. One of these is potentially the SNP described by Hamer in [The God Gene](#) – due to differences in how the DNA sequences are numbered since Hamer wrote that book, it was not possible to identify A33050C as such. However, there is a location on the VMAT2 DNA where either an A or a C may be present, and the variation occurs at the frequency that Hamer describes.

The SNPs cannot be detected by PCR in the way we will use it. There is a way to use PCR so that you have your primers stick to the region that varies, described on page 68 of [The God Gene](#). One fine-tunes conditions so that the primers don't stick if there is even one base that does not match – but this is experimentally tricky, so I decided not to use it. Our PCR is just going to make lots of copies of part of the VMAT2 gene that is known to contain the three sites of variation. We will then purify this DNA, and have its sequence analyzed in separate steps.

PCR is wonderful because you can start with a sample that has all sorts of DNA (in our case, every gene in a human, plus any from bacteria in your mouth) and make a billion copies of just one bit of it. The original DNA can be at such trace levels that it is insignificant compared to the newly synthesized DNA in most types of analyses. Thus, PCR allows one to "zero in" on a particular section of DNA. PCR is widely used in forensics as well as in research, again because only tiny amounts of sample DNA are needed to generate large amounts of the region of interest. In addition, as we shall see, DNA does not have to be very pure for this technique to work. These features are very useful when dealing with trace DNA samples that might have been scraped off a sidewalk or off ancient bones.

How will we tell if we are successful in getting amplification in the PCRs? We will run our samples on something called an agarose gel that will separate them by size. We will have ethidium bromide, which stains DNA, present in the gel. The amount of original crude cheek cell DNA used will be too little to see on the gel, but if there is a substantial amount of PCR product it should be visible. More on this later....

Methods

Warning: Because PCR is so sensitive you must be very careful not to cross-contaminate samples with DNA from other people. Many researchers have tried to amplify genes from one source and ended up amplifying genes from other DNA that was around in the lab. Use a fresh tip when you pipet a new solution!! Try to keep your hands and hair out of your partner's tubes. DNA from food or bacteria is not of such great concern as the primers bind to sequences apparently found only in human DNA.

Isolation of DNA from cheek cells – Each student should do this

- Obtain a tube containing 10 ml of salt solution (0.9% sodium chloride) and a paper cup. Label the tube with your name. Pour the solution into your mouth, and swish it around vigorously for 10 seconds. Expel the solution into the paper cup, then pour it back into the tube (the tubes are brand-new, and the salt is ordinary table salt, in case you are worried!). Cap the tube. Toss the cup.
- Spin the tube (balanced opposite another tube!) in the clinical centrifuge for 10 min. The clinical centrifuge is the "spaceship on a pedestal"-shaped thing. This will collect all of your cheek cells (as well as other gross stuff) into a pellet at the bottom of the tube.
- Carefully pour off as much as possible of the liquid into the waste container (contains bleach). The pellet should stick to the tube. If you have a lot of liquid left in the tube you can pipet some of it out. Put the tube on ice.
- Cut off the end of a blue P-1000 pipetman tip to make a wider bore hole at the end. You don't have to cut off very much, just about an eighth to a quarter of an inch. This is done so the Chelex beads used in the next step don't jam the tip. Set the P-1000 micropipettor (the one with the blue button) to 500 μ l (050 on the dial).
- Find the tube of Chelex. Using the P-1000 and the cut-off tip, pipet the Chelex in and out of the tip several times to suspend the tiny beads of Chelex, then quick quick quick before the beads settle pipet up 500 μ l of the Chelex suspension. Put the 500 μ l of Chelex suspension into your tube of cells.
- Suspend your cells in the solution by stirring with the pipetman tip and pipeting the solution in and out of the tip several times. Make sure that there are no chunks of cells left when you are done.
- Transfer 500 μ l of the cell/Chelex mixture to a labeled eppendorf tube (small plastic tube). Wipe the ejector and barrel (the white part) of the pipetman with 70% ethanol before giving it to someone else to use. This is to prevent cross-contamination of DNA samples and also to eliminate any other "cooties" you may have transferred to the barrel.
- Place a lid lock on your tube to prevent it from popping open with the heat and heat the cell/Chelex mixture in a hot block at (or near) boiling temperature for 10 min.

The Chelex binds to dissolved metals that would otherwise inhibit the PCR reaction or allow the DNA to be chopped into many pieces. In a part of your cells called the lysosome there are enzymes that cut up DNA and could ruin our experiment. However, these enzymes require certain metals when they cut DNA, so you can inactivate them by hiding all the metals in the Chelex. The solution is also very basic (the opposite of acidic), which causes the cells to rupture when they are heated, releasing the DNA into the solution.

- Carefully remove the tube from the hot block and put it on ice for a minute or so to cool it.

- Remove the lid lock and spin the sample in the microcentrifuge for about 30 sec to pellet the Chelex and cell debris.
- Transfer 200 μ l (use P200 – orange button, straight barrel, set to 2-0-0) of the cleared solution to a new tube, pipeting from the top of the liquid so that you do not get any of the Chelex or cell gunk.

This is your DNA sample. Clearly it has a lot of other junk in it that was released from the cells, but this should not harm the PCR reaction.

Setting up the PCR reaction

- Get a PCR Ready-to-Go tube and label the TOP with your name or initials.

These tubes are designed so that they don't pop open during repeated heating and cooling.

- In the tube you should see an expensive little bead. This bead contains stablized Taq DNA polymerase, buffers (which control the acidity or pH of the solution), and all four deoxyribonucleotides (A, G, C, and T). Make sure the bead is at the bottom of the tube.
- What's still missing? Primers and your DNA. First add the primers – they should be in a tube labeled PCR mix. Add 20 μ l using a P20 pipettor (yellow button, pinched barrel, set dial to 2-0-0). The bead will begin to dissolve.
- Now add your DNA – remember, this is the tube of stuff you just made in the previous part of the lab. Use 5 μ l (set 0-5-0 on the P20) of your DNA. Mix the tube contents and centrifuge for a few seconds if necessary. Ask the instructor how to spin these small tubes. Put your tube on ice until everyone is ready.
- When everyone is ready, put your tube into the PCR machine.
- The machine will heat the tubes to 95°C for two minutes to denature the DNA. Then it will cycle to the following temperatures 35 times:

92°C for 90 seconds to denature DNA

53°C for 1 minute to allow primers to stick to the DNA

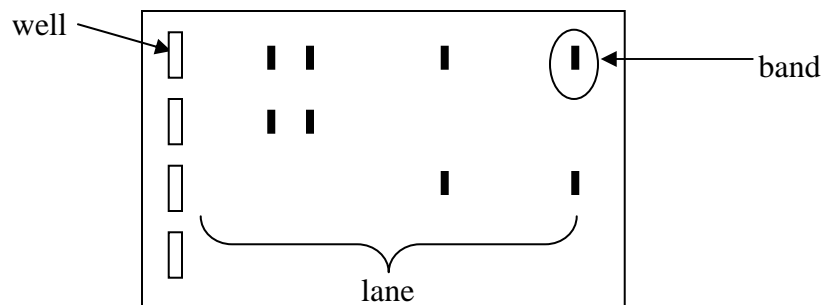
72°C for 2 minutes to allow Taq polymerase to add new nucleotides to the ends of the annealed primers.

At the end of the 35 cycles the machine is programmed to sit at 72° for 5 minutes, in case there are any single stranded ends that need to finish being replicated. The whole reaction takes a few hours, so I will put the tubes away for you. Next time we will load the samples on a gel to see if any amplification of the VMAT2 region occurred.

Agarose gel electrophoresis

To determine if your PCR was successful, agarose gel electrophoresis will be used. This technique is particularly convenient in this case as it can also be used to purify our desired PCR product away from any minor products that have formed.

Agarose gel electrophoresis reveals the number and sizes of the DNA fragments in a sample. Agarose is a jelly-like substance isolated from seaweed. An agarose gel is made by melting a solution of agarose and water (and buffers and salts) and pouring it into a mold. The mold has a piece called a **comb** that makes **wells** in the agarose block. The agarose gel is submerged in a solution, the DNA sample is placed in the wells, and an electric current is applied. DNA is negatively charged (the phosphates are negative), so it is attracted to the positive pole and moves toward it through the agarose. Small fragments of DNA move through the microscopic pores in the gel more quickly than bigger pieces because it is easier for them to fit through. Thus, the DNA fragments are separated by the gel according to their size. The DNA can be visualized because the gel contains a DNA-binding dye, **ethidium bromide**, which fluoresces in a pinkish-orange color when illuminated by UV light. Each visible "pile" of DNA is called a **band**, and the path that the DNA travels in is called a **lane**.



Since many PCR reactions, including this one, can give rise to minor products it is best to purify the particular band we are interested in. A first step in this purification will be to carefully cut out the band of interest from the gel and save it in a tube. We will later extract the DNA from this piece of agarose and send it to a DNA sequencing facility to have its sequence determined.

Methods

I'll demonstrate how an agarose gel is made. You can all use the same gels for your samples.

CAUTION!! Ethidium bromide is a mutagen, so wear gloves whenever you handle or load gels. We will be working with low levels of ethidium bromide in this experiment. In addition, **electrophoresis involves electrical current high enough to cause severe injury.** NEVER touch an electrophoresis apparatus unless the power supply is turned off and the leads are disconnected. As an extra precaution, connect gel to the power supply only when it is turned off and using only one hand.

- To your PCR tube add 6 μ l of loading dye buffer (green-blue solution) to the reaction. Mix by flicking tube or with a vortex mixer. Spin tubes briefly in a

microcentrifuge if any liquid is not on the bottom of the tube (as with any centrifuge, remember to arrange tubes in a balanced way!).

The loading buffer has a dye that migrates in an electric field, which helps you follow the progress of electrophoresis. This dye does NOT bind DNA, but moves through the gel faster than most DNA pieces will. The loading buffer also contains glycerol, which makes the sample dense so it will sink into the wells of the gel.

- Locate an empty well on an agarose gel. **Write down (or draw) the location of this well so that you can identify your samples later.** When the gel is ready to use it should be submerged in buffer (a clear liquid solution) in the gel box.
- Load your sample onto the submerged gel as follows: Wear gloves. Draw up 20 μ l of the sample carefully with a P20 Pipetman, making sure that you don't make bubbles or draw air into the tip. Steadying your pipeting hand with the other hand, or against the side of the gel box, put the tip into the first well. **DON'T put the tip all the way to the bottom of the well or you will punch a hole in the well and all the sample will leak out.** Carefully let out the sample into the well, then remove the Pipetman from the well before you let up on the plunger. If there is more room in the well, draw up another 7 microliters of sample (avoiding air bubbles - stop letting the plunger up if any air is going to go into the tip) and try loading that into the same well. The goal is to not to slop over into an adjacent well, since that would cross-contaminate samples, but to get as much of your sample as is possible into the well. **For each row of wells make sure someone loads a sample containing DNA of known sizes for comparison (molecular size standards) in its own well; five to ten microliters of this marker solution should be loaded.** Each row should be able to hold six samples in all.
- When everyone's samples are loaded, connect the gel so that positive (red) pole is farthest from the samples. Run the gel at the voltage suggested by the instructor (probably 100V).

The gel will be illuminated with UV light and photographed when it is done running. After we photograph the gel, we will want to cut out the major product for purification. If all is well, the size of this product should be 341 bases.

- Place the gel on a piece of plastic wrap on the UV light source. Wear goggles and gloves to protect yourself from the UV light and the ethidium bromide. Turn on the UV light.
- When it is your turn to cut out your band, wear a face shield to further minimize UV exposure (the closer you are to the light, the more strong the rays) and cut lightly around the band with a fresh (so as not to cross-contaminate) razor blade (I usually cut a section all the way to the edge of the gel so the piece can be slid out sideways). Trim off any agarose that is not glowing. Avoid any adjacent bands (in my samples there was a small band directly above the major band).

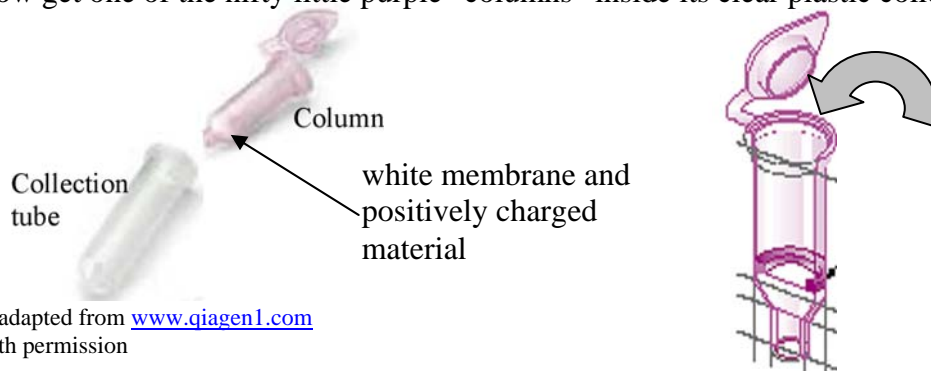
- Pick up the glowing band with the razor blade and let it fall into a clean eppendorf tube. Close and label the tube. Refrigerate until next class.

Extracting your DNA from the agarose gel

Last lab you ran your DNA through an agarose gel to see if PCR had successfully amplified your DNA. If you got a PCR product we can now extract it from the agarose gel. Then we will send it to Cal State Northridge's DNA sequencing facility, where its base sequence will be determined. It is possible to determine the base sequence in an ordinary lab, but people nowadays generally send their DNA out to a facility to be sequenced, since the process can be automated and, hence, is relatively inexpensive (we pay \$15 per sample). DNA sequencing uses PCR to incorporate nucleotides that have colored labels into new copies of the DNA. The labels can then be detected in an ordered fashion during electrophoresis (there's some more key details, but this is not the place for them, since they get a bit involved). A computer compiles the sequence and we are sent a computer file with the sequence via email.

Methods:

- Last lab you cut a band containing your DNA out of an agarose gel. If you have not already done so, determine the weight of your gel slice by taring (zeroing) the balance with an empty tube on the pan, then gently removing the empty tube from the balance pan and putting your tube with its gel slice in its place. The scale will read in grams. Convert to milligrams by multiplying by 1000 – for example if the scale read 0.25g, then your slice weighs 250 mg. See me if your gel slice weighs more than 400 mg.
- Multiply the number of mg in your gel slice by 3. Add this amount of buffer QG in microliters (all these solutions come from a kit and they all have weird two-letter names. QG is the yellow/orange solution). For example, if your slice weighs 250 mg, then add 750 μ l of buffer QG by setting a P1000 (the one with the blue tips) to read “075” on its dial.
- Put the gel slice/buffer mixture in a 50°C hot block (just to give you an idea about how hot this is, this is the temperature of hot water from the tap) for about 10 minutes. You want the gel slice to completely dissolve. You can help this along by mixing the tube every two or three minutes. Make sure there are no little bits left before proceeding. Also confirm that the solution is still the same color as it was (yellow-orange) and not deep orange or violet before proceeding.
- Recall the weight of your gel slice. Add this volume in microliters of isopropanol (rubbing alcohol) to your solution. To continue our example from above, if your gel slice weighed 250 mg, then you would add 250 μ l of isopropanol to your tube by setting a P1000 to “025.”
- Now get one of the nifty little purple “columns” inside its clear plastic collection tube.



images adapted from www.qiagen1.com
used with permission

- Label each part – the top of the purple column and the side of the collection tube are good spots – with a Sharpie marker. The white stuff at the bottom of the tube is a thin membrane that covers some positively charged material. Don't stab the membrane with a pipet tip at any point during the procedure. The DNA will stick to the positively charged material because DNA is negatively charged (all those phosphate groups). Open the top of the column and pipet your sample (up to about 700 μ l) inside. If your sample is larger than about 700 μ l, just put about half of it on the column, go on to the next step (spin in the centrifuge, empty collection tube) and then put the rest of the solution into the now-empty column and repeat the spin. What you want to accomplish is to have all of the solution be filtered through the white column material.
- Spin your column, in its clear collection tube, in the microcentrifuge for one minute. The solution will filter through the white material and be captured in the collection tube. The DNA is now stuck to the white material in the column. The other components of the gel and the buffer QG are in the collection tube. Discard this solution and replace the column in the collection tube.
- The next few steps are designed to wash the white column material and its adherent DNA to remove any traces of gel or other contaminants: Add 500 μ l of buffer QG to your column ("050" on a P1000) and centrifuge for one minute to wash buffer through the column. Discard solution from the collection tube and replace the column in the tube.
- Add 750 μ l of buffer PE to the column. Let the column sit with the buffer in it for 2-5 minutes. This will leach any salts (they might interfere with the sequencing) out of the column.
- Spin the column for 1 minute to move the PE buffer through the column into the collection tube. Empty the collection tube well and replace the column in the tube.
- Spin the empty column and collection tube again to get rid of any tiny bits of remaining buffer PE that is stuck in or on the column.
- Discard the collection tube and put the column into a clean eppendorf tube. Carefully pipet 30 μ l of EB (set a P200 to "030" and use a yellow tip) onto the center of the white membrane without stabbing the membrane. Let the buffer sit one minute on the column and then spin in the microcentrifuge for one minute. The EB buffer releases the DNA from the column, so now the DNA is in the eppendorf tube. I will send this DNA to a facility that will determine its sequence.

Analyzing Variations in VMAT2

We have successfully obtained sequence of VMAT2! Hooray for us! Now, it is time to take a look and to see what variations there are.

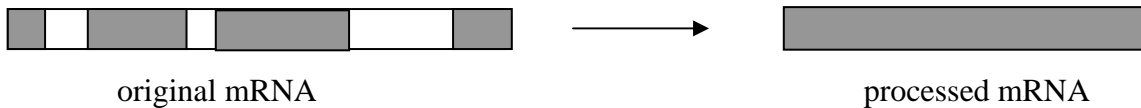
There are three sites known to vary on the piece of VMAT2 we are looking at. The section that we amplified with PCR is shown below. Only one strand is shown for simplicity and the bases are grouped into sets of 10 to make finding things easier:

```

cccaaa gccttattgg aacaaagtag agagagaaac acaagagtca
aatagatggt tctagtacag ggagagggca tgtgtcccag gggtggtgtc
cccactttct ctccctgcag tgggccccc cttcgggagt gtgctctatg
agtttgtggg gaagacggct ccgttcctgg tgctggccgc cctgggtactc
ttggatggag gtgagtgagt ccacgtgggc gccatgccat gaccttggca
tcgtgctggc acgtgcttgg gccacacctg atttctgaag aggggcttgt
cttttttatt tttatTTTT agctattcag ctctttgtgc
  
```

The parts that are double-underlined are the primers (single-stranded DNA) that we put into the PCR reaction. These define the ends of our fragment since the way the DNA was copied was by adding bases onto the ends of these primers.

Non-bacterial cells (like ours) are somewhat strange in that within a gene there are sections that are not used to encode protein. The whole gene, including these sections, is transcribed. After transcription, the non-coding parts, called introns, are cut out of the mRNA. In the sequence above, the coding segments are shown with gray shading (these are called exons).



Because the introns are not used for coding, they are often much more variable than the exons, where variation can be deleterious for the organism. The variations we will be seeing are in the intron between the two gray shaded sections. They are shown as bold/italicized/underlined characters. Their positions from the end of the first gray exon are base #52, 54, and 71, counting from the first “g” in “gtgagtgagt.” The type and frequency of the variations have been observed to be as follows:

Base number	Variation	Allele frequency
52	C or T	0.074
54	T or C	0.309
71	A or C	0.292

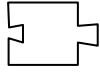
data from <http://pharmacogenetics.ucsf.edu/public-results/vmat2/VMAT2-tab.html>

What this means is that in the sequence at position 71, atttctgaag, in some people that first A is sometimes a C. This C allele is more rare, occurring in 29.2% of all alleles. The A allele, therefore comprises 71.8% of all sequences seen. While I can’t be certain this is the A to C change discussed in *The God Gene*, it does have very similar frequency in the population that was studied by UCSF to get these data.

How did the sequence data arise?

There is a completely non-intuitive, but very clever method used for determining sequence.

Let's say this is a nucleotide:

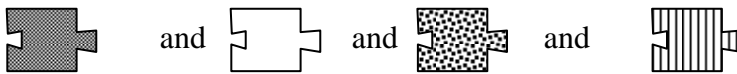


So a string of nucleotides would look like this:

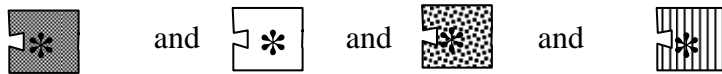


(There would also be a complementary second strand, which is not shown)

Let's say you replicate this DNA in a PCR-like reaction. The only difference is that only one primer is used and in addition to many copies of the four regular nucleotides



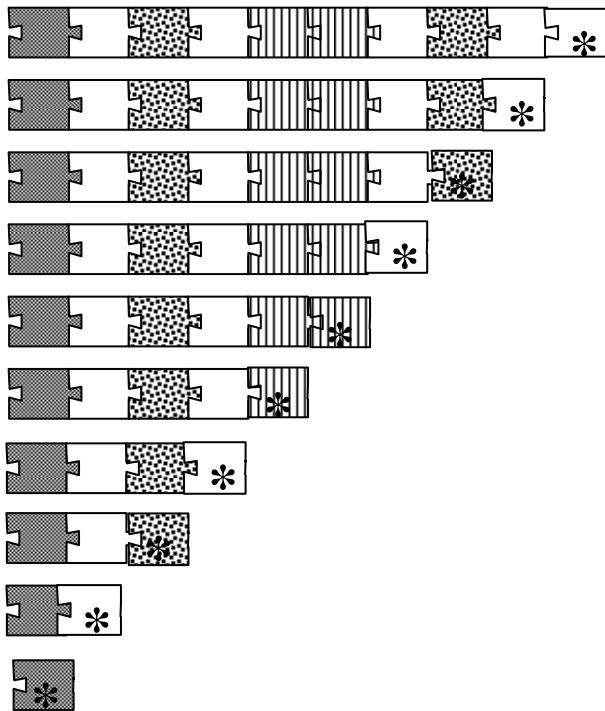
there are also some modified nucleotides



These are chemically identical to regular nucleotides, except for two things:

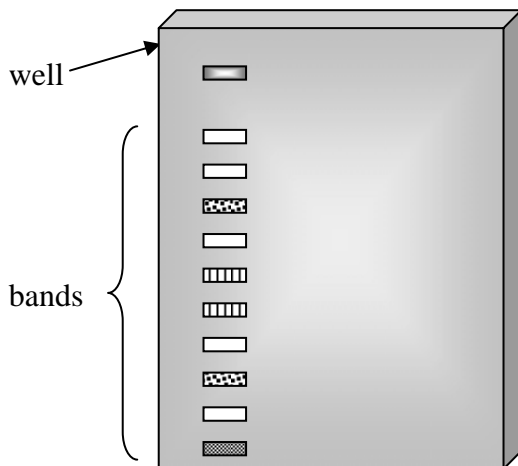
- They lack a chemical group at the 3' (sugar) end that is the normal point of attachment for the next nucleotide in the chain. This is represented in the figures above as a lack of an "outie" on the right-hand side.
- They have a colored label that glows under certain types of light (symbolized by the *). Thus, each nucleotide has a distinct color.

So as the DNA is being replicated over and over again, sometimes an ordinary nucleotide is incorporated and sometimes one of the funny ones is incorporated. Once a funny one is incorporated, the strand cannot be extended since there is no "outie" to hang the next nucleotide onto. Whether an ordinary nucleotide or an altered one is incorporated is due to random chance, so there will be all the various combinations once the many replication reactions are completed. Each strand will end in a funny nucleotide:



If these pieces are run on a gel, they will separate by size. The bands can be seen since they have colored labels on the DNA

If one looks at the gel from shortest to longest band the sequence can be read. The shortest band will be made of a shaded piece, with the appropriate color label that is attached to the shaded piece. So we read that band as “shaded.” The next larger band will have the color label that is associated with the open nucleotide, so we read that band as “open.” The next band up will have the spotted label, then next will have the open label, etc. So the sequence of bands from smallest to longest reads:



shaded, open, spotted, open, striped, striped, open, spotted, open, open.

This is exactly the sequence seen in our original piece!



Does that deserve a Nobel Prize or what?

To look at your sequence:

The gel is read by a computer and the bands are written as a series of peaks of different colors, rather than as a picture of a gel. Green is A, black is G, blue is C, and red is T. These files can be read with a program called Editview.

Open the Editview program by double-clicking on the "Editview" icon (a rainbow-colored triangle), and under the pulldown heading of "File" select "Open". Then select the file name corresponding to your sequence (usually the filename will have your initials in it). At this point you should see peaks of four different colors. There is also an X-axis that consists of a nucleotide number, and a nucleotide (designated A, C, G, or T). The latter is determined by the computer upon analysis of each peak. There are a few things to note:

- i) at many points there is only one peak and that the color of the peak determines the nucleotide indicated on the X-axis.
- ii) at some points, there is more than one peak (see bases near the beginning, for example). This is most often due to technical problems and makes it difficult or impossible to predict the base that is at that position on the DNA strand. This is why there is occasionally a "N" instead of an A, C, G, or T. "N" designates that the identity of the base at this position was not clearly determined.
- iii) it is technically more difficult to determine the base sequence near the beginning. It is also technically difficult to determine the base sequence much more than about 500 nucleotides from the start. So one tries to sequence a such that the relevant information will be somewhere in the middle of the series of peaks.

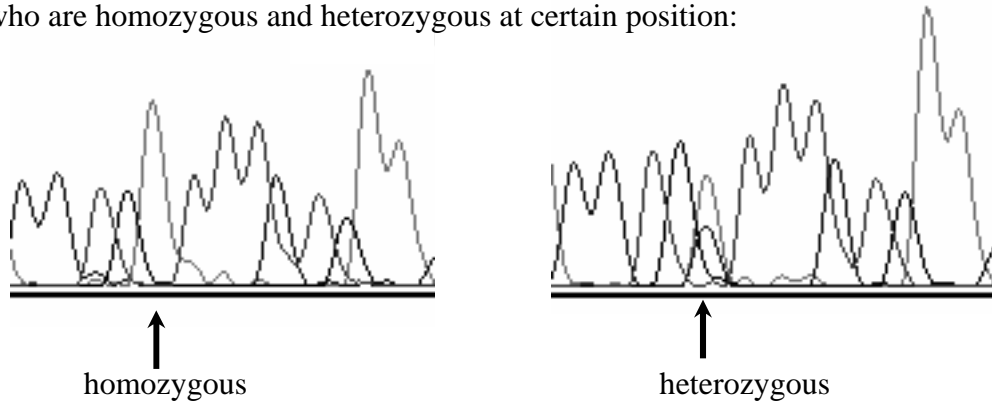
In the bottom left corner of the window there are four icons. One of them reads "CATG". Click once on this icon. This changes the information that you were looking at into text only, rather than the multicolored peaks with the various other annotations. Notice there are often several N's near the ends of the text string.

This text form of the data is searchable. Since we wish to quickly find the area around the variation, it would be convenient to search for a nearby invariant sequence. I successfully searched many of the files for the text string "gggccca," which is located in between the variable sites and is unique within the total sequence. Make sure the cursor is at the beginning of the file so that the entire file will be searched. Select Find from the Edit menu or hit the apple and F keys simultaneously to bring up a search box. Type the text string of your choice into the box and hit the "find" button. If you do not find the text string in your sequence try another nearby one in case there is an ambiguity in your sequence in the gggcca area.

Once you successfully find your text, toggle back to the peak view using the button with the peaks located on the bottom left of the sequence window. The found text will be highlighted above the relevant peaks. Scroll to the left and right to see the variable sequences.

Remember that you have two copies of every gene (except for those on the X and Y chromosomes, if you are a male). So this sequence was derived from a mixture of VMAT2 DNA, one from each of your two copies of the chromosome where VMAT2 is located. If both of your copies have the same nucleotide at a variable site, you will see just one peak at that position. For example if you

saw you had one peak for the C in “acgtg” that would mean that both of your chromosomes had a C allele. If you saw two peaks overlapping at this position, a C and a T, that would mean you are heterozygous; one chromosome has a C at this position and the other has a T (if the second peak is very small it is probably just background, not a real peak – I can help you judge, if you’d like). If you see a T peak only, then you are homozygous for T. The picture below shows an example of people who are homozygous and heterozygous at certain position:



Check each of the three variable sites and record what your genotype is at each position in the table below. Note that you should look at the peaks, not just the text, since the computer can be confused by overlapping peaks; it can choose to report just one, call the peak “N” or if they are slightly offset it may list both of them. You might want to take a look at some of the other sequences just to see what the variations look like.

Position:	52	54	71
Genotype (circle one)	CC CT TT can't tell	CC CT TT can't tell	AA AC CC can't tell

Let me know your genotype and I will compile the information into a table for the class.

A good question....

Why would a change in an intron, which does not code for a protein product, affect anything about the functioning of the gene? The short answer is that in most cases it doesn't. However, there may be other changes in the gene that tend to be seen when there are also variant forms of the intron. For example, it may be that most people who have a C at position 71 also have a change in an exon somewhere in VMAT2. This idea is referred to in *The God Gene* (p. 73) – Hamer states that the A to C change he looked at is often accompanied by several other variations and so can be used as an indicator. If these changes are in exons, they may indeed change the form and function of the encoded protein.

Designing and testing a hypothesis with a t test

Complex traits are sometimes known as quantitative traits – it's not a matter of whether you are spiritual or not – it's a matter of the degree of spirituality you have. These sorts of traits are not like purple flowers vs. white flowers, where the phenotypes fall neatly into distinct classes. Instead, the phenotype (such as spirituality, risk-taking, intelligence, or height) is variable over a spectrum. Because multiple genes contribute to these traits you will not see a 1:1 correspondence between a particular allele of one gene to a particular trait, as you do with simpler traits like flower color. So knowing whether you have identified a genetic difference that influences a quantitative trait can be difficult.

In looking to see if a particular allele is correlated to a quantitative trait, one must do a statistical test to see if the association of that trait and that allele are significant. In statistics “significant” means it is beyond what one might expect by random chance. To assess significance we will use a common statistical test called the Student's t test (so-named because its inventor published it under the pseudonym “Student”).

In a t test one divides the dataset into two classes. We could, for example, concentrate on the genotypes at position 71 – AA, AC, or CC. These data could be divided into two groups: those that have at least one “C” – Hamer's “spiritual allele” – at position 71 and those who do not. This is how Hamer divides his data. We then look to see if the average score of our quantitative trait (for example, the score on the TCI self-transcendence section) differs in a significant way between people in those two groups. Remember that Hamer contends that people with AC and CC genotypes have higher self-transcendence scores, on average, than people with AA genotypes.

Significance is increased by seeing a particular result in a larger number of cases and by seeing a dramatic difference in mean score between the two cases. That one needs to look at a reasonable number of cases to make a determination makes sense intuitively. To use an example unrelated to the case at hand, if one rolls a die three times and never sees a six, it would not be sensible to conclude that this is a significant finding and that there is no six on the die. If one rolls a die 1000 times without rolling a six, you would have serious cause to believe this finding IS significant – it's telling you that there is no six on that die or that the die is weighted or something that is skewing the results away from random chance.

In the t test one phrases the research hypothesis as a null hypothesis – a hypothesis that there will be no difference between the two groups. So in this case we would say that our null hypothesis is that a group of people with a C at position 71 will not have different average scores on the TCI than a group of people without a C. Then, if the t test tells you to reject that hypothesis, that means that the data are consistent with there actually being a significant difference between the two groups. Note that stating the hypothesis in this negative form is not necessarily reflective of the way you are thinking about the problem. Hamer asserts that there IS a spirituality difference between the people with a C and the people without a C. However, to test his idea with a t test, one must phrase it in the negative, as a null hypothesis, and see if it is rejected. If the null hypothesis is rejected, then one can say there IS a difference.

To perform a t test

Arrange your data in an Excel spreadsheet as follows:

Make two columns – the one column is genotype that will be examined and the other column is score for that person on the desired section of the TCI (or you can have three columns if you want to keep track of which person is which). Arrange the data in these columns so that the genotypes without a C are in consecutive rows and the genotypes with a C are in consecutive rows. See the “before” and “after” example below:

Original dataset (with columns of other TCI scores and genotypes omitted):

Individual	TCI ST score	Genotype at 71
1	10	aa
2	8	aa
3	7	ac
4	11	aa
5	-	aa
6	12	ac
7	8	aa
8	14	aa
9	28	ac
10	4	aa
11	27	cc
12	29	aa
13	18	aa

Rearranged data so that genotypes with and without C are grouped together, omitting any incomplete data. Make sure you keep the correct TCI score with each genotype:

Individual	TCI ST score	Genotype at 71
1	10	aa
2	8	aa
4	11	aa
7	8	aa
8	14	aa
10	4	aa
12	29	aa
13	18	aa
3	7	ac
6	12	ac
9	28	ac
11	27	cc

The order of the individuals in the rearranged data is not important, as long as all the aa's are grouped together and the ac/cc's are grouped together. You can have the additional column with the individual's number there, as I did for clarity, or not.

Now, under the Tools menu select "Data Analysis." From the pop-up box choose "t test: Two-Sample Assuming Unequal Variances." Click OK.

Now another pop-up box is confronting you. It looks gnarlier than it really is. Click inside the Variable 1 range box. In this box you want to tell Excel which boxes contain your first variable, which is the TCI scores of our first group, our group with only A alleles. Select these boxes on the spreadsheet – see shaded region in the picture below. Excel will automatically enter some numbers and letters in the Variable 1 box that correspond to the location of your data in the spreadsheet (for example, I had these boxes in the B column in rows number 17-24 so I got \$B\$17:\$B\$24 in my Variable 1 box).

Individual	TCI ST score	Genotype at 71
1	10	aa
2	8	aa
4	11	aa
7	8	aa
8	14	aa
10	4	aa
12	29	aa
13	18	aa
3	7	ac
6	12	ac
9	28	ac
11	27	cc

Now click inside the Variable 2 range box. This is where we enter the location of our second set of variables, which are the scores of the people who had at least 1 C allele. Select these scores as shown below and Excel will enter their locations in the box for you (it should look the same as Variable 1 with respect to letters and dollar signs; the numbers will be different).

Individual	TCI ST score	Genotype at 71
1	10	aa
2	8	aa
4	11	aa
7	8	aa
8	14	aa
10	4	aa
12	29	aa
13	18	aa
3	7	ac
6	12	ac
9	28	ac
11	27	cc

The hypothesized mean difference box can be left blank or you can enter a “0” - either of these tells Excel that our hypothesis is that there is NO difference between the two sets of values. Remember, we are phrasing our hypothesis as a null hypothesis, even though we are actually looking for a difference.

In a statistical test one chooses the level of significance one is looking for. The typical level is 0.05, which means in only 5% or less of all random cases you would expect a result similar to our result simply by chance. The alpha box is where we tell Excel how stringently we want to measure significance. Since a value of 0.05 is the typical level people use, we can just leave that value there.

For output, it doesn't really matter, but you can click new workbook if you want Excel to dump the t test results in a separate file than the original data. Or you can have it stick them in a new sheet (ply) in the current file, or on the same page (in the latter case, click in the Output range box, then click a box on the spreadsheet to show excel where to put the table on the page you are looking at).

When you click OK Excel will generate a table for you with all kinds of statistical information about your data. Weehaw.

Look at the means for the two variables. Which group has the higher average TCI self-transcendence score – group 1 (AA genotype) or group 2 (AC or CC genotypes?) Does this agree with Hamer's findings?

But is this difference statistically significant or could it just be a result of random chance? To figure this out look at the two rows labeled with P. Let's consider the one labeled "P(T<=t) one-tail."

The "one-tail" part means that we expect the two groups to have a certain relationship – we expect that group 1 is going to be lower than group 2 in TCI self-transcendence scores. For other tests on the dataset that you might do, you might want to consider the two-tailed test. This test is used when you don't have an idea about whether group 1 scores will be higher than group 2 or vice versa.

If the number in the appropriate "P" row is lower than 0.05 then that means that less than 5% of the time you would get our results simply by random chance. That means the results are significant, the two groups really do differ. Here you must reject the null hypothesis, reject the idea that there is no difference between people with A-only vs. people with at least one C. If it is higher than 0.05, then the differences in the averages between the two groups could be a difference based on random chance and we must accept the null hypothesis – there is not a meaningful difference in TCI score between the two groups.

- Which scenario do our data fall into?
- What is a shortcoming of our experiment that makes it difficult for us to get significant differences? (I'm actually surprised that the values are as good as they are)
- What is/are some other potential shortcoming(s) of our experimental sample?
- Repeat the above process using some other combination of variables. For example, you can look at a particular position genotype and how it correlates with a different TCI score or sub-score set. Or you can also look at some combination genotypes – people who have both a T at position 52 and a C at position 71 vs. those who don't. Remember, you need to divide people into two groups to do a t test. If you are a glutton for punishment and want to look at three or more groups of people, we can use an ANOVA test (it's not any harder than a t test, but no way am I typing up a handout for it – I can show you if you want).

Student ID#: _____

Please answer these questions to the best of your ability. Some questions will ask for your opinion and others will test working knowledge of biological concepts. Please do not consult other people or resources in answering these questions.

These questions have been designed to allow me to assess your knowledge of course content, your attitudes toward issues related to biotechnology and genetics, and the role of this course in affecting your learning or your attitudes. I plan to publish a description of certain modules used in this course and to refer to these assessment data in this publication. You will not be identified in the publication. Your instructor will only use your identification to match pre- and post-tests with each other to identify any changes; your answers will otherwise be treated anonymously. *You are not required to participate in this study, although your participation is appreciated.* Participation in this study will not require any extra time over and above that required to participate in the course itself. You will receive course credit for the pre- and post-tests regardless of whether you participate in the study. Please see your instructor if you wish to withdraw from the study or if you have concerns about the study. You may obtain the results of the study from me when they have been analyzed.

1. When I read an article in a newspaper or magazine about a scientific subject, my belief in the validity of the finding is influenced by: (please rate the following items on a 1-5 scale with 5 being very influential/true and 1 being not influential/true. Numbers may be used more than once).

- _____ scientists performing the work are at institutions that I trust
- _____ the work is said to be published in a scientific journal
- _____ the ideas agree with my personal experience or knowledge
- _____ multiple scientists concur with the finding
- _____ the article is easy to understand

Rate the following statement on the same 1-5 scale (1= not at all true, 5= very true)

_____ I don't usually read articles about scientific findings

2. I personally believe genes play a role in complex traits such as intelligence and/or behavior.

_____ (use a scale of 1-5, 1=strongly disagree, 5= strongly agree)

3. I believe science can help us understand non-scientific areas such spirituality.

_____ (use a scale of 1-5, 1=strongly disagree, 5= strongly agree)

4. A version of a particular gene is found to be correlated with smoking – heavy smokers have that version of the gene more frequently than non-smokers. Given this is true, then:

- T/F _____ The gene causes people to smoke
 _____ Genes may play a role in smoking
 _____ Environment does not play a role in smoking

5. Answer each true or false:

- _____ If a 100% correlation is seen between factors A and B, then A must cause B or B must cause A.
_____ Alleles are versions of proteins.
_____ Genes encode proteins.
_____ Some mutations cause no further effects in the cell.
_____ Most of the “jobs” in a cell are carried out by DNA
_____ Agarose gel electrophoresis can be used to determine the size of DNA and purify it from other DNA in a sample.
_____ Genes are made of protein.
_____ The order of bases holds information as to genetic traits.
_____ Sample size strongly affects the credibility of genetic studies.
_____ Some genetic traits are due to a single gene.
_____ It is not yet possible to determine the exact sequence of DNA isolated from humans.
_____ Some traits are due to multiple genes.
_____ DNA can be replicated in a test tube to produce large amounts of specific regions.
_____ Mutations are damage to proteins.
_____ The position of a mutation in a sequence affects its severity.
_____ Twin studies are useful in genetic research.
_____ DNA can be isolated from cells by using detergents and salts.
_____ It is possible to assess the relative contributions of genetics and environment to a trait.
_____ Statistical tests allow one to gauge the significance of a particular result.
_____ Mutations may alter a protein that is produced by the cell.

6. People often refer to BRCA1 as “the breast cancer gene.” This shorthand means:

(T/F)

- _____ If one has the BRCA1 gene, one is more likely to get breast cancer
- _____ Some versions of BRCA1 make it more likely that one gets breast cancer.
- _____ Other genes do not influence breast cancer
- _____ BRCA1 does not play a role in normal cells

7. Consider the following two scenarios:

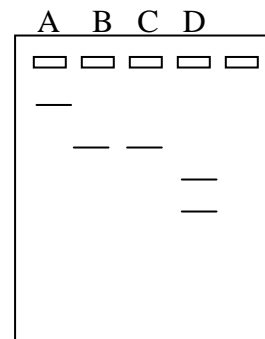
i) A gene is examined in six students and is found to have two forms, A and B. Three of the students have brown hair and three have blond hair. The three with brown hair each have version B of the gene, and the three with blond hair each have version A.

ii) A gene is examined in 10,000 students and is found to have two forms, A and B. Three thousand of the students are smokers; the rest do not smoke. Amongst the smokers, 75% of the students have the A form of the gene. Amongst the non-smokers, 19% have the A form of the gene.

Which conclusion is more compelling to you: (choose one)

- a) In scenario i, A probably makes a person have blond hair.
- b) In scenario ii, A probably makes a person more likely to smoke

8. Answer each true or false, using the diagram as needed:



- _____ The band marked A is bigger than the band marked B.
- _____ The sequence of DNA in lane B can be determined by the information shown.
- _____ The sequence of DNA in lane C must be the same as that in lane B.
- _____ It is not possible for the two forms of DNA in lane D to be isolated from one another.
- _____ Someone remembers that the DNA in lanes A and B are related in that one was derived from the other by PCR. S/he can't remember which was the starting DNA and which was the PCR product. It is more plausible that the DNA in lane A was derived via PCR from the DNA in lane B than vice versa.

9. Use a 1-5 scale to respond to each of the following statements (1=strongly disagree, 5=strongly agree).

_____ The VMAT2 project helped me to understand the use of statistics in biology.

_____ The VMAT2 project helped me to evaluate the role of genetics in behavior.

_____ The VMAT2 project helped me to critically evaluate science as reported by the press

_____ The VMAT2 project convinced me that genes do not play a role in behavior.

_____ The VMAT2 project helped me to understand how variations in DNA sequence can alter phenotype.

_____ The VMAT2 project helped me to understand the relationship between genes and proteins.

_____ The VMAT2 project increased my interest in genetics.

_____ I think it is harder to analyze a DNA sequence than I thought it would be before I took this course.

_____ I think it is harder to link DNA variations to a complex trait than I thought it would be before I took this course.

10. What two things did you find most effective or interesting in the VMAT2 project?

11. What two things were least effective or uninteresting in the VMAT2 project?

Questions for discussion - *The God Gene* and the Temperament and Character Inventory

What is your general reaction to idea of a God gene?

Does reductionist study of something diminish it or demean it or enhance it?

Regarding the TCI, what do you think about the connection between self transcendence and spirituality? Does this strike you as a valid relationship?

Individual components are: self forgetfulness
transpersonal identification
openness to things not literally provable

After reading the book, the *Time* article, and the reviews, what are your impressions as to the scientific acceptance of Hamer's study?

Do you accept there are distinctions between spirituality, belief in a deity, and religious practice?

What are your reactions to twin studies – do they convince you of a genetic component to spirituality?

What do you think about the argument that spirituality is adaptive?