Part A. VISUAL SEARCH ONLINE with NCBI Name_____

Uses the NCBI website, with its great visual maps showing locations from which the many cloned segments were taken, along with detailed DNA sequences arranged in rows (as used in this lesson):

1. Go to **http://www.ncbi.nlm.nih.gov/mapview/maps.cgi**. See "***Homo sapiens*** **Build 36.1** near the top. Just beneath this, see "Chromosome:" followed by numbers 1-22, X, Y, and MT.
2. Click on the **2**. This will display chromosome #2, along with many details of its contents (**Fig. 2**).
3. Find the centromere (constriction almost 1/2 way down from top end). See detail of the region in **Fig. 3**.
4. Note: the dark and light bands in the *upper* portion (above the centromere constriction) are labeled with a "p" (e.g. in 2p16.3); the bands in the *lower* portion are labeled with a "q" (e.g. in 2q24.1).
5. Find the white band called **2q13**. Look just to the right to see the vertical blue lines. The blue lines are called "contigs" (= contiguous short portions studied in detail). Look carefully (in Fig. 3) for the ID number of the contig that includes all of **2q13**. What is that number? NT-_____
6. Click on that NT number, then click on the enlarged "pop-up" version of that number, which looks like:

   symbol: NT 022135.15
   Orientation: +
7. This brings up the detail page for that contiguous sequence ("contig").

8. On the "**LOCUS**" line, notice that the NT_022135 contig has **38,390,280 bp**. That means there are over 38 million basepairs in that region of chromosome #2!
9. Scroll down to where all the CONTIGs are listed, mostly in 3 columns. These are the many short "sub-contigs" contained within the larger one. Most have blue "AC…" numbers, but search down the *left* column until you come to AL078621.19 (on the 10th line down).
10. Clicking on this number brings up the detail page for that short sub-contig "AL078621."

11. On the **LOCUS** line, notice that there are **176,734 basepairs** in this sub-contig.
12. Then the **DEFINITION** paragraph says there are **6** genes or Pseudogenes in this sub-contig. Interesting!
13. Scroll quickly down through other information and a *long* list of "FEATURES" (with various titles as links), until you see the ORIGIN of this nucleotide (bp) sequence (about 1/3rd of the scrolling range from the top). Here's a sample of the first few lines:

```
ORIGIN
        1 gaattcttgt tctgtattta gaaacccact cacgttactt gatatttggg tatttaagtc
       61 atgaaaggta tttcttctag gaagcagtga ttctaaagtg tatgcttaac cagtcagttg
      121 agtgtctact cttgtgtgtt cacaagtgtg cacaaagttt ttggtaaatt aagaatatta
```

14. Each number represents the first base (bp) on that line, and on each line there are 6 sets of bases (bp), with 10 bp in each set (total of 60 bp per line): bp#1 is **g**; bp#71 is **t**; what is bp#121?___; bp#150?___

15. What number would you expect to find for the *last* bp in the entire sequence? _____(see #11 above). Scroll down to the very bottom of the page to see if that number is correct. What is that letter?___
16. Start searching for the fusion-transition sequence expected where the telomeres had to meet and join. Start with a quick scroll down from the top line, looking for any repeat pattern that stands out, and getting a feel for the massive amount of DNA represented here. (The fusion region is shown below. Notice the many **multiples of g**, followed suddenly by many **multiples of c**.) Don't spend too much time here, maybe 5 minutes, max (it could take hours!) If you find this region, record the bp number where that change occurs:_____. Now go on to Part B to see if you are right.

```
agggttgggg ttggggttag ggttaggggt tagggttggg gttggggttg gggttggggt
tggggttggg gttagggtta gctaaaccta accctaaccc ctaaccccaa ccccaacccc
aaccctaccc ctacccctac ccctaacccc aacccccacc cttaaccctt aacccttacc
ctaaccctaa cccaaaccct aaccctaccc taaccctaac ccaaccctaa ccctaaccct
```

16. **Don't feel bad if you can't find it!** Very few people do. Instead, do what scientists do: a "BLAST2" search, in which you ask the online computer to search the entire sequence using a probe (short DNA sequence desired). It then shows you the areas that match, and how closely they match. Let's try it. [Go on to **Part B: BLAST2 search**; or, if assigned, go to Part D for the paper search].

Part B. SEARCH WITH the "BLAST 2" SEARCH TOOL Name_____

This is one of the internet tools that scientists use to look for specific sequences. It involves entering a short search sequence "probe" and specifying the DNA region to search.

1. To find the head-to-head telomere repeats, go to the **BLAST 2 SEQUENCES** program at:
   http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi
2. In the **Sequence 1** entry box, insert the following array listed below (exactly as shown, including ">array"):

```
>array
ttagggttagctaaacctaaccc
```

3. For **Sequence 2**: Enter: **AL078621** [type that accession number in the large box below "**Sequence 2**"]
4. Click on "**Align**".
5. This will take you to a detail page: "**Blast 2 Sequences results**". Scroll down to "**`Query: 1`**," where the sequence entered is displayed (**23** bp), followed by "**`Sbjct:`**" and the bp number range in the searched array where the sequence appears.
   That number range is 108552 to _____
6. Scroll back to the top of the page and find **Sequence 2:** gi with a blue-linking number following it (6013067). Click on that number. This will take you to the detail page again. Scroll down (about 1/3rd of total scroll distance) to where the DNA sequence for the AL078621 clone is displayed (all 176,734 bp).
7. Scroll down further until you come to a number close to the bp number recorded in step 5 above (108574)
8. From here, search backward carefully until the many **multiple-g** sets seem to **start**, at about _____bp
9. Search forward through the ttaggg tandem sequences (or slight variations) until the **multiple-g** sequences ***end,*** and the **multiple-c** sequences ***begin***. What is the bp number at that point?_____
10. Continue forward until the **multiple-c** sequences seem to **end**. What's the bp number there?_____
11. About how many bp long was the region with many sets of **multiple-g**? (#9 minus #8) _____bp
12. About how many bp long was the region with many sets of **multiple-c**? (#10 minus #9) _____bp
13. As you look over this region, with the remains of the telomeres of two chromosomes appearing head-to-head near the middle of chromosome #2, what does this suggest must have happened to those two chromosomes at some time in the distant past?

14. Which number represents the actual head-to-head fusion point?_____
15. As you may have learned in another lesson, the number of changes in DNA (or proteins), between corresponding molecular sequences from different living species, can serve as an approximate measure of the time of branching, especially when it can be matched with fossils. Using this "molecular clock" technique, studies have shown that this fusion very likely happened about 3 million years ago. Molecular and fossil studies indicate that our closest relatives, chimps, branched off our line about 6-7 million years ago (mya). So, which scenario seems most likely:
    a. Our chromosome #2 formed before any of the apes branched off.
    b. Our chromosome #2 formed before chimps and humans branched off.
    c. Our chromosome #2 formed after chimps and humans branched off.
    d. Our chromosome #2 formed around the time modern humans first appeared (~0.2 mya)
    e. Our chromosome #2 split to form two shorter chromosomes in the apes.
16. If time (or for homework), go to **Part C: Search the Sanger Genome Browser**. This will take you though a somewhat different format showing essentially the same evidence for head-to-head fusion. In addition, an interesting extension (#19) will give you a very compelling visual indication of exactly which chimp chromosomes were combined to create the human #2 (based on detailed DNA sequences.)

Part C. SEARCH on the SANGER GENOME BROWSER Name_____

The Sanger Human Genome Browser: **e! Ensembl Human**, displays our DNA in a horizontal linear arrangement, with each nucleotide base in a different color, plus the complementary matching sequence, and all the possible amino acids that they code for. Very colorful, and easy to use.

1. Access at: **http://www.ensembl.org/Homo_sapiens/index.html**.
2. Under "**Karyotype**", (middle column) click on the diagram of Chromosome 2. When the detail page of Chromosome 2 appears, look under the "**Chromosome 2**" title on the right:
   What's the **Length** (total number of bps in chrom. 2)? _____bps
   As a matter of interest, how many **Known Genes** are in chromosome 2? _____
   How many **Pseudogenes** in chromosome 2? _____
3. Click on the **q13** band in the chromosome diagram. This takes you to the **Human Contig View**.
4. Note the horizontal diagram of Chromosome 2 near the top, with the **red marker** in the q13 band (just to the right of the centromere constriction).
   Estimate how far that **red marker** is from the *left* end, relative to total length: _____% of distance.
5. Go down to the Detailed View area. You will need to select a few items to be displayed.
   In the yellow bar below that title, notice the several categories:
   Features▼ Comparative▼ DAS Sources▼ Repeats▼ Decorations▼ etc.
   Click on Repeats to open list of check boxes; click on *Tandem Repeats*; click on Repeats again to close the drop down list. Click on Decorations, then click on boxes for *Sequence, Codons, Contigs, Ruler, Scale Bar,* and *Show Register Lines* if their boxes are not checked; then click on Decorations again, to close the list.
6. In the Jump to region 2: boxes (below the "Detailed View" title), double-click on existing numbers and replace them with these:
   Jump to region 2 : 114076620 114076700 Then click on Refresh.
7. Notice in the Chromosome 2 diagram (near top) that the **red marker** has shifted slightly into the gray band (q14), just to the right of q13.
8. Scroll further down under the Detailed View where you will see multi-colored diagrams. Notice the broad blue band [labeled "DNA(contigs)"] running left to right through the middle of the multicolored strands above and below it. Note especially the upper "Sequence" strand (with AGAGACCAGTT…etc) just above that broad blue band, and the other (matching) "Sequence" strand below it. The upper one, above the *left*-to-right arrow, goes from the 5' toward 3' end. The lower arrow runs from *right* to left (5' to 3'), as do the nucleotides below it. Also, see the three rows of possible "Amino acids" shown above (and three below) their respective nucleotide sequences, each depending on which triplet-frame of nucleotides (codon) is coding for it. Notice that each base letter (in the two "Sequence" strands) has its own unique color. Notice the rather random sequence of As, Gs, Cs and Ts across the screen (left to right).
9. Near the top of this diagram box, just above the "Amino acids" 80-bp arrow, notice the Chr. 2 Length numbers. The first number (114,076,620) refers to the first darker vertical line touching that number on the left. The next darker vertical line is *not* numbered, but the one following that *is*: 114,076,640. 114,076,640 – 114,076,620 = ?____ How many nucleotides (bases) fit between the two dark lines?_____. So what are the numbers counting?_____.

These numbers are counting the bases (bp) from the left end of the *entire* chromosome to the right end.

continue overpage

10. Near the top of the **Detailed View** box, notice the **Window >** button with the arrow pointing to the *right*. Click it. When the new screen appears, scroll back down to the **Detailed View**, and you will notice that you have shifted 80 base pairs (bp) to the *right*. View the *upper* nucleotide **Sequence** again. What do you observe, regarding multiple sets of Cs and Gs?

    multiple sets of Cs?: _____; multiple sets of Gs?: _____

11. At about what nucleotide number do the multiple Gs begin (without any Cs)?_____

12. Repeat step 10. Again, describe what you see in the top sequence regarding Gs and Cs:

13. Repeat step 10 **three** more times. You should now see something strikingly different from the sequences in the previous few windows. What is it?

14. At about what point (nucleotide number) do you see the change?_____. Continue step #10 several more times to see all the AATCCC tandem repeats (and slight variations). About how many Window-Shifts did it take before the tandem multiple Cs stopped? _____ At about what nucleotide number did the multiple Cs end? _____

15. In what parts of any chromosome do we normally find those many tandem repeats of "TTAGGG" or "AATCCC" (with slight variations)?

16. What does finding these two sets of head-to-head tandem repeats here, near the middle of chromosome #2, suggest must have happened at some time in the past?

17. What can we call that point (#14) where the base sequences suddenly changed? _____

18. What percentage of the DNA lies to the left of that fusion point? [Divide the basepair number at the fusion point (#14) by the total number of basepairs in the chromosome (see item #2), and multiply that by 100 =] _____%. Compare this number with your earlier estimate (#4):

19. **Interesting Extension**: Return to the home page for Chromosome 2. Do this by scrolling to top of page, and clicking on "**View of chromosome 2**" near top of column on left side of the page. On that page, mouse over the next title in that column: "**View Chr 2 Synteny**", and down the pop-up box to "vs Chimp (Pan troglodytes)", click on the "**Chimp**." Be patient, as it may take about 30-60 seconds for the next page to display. Scroll down that same left column to "Other Links", then mouse over "View previous releases…" and click on Release 36 in the pop-up menu. The diagram will show how our chromosome 2 is composed entirely of two chimp chromosomes (based on their DNA comparisons, not their banding patterns). Notice that that the entire green chimp chromosome matches the upper part of our #2, while the entire lavender chimp chromosome matches the lower part of our #2.
    What are the numbers of those chimp (*Pan troglodytes*) chromosomes?____ and ____.

    The small white areas represent the centromeres. Which one of the two chimp chromosomes has the centromere that serves as the centromere in *our* chromomome #2? ____. If you have time, look for evidence of old centromere from Chimp chromsome 2B in our #2 chromosome.

Part D. PAPER SEARCH

Name _____

If internet resources are not available to you, use the provided page (**Fig. 4**) showing the DNA sequence copied from the tiny region in a section from the NCBI Genome Catalogue (accession number AL078621) that includes the **2q13** region where telomere fusion should be found. **Fig. 2** shows the human chromosome #2 as it appears on the NCBI website. **Fig. 3** shows an enlarged view of the area of interest in that chromosome. Notice band # 2q13, and the catalog number for the DNA segment from that area: NT_022135. A small portion of that segment is given the number AL078621.

The nucleotide bases in that entire portion (AL078621) are numbered from 1 to 176,734. The tiny selected sequence of that portion that you will search fills the page titled: "**Fig. 4: Chromosome Region 2q13: Portion of Accession No. AL078621**" The first few lines look like this:

```
106921 taagttacat gcagacaaca ggggccagaa gatgaacaat ggcccatccc actctaggca
106981 tggctcctct ccacaggaaa actccactcc agtgctcagc ttgcaccctg gcacaggcca
107041 gcagttgctg gaagtcagac acctgcagat caagaccaca gcatcaagac cctgtgacct
```

1. The number at the beginning of each line represents the first base on that line, and each line has 6 sets of bases (bp), with 10 bp in each set (total of 60 bp per line):
   so, bp#106921 is **t**; bp#106991 is **c**; what is bp#107041?____; bp#107060?____

As you can see, the page shows basepairs (bp) numbering from 106921 to 110341.

2. Visually scan the sequence, searching for the tandem repeats associated with telomeres. It's easiest if you just look for lots of sets of multiple **g**'s, followed by lots of sets of multple **c**'s When you find that region, search backward for the bp number where the **multiple-g** sets seem to **start**, at about _____bp

3. Search forward through the ttaggg tandem sequences (or slight variations thereof) until the **multiple-g** sequences ***end,*** and the **multiple-c** sequences ***begin***. What is the bp number there? _____bp

4. Continue forward until the **multiple-c** sequences seem to **end**. What's the bp number there? _____bp

5. About how many bp long was the region with many sets of **multiple-g's**? (#3 minus #2) _____bp

6. About how many bp long was the region with many sets of **multiple-c**? (#4 minus #3) _____bp

7. As you look over this region, with the remains of the telomeres of two chromosomes appearing head-to-head near the middle of chromosome #2, what does this suggest must have happened to those two chromosomes at some time in the distant past?

8. Which number represents the actual head-to-head fusion point? _____bp

9. As you may have learned in another lesson, the number of changes in DNA (or proteins), between corresponding molecular sequences from different living species, can serve as an approximate measure of the time of branching, especially when it can be matched with fossils. Using this "molecular clock" technique, studies have shown that this fusion very likely happened about 3 million years ago. Molecular and fossil studies indicate that our closest relatives, chimps, branched off our line about 6-7 million years ago (mya). So, which scenario seems most likely:
   a. Our chromosome #2 formed before any of the apes branched off.
   b. Our chromosome #2 formed before chimps and humans branched off.
   c. Our chromosome #2 formed after chimps and humans branched off.
   d. Our chromosome #2 formed around the time modern humans first appeared (~0.2 mya)
   e. Our chromosome #2 split to form two shorter chromosomes in the apes.

EXTENSION:

If at all possible, go to one or both of the online sites where you can see the DNA sequences for yourself. Think of it as a treasure hunt as you search for a particular pattern/sequence. Ask your teacher for the directions for each part: Parts A, B, and C.