Bioinformatics/Evolutionary Genomics Lab

**Human Variation and Long-Term Evolution at the *BRCA1* Locus**

**by Marc Tollis, Ph.D.**

Observation: Breast cancer 1 (*BRCA1)* is a tumor-suppressor gene expressed in human breast tissue that repairs damaged DNA or destroys cells when the DNA cannot be repaired. Some inherited mutations in the *BRCA1* gene can lead to a very high lifetime risk of breast cancer.

Hypothesis: Since *BRCA1* is so important for a functioning multicellular organism, it should have evolved under very strong purifying selection, and mutations that cause human breast cancer should be evolutionarily very new (*i.e.*, only occurring in human populations).

Prediction: Disease-associated single nucleotide polymorphisms (SNPs) at *BRCA1* should occur in regions that have been highly conserved throughout the evolution of mammals.

Tasks for this exercise:

1. Set up the UCSC Genome Browser display to study the *BRCA1* human gene.
2. Turn on the SNPs track to see the human SNPs that have been found in *BRCA1*.
3. Compare *BRCA1* sequence across mammals.
4. Visualize evolutionary conservation for clinically relevant SNPs in *BRCA1*.

*Expected results are shown in italics.*

**Questions are in bold.**

Set up the Genome Browser display to see the *BRCA1* human gene.

1) Go to <http://genome.ucsc.edu>

2) You can find the hg19 under the Genomes Tab in the upper left.

5) Click submit.

*Here you see the Browser, with some tracks, including gene annotations, allele variants, conservation and repeat elements.*

6) Scroll down to customize which tracks will be displayed.

1. Make sure all tracks are hidden. Click the “hide all” button just above the Mapping and Sequencing bar.
2. Display the “UCSC Genes” in “Genes and Gene Predictions”, which should be set to “pack”. Click a “refresh” button.
   1. *Now you should only see a track with gene annotations.*
      1. *Small rectangles = UTRs*
      2. *Thicker rectangles = coding exons*
      3. *Thin lines = introns*
      4. *Arrows indicate plus or minus strand.*

7) In the empty field at the top, type “brca1”, select from the pull-down menu and click “go”.

1. *Browser should now display all the transcripts for BRCA1.* 
   1. **What chromosome is BRCA1 on?  
      What strand is BRCA1 on?  
      What are some major differences/similarities in the different transcripts for BRCA1?**

Turn on the SNPs track to see the SNPs in *BRCA1*.

1) Scroll down to customize the “Variation” track.

2) Set “All SNPs (144)”, “Common SNPs (144)”, and “Flagged SNPs (144)” to “pack”. Hit a “refresh” button.

1. *You should now see on the browser all the SNPs for the tracks and their rs numbers.*
   1. *Red = non-synonymous nucleotide changes*
   2. *Green = synonymous nucleotide changes*
   3. *Bule = splice sites or untranslated regions*
   4. *Black = intronic regions*

3) On each visible SNP track in the browser, right-click and select “dense” for a more human-readable display.

1. *The SNP tracks now are in single rows, where you can see the position of all SNPs. Some tracks are denser than others.*
   1. **What are the differences between the “All”, “Common” and “Flagged” SNP tracks?**
   2. **Which type of SNP is the most commonly flagged as clinically significant?**

4) Click, hold and drag the area to the left of each track to change the order/position on the browser.

5) Scroll down to the “Variation” track choices and click on the link for “Flagged SNPs”.

1. **What is the useful information found on this page?**

Compare genomic sequence between mammals.

1) Scroll to the “Comparative Genomics” section, and click on “Conservation”

1. *Here you see the settings for the 100-Vertebrate MultiZ whole genome alignment track and sub-tracks. The phylogenetic relationships of all the included taxa are shown.*

2) Set “Maximum display mode” to “full” and set “MultiZ alignments” to “full”.

3) For “Species selection”, press the minus button to uncheck all species.

4) Select only eutherian mammals: press the plus buttons for “Primate”, “Euarchontoglires”, “Laurasiatheria”, and “Afrotheria”.

5) Scroll down further to “List subtracks” and make sure only “MultiZ Align” is checked.

6) Scroll back to the top of the page and click “submit” to return to the browser.

1. *Now you should see each species which has been aligned to the human genome.*
   1. *Green bar = height indicates alignment quality*
   2. *Yellow = indicates Ns in the query assembly*
   3. *Double lines = indicate large gaps*
2. ***Compare the gene annotation to the conservation track. Which regions tend to be more conserved?***
3. ***Which species show more conservation in terms of aligned sequence to the human BRCA1 gene? Why is that?***

7) At the top of the browser region where you see the scale and the chromosome coordinates, you can click and drag the cursor to highlight a region you would like to zoom in on. Try to zoom in on a region that includes part of a conserved exon and some flanking intronic sequence. A “Drag-and-select” window should open up. Select “Zoom In”.

8) Repeat this process until you get to the single base level in the browser (you can see the individual nucleotides for each base).

1. *Now you can see individual nucleotide differences between species, amino acid changes, single-base gaps and multiple gaps are given (gold type)*
2. **What causes these differences?**

9) Scroll back to the top and enter “brca1” in the empty field and select from the drop down menu. *This should return you to the browser viewing the entire BRCA1 gene.*

Retrieve conservation information for the *BRCA1* gene.

1) Scroll to the “Comparative Genomics” section, and click on “Conservation”.

2) for “MultiZ Alignments” and “phyloP” select “hide”. For “Element Conservation (phastCons)”, select “full”.

3) Scroll down further to “List subtracks” and make sure only “100 Vert Cons” is checked and indicates “full”.

4) Select “submit” to return to the browser.

1. *Now you should see the “100 Vertebrate Conservation” track in the browser.*
2. *Bars indicate the probability that each nucleotide belongs to a conserved element.*
3. ***On average, are clinically flagged SNPs more likely to be found in conserved or non-conserved regions?***
4. ***Are these otherwise conserved SNPs more likely to be synonymous, non-synonymous, intronic or in UTRs?***

**Explain using about 100 words the main results of this lab and what it might mean for human medicine.**