Class 14: Whole exome sequencing and disease gene identification

Reading assignment

- Glazov et al. Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. PLOS Genetics 2011 7:e1002027

Classroom activity (limit 45 minutes)

1. What are the symptoms of the inherited disease anauxetic dysplasia? What gene had been identified as causing this disease prior to this publication, what is the product of that gene and what functions(s) does it play in the cell?

   Anauxetic dysplasia causes skeletal deformities of the limbs and spine as well as short stature. This particular form of dwarfism has been attributed to mutations of the RMRP *gene, a non-translated RNA that forms part of RNAse MRP, an essential ribonucleoprotein complex involved in processing the 5.8S rRNA of the ribosome.*

2. How did the authors show that the previously identified gene was not the cause of the disease in the affected family studied? What is "whole-exome sequencing" and why do you think it was a good choice to identify he disease causing gene?

   They sequenced the RMRP *gene in the affected individuals and found no mutations, which showed that some other gene must be the cause. Without any genetic evidence about the location of the mutation (it was novel and so no data could have been available) they chose whole-exome sequencing to find mutation(s) responsible for the disease. They assumed the mutation would be in the coding region [they could have been wrong] so the limited sequencing to the expressed parts (exons) of the genes. The "exome" is the sum of all exons in the genome. That would reduce the amount of sequence needed to be determined—only 1.5% of the genome or about 45 million bases.*

3. Why did the results of whole-exome sequencing reveal so many single nucleotide polymorphisms (SNPs) in the four family members? Why did the authors assume that mutations of a single gene would be the cause of the disease and that it was likely to involve recessive alleles carried by each parent? What does "autosomal recessive compound heterozygous" inheritance mean and why did the authors assume that the alternative of "autosomal recessive homozygous" inheritance was unlikely? (If you didn't do it already—and you should have—you may have to Google these terms.)

   SNPs are polymorphisms between individuals, so the number found (~17,000) is about what is expected. Because both parents are unaffected, the disease must be recessive and both parents are carriers (Aa). Autosomal recessive compound heterozygous describes a situation where the affected person carries two dissimilar allelic mutations. Because the disease is so rare, and the two parents were unrelated they assumed that they would carry different mutations, which turned out to be true. The mutations changed an Arg codon to a termination codon and a Gly to a Glu codon (in Figure 2A).

4. How did the authors narrow their search down to only four candidate genes? What is the "SIFT algorithm" (Google again?) and how does it identify the likely disease gene? How many candidate genes did the authors identify and how did they narrow the candidate down to the *POP1* gene? What is the function of the *POP1* gene product and how does it relate to the previously identified gene known to mutate to produce the diseas?

   They identified SNPs that fell in the same gene in both parents and satisifed the autosomal recessive compound heterozygous prediction. The SIFT algorithm looks for mutations at evolutionarily highly conserved positions on the assumption that the conservation implies an important function for the amino acid. Both of the POP1 *mutations satisfy this criterion while those in the other candidates did not. The Pop1 protein is part of the RNAse MRP and P complexes, both of which include a catalytic RNA, the RMRP RNA being part of MRP, suggesting that the* POP1 *mutation might cause the same molecular defect as the* RMRP *mutations.*

5. Describe the results of research in yeast on the *POP1* homologue that the authors quote to suggest a probable model for disease etiology. Why do you think that the effect of the studied yeast and Drosophila mutations were so much more severe than those in the human subjects?

   The POP1 *mutations are lethal in the model systems (they interfere with proper ribosome assembly, though the paper doesn't explain that). Mutations in* RMRP *vary in severity suggesting that they affect activity of RNase MRP to different extents. The lethal phenotype of the* POP1 *mutations in yeast and Drosophila result from true null mutants (loss of function) whereas the mutations in humans are not lethal. This suggests that the mutations in this affected family reduce but don't eliminate RNase MRP function. [I didn't ask this, but the molecular defect in the affected individuals is substantial reduction in RMRP levels but not total loss, which could be consistent with this explanation.]*

6. Why did the authors test cell proliferation in the affected and unaffected members of the family? CFSE (carboxyfluorescein succinimyl ester) was used in this experiment. Why do you think this molecule allows you to measure cell proliferation and what results did the authors obtain? Is this result the one you would expect for this disease?

   Cells defective in the essential RNase MRP should grow slowly. CFSE covalently labels proteins in the cell and is reduced in intensity as cells divide. They saw increased reduction ("dilution") in the parents compared to the affected children. This is the expected result.

7. What general principle of the genetics of diseases like anauxetic dysplasia does this experiment suggest and how does it explain the frequent occurrence of novel forms of orphan diseases like it? What is the clinical importance beyond this disease does the experiment suggest?

   By showing that two components of the complex can have the same affect on the cell and induce similar (they weren't identical) symptoms they demonstrated the possibility for multiple genes encoding complex components to underlie diseases, which explains why not all individuals can be identified as carrying a mutation in an identified disease gene. They showed that whole-exome sequencing could solve the mystery of such diseases; the approach is expensive but not significantly greater than other diagnostics currently used.