BSC 4934: Projects List

1. **SNPs and the HapMap Project**: The International HapMap Project is a multi-country effort to identify and catalog genetic similarities and differences in human beings. Read about it at: http://snp.cshl.org/thehapmap.html.en. You can find a wonderful list of publications at: http://snp.cshl.org/publications.html.en. You will start by understanding (a) what is the HapMap, (b) how will the HapMap benefit human health and (c) what is the SNP Consortium? You will then read the paper by Sabeti et al. (Sabeti *et al.*, 2007) from the list of publications.
2. **Metagenomics**: Microbes live in communities. In general, it is hard to isolate and study individual species/strains of microbes. Current technologies focus on the study of the community of microbes as a whole. Such a study is referred to as metagenomics. Craig Venter in a pioneering study looked at microbial communities in the ocean and discovered about 150 new bacterial species and over a million new genes. You will read Venter's paper (Venter *et al.*, 2004) and a paper on how to analyze metagenomic data (Wooley *et al.*).
3. **DNA and RNA Editing:** Proteins from some families are known to be able to "edit" DNA and RNA molecules. Using data from the NCBI Trace Archive, researchers from George Church's laboratory have identified one systematic sequencing error (which may have crept into HapMap data) and evidence of DNA and RNA editing (Zaranek *et al.*).
4. **Human Methylome**: Epigenetics refers to changes in phenotype (appearance) or gene expression caused by mechanisms other than changes in the underlying DNA sequence. One of these mechanisms is DNA cytosine methylation, which has been proposed as a stably inherited modification that can affect gene regulation and cellular differentiation. The Human Epigenome Project (HEP) aims to study DNA methykation in a comprehensive manner. See http://www.epigenome.org/ and publications at http://www.epigenome.org/index.php?page=publications. You will present the work of Lister et al. published in Nature 2009 on high resolution methylome maps (Lister *et al.*, 2009).
5. **Phenologs:** Orthologous proteins in two different organisms often have the same molecular function, but may affect different phenotypes. Such functionally equivalent phenotypes between two organisms related by the orthology of the associated genes are called "phenologs". Thus instead of studying genes affecting human breast or eye cancer, an acceptable model system to study specific sets of genes in *C. elegans*. McGary et al. show a systematic way to identify phenologs in order to discover human disease models in organims that are easier to study (McGary *et al.*).
6. **Alternative Splicing in Human Genomes**: The paper by Sultan et al. from Science 2008 used deep sequencing to show that exon skipping is the most prevalent form of alternative splicing in the human genome (Sultan *et al.*, 2008).
7. **Applications of Next Generation Sequencing (NGS) Technologies**: Start with a review of next generation sequencing techniques (Shendure & Ji, 2008) and a review of ChIP-Seq, which is an application of NGS to do a genome-wide profile of transcription factor binding (Mardis, 2007). Also see (Johnson *et al.*, 2007).

References

Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502.

Lister, R., Pelizzola, M., Dowen, R. H. & other authors (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322.

Mardis, E. R. (2007). ChIP-seq: welcome to the new frontier. *Nature methods* **4**, 613-614.

McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B. & Marcotte, E. M. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci U S A* **107**, 6544-6549.

Sabeti, P. C., Varilly, P., Fry, B. & other authors (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918.

Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nature biotechnology* **26**, 1135-1145.

Sultan, M., Schulz, M. H., Richard, H. & other authors (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956-960.

Venter, J. C., Remington, K., Heidelberg, J. F. & other authors (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74.

Wooley, J. C., Godzik, A. & Friedberg, I. A primer on metagenomics. *PLoS Comput Biol* **6**, e1000667.

Zaranek, A. W., Levanon, E. Y., Zecharia, T., Clegg, T. & Church, G. M. A survey of genomic traces reveals a common sequencing error, RNA editing, and DNA editing. *PLoS Genet* **6**, e1000954.