

Genomic Dark Matter

17.2.09

Objective: To give a presentation of about 40-90 minutes at the end of the week covering the key aspects of genomic dark matter – the part of the genome under purifying selection, but with no functional characterisation.

Characterizing and understanding the genome is very important. The last decade has been faced with a challenge in relationship to this. Protein coding genes constitute 1.5%, RNA coding genes probably less than this. The amount of DNA under selection is at least 5% and would be declared functional, but what does the remaining percentage do?

The Big Questions:

- What is the overall structure of the human genome?
- How do you classify the different components of the human genome?
- How do you detect purifying selection?
- How does the fraction under negative selection change over evolutionary time?
- Does a sequence lose/gain function?
- How is the percentage of functional DNA
- How is function assigned to a sequence?
- Find special classes of sequences that have been assigned function in recent years.

Recommended literature is only meant to get you started. You might very well be able to find papers more suited for your purpose.

Boffelli D, Nobrega MA, Rubin EM: Comparative genomics at the vertebrate extremes. *Nat Rev Genet* 2004, 5:456-465.

Brent, M.R. Steady progress and recent breakthroughs in the accuracy of automated genome annotation. *Nature Rev. Genet.* 9, 62 - 73 (2008)

Chiaromonte F, Weber RJ, Roskin KM, Diekhans M, Kent WJ, et al. (2004) The share of human genomic DNA under selection estimated from human mouse genomic alignments. *Cold Spring Harb Symp Quant Biol* 68:245-254.

Clamp, M. et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. USA* (2007)

Glazov EA, Pheasant M, McGraw EA, Bejerano G, Mattick JS. Ultraconserved elements in insect genomes: a highly conserved intronic sequence implicated in the control of homothorax mRNA splicing. *Genome Res.* 2005 Jun;15(6):800-8

Hoskins et al (2007) "Sequence mining and mapping of *Drosophila melanogaster* heterochromatin," *Science*

Keith JM, Adams P, Stephen S, Mattick Delineating slowly and rapidly evolving fractions of the *Drosophila* genome. *JS. Symp. Quant. Biol.* 68, 245 - 254 (2008)

Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* 409, 680-921 (2001)

Lunter, G. et al. Genome-wide identification of human functional DNA using a neutral indel model. *PLoS Comput. Bio.* 2, e5 (2006)

Ponting, C.P. The functional repertoires of metazoan genomes. *Nature Reviews Genetics* 9, 689-698 (2008)

Rymarquis, L.A. et al. Diamonds in the rough: mRNA-like non-coding RNAs. *Trends in Plant Science* 13, no.7 (2008)

Venter, J.C. et al. The sequence of the human genome. *Science* 291, 1304-1351 (2001)

Siepel, A. et al. Evolutionary conserved elements in vertebrate, insect, worm and yeast genomes. *Genome Res.* 15, 1034 - 1050 (2005)

Smith et al. (2007) "The release 5.1 annotation of *Drosophila melanogaster* heterochromatin," *Science*.

Vavouri et al. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans.

Walter K, Abnizova I, Elgar G, Gilks WR. Striking nucleotide frequency pattern at the borders of highly conserved vertebrate non-coding sequences. *Trends Genet.* 2005 Aug;21(8):436-40s

Woolfe et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 3: e7. DOI: 10.1371/journal.pbio.0030007.