

5th Bioinformatics Workshop 21st April 2005
Oxford Centre for Gene Function

“Regulatory Signals in Eukaryote Genomes”

- 2.00 Introduction: Gerton Lunter and Jennifer Taylor
Regulatory signals: a beginner's guide
- 2.30 Talk 1: Martin Taylor
Mammalian transcription: the scattergun effect.
- 3.00 Talk 2: Manolis Dermitzakis
Cis regulatory variation in the human genome.
- 3.30 Coffee Break
- 4.00 Talk 3: Irina Abnizova
Some statistical properties of regulatory DNA and association of regulatory elements and over-represented motifs.
- 4.30 Talk 4: Anton Enright
Computational prediction of microRNA targets.
- 5.00 Talk 5: John Hancock
TBA
- 5.30 Conclusion & Summary: Jotun Hein
Excursion to pub: ‘The Turf’
-
-

Light refreshments provided.

RSVP: Jennifer Taylor, taylor@stats.ox.ac.uk

Directions to the Henry Wellcome Building for Gene Function:

<http://www.stats.ox.ac.uk/~taylor/Seminars.htm#Directions>

Organised by: Jennifer Taylor¹, Gerton Lunter¹, Chris Holmes^{1,2}, Jotun Hein¹

¹Oxford Bioinformatics Group, Department of
Statistics, University of Oxford.

²MRC Mammalian Genetics Unit
Harwell

Abstracts

MAMMALIAN TRANSCRIPTION: THE SCATTERGUN EFFECT.

Martin Taylor

Bioinformatics and Statistical Genetics, Wellcome Trust Centre for Human Genetics, Headington, Oxford. (<http://www.well.ox.ac.uk/%7Emst/>)

There is increasing evidence that a large portion of mammalian genomes are transcribed into a hugely diverse population of RNA species, only a fraction of which appear to encode functional proteins. This diversity represents both alternate RNA processing and diversity in transcription initiation. I will present results from whole genome surveys of transcription initiation in both mouse and man, illustrating the frequent use of alternate transcription start sites, leaky promoters and the conservation of transcription initiation between species.

CIS REGULATORY VARIATION IN THE HUMAN GENOME

Barbara E. Stranger¹, Matthew Forrest¹, Simon Tavaré², Andrew G. Clark³, Samuel Deutsch⁴, Robert Lyle⁴, Brenda Kahl⁵, Stylianos E. Antonarakis⁴, Panagiotis Deloukas¹, Emmanouil T. Dermitzakis¹

¹The Wellcome Trust Sanger Institute, Hinxton, UK; ² Dept of Oncology, Cambridge University, UK; ³ Department of Molecular Biology and Genetics, Cornell University, NY, USA; ⁴Department of Genetic Medicine, University of Geneva Medical School, Geneva, Switzerland; ⁵Illumina, Inc., San Diego, CA, USA.
(<http://www.sanger.ac.uk/Teams/Team16/>)

The goal of this work is to identify and characterize functionally variable regulatory regions that are likely to contribute to complex phenotypes and disorders in human populations, through effects on regulation of gene expression. We surveyed gene expression levels for ~ 700 genes in a sample of immortalized lymphoblastoid cell lines from 60 unrelated humans of the CEPH pedigrees, and used the publicly-available HapMap SNP genotypes of the same individuals to perform association analyses, in an attempt to localize the genetic determinants of these quantitative traits. Approximately 300 of the 700 genes gave a detectable expression signal relative to background, and for most of those loci, we observed significant gene expression variation among individuals. We identified loci that exhibited highly significant associations between gene expression and SNP variants located *cis*- to the coding locus. We are finding many regulatory haplotypes several Mb away from the target gene suggesting that the regulatory landscape may be different from what has been hypothesized. For other genes, a *cis*- signal was detected, but its effect was spread over many SNPs and thus did not meet our significance threshold. By working with the complete genomic set of HapMap SNP genotypes, we were also able to identify significant *trans*-acting SNPs influencing expression variation.

SOME STATISTICAL PROPERTIES OF REGULATORY DNA, AND ASSOCIATION OF REGULATORY ELEMENTS AND OVER-REPRESENTED MOTIFS.

Irina Abnizova, Klaudia Walter, Rene te Boekhorst and Walter R. Gilks

MRC Biostatistics Unit, Cambridge, UK

(<http://www.mrc-bsu.cam.ac.uk/BSUsite/AboutUs/People/irina.shtml>)

A main goal in interpreting DNA sequences is to understand how the information that specifies time and place of gene expression is encoded. An important step in this process is the recognition of gene expression regulatory elements and regions. Experimental procedures for this are slow and expensive. We present a statistical approach to show the association of experimentally verified regulatory elements with over-represented motifs within regulatory regions, together with a way to recognise these regulatory regions.

Recently, we have developed and applied a new statistical method to characterise regulatory DNA. The method exploits the fundamental property of the abundance of over-represented transcription factor binding motifs within regulatory regions.

In our data the method detected significant statistical differences between the probability distributions of similar words in regulatory and other DNA. The over-abundance of similar words should show up as outliers in the right tail of the distribution of similar word lists of variable length. The “fluffy tail test”, we proposed in this work, is designed to identify such outliers and is a useful technique when data from multiple genes and genomes are lacking.

An additional benefit of the method is exact way to find the over-represented motifs in the form of exceptionally large lists of similar words. We show that, in the data used, our method is able to establish the statistical association of over-represented motifs with experimentally confirmed TFBS. This association allows the method to be potentially used as complementary tool for motif discovery.

We apply the recognition method to a set of highly conserved non-coding elements (CNEs) derived by cross-vertebrate-genomic comparison, believed to be enhancers. Our aim is to prove that CNEs contain more instances of putative regulatory elements than by chance, and to discover putative binding motifs within them.