

9th Bioinformatics Day on "Genetics of Disease – problems and new methods"

April 26th 2007 - 1pm - 5pm in the Oxford Centre for Gene Function, Oxford

On April 26th a Bioinformatics Day with focus on genetics of disease will take place. The day will be dedicated to interesting topics and will also be used to announce seminars, study group and other meetings in Genomics/Bioinformatics in the following term. If you wish to attend this event, please send an email to Madeline Mitchell (mitchell@stats.ox.ac.uk) with your name and affiliation. This is necessary, so we can arrange coffee and biscuits appropriately and since there is a limit of a 100 in the seminar room in OCGF. Information on bioinformatics days can be found at <http://www.stats.ox.ac.uk/mathgen/bioinformatics/> under activities.

1.00 **Professor Jotun Hein:** *Short Welcome*

1.10 **Professor Bill Cookson (Imperial College London):** *"Asthma as test case for the study of genetic diseases"*

The research group of Cookson and Moffatt have been completing a high-throughput genomics programme to identify susceptibility alleles for asthma and AD. Datasets; 1. Gene expression of stimulated epithelial cells: The global gene expression response of keratinocytes, dendritic cells and airway epithelial cells to immunogenic substances such as bacterial proteins and whole bacteria was observed. 2. Global gene expression levels in EBV transformed cell lines from children with asthma and AD and their siblings (600 sibpairs) have been profiled (eQTL mapping dataset). 3. Whole genome association study of 3000 subjects in nuclear families identified through probands with asthma or AD. The families include 910 sibpairs selected to provide extreme trait distributions. This dataset contains all the subjects from the eQTL dataset described above. The methodology applied consist of: Disease Loci and eQTL Mapping will define genomic regions linked to disease, Expression analysis will identify gene expressions of interest to disease and epithelial cell biology. Regulatory Signals will further propose regions of regulatory significance and interactions between genes.

1.30 **Dr Jonathan Marchini (Dept Statistics, Oxford):** *The Wellcome Trust Case Control Consortium : "A Genomewide association study of 14,000 cases of seven common diseases and 3,000 shared controls"*

The Wellcome Trust Case-Control Consortium (WTCCC) is a large collaborative study that has carried out genomewide association studies for 7 common human diseases (Type 1 Diabetes, Type 2 diabetes, Coronary Artery Disease, Hypertension, Bipolar Disorder, Rheumatoid Arthritis and Crohn's Disease) based on 3,000 UK controls and 2,000 case subjects of each disease (<http://www.wtccc.org.uk/>). The study has been successful in uncovering numerous novel disease genes, many of which have already been replicated. I will present a selection of the results from the study, together with a brief overview of the methodological developments and lessons learnt from the study.

1.50 **Dr Richard Mott (Wellcome Trust Centre For Human Genetics, Oxford):** *"QTL Networks"*:

We present results suggesting that quantitative trait loci (QTL) for some phenotypes can be linked together by genes under the QTLs with correlated expression levels. This suggests a method for identifying the causative genes. However, some caution must be applied since not all phenotypic variation is caused by transcript level variation, and linkage disequilibrium effects may produce apparent expression level correlations that are unrelated to the causative gene.

2.10 **Dr Jennifer Taylor (WTCHG, Oxford):** *"Correlation between expression and shared physical genome location"*

Transcriptional units that lie physically close to each other in the genome may share regulatory machinery, evolutionary history and functional properties. While several famous examples of this phenomenon exist in the genome, for example the MHC, the generality and utility of this concept beyond such examples is not clear. An increased understanding of how multi-genic regions of the genome are arranged and regulated may assist in our ability to detect regulatory processes such as chromatin modelling and shared regulatory motifs. We present the analysis of several whole genome expression datasets within which we looked closely at relationships between transcript abundance, functional annotations and shared physical location.

2.30 **Dr Julian Knight (WTCHG, Oxford):** *"Functional consequences of DNA sequence polymorphism for gene expression"*

In this talk I will discuss the accumulating evidence that genetic diversity in non-coding DNA may modulate gene expression. Functional analysis of polymorphic variants is an essential tool to aid fine mapping of disease association. I will review approaches to analysing gene expression as a quantitative trait and strategies for analysing allele-specific gene expression. Data from the TNF locus illustrates how gene expression may be associated with specific haplotypes and single nucleotide polymorphisms.

2.50 -3.20 **BREAK**

3.20 **Jotun Hein, Professor Lior Pachter and Rahul Satija (Dept Statistics, Oxford):** *"Statistical Alignment and Regulatory Signals"*

This very short presentation will summarize some recent progress we have done on a very natural way of finding regulatory signals. Phylogenetic footprinting has proven popular and successful in finding regulatory signals by comparison, but has the flaw

of being alignment dependent. The methodology of statistical alignment has the ability of removing that flaw. New algorithms has been designed and implemented by Rahul Satija. Much remains to be done, but the approach is very promising.

3.40 **Jo Davies (Dept Statistics, Oxford): “Expression Data Analysis”**

I present work (in progress!) on a Bayesian approach to detecting differential gene expression motivated by an asthma case-control expression data set. I show how prior information from previous high-throughput analyses can be used to yield posterior probabilities for differential expression and identify genes which may be missed using standard blind testing methods.

4.00 **Dr Chris Holmes (Dept Statistics, Oxford): “Statistical analysis of Illumina BeadChip data for genotyping and inference of copy number variation in the human genome”**

There is growing evidence that regions of copy number variation (CNVs) in the human genome are more widespread than previously thought. A CNV is defined as a segment of DNA >1 kb that is present at a variable copy number in comparison to a reference genome. It has been suggested that up to 10% of the human genome might be copy number variable (contributing to around 17% of genetical transcription variation) and copy number polymorphisms have previously been linked to a number of diseases. In recent work we have developed an objective Bayesian Hidden Markov model to detect regions of copy number variation from Illumina BeadChip (SNP) data. In our model the hidden states refer to unobserved copy number variants at a (SNP) locus and the transitions between states capture the persistence within CNV states across chromosomal regions. Parameters in the hyper-priors of the model are set "objectively" using prior training data of known CNV and Bayes factor thresholds (used for calling regions of CNV) are calibrated, via a simulation stage, to user set false positive rates. This is joint work with the Wellcome Trust Centre for Human Genetics.

4.20 **Dr Doug Higgs (WIMM, Oxford): “A promoter mutation identified as disease cause in α -thalassaemia”**

We describe a new mechanism underlying human genetic disease by identifying a gain of function regulatory SNP (rSNP) that causes a form of alpha thalassaemia which occurs at polymorphic frequencies in Melanesia. Association studies localised the mutation to a 168kb segment of the genome including the alpha globin locus but conventional analyses failed to detect any molecular defect. After re-sequencing this region and using a combination of chromatin immunoprecipitation and expression analysis on a tiled oligonucleotide array, a regulatory SNP (rSNP) was identified in a nondescript region of the genome lying between the alpha globin genes and their highly conserved, remote, upstream regulatory elements. The rSNP creates a new promoter-like element which interferes with normal activation of all downstream alpha-like genes. This not only demonstrates a new mechanism of human genetic disease but also illustrates an important general strategy for distinguishing between neutral and functionally important rSNPs.

4.40 **Dr Charlotte Dean (MRC, Harwell): “Mouse as model organism for Asthma”**

We are taking two complimentary approaches to understanding the pathobiology of asthma using mouse models. Firstly we analyse mouse mutants of previously identified human asthma susceptibility genes. Secondly we identify mice carrying mutations that result in developmental lung defects; we then assess these mice using inducible mouse models of asthma and fibrosis to see if the mutations result in asthma susceptibility and/or enhanced induction of fibrosis. I will present an overview of these approaches.

5.00 **Jotun Hein and Chris Holmes: Open Discussion and Summary**

Beers at Lamb and Flag