

HGM2005



Plenary Abstracts

PLENARY SESSION I THE HUMAN GENOME

Date and time: Monday 18TH APRIL 2005

Place: MAIN HALL(1ST FLOOR)

PLENARY ABSTRACT NO. 1

Exon sequencing to identify variants in human genomes

Jane Rogers, Andy Dunham, Sarah Hunt, David Niblett, Steve Leonard, Robert

Davies, James Bonfield, John Burton, Simon Holroyd, Kirsten McLay, Claire Bryan, Andrew Bentley, Angus Clayton, Tom Martin, Mark Earthrowl, Alison Coffey, David Bentley

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs, England, CB10 1SA UK

Understanding human genetic variation and its relationship to disease susceptibility is a major goal of the next phase of the human genome project. As the cost of sequencing multiple humans at the whole genome level with existing technologies is currently prohibitive, the Sanger Institute has initiated a project that aims to identify sequence variants that are common in the population in the exons of all known human protein coding genes. In the first phase of the project, each exon is being amplified and sequenced in a panel of 48 Caucasian DNAs. Sequence data are analysed for variation using automated SNP detection software developed at the Sanger Institute and SNPs and other variants will be deposited in dbSNP. A new web browser, Glovar, based on Ensembl, will also provide views of the position of SNPs in genes and their consequences on expressed proteins.

To undertake this project in a systematic way, effort has been invested in developing a robust pipeline tracked in an Oracle database. It has also been necessary to develop software capable of automatic sequence analysis and SNP calling (Exotrace), which can be coupled with manual review (using Gap4), to be able to process the anticipated throughput of 50,000 traces per day.

In addition to the core human genetic variation project, working primer sets are available to study genes of medical interest in great depth by sequencing DNA from patients with particular clinical phenotypes. To date nearly 1,000 genes of medical interest have been prioritised for sequencing in a variety of disease studies.

PLENARY ABSTRACT NO. 2

From Elements to Systems

Yoshiyuki Sakaki

RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

The completion of the human genome followed by the experimental and computational annotation revealed the presence of 20000 to 25000 protein-coding genes and a large amount of non-coding RNA genes in the human genome. An obvious next question to be addressed is what function those genes have and how those genes co-operatively work to generate complex biological systems. Many studies on gene function have been done in the past, but in many cases only from the viewpoint of specific biological process. Now the completed sequence of the human genome enables us to see the function of the genes from genome-wide viewpoints by the aid of high-throughput technologies and computational tools. In other words, we are at the stage to be able to understand the biological phenomena as a system operated by a cooperative work of many molecular elements. Taking the progress of the human genome researches into consideration, Japan initiated a national project designated 'Genome Network Project (GNP)' in 2004. The first goal of GNP is to construct a protein-DNA, protein-protein interaction map of nearly 2000 transcription regulatory proteins and the regulatory elements of all the human genes to elucidate the regulatory network of gene expression. GNP can be characterized by systematic cooperative studies of a large scale genome center (RIKEN GSC) and many university laboratories. The large scale center will produce a large amount of data by using its high-throughput facilities to construct a framework of the interaction map and the university laboratories shape up the map based on their own biology and medicine. GNP has already made significant progress in the first year. I will, as the principal coordinator of the project, present an outline and some progress of the Genome Network Project.

PLENARY ABSTRACT NO. 3

From Genomics Back to Genetics

Richard Gibbs

Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA

The expertise developed in the last decade has enabled the generation of multiple genome sequences. We have deciphered the human, rat, *Dictyostelium discoideum* and *Drosophila melanogaster* genomes in collaborative efforts, and undertaken the sequencing of *Drosophila pseudoobscura*, the honey bee, the sea urchin and the bovine genome. The Rhesus macaque is also being analyzed in collaboration with the Washington University Genome Sequencing Center (GSC) and TIGR, and the Orangutan with the GSC. The Tammar Wallaby will be sequenced in collaboration with the Australian AGRF. An original and central aim of the human genome project was to improve the understanding of the biology of individual diseases by discovery of specific alleles that underlie different human disease conditions, and both data and methods from genome projects are being applied to this task. We have established a pipeline that allows sequencing of thousands of PCR products, simultaneously screening large numbers of candidate disease genes in several affected individuals for both putative functional mutations and associated polymorphisms. The system was calibrated by resequencing a 2.5Mb genomic region and has about a 7%/7% false +/-ve rate for mutation detection - or near 100% detection if humans analyze the data. The pipeline is being applied to re-sequence several genes and regions, including behavioral disorder candidates, and West Nile Virus susceptibility regions, and all ~300 ion channels in patients with Idiopathic Epilepsy. The epilepsy study has already revealed multiple coding changes from the first few genes and patients, and the current challenge is to attribute functional significance. This project, which is a collaboration with Jeffery Noebels and Dan Burgess, is a model for approaching other disorders with complex genetics.

HGM2005



Plenary Abstracts

PLENARY SESSION II DISEASE-ORIENTED LARGE-SCALE POPULATION STUDIES

Date and time: Tuesday 19TH APRIL 2005

Place: MAIN HALL(1ST FLOOR)

PLENARY ABSTRACT NO. 4

Genetics of Common Complex Traits: Diseases and Lifespan

Kari Stefansson

deCODE genetics, Reykjavík, Iceland

The genetics of complex traits is not only a field that deals with traits that are complex; the field itself is rather complex in its own right. The complexity of the traits is rooted in many things, including the relationship between variations in the sequence and variations in phenotype; interactions between polymorphisms in the genome and their impact on the phenotype; the influence of environmental factors on the phenotype; and the impact of interactions between the environmental factors and polymorphisms on the phenotype. The complexity of the field could be looked upon simply as the sum total of the complexity of the traits, but there is probably even more to it than that. There are fundamental issues that muddy the water, such as the ambiguity in all of the relationships mentioned above. This ambiguity is rooted in the fact many of the relationships do not exist independent of the others and in the lack of consensus on methodological issues such as how and how much to correct for multiple testing; how to select populations for replication studies; and the definition of what constitutes a replication. To the untrained eye it may seem as though the field of statistics has lost some of its credibility when it comes to complex traits. For example, the best of journals will not publish papers on the cloning of genes with variants that confer risk of complex traits in man without a replication in a second population. It appears that they are reluctant to bet on the reliability of the methods of statistical genetics to establish whether an association of a sequence variant to a phenotype is more than a chance association. One fundamental reason to ask for a replication is that if the same sequence variant is associated with the phenotype in the second population, then association has been established without multiple testing. Hence, the concern over how and how much to correct for it evaporates. That said, it will probably often be the case that the sequence variants that associate with a trait will differ between populations. At deCODE genetics we have over the past eight years gathered considerable

experience in analysing the genetics of complex traits. In my talk I will try to use this experience to shed some light on the problems discussed above and in the process tell a few short stories about schizophrenia, stroke, myocardial infarction, peripheral arterial occlusive disease, osteoporosis, type 2 diabetes, and longevity.

PLENARY ABSTRACT NO. 5

From families to populations: impact of disease genes

1, 2 Leena Peltonen

1. Department of Molecular Medicine, National Public Health Institute and Department of Medical Genetics, University of Helsinki, Helsinki, Finland,

2. Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA

The Human Genome Project has produced a high number of catalogued sequence variants enabling genome-wide studies of genetic loci behind rare- and common disease-related phenotypes. However, the strategy of disease gene hunt is highly dependent on the accessible population and study samples. Multiple uncertainties must be solved before the best possible strategy for the gene hunt can be designed and large-scale genome-wide investigations undertaken. Further, rapidly increasing information of the structural or functional variability within the genome will greatly affect the interpretation of the impact of identified variants.

European population isolates like Finland have been very useful for mapping and cloning genes for rare disorders; in such isolates genetic drift leads to an overabundance

of disease-alleles for particular disorders, and a high proportion of patients share these alleles, identical by descent. The concept that the isolates are similarly advantageous for genetic studies of common diseases has been challenged,

and only few samples exist to prove if it really would be more straightforward in such populations to detect disease-related haplotype signatures through association studies. Despite these reservations, detailed information of the population

history is increasingly understood as one crucial factor of success in genetic studies of common diseases. I will describe the features of the Finnish population and our efforts to search for disease genes for rare and common phenotypes. Our studies will exemplify the strategies used to identify disease genes in various Finnish data sets and study samples. I will also demonstrate how the impact of specific allelic variants of disease genes, identified in families, can be addressed in epidemiological cohorts containing excessive amount of quantitative phenotype information. I will also describe our recent efforts in the EU-funded multinational twin study, Genomeutwin (<http://www.genomeutwin.org/>).

PLENARY ABSTRACT NO. 6

UK Biobank: how, what, why and when?

Bill Ollier, John Newton, Tim Sprosen, Tim Peakman
UK Biobank, Incubator Building, Grafton Street, Manchester, M13 9XX, UK

A major challenge for the 21st century is to understand how variation in genome sequence and environmental exposures interact to determine health and disease at both the individual and population level. UK Biobank is a major research initiative to support long-term prospective studies of genes and environment. It will follow the health of 500,000 volunteers aged 40-69 years across the UK for many years. Information on demographic, environmental and lifestyle factors is being collected and linked with stored biological samples and medical records.

A wide range of longitudinal health data will also be collected, and biological samples will be used for biochemical and genetic analyses. UK Biobank will provide an important resource for researchers to investigate a range of important common complex disorders and generate hypotheses that can be tested in other populations and case-control studies. The outcomes will be improved and coordinated public health initiatives, diagnoses and treatments.

The project is on course to start full recruitment of participants in January 2006. Phase 1 pilots will start in February 2005 and will involve 300-600 volunteers at various locations across the UK. They are testing the participant experience at the appointment centre and the ability to complete the questionnaire and physical measurements. Integrated phase 2 pilots involving up to 3000 participants are expected to start in September 2005 and will test the entire process, from invitation letter and consent through to sample collection, storage and retrieval, as well as the IT systems that will underpin the project. An independent Ethics and Governance Council has been established which will act as an independent guardian of the project's Ethics and Governance Framework.