" A stage for shaping the next generation of genome-wide association studies (GWAS)."

Workshop:

GWAS mega-analysis for complex diseases

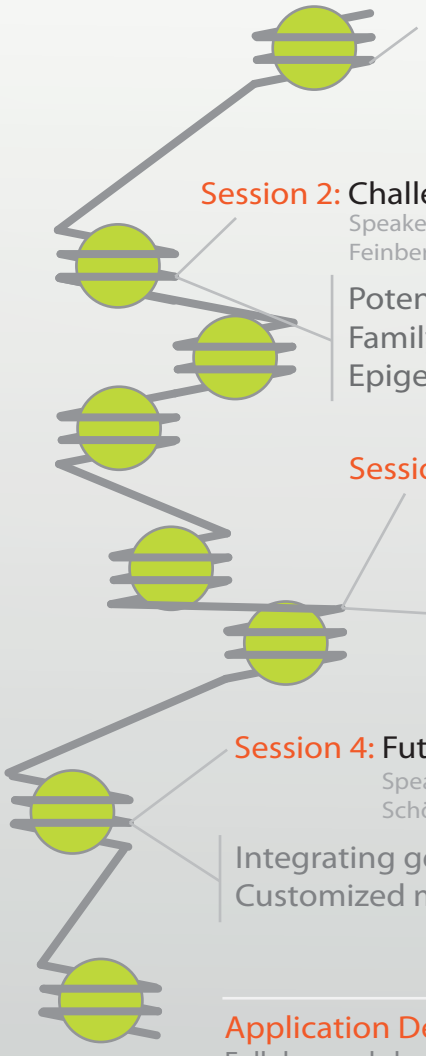
Part of the International Conference on Systems Biology (ICSB2010)

The workshop is aimed at statistical geneticists looking for improved tools for GWAS as well as researchers in machine learning and statistics with an interest in complex diseases, such as schizophrenia.

15. October 2010
Edinburgh, Scotland, UK

www.ICSB2010.org/workshop12

Program Highlights



Session 1: Introduction and showcase of GWAS potential

Speakers: Mark Daly, Emmanouil Dermitzakis, Leonid Kruglyak, Frank Dudbridge, Naomi Wray

" Together we can identify the most promising avenues of future research. "

Session 2: Challenges of GWAS

Speakers: Michael Metzker, Douglas Blackwood, Andrew Feinberg, Nicholas Schork, Benjamin Pickard

- Potential of sequencing technologies (rare SNPs, CNVs)
- Family studies for causal variants
- Epigenetics and the environment

Session 3: Bioinformatics and the pursuit of the silver bullet

Speakers: Jason H. Moore, Hakon Kakonarson, Mihai Pop, Silke Szymczak , Michael Brudno

- Bioinformatic approaches for next-gen sequencing data
- Biostatistics vs. Bioinformatics
- Practical limitations

Session 4: Future directions of GWAS analysis

Speakers: Hans van Houwelingen, Geoff McLachlan, Bernhard Schölkopf, Stephan Ripke, Heather Cordell

- Integrating gene function and disease pathways
- Customized machine learning approaches

See webpage for the full program

Application Details:

Full day workshop on the 15.10.2010
Application deadline: dd.mm. 2010
Registration Fee: £ xx (includes lunch)
Registration via webpage

Workshop Organizers:

Professor Bryan J. Mowry and
Denis C. Bauer,
The University of Queensland, Australia.

The successes, challenges and prospects for GWAS mega-analyses for complex diseases such as schizophrenia.

Purpose of the workshop

Genome-wide association studies (GWAS) are becoming the method of choice for studying disease etiology with an increasing number of GWAS studies reporting rapid progress towards uncovering the genetic markers for complex diseases such as schizophrenia[1]. This workshop is aimed at introducing an interdisciplinary audience to the potential of GWAS. It will cover the fundamental assumptions, showcase recent successes and discuss limitations of current GWAS approaches in the field of complex diseases. It will provide a stage for shaping the next generation of GWAS by drawing on the audience's interdisciplinary expertise in e.g. statistics and machine learning to overcome present challenges and identify the most promising avenues of future research.

Background

GWAS has recently made possible by the generation of human genetic variation maps (e.g. HapMap project[2]), the large-scale availability of clinically-phenotyped samples and the development of statistical methods to identify significantly associated genetic variants without bias[3]. Despite the success for some diseases where the identified markers are sufficient to reliably predict disease risk, for complex diseases such as schizophrenia[4], the single nucleotide polymorphisms (SNPs) reported using current GWAS methods describe only as small proportion of the observed heritability. Studies aiming to explain this "missing" heritability by analysing SNPs of nominal statistical significance appear speculative[3]. Thus, reliable methods for interrogating these types of data are needed to move GWAS forward and identify markers of medical relevance for complex diseases.

Incorporating expression of gene transcripts into the analysis is one way of increasing predictive power[5]. Polymorphisms in regulatory elements can have a direct influence of the abundance of gene transcripts. They are quantitative traits that can be mapped to the genome with high resolution and are termed expression quantitative trait loci (eQTLs)[6]. The evidence of SNPs, identified via GWAS, can hence be strengthened by requiring the same genetic markers to also be associated with quantitative transcription levels of one or more genes (eSNPs)[5]. However, this method strongly depends on the availability of appropriate expression data, e.g. appropriate tissue, which, in the case of psychiatric disease, is problematic. Other strategies are needed to overcome the issue of "missing" heritability.

Recent studies suggest that hundreds or even thousands of variants contribute to the risk of complex diseases and that other effects play a role, such as epigenetics, copy number variants (CNVs), insertions and deletions, short tandem repeats and single amino acid repeats[1]. The influence of these genetic variants cannot be studied using common SNP libraries averaged over a population (meta-analysis) but calls for the sequence data from individuals (mega-analysis)[7]. The sheer size and complexity of such data highlights the need to develop increasingly sophisticated statistical and machine learning methods for studying the genetic origins of complex diseases.

Audience and Outcome

By bringing together researchers with an interest in augmenting GWAS with novel statistical approaches and machine learning technologies to study the etiology of complex diseases such as schizophrenia, the workshop aims to bridge the gap between statistics, bioinformatics and genetics.

We envisage that researchers in machine learning and statistics who have an interest in schizophrenia, as well as statistical geneticists looking for improved tools for GWAS will gain from participating in this workshop. The participants might also welcome the co-authorship on the editorial or opinion paper, which will be written on the discussion points and the identified avenues of future research.

Format and preliminary program

The one-day workshop will have a program of invited contributions seen to fulfil the purpose of the workshop: to nurture collaboration between active researchers and to showcase significant and representative efforts, advancing GWAS and complex disease etiology.

Time	Topic	Speaker
9:00 am	Welcome	Bryan J. Mowry
	Session 1: Introduction and showcase of GWAS potential	
9:10 am	The theory and practice of GWAS	Mark Daly
9:30 am	Successful case studies and the type of knowledge to be gained from GWAS	Emmanouil Dermitzakis
9:50 am	The nature of complex disease genetics: large number of variants of small non-linear effects	Leonid Kruglyak
10:10 am	Discussion • Power/sample size implications • How to link genomic variants with disease genes	Frank Dudbridge, Naomi Wray
10:40 am	Break	
	Session 2: Challenges of GWAS: from genome architecture to environmental influences	
11:00 am	The potential of sequencing technologies to identify: (a) low frequency and rare SNPs; (b) structural variants (CNV)	Michael Metzker
11:20 am	The utility of family studies to identify causal variants	Douglas Blackwood
11:40 am	The influence of non-DNA sequence: epigenetics and the environment	Andrew Feinberg
12:00 am	Discussion • Strength and weaknesses of arrays and sequencing platforms • Strength and weaknesses of targeted re-sequencing	Nicholas Schork/ Benjamin Pickard

	and the appropriate reference sequence • Anticipated improvements of whole genome sequencing • How best to seek gene-environment interactions • How to study epigenetic influences in inaccessible tissue (brain)	
12:30 pm	Lunch	
	Session 3: Bioinformatics and the pursuit of the silver bullet	
1:30 pm	Biostatistics vs. Bioinformatics from regression to machine learning data	Jason H. Moore
1:50 pm	Practice and outcome of bioinformatics analysis of GWAS data: success and limitations	Hakon Kakonarson
2:10 pm	Computational methods for discovering disease-relevant SNPs and CNVs using next-gen sequencing data	Mihai Pop
3:30 pm	Discussion • The importance of epistasis (gene-gene interaction) • Filtering, wrapping and selecting which method to choose • *De novo* genome assembly benefits and drawbacks • Non-computational improvements in next-gen sequencing data (paired-end and length increase) • Computational infrastructure: are we prepared for the flood?	Silke Szymczak , Michael Brudno
4:00 pm	Break	
	Session 4: Future directions for GWAS analysis with customized machine learning approaches and data integration	
4:20 pm	Integrating biological expert knowledge of gene function and disease pathways in GWAS	Hans van Houwelingen
4:40 pm	Few samples with many features – lessons already learned from micro-array analyses	Geoff McLachlan
5:00 pm	Multi-instance learning and the benefit of sophisticated and customized machine learning approaches	Bernhard Schölkopf
5:20 pm	Discussion • Signal to noise ratio in public databases • Selection bias with filter- and wrapper-based feature selection • Robustness and over-fitting control • Risks of using “black-box” approaches • How to verify the accuracy of our findings	Stephan Ripke, Heather Cordell
5:50 pm	Forums Discussion and Wrap-up	Bryan J. Mowry

Details of invited speakers

Bryan J. Mowry (University of Queensland, Brisbane, Australia)
Mark Daly (Massachusetts General Hospital/Harvard University, USA)
Emmanouil Dermitzakis (University of Genf, Switzerland)
Leonid Kruglyak (Princeton University, USA)
Frank Dudbridge (MRC Biostatistics Unit, Cambridge, UK)
Naomi Wray (Queensland Statistical Genetics Laboratories, Brisbane, Australia)
Michael Metzker (Baylor College of Medicine, Houston, USA)
Douglas Blackwood (University of Edinburgh, UK)
Benjamin Pickard (University of Strathclyde, UK)
Andrew Feinberg (Johns Hopkins University School of Medicine, Baltimore, USA)
Nicholas Schork (UCSD, USA)
Paul Medvedev (University of Toronto, Canada)
Jason H. Moore (Dartmouth Medical School, Hanover, USA)
Hakon Kakonarson (The Children's Hospital of Philadelphia, USA)
Mihai Pop (University of Maryland, USA)
Silke Szymczak (Universität zu Lübeck, Germany)
Michael Brudno (University of Toronto, Canada)
Hans van Houwelingen (Leiden University, Netherlands)
Geoff McLachlan (University of Queensland, Brisbane, Australia)
Bernhard Schölkopf (MPI Kyp, Tübingen, Germany)
Stephan Ripke (MPI-P Munich, Germany)
Heather Cordell (Newcastle University, UK)

Additional information

Please note that this workshop is **not** overlapping but following the World Congress on Psychiatric Genetics (3-7. Oct 2010) http://www.ispg2010.org/

[1] Purcell, S. M. et al., Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460** (7256), 748 (2009).

[2] Frazer, K. A. et al., A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449** (7164), 851 (2007).

[3] Nica, A. C. and Dermitzakis, E. T., Using gene expression to investigate the genetic basis of complex disorders. *Hum Mol Genet* **17** (R2), R129 (2008).

[4] Cichon, S. et al., Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry* **166** (5), 540 (2009).

[5] Cookson, W. et al., Mapping complex disease traits with global gene expression. *Nat Rev Genet* **10** (3), 184 (2009).

[6] Schadt, E. E. et al., Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422** (6929), 297 (2003); Rockman, M. V. and Kruglyak, L., Genetics of global gene expression. *Nat Rev Genet* **7** (11), 862 (2006).

[7] Psychiatric GWAS Consortium Steering Committee, A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry* **14** (1), 10 (2009).