



HAL
open science

HGV2009 Meeting: Bigger and better studies provide more answers and more questions

Katherine Reekie, Andres Metspalu, Stephen J Chanock, Edison T Liu, Elaine Mardis, Stephen Scherer, Pui-Yan Kwok, Anthony J Brookes

► To cite this version:

Katherine Reekie, Andres Metspalu, Stephen J Chanock, Edison T Liu, Elaine Mardis, et al.. HGV2009 Meeting: Bigger and better studies provide more answers and more questions. *Human Mutation*, 2010, 31 (7), pp.886. 10.1002/humu.21270 . hal-00552389

HAL Id: hal-00552389

<https://hal.science/hal-00552389>

Submitted on 6 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HGV2009 Meeting: Bigger and better studies provide more answers and more questions

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0139
Wiley - Manuscript type:	Meeting Report
Date Submitted by the Author:	24-Mar-2010
Complete List of Authors:	Reekie, Katherine; University of Leicester, Department of Genetics Metspalu, Andres; <none> Chanock, Stephen; National Cancer Institute, Pediatric Oncology Branch Liu, Edison; National University of Singapore, Genome Institute of Singapore Mardis, Elaine; <none> Scherer, Stephen; The Hospital for Sick Children, The Centre for Applied Genomics and Programin Genetics Kwok, Pui-Yan; University of California, San Francisco, Cardiovascular Research Institute and Department of Dermatology Brookes, Anthony; University of Leicester
Key Words:	HGV2009, SNP, Variation, GWAS, CNV



Humu-2010-0139

Meeting Report

HGV2009 Meeting: Bigger and better studies provide more answers and more questions

Katherine Reekie(1), Andres Metspalu(2), Stephen J. Chanock(3), Edison T. Liu(4), Elaine R. Mardis(5), Stephen W. Scherer(6), Pui-Yan Kwok(7), Anthony J. Brookes(1)

- (1) Department of Genetics, University of Leicester, Leicester, UK;
- (2) Estonian Biocentre & University of Tartu, Tartu, Estonia;
- (3) Division of Cancer Epidemiology and Genetics and Center for Cancer Research, National Cancer Institute, Bethesda, Maryland, USA;
- (4) National University of Singapore, Genome Institute of Singapore, Singapore;
- (5) The Genome Center, Washington University, 4444 Forest Park Blvd, St. Louis, USA;
- (6) The Centre for Applied Genomics, Program in Genetics and Genomic Biology, Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada;
- (7) Dermatology, Cardiovascular Research Institute, & Institute for Human Genetics, University of California, San Francisco, San Francisco, California, USA.

1
2
3
4
5
6
7 **Corresponding Author:**
8

9
10 Anthony J. Brookes

11
12 Phone: +44 (0)116 2523401

13
14 Fax: +44 (0)116 2523378

15
16
17 Email: ajb97@le.ac.uk
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

Abstract

The 11th International Meeting on Human Genome Variation and Complex Genome Analysis (HGV2009: Tallinn, Estonia, 11th-13th September 2009) provided a stimulating workshop environment where diverse academics and industry representatives explored the latest progress, challenges, and opportunities in relating genome variation to evolution, technology, health, and disease. Key themes included Genome-Wide Association Studies (GWAS), progress beyond GWAS, sequencing developments, and bioinformatics approaches to large-scale datasets.

For Peer Review

Introduction

The 11th International Meeting on Human Genome Variation and Complex Genome Analysis was held in Tallinn, Estonia on 11th-13th September 2009 at the Sokos Hotel Viru Conference Centre. Over three days there were seven sessions and a total of 35 speakers, interspersed with poster sessions. The workshop format, involving over 120 academic and commercial attendees with diverse expertise, encouraged vibrant discussion and collaboration.

A core theme during the meeting concerned the “next step after GWAS”, including issues such as identifying causal genes and their role in disease, managing large volumes of GWAS data, meta-analyses and combining resources, and approaches to data sharing. Another main topic was technology development and the emergence of sequencing as an increasingly powerful and universal tool for large-scale genetic variation research. The increasing number of complete genomes available continues to emphasize the degree of variation that exists between individuals and between populations. And the massive amounts of data generated by modern techniques stresses the need for lab scientists to have a good knowledge of bioinformatics. More generally, investigations focussed upon the human exome, comprising 1% of all human DNA, were also a main feature of the meeting. Compared to previous meetings in this series, there was less focus on animal models of disease but more emphasis upon understanding and predicting a broad range of human phenotypes (Box 1).

What Next After GWAS?

The power of GWAS was illustrated in the meeting's first session by Pui-Yan Kwok, who spoke about moving on from GWAS given that such studies have now yielded many positive genome regions with encompassed candidate genes that can be variously explored. The success of GWAS to date was illustrated by Stephen Chanock in the realm of identifying multiple novel cancer susceptibility regions, most having a moderate effect on risk (odds ratios 1.1 – 1.3), and by Laura Zagato who described hypertension studies which started with GWAS and progressed to ultimately confirm disease association with a single gene. However, many regions identified by GWAS do not contain any, or any obvious, genes or regulatory elements, and this makes it very difficult to move forward. This was discussed by Rick Kittles, who described the 8q24 locus that contains few known genes but many SNPs which have been strongly and repeatedly associated with prostate cancer.

GWAS research will often suggest many regions of association, each having relatively small effect and containing many genes that could be involved in the disease risk modification. To help prioritise these for analysis, Yves Moreau described ENDEAVOUR - a tool that ranks genes based on their similarity to genes already implicated in a particular trait or process. Alternatively, one might prefer to follow a pathway-based strategy, as exemplified by Edison Liu, who showed how well this has worked with regard to breast and endometrial cancer.

Ultimately, it is hoped that GWAS and related research will enable the prediction of phenotypes based on genetic data. Beyond obvious uses in medicine, Manfred Kayser showed how an understanding of the genetics behind human phenotypes such as eye colour, hair colour and height can be used to help identify individuals. For example, eye colour is predictable with high

1
2
3 accuracy using only a small number of genetic variants, suggesting a role in forensics and the
4
5 profiling of crime suspects.
6
7

8 9 **Diagnostics and Pharmacogenetics**

10
11 GWAS has great potential in the area of pharmacogenetics and several examples of this were
12
13 presented at the conference, with one session focused entirely on this subject. Eileen Dolan
14
15 reported the discovery of “genetic signatures” contributing to susceptibility to chemotherapeutic–
16
17 induced cytotoxicity using cell-based models. Some of these SNPs translated into clinical
18
19 outcomes, highlighting SNP alleles significantly associated with survival in ovarian cancer
20
21 patients receiving carboplatin and paclitaxel.
22
23
24

25
26
27 Another example came from Lorena Citterio, who showed that SNPs identified from candidate
28
29 gene and GWAS studies could predict patients most likely to be responsive to a novel anti-
30
31 hypertensive drug. Such progress will steadily open the door to patient sub-classification by
32
33 genotype, enabling smaller and more cost-effective clinical trials. To progress in this direction,
34
35 Michael Cariaso described an analysis of five complete human genomes that involved mining
36
37 public databases (e.g., dbSNP, PubMed, and SNPedia) via a tool called Promethease that outputs
38
39 a report listing, for example, medical issues related to drug metabolism specific to each
40
41 individual genome.
42
43
44

45
46 Similarly, structural variation profiles of cancer can provide a useful diagnostic and prognosis
47
48 indicator. Bauke Ylstra reported upon the use of array-CGH to distinguish between cancers that
49
50 are newly arising, and those resulting from metastasis of a previous tumour. Furthermore, some
51
52 different types of cancer have been shown to have characteristic array-CGH profiles and
53
54
55
56
57
58
59
60

1
2
3 distinctive methylation patterns that can be prognostic, as in the case of childhood leukaemias, as
4 explained by Lili Milani.
5
6

7
8
9 As such possibilities grow for the emergence of personalised and predictive medicine based upon
10 genome data, many ethical questions arise. The importance of good genetic testing practice and
11 considerations involved with the handling of such information was discussed by Jean-Jacques
12 Cassiman. He also noted the need for good genetic counselling facilities, and made a point of
13 asking who would be responsible for providing such testing and advice services in the future.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Currently, commercial providers are forging ahead in this new arena.

Data Sharing and Availability

27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
The issue of identifying individuals from aggregate level GWAS data was addressed head-on by
Kevin Jacobs, who provided guidance on how much information could be safely made public
without risking the privacy of research participants. The scope for individual identification was
shown to very much depend on factors such as the size of the study, and the number and
characteristics of the variants under investigation, with no identification being possible if any
number of p-values were reported without allele frequency data or directionality. If rational
guidelines can now be based upon such new insights, then researchers (and databases) will be
able to make openly available much more summary-level GWAS data than is currently being
shared, helping to accelerate biomedical research substantially.

49
50
51
52
53
54
55
56
57
58
59
60
Nevertheless, many data components that relate to individuals and their phenotypes should
probably never be openly shared, due not least to privacy concerns, commercial considerations,
and legal and financial issues. Therefore, increasingly standardised and unified methods for
controlling data access that do not unduly burden research progress must be devised. To this end,

1
2
3 Anthony Brookes reported on the EU funded GEN2PHEN project (<http://www.gen2phen.org>),
4
5 which is working towards effective data publication and data sharing enabled by digital user
6
7 identities and federated database solutions. Such developments will help move biomedical
8
9 databasing into the 'web 2.0' era, as exemplified by the current status and ongoing development
10
11 plans of the HGVbaseG2P genetic association database (<http://www.hgvbaseg2p.org>), as
12
13 described by Gudmundur Thorrisson. Furthermore, related ideas could be used to great benefit in
14
15 clinical settings, as discussed by Andrew Devereau, with particular focus on the experiences of
16
17 the National Genetics Reference Laboratory in the UK.
18
19
20
21
22

23 **Spotlight on Populations**

24
25
26 In the last year, the number of complete and semi-complete human genome sequences has
27
28 increased from two Caucasian genomes to several fold more individuals from a diverse range of
29
30 populations including Korean, Chinese, and African. Similarly, most previous GWAS have been
31
32 conducted upon Caucasian individuals, but Jianjun Liu described large-scale GWAS research
33
34 now in progress in China using Chinese population samples. This initiative has already
35
36 uncovered differences in psoriasis susceptibility between Chinese and Caucasian populations,
37
38 and dramatically increased the amount of genetic information available for the Chinese
39
40 population. One clear message from HGV2009 was the need for further GWAS projects in other
41
42 populations, to examine how disease susceptibility varies between populations.
43
44
45
46
47

48
49 Much is already known about the main differences in patterns of genetic variation between
50
51 populations, and even between different regions within countries. Given this, individual
52
53 countries have begun to create their own biobanks to enable their own population's genetic
54
55 aetiology to be optimally explored. A leading example of this is the Estonian Biobank, presented
56
57
58
59
60

1
2
3 by Andres Metspalu. Such resources will not only provide population-specific disease materials,
4
5 but also supply properly matched control samples for future GWAS investigations.
6
7

8 9 **Latest Technologies**

10
11 Many new technologies are being developed to increase the speed and reduce the cost of
12
13 studying genome variation. Han Cao described a novel method for the analysis of structural
14
15 variation at the single molecule level, based upon a nano-channel array. This technique uses
16
17 fluorescent labelling of specific motifs to “barcode” genomic DNA fragments, with these motifs
18
19 then being detected as they pass down a nano-tube. Andreas Gnirke described target selection
20
21 technologies in support of next-generation sequencing, wherein DNA regions such as sets of
22
23 exons, or even complete exomes, are selectively recovered by DNA hybridisation and
24
25 enrichment using RNA “baits”.
26
27
28
29

30
31 The direct application of advanced sequencing technologies was very much in evidence at the
32
33 meeting. Deep sequencing was used by Allen Roses to identify novel *TOMM40* gene variants,
34
35 adjacent to *APOE* genotypes but able to distinguish two evolutionary groups of *APOE3* strands
36
37 which associate with age of onset for Alzheimer disease. Previous age of onset curves were only
38
39 accurate for *APOE4* which is associated with “long” *TOMM40* fragments 98% of the time.
40
41 Martin Kerick also reported on 2nd generation sequencing studies, showing how complete
42
43 exome analysis in colon cancer patients suggests a multi-hit strategy for tumour development.
44
45 Complete resequencing of whole genomes is also starting to be evaluated as a disease research
46
47 tool, e.g., for the analysis of cancer tissue versus normal tissue comparisons as presented by
48
49 Elaine Mardis. Other sequencing work reported by Kelly Frazer used high-throughput
50
51 technologies to study the 9p21 interval which has been signalled by GWAS association data to
52
53 be involved in both Coronary Artery Disease and type 2 diabetes.
54
55
56
57
58
59
60

1
2
3 Second generation sequencing could also start to replace array-CGH and microarrays as a means
4 for structural variation analysis. Snehit Prabhu compared sequencing and arrays in this context,
5
6 concluding that limitations such as sequence fidelity and data storage costs may limit the degree
7
8 to which sequencing will soon completely replace array-based methods. One factor here will be
9
10 how effectively large-scale sequencing data can be analysed, and progress in this direction was
11
12 reported by Gabor Marth. He explained algorithms that could be applied to next-generation
13
14 sequencing data from hundreds of samples to extract genetic variation profiles that span from
15
16 single base pair differences right through to large structural variants.
17
18
19
20
21
22

23 **New Avenues for the Study of Genetic Variation**

24
25
26 Iuliana Ionita-Laza raised the question of how many new variants of various frequencies might
27
28 still be unknown at any stage of research progress, and described statistical methods designed to
29
30 help answer this question. Findings for rare variants revealed a clear distinction between the
31
32 Chinese and the Japanese populations. Michael Eberle reported on the 1000 genomes project,
33
34 which aims to directly discover rare SNPs in several populations, using low-depth and high-
35
36 depth sequencing of many DNAs.
37
38
39
40

41 Beyond SNPs, there is a need to study copy-number variants (CNVs), and how this class of
42
43 variation impacts phenotypes. But this goal is very challenging, not least because sequence
44
45 differences exist between CNV unit copies, with potential allelic variations at these and other
46
47 positions within each unit copy in the population. This complexity is termed multi-site variation
48
49 (MSV), and it is very difficult to experimentally decipher. Marco Groth, nevertheless, showed
50
51 that CNVs and MSVs may act together to result in an end phenotype, for example expression
52
53 levels of the defensin gene *DEFB4*. Furthermore, Jonathan Mudge reported that gene
54
55
56
57
58
59
60

1
2
3 transcription rates vary between individuals, and reported on efforts to determine why this is so,
4
5 and whether it contributes to phenotypes. Such examples emphasize the huge gaps remaining in
6
7 our understanding of genetic variation and the mechanisms of gene expression and phenotype
8
9 determination.
10
11

12 13 **Transmission and Inheritance**

14
15
16 Mendelian disease, wherein phenotypes are directly passed down across generations, is generally
17
18 taken to be the result of the inheritance of single pathogenic gene mutations. However, several
19
20 presentations showed that this is not always (or even usually) truly the case. Pablo Sandro
21
22 Carvalho Santos described transmission distortion in the human genome (unequal chance of
23
24 passing on two diploid alleles), presenting confirmed examples at two loci using SNP genotyping
25
26 of family trios. The underlying mechanism remains unknown. As well as being important in its
27
28 own right, this finding has implications for family-based disease association studies that assume
29
30 the absence of transmission distortion. Doron Behar discussed the advantages and disadvantages
31
32 of using uniploid and diploid genomes for tracing genetic variation.
33
34
35
36

37
38 Naisha Shah described how there may be a parental origin bias for de novo CNVs, with males
39
40 showing an increased frequency of deletions. Joris Vermeesch then further complicated the story
41
42 for CNV inheritance; emphasizing that phenotypic effects entail many forms of genetic
43
44 inheritance, including quantitative effects and the interplay of X-inactivation. He also raised the
45
46 question of what should constitute a CNV reference map, since individual CNV profiles vary so
47
48 dramatically. Jonathan Mudge, however, argued that the wealth of data describing large and
49
50 small scale variation does not mean that the idea of a single human reference sequence is
51
52 obsolete, but that this data can be used to improve the fidelity of the assembly – as is being done
53
54 in the HAVANA project.
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Extrapolating from the totality of work presented at HGV2009 and looking to the future, the next few years are certain to bring rapid progress in understanding complex disease aetiologies, revealing roles for both single base variants and structural variation alleles. This will be enabled by continued technology development and the support of new and powerful databasing and bioinformatics strategies. Clinical utility will be increasingly emphasized, with ethics and data sharing challenges being navigated. These dimensions, and many more, will be thoroughly explored at the HGV2010 meeting, being planned for September in the USA.

Box 1. HGV2009 Meeting Sessions

HGV2009 Meeting Sessions:

1. Disease (I)
2. Bioinformatics and Large Scale Studies
3. Population Genetics
4. Methods
5. Medicine and Pharmacogenetics
6. Structural Variation
7. Disease (II)

Peer Review