THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Section ON Statistical Genetics

Metabolomics meets Genomics

Hemant K. Tiwari, Ph.D.
Professor and Head

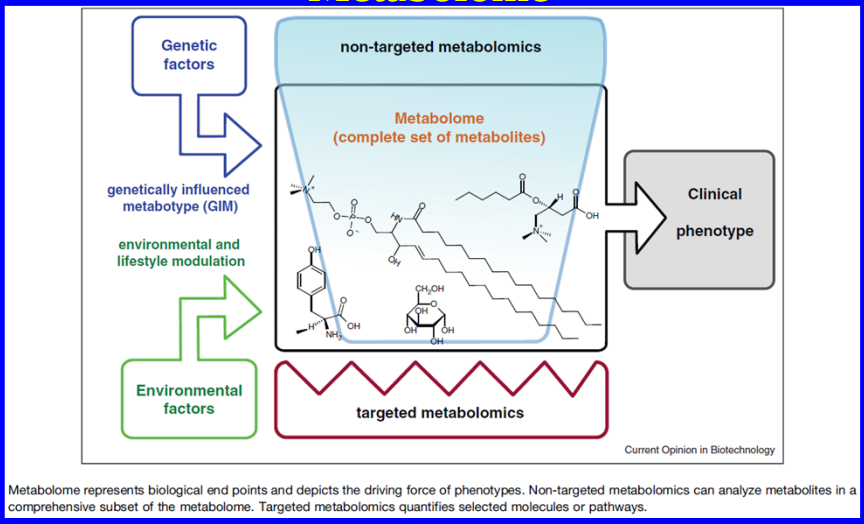
Section on Statistical Genetics
Department of Biostatistics
School of Public Health

Metabolomics: Bench to Bedside

Workflow	Considerations	Choices
Study design	Population	For example, Caucasian, Asian, African
	Study type	For example, population-based, twins study, clinical studies
	Sample type	For example, blood, urine, saliva
Sample collection	Standard operating protocols	Compatibility between study centres
	Fasting state	For example, fasting, non-fasting, controlled nutritional challenges
	Sample quantities	Serum, plasma, small volumes to avoid thawing
Sample storage	Temperature	−80°C, liquid nitrogen
	Aliquoting	200 µl for mass spectrometry, 1 ml for NMR, avoid thawing cycles
	Biobanking	Manual, automated
Sample preparation	Metabolite extraction	For example, polar, charged
	Derivatization	Changing biochemical properties for better measurement
Sample analysis	Method	^{1}H NMR, LC–MS/MS, GC–MS/MS
	Identification	Targeted, non-targeted, quantitative
	Provider	Proprietary, core facility, fee-for-service
Data analysis	Covariates	Age, gender, body mass index, medication, lifestyle
	Statistical analysis	For example, linear model, using ratios, advanced statistics
	Initial data processing	Log-normal scaling, principal-component transformation
Data interpretation	Functional	For example, GRAIL, overlay with eQTL data
	Biochemical	KEGG, HMDB
	Medical	GWAS catalogue, pharmacogenomics database

Suhre and Gieger (Nature Review Genetics, Vol 13, Nov 2012)

Linkage between Genome to Metabolome



Metabolome represents biological end points and depicts the driving force of phenotypes. Non-targeted metabolomics can analyze metabolites in a comprehensive subset of the metabolome. Targeted metabolomics quantifies selected molecules or pathways.

Figure 1 from Adamski and Suhre (2013). Current Opinions in Biotechnology

Several Statistical Approaches for Metabolomics

- Unsupervised (uses only metabolites)
  - Hierarchical clustering
  - Principal Component analysis
  - Kohonen neural network
- Supervised (Uses both the metabolites and traits)
  - Artificial neural networks
  - Discriminant analysis
  - Regression analysis
  - Regression trees
  - Inductive logic programming

Metabolites as intermediate phenotypes

- Metabolites represent intermediate phenotypes leading to clinical phenotypes. We want phenotype to be as “close” to molecular products as possible
- We have been using GWAS for intermediate phenotypes to detecting the genes for diseases or traits
- Examples: blood glucose levels, numerous hormones, cholesterol, triglyceride levels, lipids, etc.

Metabolites as intermediate phenotypes

- We already know many endogenous human metabolite pathways
- There are 2,200 enzyme coding genes annotated in the human genome
- The SNPs in the genes that are related to enzymatic or transport activities are prime candidates for harboring the causative variance

First Genome-wide association studies with metabolites (mQTL analysis)

Genetics Meets Metabolomics: A Genome-Wide Association Study of Metabolite Profiles in Human Serum

Christian Gieger[1,2], Ludwig Geistlinger[1], Elisabeth Altmaier[3,4], Martin Hrabé de Angelis[5,6], Florian Kronenberg[7], Thomas Meitinger[8,9], Hans-Werner Mewes[3,10], H.-Erich Wichmann[1,2], Klaus M. Weinberger[11], Jerzy Adamski[5,6], Thomas Illig[1], Karsten Suhre[3,4]*

Abstract

The rapidly evolving field of metabolomics aims at a comprehensive measurement of ideally all endogenous metabolites in a cell or body fluid. It thereby provides a functional readout of the physiological state of the human body. Genetic variants that associate with changes in the homeostasis of key lipids, carbohydrates, or amino acids are not only expected to display much larger effect sizes due to their direct involvement in metabolite conversion modification, but should also provide access to the biochemical context of such variations, in particular when enzyme coding genes are concerned. To test this hypothesis, we conducted what is, to the best of our knowledge, the first GWA study with metabolomics based on the quantitative measurement of 363 metabolites in serum of 284 male participants of the KORA study. We found associations of frequent single nucleotide polymorphisms (SNPs) with considerable differences in the metabolic homeostasis of the human body, explaining up to 12% of the observed variance. Using ratios of certain metabolite concentrations as a proxy for enzymatic activity, up to 28% of the variance can be explained (*p*-values 10^{-16} to 10^{-21}). We identified four genetic variants in genes coding for enzymes (FADS1, LIPC, SCAD, MCAD) where the corresponding metabolic phenotype (metabotype) clearly matches the biochemical pathways in which these enzymes are active. Our results suggest that common genetic polymorphisms induce major differentiations in the metabolic make-up of the human population. This may lead to a novel approach to personalized health care based on a combination of genotyping and metabolic characterization. These genetically determined metabotypes may subscribe the risk for a certain medical phenotype, the response to a given drug treatment, or the reaction to a nutritional intervention or environmental challenge.

PLoS Genet. 2008 Nov;4(11):e1000282. doi: 10.1371/journal.pgen.1000282. Epub 2008 Nov 28.

Some more examples

PLoS Genet. 2009 Jan;5(1):e1000338. doi: 10.1371/journal.pgen.1000338. Epub 2009 Jan 16.

Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study.

Tanaka T, Shen J, Abecasis GR, Kisialiou A, Ordovas JM, Guralnik JM, Singleton A, Bandinelli S, Cherubini A, Arnett D, Tsai MY, Ferrucci L.

Medstar Research Institute, Baltimore, MD, USA. tanakato@mail.nih.gov

PLoS Genet. 2009 Oct;5(10):e1000672. doi: 10.1371/journal.pgen.1000672. Epub 2009 Oct 2.

Genetic determinants of circulating sphingolipid concentrations in European populations.

Hicks AA, Pramstaller PP, Johansson A, Vitart V, Rudan I, Ugocsai P, Aulchenko Y, Franklin CS, Liebisch G, Erdmann J, Jonasson I, Zorkoltseva IV, Pattaro C, Hayward C, Isaacs A, Hengstenberg C, Campbell S, Gnewuch C, Janssens AC, Kirichenko AV, König IR, Marroni F, Polasek O, Demirkan A, Kolcic I, Schwienbacher C, Igl W, Biloglav Z, Witteman JC, Pichler I, Zaboli G, Axenovich TI, Peters A, Schreiber S, Wichmann HE, Schunkert H, Hastie N, Oostra BA, Wild SH, Meitinger T, Gyllensten U, van Duijn CM, Wilson JF, Wright A, Schmitz G, Campbell H.

Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy.

Nat Genet. 2010 Feb;42(2):137-41. doi: 10.1038/ng.507. Epub 2009 Dec 27.

A genome-wide perspective of genetic variation in human metabolism.

Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmüller G, Kato BS, Mewes HW, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann HE, Spector TD, Adamski J, Suhre K.

Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

Nature. 2011 Aug 31;477(7362):54-60. doi: 10.1038/nature10354.

Human metabolic individuality in biomedical and pharmaceutical research.

Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wägele B, Altmaier E; CARDIoGRAM, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmüller G, Köttgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes HW, Milburn MV, Prehn C, Raffler J, Ried JS, Römisch-Margl W, Samani NJ, Small KS, Wichmann HE, Zhai G, Illig T, Spector TD, Adamski J, Soranzo N, Gieger C.

Nat Genet. 2011 Jun;43(6):565-9. doi: 10.1038/ng.837. Epub 2011 May 15.

A genome-wide association study of metabolic traits in human urine.

Suhre K, Wallaschofski H, Raffler J, Friedrich N, Haring R, Michael K, Wasner C, Krebs A, Kronenberg F, Chang D, Meisinger C, Wichmann HE, Hoffmann W, Völzke H, Völker U, Teumer A, Biffar R, Kocher T, Felix SB, Illig T, Kroemer HK, Gieger C, Römisch-Margl W, Nauck M.

Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. karsten@suhre.fr

How do we relate metabolites to SNP data?

- Metabolite can be modeled as an outcome & SNPs then used as a predictor
- Type of Analysis: Whether to do univariate analysis, use ratio of metabolites, or use multivariate analysis?
- Selection of covariates: Which covariates to model? For example, some metabolic traits vary with BMI and fasting state, so should be included as covariates.

Overview of GWAS

- Well established Quality Control (QC) protocols
- Validated statistical methods exit
- Software programs are available to analyzed data, e.g. PLINK
- For QC see
  - Laurie, C. C. et al. (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. Genetic Epidemiology, 34: 591-602
  - Turner et al. (2011) Quality Control Procedures for Genome-Wide Association Studies. Current Protocols in Human Genetics. 68:1.19.1-1.19.18

Genome-wide Association Studies (GWAS)

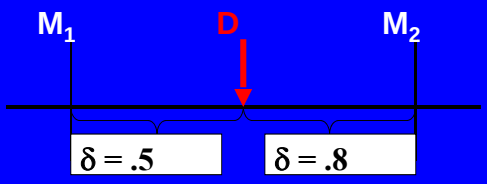
- To scan 1 to 2.5 M SNPs of many people to find genetic variations associated with a disease
- GWAS are particularly useful in finding genetic variant that contribute to common, complex diseases, such as asthma, cardiovascular diseases, cancer, diabetes, obesity, and mental disorders.

Source: http://www.genome.gov/20019523#1

http://www.genome.gov/26525384

Why GWAS will enable us to find disease genes?

- It utilizes linkage disequilibrium between SNPs and putative gene loci.



- The coverage of the genome by SNPs has to be excellent
- Availability of genome-wide SNPs chip

First Successful GWAS on Age-Related Macular degeneration
Science: March 10, 2005

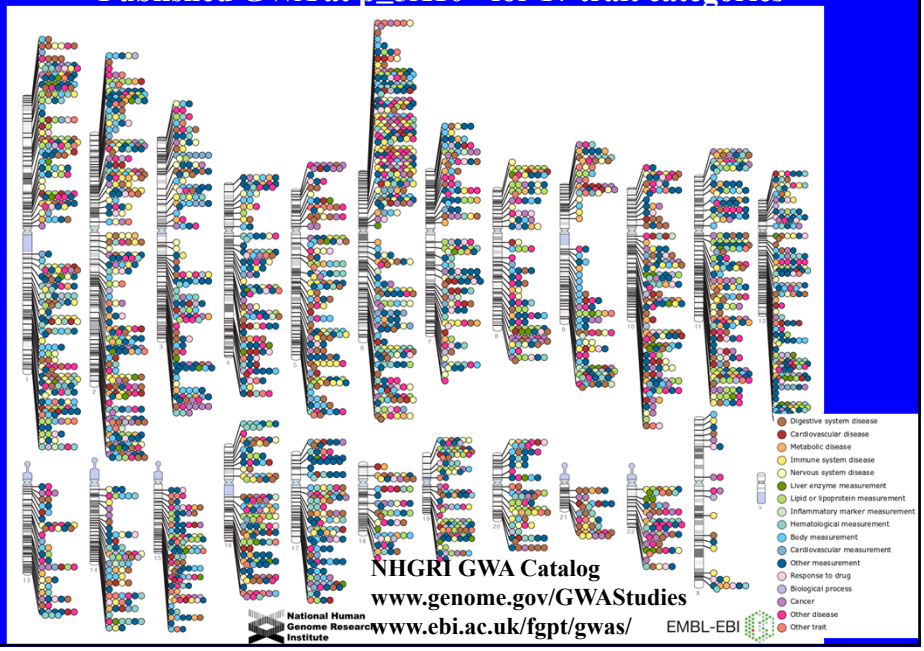
Complement Factor H Polymorphism in Age-Related Macular Degeneration

Robert J. Klein,[1] Caroline Zeiss,[2*] Emily Y. Chew,[3*] Jen-Yue Tsai,[4*] Richard S. Sackler,[1] Chad Haynes,[1] Alice K. Henning,[5] John Paul SanGiovanni,[3] Shrikant M. Mane,[6] Susan T. Mayne,[7] Michael B. Bracken,[7] Frederick L. Ferris,[3] Jurg Ott,[1] Colin Barnstable,[2] Josephine Hoh[7†]

Age-related macular degeneration (AMD) is a major cause of blindness in the elderly. We report a genome-wide screen of 96 cases and 50 controls for polymorphisms associated with AMD. Among 116,204 single-nucleotide polymorphisms genotyped, an intronic and common variant in the complement factor H gene (*CFH*) is strongly associated with AMD (nominal *P* value $<10^{-7}$). In individuals homozygous for the risk allele, the likelihood of AMD is increased by a factor of 7.4 (95% confidence interval 2.9 to 19). Resequencing revealed a polymorphism in linkage disequilibrium with the risk allele representing a tyrosine-histidine change at amino acid 402. This polymorphism is in a region of CFH that binds heparin and C-reactive protein. The *CFH* gene is located on chromosome 1 in a region repeatedly linked to AMD in family-based studies.

Using 96 cases and 50 controls Klein et al. (2005) found CFH gene on chromosome 1 ($p=4x10^{-8}$, OR=4.60) using 100K affy chip

Published Genome-Wide Associations through 12/2012
Published GWA at $p \leq 5X10^{-8}$ for 17 trait categories



NHGRI GWA Catalog
www.genome.gov/GWAStudies
www.ebi.ac.uk/fgpt/gwas/

National Human Genome Research Institute

EMBL-EBI

What steps needed for GWAS

- Use appropriate design
  - Pedigrees, case-control, unrelated individuals
- Determine the sample size
  - Power
- Choose SNP genotyping platform
  - Affy, Illumina, Perlegen
- Perform QC (HWE, Mendelian errors, outliers, etc.)
- Imputation
- Choose appropriate Association test

Quality Control (QC)

- The first step of GWAS analysis is the quality control of the genotypic and phenotypic data. There are number of procedures needed to ensure the quality of genotype data both at the genotyping laboratory and after calling genotypes using statistical approaches.
- The QC and association analysis of GWAS data can be performed using the robust, freely available, and open source software PLINK developed by Purcell *et al.* (2007)

Quality Control (QC)

- **Sex Inconsistency:** It is possible that self-reported sex of the individual is incorrect. Sex inconsistency can be checked by comparing the reported sex of each individual with predicted sex by using X-chromosome markers' heterozygosity to determine sex of the individual empirically.
- **Relatedness and Mendelian Errors:** Another kind of error that can occur in genotyping is due to sample mix-up, cryptic relatedness, duplications, and pedigree errors such as self-reported relationships that are not accurate. The relationship errors can be corrected by consulting with the self-reported relationships and/or using inferred genetic relationships.

Quality Control (QC)

- **Batch Effects:** For GWAS, samples are processed together for genotyping in a batch. The size and composition of the sample batch depends on the type of the commercial array, for example, an Affymetrix array can genotype up to 96 samples, and an Illumina array can genotype up to 24 samples. To minimize batch effects, samples should be randomly assigned plates with different phenotypes, sex, race, and ethnicity.
- The most commonly used method is to compare the average minor allele frequencies and average genotyping call rates across all SNPs for each plate. Most genotyping laboratories perform batch effect detection and usually re-genotype the data if there is a batch effect or a plate discarded when there is a large amount of missing data.

Quality Control (QC)

- **Marker and sample genotyping efficiency or call rate:** Marker genotyping efficiency is defined as the proportion of samples with a genotype call for each marker. If large numbers of samples are not called for a particular marker, that is an indication of a poor assay, and the marker should be removed from further analysis. A threshold for removing markers varies from study to study depending on the sample size of the study. However, usual recommended call rates are approximately 98% to 99%.

Quality Control (QC)

- **Population stratification:** There are a number of methods proposed to correct for population substructure. Three commonly used methods to correct for the underlying variation in allele frequencies that induces confounding due to population stratification:
  - genomic control
  - structured association testing
  - principal components (Most Commonly Used Method)

Population Stratification

- Population stratification: Sample consists of divergent populations
- Case-control studies can be affected by population stratification

Quality Control (QC)

- Principal components analysis (PCA) uses thousands of markers to detect population stratification and Principal Components (PCs) then can be used to correct for stratification by modeling PCs as covariates in the model
- PCs can be calculated using a program Eigenstrat (Patterson et al., 2006; Price et al., 2006). There are two issues with using PCA, (1) how many SNPs to use, and (2) how many PCs should be included as covariates in the association analysis.

Quality Control (QC)

- **Hardy-Weinberg equilibrium (HWE) filter:** The HWE test compares the observed genotypic proportion at the marker versus the expected proportion. Deviation from HWE at a marker locus can be due to population stratification, inbreeding, selection, non-random mating, genotyping error, actual association to the disease or trait under study, or a deletion or duplication polymorphism. However, HWE is typically used to detect genotyping errors. SNPs that do not meet HWE at a certain threshold of significance are usually excluded from further association analysis.

Statistical Methods & Software for Genetic Association Studies

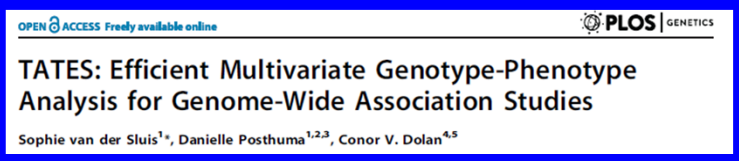
	Approach	Reference	Software	URL
Logistic regression	Model log odds of disease as linear function of underlying genotype variables	20, 74, 20	Standard statistical package (eg, Stata, SAS, S-Plus, R)	http://www.stata.com/ http://www.sas.com/ http://www.insightful.com/products/splus/ http://www.r-project.org/
χ^2 test of association	Test for independence of disease status and genetic risk factor	20	Standard statistical package	See above
Linear regression	Model quantitative trait as linear function of underlying genotype variables	75	Standard statistical package	See above
Survival analysis	Model survivor function or hazard as function of underlying genotype variables	20, 52	Standard statistical package	See above
Transmission/ disequilibrium test	Test departure of transmission of alleles from heterozygous parents to affected offspring from null hypothesis of half	71, 76–78	Various (eg, Genehunter, RC-TDT, Genassoc, Transmit, Unphased	http://fhcrc.org/labs/kruglyak/Downloads/index.html http://www.uni-bonn.de/~umt70e/soft.htm http://www-gene.cimr.cam.ac.uk/clayton/software/ http://www.mrc-bsu.cam.ac.uk/personal/frank/
Conditional logistic regression	Calculate conditional probability of affected offspring genotypes, given parental genotypes	54, 60, 79, 80	Genassoc Unphased	http://www-gene.cimr.cam.ac.uk/clayton/software/ http://www.mrc-bsu.cam.ac.uk/personal/frank/
Log linear models	Model counts of genotype combinations for mother, father, and affected offspring	57, 58, 59	Standard statistical package	See above
Pedigree disequilibrium test	Test departure of transmission of alleles to affected pedigree members from null expectation	81, 82	Pedigree disequilibrium test Unphased	http://www.chg.duke.edu/software/pdt.html http://www.mrc-bsu.cam.ac.uk/personal/frank/
Family-base association test	Tests for association or linkage between disease phenotypes and haplotypes by utilising family-based controls	83–86	Family-based association test	http://www.biostat.harvard.edu/~fbat/fbat.htm
Quantitative transmission/ disequilibrium test	Linkage disequilibrium analysis of quantitative and qualitative traits based on variance components	87, 88	Quantitative transmission/ disequilibrium test	http://www.sph.umich.edu/csg/abecasis/QTDT/
DNA pooling	Test for differences in allele frequencies in different pooled samples while estimating components of variance due to experimental error	61, 89–91	Standard statistical package	See above

The references are those from the following paper:
HJ Cordell, DG Clayton. Genetic association studies. Lancet 2005; 366: 1121-31

Commonly Used Software

- FBAT
  - Family based association analysis
- PLINK
  - Whole genome association analysis toolset
- SAGE (ASSOC)
  - Statistical Analysis for Genetic Epidemiology
- LMEKIN in R
  - Mixed-model procedure to analyze familial data
- STRUCTURE
  - Population structure inference
- EIGENSTRAT
  - Detects and corrects for population stratification in genome-wide association studies

Some new methods to analyze multivariate metabolomic data in GWAS framework


OPEN ACCESS Freely available online

PLOS | GENETICS

TATES: Efficient Multivariate Genotype-Phenotype Analysis for Genome-Wide Association Studies

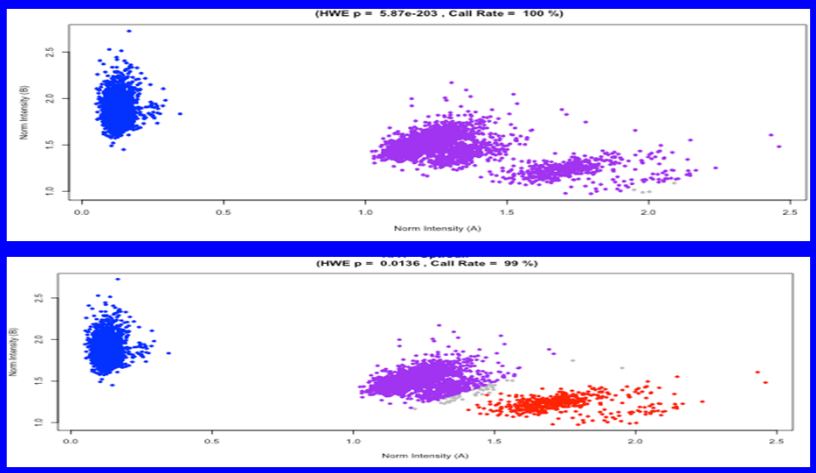
Sophie van der Sluis[1]*, Danielle Posthuma[1,2,3], Conor V. Dolan[4,5]


Genetic Epidemiology 36:244–252 (2012)

PSEA: Phenotype Set Enrichment Analysis—A New Method for Analysis of Multiple Phenotypes

Janina S. Ried,[1] Angela Döring,[2,3] Konrad Oexle,[4] Christa Meisinger,[3] Juliane Winkelmann,[4,5,6] Norman Klopp,[7,8] Thomas Meitinger,[4,6] Annette Peters,[3] Karsten Suhre,[9,10,11] H.-Erich Wichmann,[2,12,13] and Christian Gieger[14]*

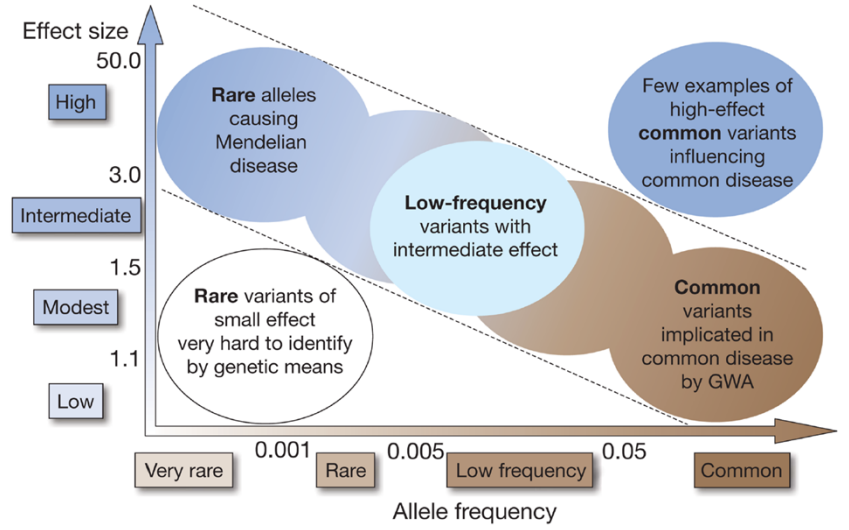
After Association Analysis QC (Cluster Plots)



Why do we stop at SNPs?

- EXOME data
- Gene Expression data
- Methylation data

Manolio et al. Finding the missing heritability of complex diseases. Nature. 2009: 747-753



Exome Data

- GWAS is good for common variants (Allele frequency ≥ 0.05)
- Exome chip or exome sequencing provides data on coding variants contains lots of rare variants (<0.05)
- Exome = Protein Coding Genome

Some Exome data analysis methods

- Cohort allelelic sum test (CAST): collapses over the rare variants and then compares the total rare variant frequency between cases and controls (Morgenther et al., 2010)
- Combined multivariate and collapsing (CMC): collapsing is done within different subgroups defined by allele frequencies and combined using a multivariate distance-based statistic (Li and Leal, 2008)
- Madsen and Browning (2009) proposed a method includes variants of any frequency, but the variants are weighted according to their frequencies
- Price et al. (2010) proposed a variable threshold approach and showed that this method can be more powerful compare to fixed threshold.

Some more Exome data analysis methods

- Hoffmann et al. (2010) method models weights, incorporates directionality (deleterious or protective) and threshold
- Wu et al. (2011) proposed the sequence kernel association test (SKAT), a supervised, flexible, computationally efficient regression method to test for association between genetic variants (common and rare) in a region and a continuous or dichotomous trait while easily adjusting for covariates.
- There are several other methods such as Lin et al. (2011), Zhu et al. (2010) (for both unrelated and family data), Ionita-Laza et al. (2011), Neale et al. (2011), etc.

How about integrating all omics data?

- Genome (G)
- Epigenome (E)
- Transcriptome (T)
- Proteome (P)
- Metabolome (M)
- Phenome (F)
- There are others lipidome, glycome, …

Example: Integrated analysis of phenotype with at least two other sources of data

An integrative genomics approach to infer causal associations between gene expression and disease

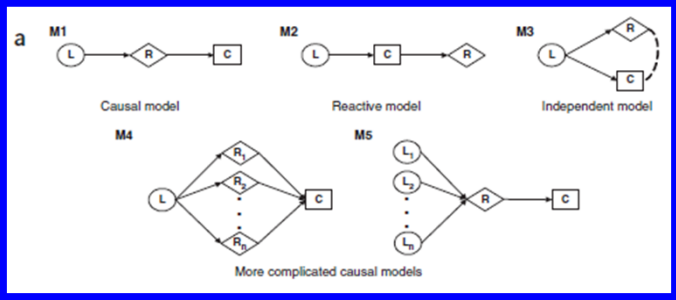
Eric E Schadt[1], John Lamb[1], Xia Yang[2], Jun Zhu[1], Steve Edwards[1], Debraj GuhaThakurta[1], Solveig K Sieberts[1], Stephanie Monks[3], Marc Reitman[4], Chunsheng Zhang[1], Pek Yee Lum[1], Amy Leonardson[1], Rolf Thieringer[5], Joseph M Metzger[6], Liming Yang[6], John Castle[1], Haoyuan Zhu[1], Shera F Kash[7], Thomas A Drake[8], Alan Sachs[1] & Aldons J Lusis[2]

A key goal of biomedical research is to elucidate the complex network of gene interactions underlying complex traits such as common human diseases. Here we detail a multistep procedure for identifying potential key drivers of complex traits that integrates DNA-variation and gene-expression data with other complex trait data in segregating mouse populations. Ordering gene expression traits relative to one another and relative to other complex traits is achieved by systematically testing whether variations in DNA that lead to variations in relative transcript abundances statistically support an independent, causative or reactive function relative to the complex traits under consideration. We show that this approach can predict transcriptional responses to single gene–perturbation experiments using gene-expression data in the context of a segregating mouse population. We also demonstrate the utility of this approach by identifying and experimentally validating the involvement of three new genes in susceptibility to obesity.

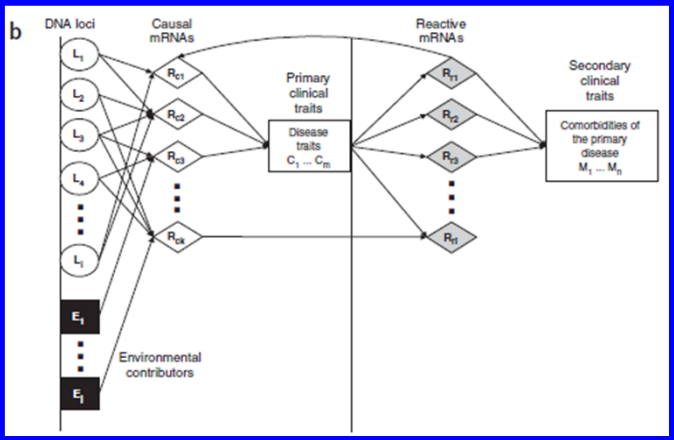
Schadt et al. (2005): Relationship among QTL, RNA levels (gene expression) and Complex traits

Define 5 models where L= QTL, R=gene expression, and C = complex trait, e.g. obesity

M1: Causal Model, M2: Reactive Model, M3: Independent model, M4: Causal model with many RNAs, and M5 Independent model with one RNA expression



Hypothetical gene network for disease traits and related comorbidities (Schadt et al., 2005)



Method used in Schadt et al. (2005)

- **Likelihood-based causality model selection (LCMS) test** : Uses conditional correlations to determine which relationship among traits is best supported by the data.
- Likelihoods associated with each of the models are constructed and maximized with respect to the model parameters, and the model with the smallest Akaike Information Criterion (AIC) value is identified as the model best supported by the data.
- If two gene-expression traits are each driven by a strong cis-acting eQTL, and these eQTLs are closely linked, they will induce a correlation structure between the two traits.

A multistep procedure to identify causal genes for obesity in mice (Schadt et al., 2005)

- Used the LCMS procedure to the omental fat pad mass (OFPM) and liver gene-expression data in the mice data. First, Identified most significant expression traits for OFPM
- Step 1: Build a genetic model for the omental fat pad mass (OFPM) trait, identifying the underlying QTLs that reflect the initial perturbations that give rise to the genetic components of the trait.
- Step 2: For each overlapping expression-OFPM QTL in the set of genes, they fit the corresponding QTL genotypes, gene-expression data and OFPM data to the independent, causal and reactive likelihood models.
- Step 3: Rank-ordered the genes according to the percentage of genetic variance in the OFPM trait that was causally explained by variation in their transcript abundances

Schadt et al. (2005)

- 90 genes tested as causal for OFPM traits at one or more QTLs
- Of these genes, Hsd11b1 was one of the best candidates. Causal model fitted the best.
- *C3ar1* and *Tgfbr2* were new susceptibiltity genes causal for obesity
- These results indicate that integrating genotypic and expression data may help the search for new targets for common human diseases

Example of Integration of SNPs, methylation, gene expression

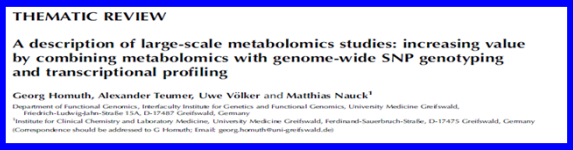
Bell *et al. Genome Biology* 2011, **12**:R10
http://genomebiology.com/2011/12/1/R10

RESEARCH Open Access

DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines

Jordana T Bell[1,3*], Athma A Pai[1], Joseph K Pickrell[1], Daniel J Gaffney[1,2], Roger Pique-Regi[1], Jacob F Degner[1], Yoav Gilad[1*], Jonathan K Pritchard[1,2*]

Nice Review paper on Integration of Genome, Transcriptome, and Metabolome



THEMATIC REVIEW

A description of large-scale metabolomics studies: increasing value by combining metabolomics with genome-wide SNP genotyping and transcriptional profiling

Georg Homuth, Alexander Teumer, Uwe Völker and Matthias Nauck[1]

Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Friedrich-Ludwigjahn-Straße 15A, D-17487 Greifswald, Germany
[1]Institute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Ferdinand-Sauerbruch-Straße, D-17475 Greifswald, Germany
(Correspondence should be addressed to G Homuth; Email: georg.homuth@uni-greifswald.de)

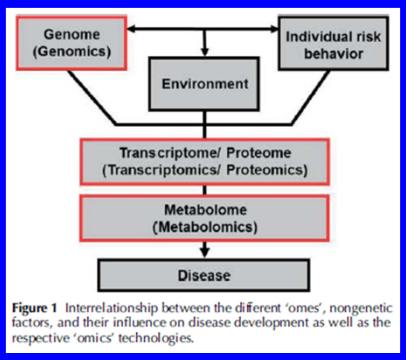
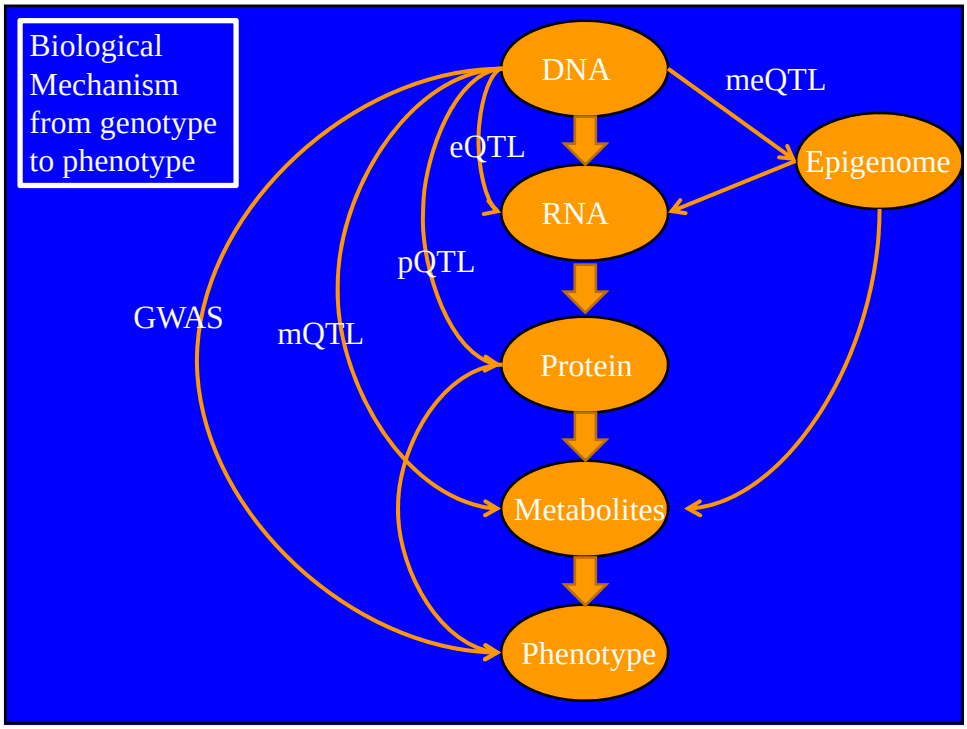


Figure 1 Interrelationship between the different 'omes', nongenetic factors, and their influence on disease development as well as the respective 'omics' technologies.



Challenges

- Database integration is a holy grail of systems biology
  - Genomic data base (dbGap, NCBI)
  - Transcriptome data base (GEO)
  - Metabolomics data base (HMDP, METLIN, KEGG)
- Not all databases can be easily integrated to visualize the results

Future: Integration of “omics” to solve the puzzle to understand genetic variation in human ?