Lecture 8: Genetic variants and their interpretation

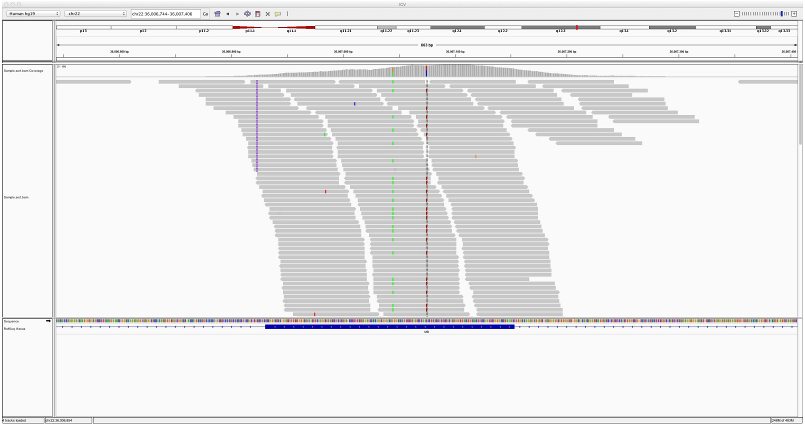
Paolo Provero

NGS and genetic variants

NGS techniques allow us to detect the *genetic variants* that characterize each individual

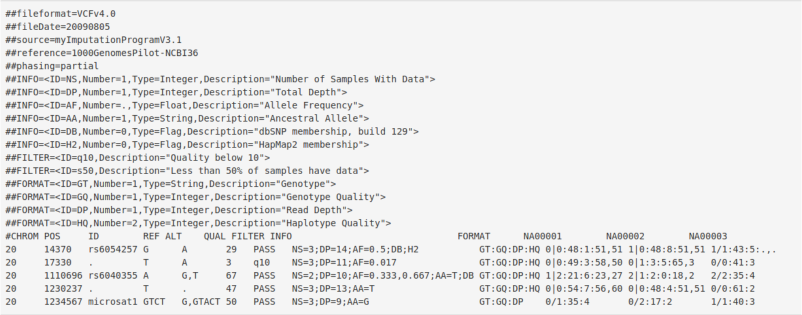
- *Targeted sequencing*: detect variants in selected regions
  - Whole Exome Sequencing
  - Sequencing of candidate genes or exons
  - The smaller the target the greater the sequencing *depth* with the same number of reads
- *Whole genome sequencing*

Variant detection



VCF format

To save space, the variants are expressed as *differences* from the reference sequence



```
##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF ALT    QUAL FILTER INFO                              FORMAT      NA00001        NA00002        NA00003
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3   0/0:41:3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2   2/2:35:4
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20     1234567 microsat1 GTCT   G,GTACT 50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2       1/1:40:3
```

Variant interpretation

Problem: determine which variants are causal of a phenotype/disease

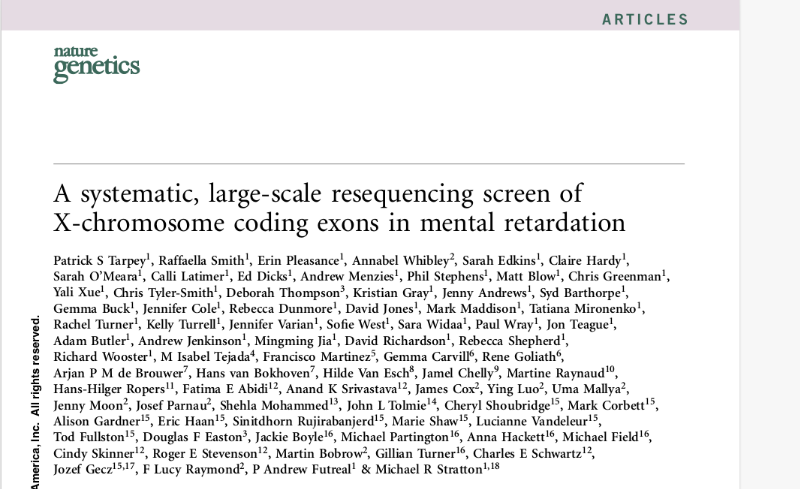
- *Mendelian* phenotypes
  - *rare* variants
  - high *penetrance*
  - in coding regions
- *Complex* phenotypes
  - *common* variants
  - non-coding regions

Effect on the phenotype

The likelihood that a variant will affect the phenotype depends on where the variant is located (coding exon, UTR, intron, intergenic)

- Coding variants can be classified based on their effect on the protein sequence
    - synonymous: no change in protein sequence (possible due to the degenerate nature of the genetic code)
    - missense: changes one aminoacid into another
        - conservative: new aa is chemically similar to old one
        - non-conservative: new aa is chemically different – likely change in structure
    - nonsense: introduces a stop codon
- Non-coding variants are assumed to affect phenotype by changing gene regulation
    - e.g. by creating/destroying a transcription factor binding site
    - much more difficult to classify/interpret

Coding variants: the XLMR example



ARTICLES

nature genetics

A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation

Patrick S Tarpey[1], Raffaella Smith[1], Erin Pleasance[1], Annabel Whibley[2], Sarah Edkins[1], Claire Hardy[1], Sarah O'Meara[1], Calli Latimer[1], Ed Dicks[1], Andrew Menzies[1], Phil Stephens[1], Matt Blow[1], Chris Greenman[1], Yali Xue[1], Chris Tyler-Smith[1], Deborah Thompson[3], Kristian Gray[1], Jenny Andrews[1], Syd Barthorpe[1], Gemma Buck[1], Jennifer Cole[1], Rebecca Dunmore[1], David Jones[1], Mark Maddison[1], Tatiana Mironenko[1], Rachel Turner[1], Kelly Turrell[1], Jennifer Varian[1], Sofie West[1], Sara Widaa[1], Paul Wray[1], Jon Teague[1], Adam Butler[1], Andrew Jenkinson[1], Mingming Jia[1], David Richardson[1], Rebecca Shepherd[1], Richard Wooster[1], M Isabel Tejada[4], Francisco Martinez[5], Gemma Carvill[6], Rene Goliath[6], Arjan P M de Brouwer[7], Hans van Bokhoven[7], Hilde Van Esch[8], Jamel Chelly[9], Martine Raynaud[10], Hans-Hilger Ropers[11], Fatima E Abidi[12], Anand K Srivastava[12], James Cox[2], Ying Luo[2], Uma Mallya[2], Jenny Moon[2], Josef Parnau[2], Shehla Mohammed[13], John L Tolmie[14], Cheryl Shoubridge[15], Mark Corbett[15], Alison Gardner[15], Eric Haan[15], Sinitdhorn Rujirabanjerd[15], Marie Shaw[15], Lucianne Vandeleur[15], Tod Fullston[15], Douglas F Easton[3], Jackie Boyle[16], Michael Partington[16], Anna Hackett[16], Michael Field[16], Cindy Skinner[12], Roger E Stevenson[12], Martin Bobrow[2], Gillian Turner[16], Charles E Schwartz[12], Jozef Gecz[15,17], F Lucy Raymond[2], P Andrew Futreal[1] & Michael R Stratton[1,18]

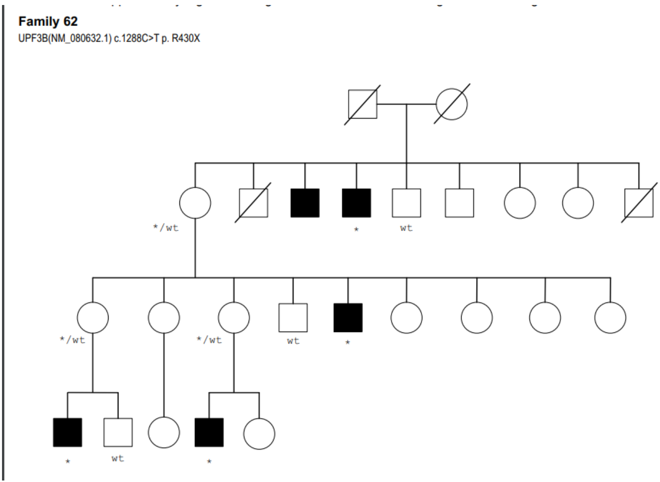
Nature Genetics 2009

Coding variants: the XLMR example

- 208 families with MR and pattern of transmission compatible with X linkage
- Exome sequencing of $\sim$700 genes in X chromosome
- Most differences in coding sequence between individuals are recurring and found in dbSNP
- Consider *truncating* variants only (found in 30 genes)

Segregation

A variant can be *causal* of the disease only if it *segregates* - presence of the variant in all affected members and only in affected members



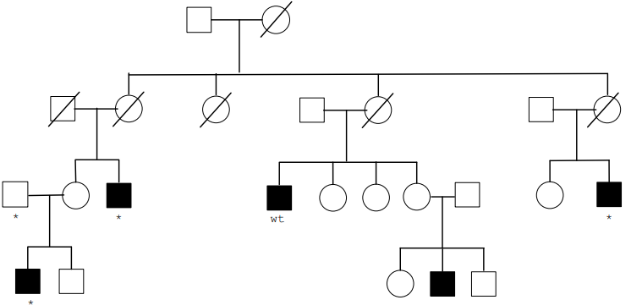
Family 62

UPF3B(NM_080632.1) c.1288C>T p. R430X

Non-segregation

Family 74

ITIH5L (NM_198510.1) IVS7+1ins T



Message

- Exome sequencing can help finding causal variants of Mendelian diseases
- However even truncating variants can be compatible with normal phenotype
  - "loss of function of 1% of the genes in the X chromosome is compatible with apparently normal existence"
- Another recent study found an average of ∼40 homozygous LOF mutations in normal individuals
- Therefore when sequencing the exome of a *proband* with a genetic disease we expect to find many candidate mutations
  - Need for *variant prioritization*

Variant prioritization

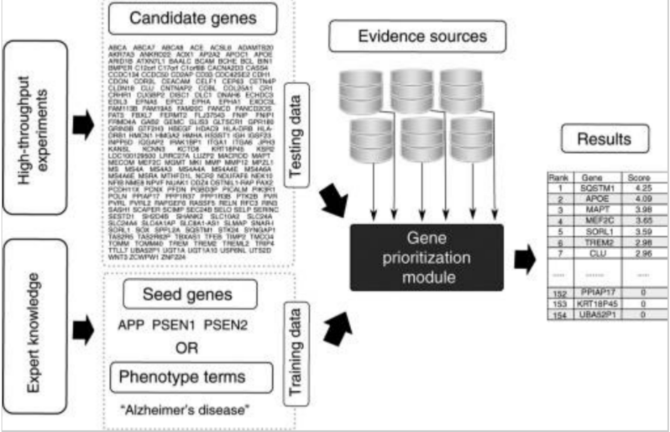
Problem:

- Among all the variants found in a proband with a genetic disease, find the ones most likely to be causative

The following criteria increase the probability that a variant is causative

- very low frequency in the population
- predicted strong effect on the protein
- gene involved in biological process relevant to the disease
- gene interacts with genes involved in the same/similar diseases

Variant prioritization


High-throughput experiments
Candidate genes
ABCA ABCA7 ABCA8 ACE ACSL6 ADAMTS20
AKR7A3 ANKRD22 AOX1 AP2A2 APOC1 APOE
ARID1B ATXN7L1 BAALC BCAM BCHE BCL BIN1
BMPER C12orf C17orf C1orf68 CACNA2D3 CASS4
CCDC134 CCDC50 CD2AP CD33 CDC42SE2 CDH1
CDON CDR2L CEACAM CELF1 CEP63 CETN4P
CLDN18 CLU CNTNAP2 COBL COL25A1 CR1
CRHR1 CUGBP2 DISC1 DLC1 DNAH6 ECHDC3
EDIL3 EPHA5 EPC2 EPHA EPHA1 EXOC3L
FAM113B FAM19A5 FAM20C FANCD FANCD2OS
FAT3 FBXL7 FERMT2 FLJ37543 FNIP FNIP1
FRMD4A GAB2 GEMC GLIS3 GLTSCR1 GPR180
GRIN3B GTF2H3 HBEGF HDAC9 HLA-DRB HLA-
DRB1 HMCN1 HMGA2 HMHA HS3ST1 IGH IGSF23
INPP5D IQGAP2 IRAK1BP1 ITGA1 ITGA6 JPH3
KANSL KCNN3 KCTD8 KRT18P45 KSR2
LOC100129500 LRRC37A LUZP2 MACROD MAPT
MECOM MEF2C MGMT MKI MMP MMP12 MPZL1
MS MS4A MS4A3 MS4A4A MS4A4E MS4A6A
MS4A6E MSRA MTHFD1L NCR2 NDUFAF6 NEK10
NFIB NME8 NPVF NUAK1 CD24 DSTNL1-RAP PAX2
PCDH11X PCNX PFDN PGBD3P PICALM PIK3R1
POLN PPIAP17 PPP1R37 PPP1R3B PTK2B PVR
PVRL PVRL2 RAPGEF6 RASSF5 RELN RFC3 RIN3
SASH1 SCAPER SCIMP SEC24B SELO SELP SERINC
SESTD1 SH2D4B SHANK2 SLC10A2 SLC24A
SLC24A4 SLC4A1AP SLC8A1-AS1 SLMAP SNAR-I
SORL1 SOX SPPL2A SQSTM1 STK24 SYNGAP1
TAS2R5 TAS2R62P TBXAS1 TFEB TIMP2 TMCO4
TOMM TOMM40 TREM TREM2 TREML2 TRIP4
TTLL7 UBA52P1 UGT1A UGT1A10 USP6NL UTS2D
WNT3 ZCWPW1 ZNF224
Testing data
Expert knowledge
Seed genes
APP PSEN1 PSEN2
OR
Phenotype terms
"Alzheimer's disease"
Training data
Evidence sources
Gene prioritization module
Results

Rank	Gene	Score
1	SQSTM1	4.25
2	APOE	4.09
3	MAPT	3.98
4	MEF2C	3.65
5	SORL1	3.59
6	TREM2	2.98
7	CLU	2.96
----	------	----
152	PPIAP17	0
153	KRT18P45	0
154	UBA52P1	0

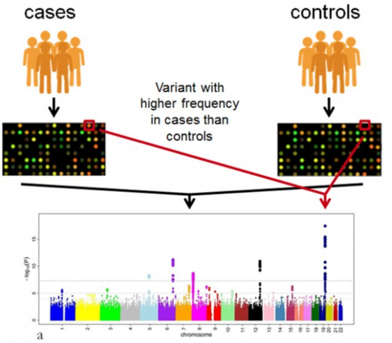
From Zolotareva & Kleine, J Integr Bioinform 16:20180069 (2019).

Complex phenotypes

- Complex trait: determined by both genetic and non-genetic (e.g. environmental, behavioral, . . .) factors
    - Diabetes
    - Rheumatoid arthritis
    - Crohn disease
    - Height
    - Blood pressure
    - Lymphocyte count
- Genetic determinants are usually
    - Many variants of small effect
    - Non-coding (hence presumably regulatory)

GWAS

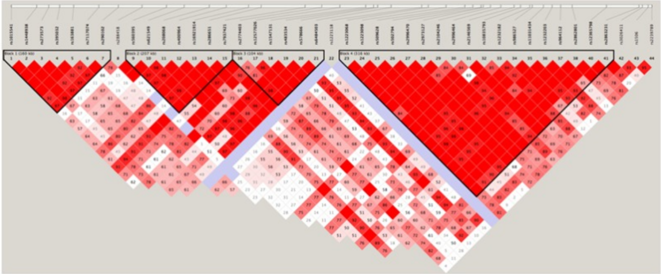
Genome-wide associations studies look for variants with significantly different frequency in cases and controls



Predicting vs understanding complex traits

- GWAS hits can be used to build *polygenic scores* predicting disease risk
- To understand disease (hence possibly find cures) one needs to understand the mechanism leading from variant to disease
- GWAS hits are mostly non-coding → regulatory effect
    - What is the *target* of a regulatory variant (gene whose expression is altered by the variant)?
    - What is the effect of the variant?
        - Regulatory code much less understood than genetic code
    - Which one is the causal variant?
        - *Linkage disequilibrium* implies that many neighboring variants are inherited together, thus showing the same correlation with the disease although presumably only one or a few are causative

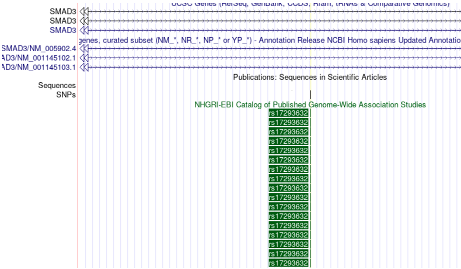
Linkage disequilibrium



Strong coorelation between nearby SNPs indicate that they are inherited together (no recombination events)

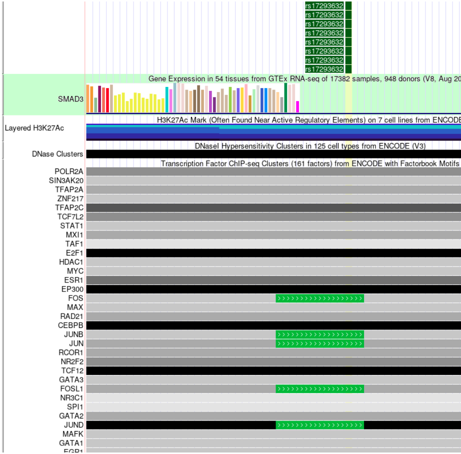
An example: rs17293632

- Associated by GWAS to many complex traits, including Crohn's disease
- Located in an intron of SMAD3

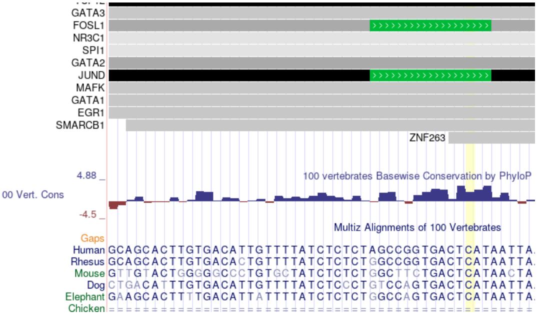

UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
SMAD3
SMAD3
SMAD3
genes, curated subset (NM_*, NR_*, NP_* or YP_*) - Annotation Release NCBI Homo sapiens Updated Annotatio
SMAD3/NM_005902.4
AD3/NM_001145102.1
AD3/NM_001145103.1
Publications: Sequences in Scientific Articles
Sequences
SNPs
NHGRI-EBI Catalog of Published Genome-Wide Association Studies
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632

An example: rs17293632

- Colocalizes with several transcription factor binding peaks from ChIP-seq
- In particular, is within the core binding motif of AP-1

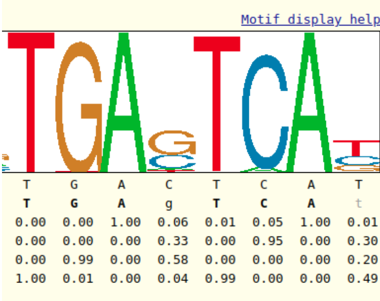

rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
rs17293632
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 20
SMAD3
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K27Ac
DNaseI Hypersensitivity Clusters in 125 cell types from ENCODE (V3)
DNase Clusters
Transcription Factor ChIP-seq Clusters (161 factors) from ENCODE with Factorbook Motifs
POLR2A
SIN3AK20
TFAP2A
ZNF217
TFAP2C
TCF7L2
STAT1
MXI1
TAF1
E2F1
HDAC1
MYC
ESR1
EP300
FOS
MAX
RAD21
CEBPB
JUNB
JUN
RCOR1
NR2F2
TCF12
GATA3
FOSL1
NR3C1
SPI1
GATA2
JUND
MAFK
GATA1
EGR1

An example: rs17293632


GATA3
FOSL1
NR3C1
SPI1
GATA2
JUND
MAFK
GATA1
EGR1
SMARCB1
ZNF263
4.88
100 vertebrates Basewise Conservation by PhyloP
00 Vert. Cons
-4.5
Multiz Alignments of 100 Vertebrates
Gaps

```
Human    GCAGCACTTGTGACATTGTTTTATCTCTCTAGCCGGTGACTCATAATTA
Rhesus   GCAGCACTTGTGACACTGTTTTATCTCTCTGGCCGGTGACTCATAATTA
Mouse    GTTGTACTGGGGCCCTGTGCTATCTCTCTGGCTTCTGACTCATAACTA
Dog      CTGACATTTGTGACATTGTTTTATCTCCCTGTCCAGTGACTCATAATTA
Elephant GAAGCACTTTTGACATTATTTTATCTCTCTGGCCAGTGACTCATAATTA
Chicken  ================================================
```

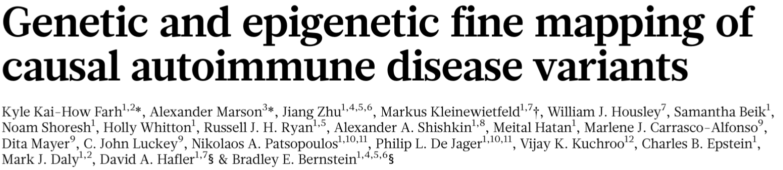
The variant disrupts a highly conserved AP-1 motif



T	G	A	C	T	C	A	T
T	**G**	**A**	g	**T**	**C**	**A**	t
0.00	0.00	1.00	0.06	0.01	0.05	1.00	0.01
0.00	0.00	0.00	0.33	0.00	0.95	0.00	0.30
0.00	0.99	0.00	0.58	0.00	0.00	0.00	0.20
1.00	0.01	0.00	0.04	0.99	0.00	0.00	0.49

Interpreting rs17293632

- The variant rs17293632 disrupts a highly conserved (hence, probably functional) AP-1 binding site
- Thus altering gene expression, presumably of SMAD3
- Thus altering the phenotype, e.g. by conferring susceptibility to Crohn's disease



Genetic and epigenetic fine mapping of causal autoimmune disease variants

Kyle Kai-How Farh[1,2]*, Alexander Marson[3]*, Jiang Zhu[1,4,5,6], Markus Kleinewietfeld[1,7]†, William J. Housley[7], Samantha Beik[1], Noam Shoresh[1], Holly Whitton[1], Russell J. H. Ryan[1,5], Alexander A. Shishkin[1,8], Meital Hatan[1], Marlene J. Carrasco-Alfonso[9], Dita Mayer[9], C. John Luckey[9], Nikolaos A. Patsopoulos[1,10,11], Philip L. De Jager[1,10,11], Vijay K. Kuchroo[12], Charles B. Epstein[1], Mark J. Daly[1,2], David A. Hafler[1,7]§ & Bradley E. Bernstein[1,4,5,6]§

eQTLs

To understand GWAS hits we need a systematic analysis of the effect of variants on gene expression.

An eQTL (expression quantitative trait locus) is a variant significantly associated with the expression of a gene
- gene expression considered as a *quantitative trait*, like height or lymphocyte count

An eQTL study needs a large cohort (hundreds of individuals) for which we have

- expression data
- genotyping (i.e. *dosage* [0,1, or 2] of each genetic variant)

The analysis is performed by *linear regression*

Regression

- *Regression* is the main statistical method for the study of the *dependence* among two (or more) variables x and y
- Assumes the existence of a *regression function* $f(x)$ which represents the true dependence of y on x as

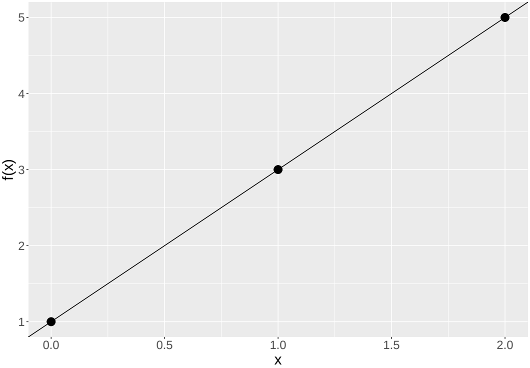
$$y = f(x) + \epsilon$$

- In our case:
  - x: dosage of a variant ($x \in \{0, 1, 2\}$)
  - y: expression of a gene
- The *error term* ϵ represents random fluctuations, or the effects of variables that we do not consider
  - e.g. environmetal effects on the expression y of the gene

Linear regression

In *linear regression* we assume $f(x)$ to be a straight line:

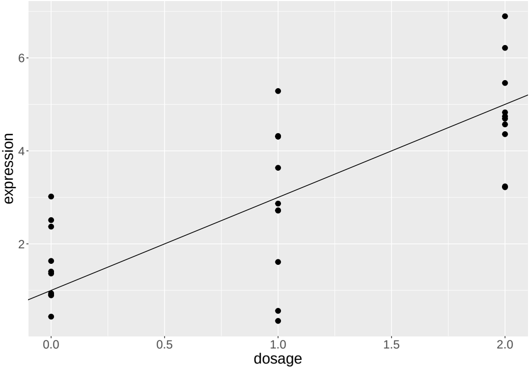
$$f(x) = \beta_0 + \beta_x x$$



In our case this means that we consider the effects of the paternal and maternal alleles on gene expression as independent and additive.

Error term

If the regression function is linear we expect, considering the error term ϵ, the actual measurements to look like this



Estimating β_0 and β_x

Given the data:

- dosage x_i $(i = 1 \dots N)$ and expression y_i $(i = 1 \dots N)$ for a large number N of individuals

our first goal is to estimate the regression function, i.e. the values of β_0 and β_x.

This is done by choosing the values that minimize the *mean square error*

$$MSE = \sum_{i=1}^{N} (y_i - f(x_i))^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_x x_i)^2$$

Is this an eQTL?

Regression function:

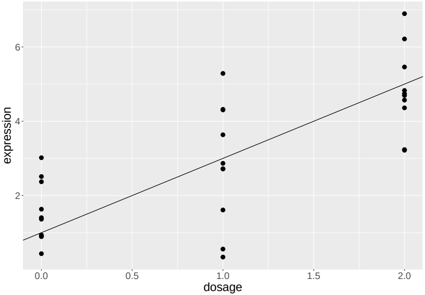
$$f(x) = \beta_0 + \beta_x x$$

The variant is called an eQTL for the gene if $\beta_x \neq 0$ (if $\beta_x = 0$ expression does *not* actually depend on the variant dosage).

- Null hypothesis: $\beta_x = 0$
- Regression algorithms provide us not only with the value of β_x (and β_0) but also with their *uncertainties*
- These can be used to *test* the null hypothesis and obtain a *P*-value
- Small *P*-values indicate that
    - the null hypothesis is probably false
    - that is, $\beta_x \neq 0$
    - that is, the variant dosage does indeed affect the expression of the gene
    - that is, the variant is an eQTL of the gene

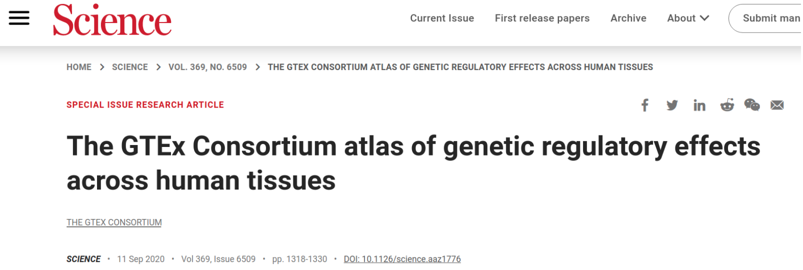
Example

For our (fake) data



- the estimated β_x is 1.64
- with an uncertainty (standard error) of 0.277
- so that it is very unlikely that the true β_x is 0
- indeed the P-value is $2.36\ 10^{-6}$.

The GTEx project


Science
Current Issue
First release papers
Archive
About
Submit man
HOME > SCIENCE > VOL. 369, NO. 6509 > THE GTEX CONSORTIUM ATLAS OF GENETIC REGULATORY EFFECTS ACROSS HUMAN TISSUES

SPECIAL ISSUE RESEARCH ARTICLE

The GTEx Consortium atlas of genetic regulatory effects across human tissues

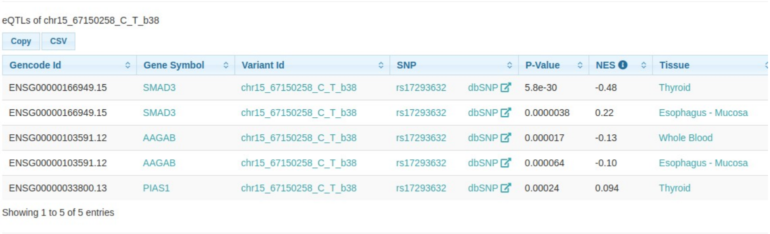
THE GTEX CONSORTIUM

SCIENCE • 11 Sep 2020 • Vol 369, Issue 6509 • pp. 1318-1330 • DOI: 10.1126/science.aaz1776

- eQTLs can be tissue-specific
- e.g. by altering the binding site of a brain-specific TF
- GTEx: eQTL analysis in ~50 human tissues
    - analyze all variants within 1 Mb of each gene

Back to rs17293632

GTEx confirms that rs17293632 is an eQTL of SMAD3 in two tissues (but also of other neignboring genes)



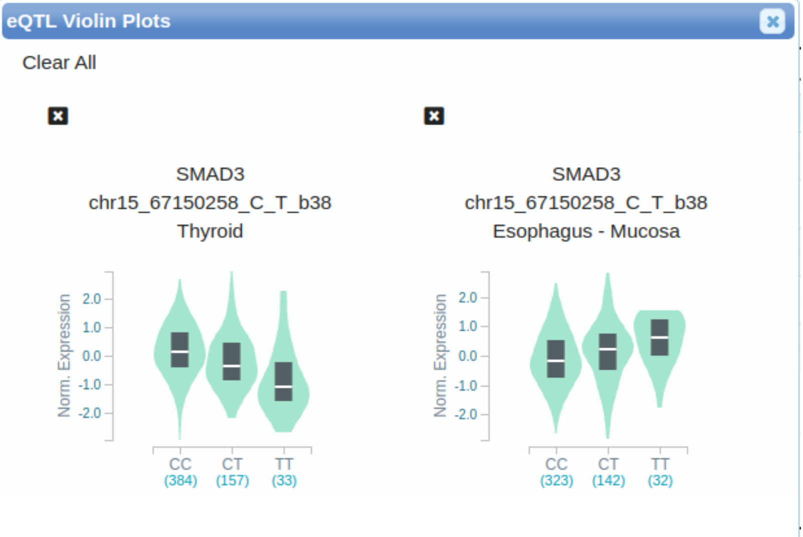
eQTLs of chr15_67150258_C_T_b38

Gencode Id	Gene Symbol	Variant Id	SNP		P-Value	NES	Tissue
ENSG00000166949.15	SMAD3	chr15_67150258_C_T_b38	rs17293632	dbSNP	5.8e-30	-0.48	Thyroid
ENSG00000166949.15	SMAD3	chr15_67150258_C_T_b38	rs17293632	dbSNP	0.0000038	0.22	Esophagus - Mucosa
ENSG00000103591.12	AAGAB	chr15_67150258_C_T_b38	rs17293632	dbSNP	0.000017	-0.13	Whole Blood
ENSG00000103591.12	AAGAB	chr15_67150258_C_T_b38	rs17293632	dbSNP	0.000064	-0.10	Esophagus - Mucosa
ENSG00000033800.13	PIAS1	chr15_67150258_C_T_b38	rs17293632	dbSNP	0.00024	0.094	Thyroid

Showing 1 to 5 of 5 entries

Tissue-dependent effects

The effect of the variant on SMAD3 expression depends on the tissue:



eQTL Violin Plots

Clear All

SMAD3
chr15_67150258_C_T_b38
Thyroid

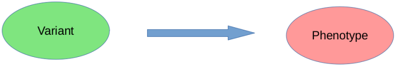
SMAD3
chr15_67150258_C_T_b38
Esophagus - Mucosa

GWAS and eQTLs

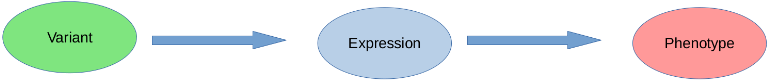
- eQTLs are useful in interpreting variants found by GWAS to be associated to complex traits/diseases
- Enrichment of eQTLs has been found among GWAS hits
- Recently a more global approach to the use of eQTLs has emerged: intermediate molecular phenotypes
    - In particular, *transcriptome-wide association studies* (TWAS)

Gene expression as an intermediate molecular phenotypes

GWAS:

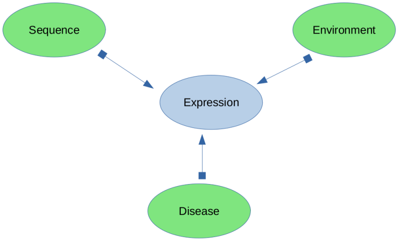


As most hits are non-coding, we actually believe

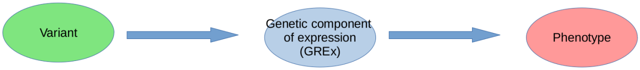


Gene expression is an *intermediate molecular phenotype*

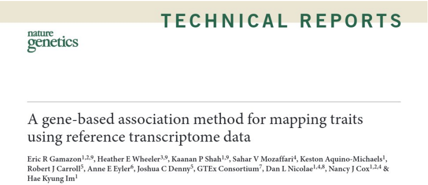
Components of gene expression



- Gene expression is not determined solely by genetics (sequence)
- However only the *genetic component of gene expression* (GREx) can mediate between variants and disease



Transcriptome-wide association studies



TECHNICAL REPORTS

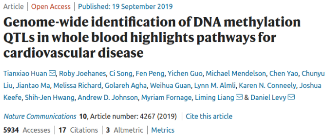
nature genetics

A gene-based association method for mapping traits using reference transcriptome data

Eric R Gamazon[1,2,9], Heather E Wheeler[3,9], Kaanan P Shah[1,9], Sahar V Mozaffari[4], Keston Aquino-Michaels[1], Robert J Carroll[5], Anne E Eyler[6], Joshua C Denny[5], GTEx Consortium[7], Dan L Nicolae[1,4,8], Nancy J Cox[1,2,4] & Hae Kyung Im[1]

- Use eQTL data *from control individuals* (e.g. GTEx) to build a model predicting expression from genotype
- Use the models to compute the *GREx* of all genes for the GWAS diseased and control individuals
- Look for genes whose GREx correlates with the disease
- *Genetic variants affect the disease through the expression of these genes*
- The genes might be *therapeutic targets*

Other molecular QTLs



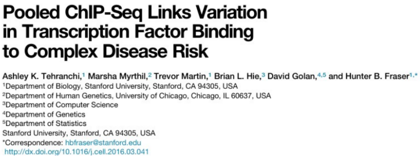
Article | Open Access | Published: 19 September 2019

Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease

Tianxiao Huan ✉, Roby Joehanes, Ci Song, Fen Peng, Yichen Guo, Michael Mendelson, Chen Yao, Chunyu Liu, Jiantao Ma, Melissa Richard, Golareh Agha, Weihua Guan, Lynn M. Almli, Karen N. Conneely, Joshua Keefe, Shih-Jen Hwang, Andrew D. Johnson, Myriam Fornage, Liming Liang ✉ & Daniel Levy ✉

Nature Communications **10**, Article number: 4267 (2019) | Cite this article

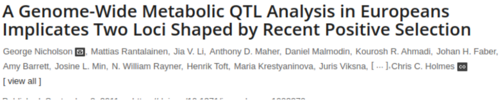
5934 Accesses | **17** Citations | **3** Altmetric | Metrics



Pooled ChIP-Seq Links Variation in Transcription Factor Binding to Complex Disease Risk

Ashley K. Tehranchi,[1] Marsha Myrthil,[2] Trevor Martin,[1] Brian L. Hie,[3] David Golan,[4,5] and Hunter B. Fraser[1,*]

[1]Department of Biology, Stanford University, Stanford, CA 94305, USA
[2]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
[3]Department of Computer Science
[4]Department of Genetics
[5]Department of Statistics
Stanford University, Stanford, CA 94305, USA
*Correspondence: hbfraser@stanford.edu
http://dx.doi.org/10.1016/j.cell.2016.03.041

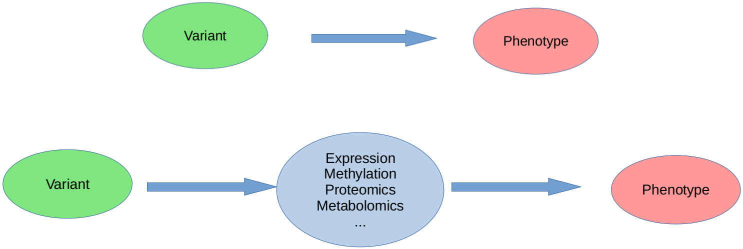


A Genome-Wide Metabolic QTL Analysis in Europeans Implicates Two Loci Shaped by Recent Positive Selection

George Nicholson ✉, Mattias Rantalainen, Jia V. Li, Anthony D. Maher, Daniel Malmodin, Kourosh R. Ahmadi, Johan H. Faber, Amy Barrett, Josine L. Min, N. William Rayner, Henrik Toft, Maria Krestyaninova, Juris Viksna, [...], Chris C. Holmes ✉

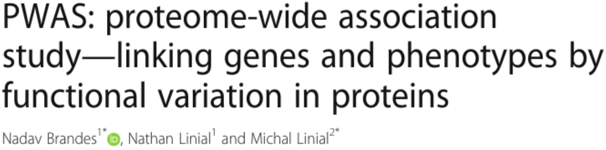
[view all]

Intermediate molecular phenotypes



TWAS generalizations

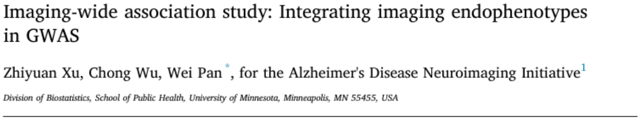
Use other molecular phenotypes as intermediate phenotypes. . .



PWAS: proteome-wide association study—linking genes and phenotypes by functional variation in proteins

Nadav Brandes[1*], Nathan Linial[1] and Michal Linial[2*]

. . . or even macroscopic phenotypes



Imaging-wide association study: Integrating imaging endophenotypes in GWAS

Zhiyuan Xu, Chong Wu, Wei Pan *, for the Alzheimer's Disease Neuroimaging Initiative[1]

Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA

Summary

Understanding the relationship between variants and phenotypes/diseases presents different challenges for Mendelian and complex phenotypes

- Mendelian diseases: find the causal coding variant through *variant prioritization*
- Complex diseases: understand the effect of many variants on gene regulation

In both cases the understanding of the causal relationship can in principle lead to new therapeutic strategies