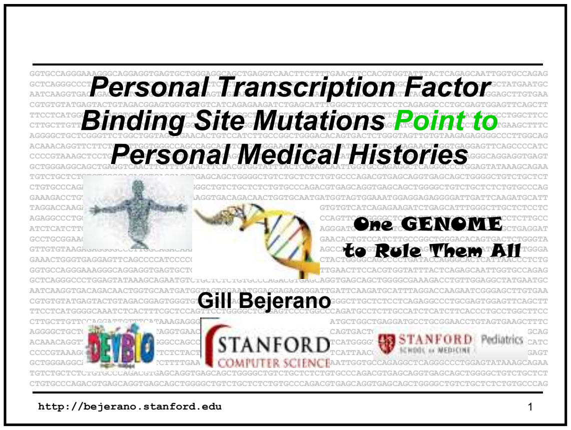
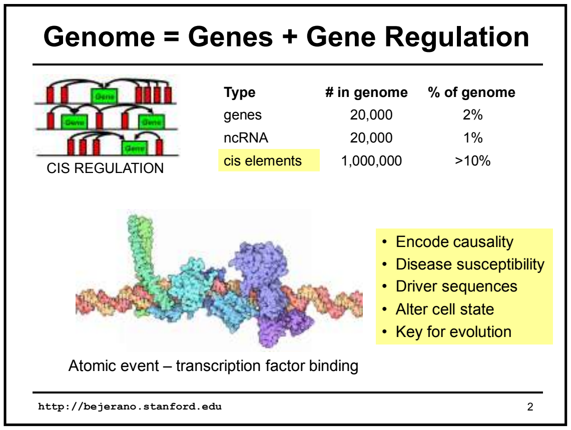


Personal Transcription Factor Binding Site Mutations Point to Personal Medical Histories

Gill Bejerano

http://bejerano.stanford.edu 1



Genome = Genes + Gene Regulation

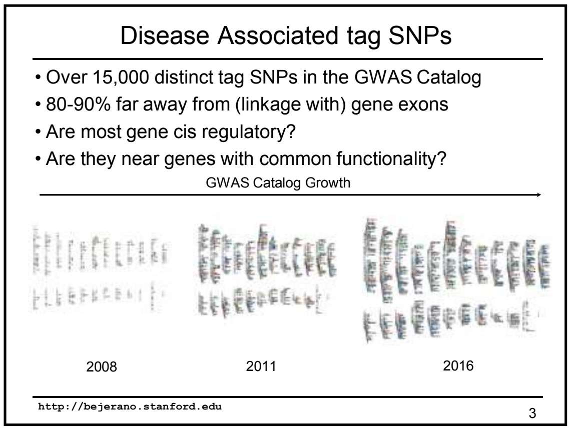
CIS REGULATION

Type	# in genome	% of genome
genes	20,000	2%
ncRNA	20,000	1%
cis elements	1,000,000	>10%

- Encode causality
- Disease susceptibility
- Driver sequences
- Alter cell state
- Key for evolution

Atomic event – transcription factor binding

http://bejerano.stanford.edu 2

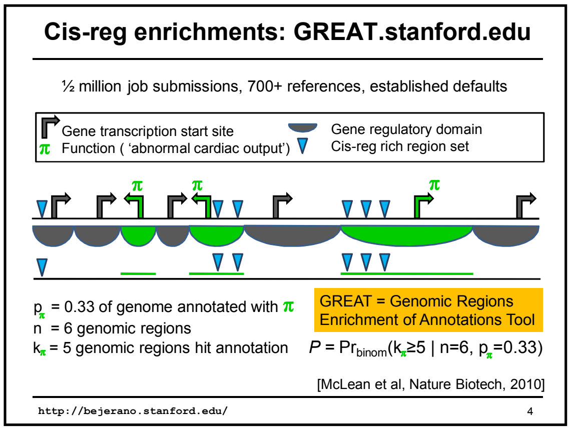


Disease Associated tag SNPs

- Over 15,000 distinct tag SNPs in the GWAS Catalog
- 80-90% far away from (linkage with) gene exons
- Are most gene cis regulatory?
- Are they near genes with common functionality?

GWAS Catalog Growth

http://bejerano.stanford.edu 3



Cis-reg enrichments: GREAT.stanford.edu

½ million job submissions, 700+ references, established defaults

Gene transcription start site	Gene regulatory domain
π Function ('abnormal cardiac output')	▼ Cis-reg rich region set

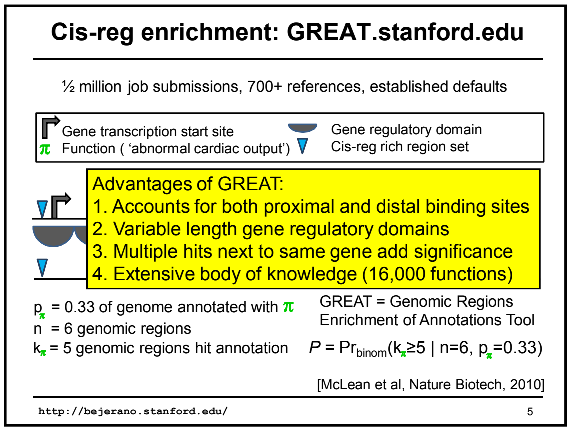
p_π = 0.33 of genome annotated with π
n = 6 genomic regions
k_π = 5 genomic regions hit annotation

GREAT = Genomic Regions Enrichment of Annotations Tool

$P = \Pr_{binom}(k_\pi \geq 5 \mid n=6, p_\pi=0.33)$

[McLean et al, Nature Biotech, 2010]

http://bejerano.stanford.edu/ 4



Cis-reg enrichment: GREAT.stanford.edu

½ million job submissions, 700+ references, established defaults

Gene transcription start site	Gene regulatory domain
π Function ('abnormal cardiac output')	▼ Cis-reg rich region set

Advantages of GREAT:
1. Accounts for both proximal and distal binding sites
2. Variable length gene regulatory domains
3. Multiple hits next to same gene add significance
4. Extensive body of knowledge (16,000 functions)

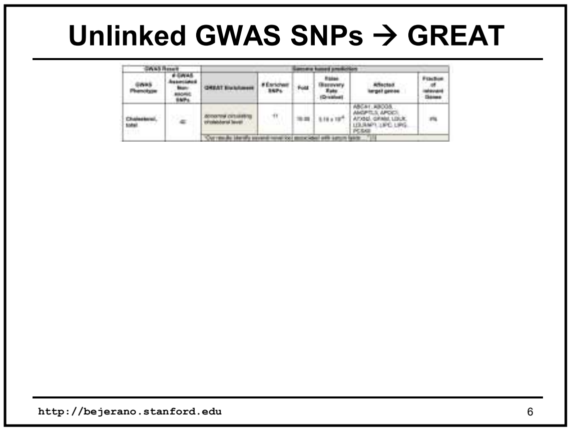
p_π = 0.33 of genome annotated with π
n = 6 genomic regions
k_π = 5 genomic regions hit annotation

GREAT = Genomic Regions Enrichment of Annotations Tool

$P = \Pr_{binom}(k_\pi \geq 5 \mid n=6, p_\pi=0.33)$

[McLean et al, Nature Biotech, 2010]

http://bejerano.stanford.edu/ 5

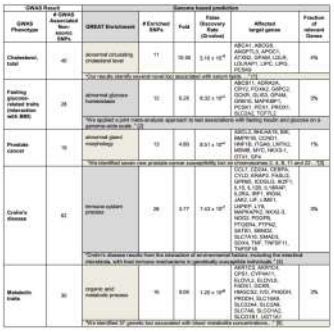


Unlinked GWAS SNPs → GREAT

GWAS Result		Genome based prediction					
GWAS Phenotype	# GWAS Associated Non-exonic SNPs	GREAT Enrichment	# Enriched SNPs	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant Genes
Cholesterol, total	42	abnormal circulating cholesterol level	11	10.00	5.18×10^{-4}	ABCA1, ABCG5, ANGPTL3, APOC1, ATXN2, GFAP, LDLR, LDLRAP1, LIPC, LIPG, PCSK9	4%
		"Our results identify several novel loci associated with serum lipids ..." [1]					

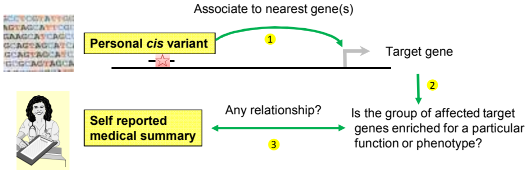
http://bejerano.stanford.edu 6

Unlinked GWAS SNPs → GREAT



GWAS Result		Genome based prediction					
GWAS Phenotype	# GWAS Associated Non-exonic SNPs	GREAT Enrichment	# Enriched SNPs	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant Genes
Cholesterol, total	40	abnormal circulating cholesterol level	11	[illegible]	[illegible]	ABCA1, ABCG8, ANGPTL3, APOC1, ATXN2, GPAM, LDLR, LDLRAP1, LIPC, LIPG, PCSK9	4%
		"Our results identify several novel loci associated with serum lipids…" [1]					
Fasting glucose-related traits (interaction with BMI)	28	abnormal glucose homeostasis	12	6.29	[illegible]	ABCB11, ADRA2A, CRY2, FOXA2, G6PC2, GCKR, GLIS3, GPAM, GRB10, MAPK8IP1, PCSK1, PDX1, PROX1, SLC2A2, TCF7L2	3%
		"We applied a joint meta-analysis approach to test associations with fasting insulin and glucose on a genome-wide scale." [2]					
Prostate cancer	16	abnormal gland morphology	13	4.80	[illegible]	[illegible], BHLHA15, BIK, BMPR1B, CCND1, HNF1B, ITGA6, LMTK2, MSMB, MYC, NKX3-1, OTX1, SPx	1%
		"We identified seven new prostate cancer susceptibility loci on chromosomes 2, 4, 8, 11 and 22…" [3]					
Crohn's disease	62	immune system process	28	3.77	[illegible]	CCL7, CD244, CEBPA, CYLD, [illegible], GPR65, [illegible], IL15, IL12B, IL18RAP, IL2RA, IRF1, [illegible], JAK2, LIF, LIME1, [illegible], LYN, MAPKAPK2, [illegible], NOD2, PDGFB, PTGER4, PTPN2, SATB1, [illegible], SLC7A10, SMAD3, SOX4, TNF, TNFSF11, TNFSF18	3%
		"Crohn's disease results from the interaction of environmental factors, including the intestinal microbiota, with host immune mechanisms in genetically susceptible individuals." [4]					
Metabolic traits	30	organic acid metabolic process	16	[illegible]	[illegible]	AKR1C3, AKR1C4, CPS1, CYP4A11, ELOVL2, ELOVL6, FADS1, [illegible], HMGCS2, IVD, PHGDH, PRODH, SLC16A9, SLC22A4, SLC2A9, SLC7A6, SLCO1A2, SLCO1B1, UGT1A1	2%
		"We identified 37 genetic loci associated with blood metabolite concentrations…" [5]					

Individual Genomes → GREAT ?

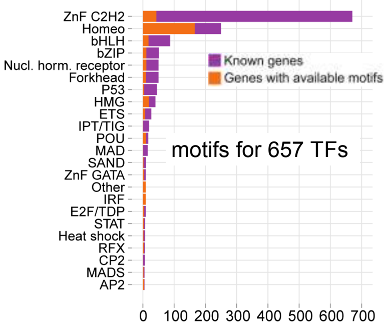

Associate to nearest gene(s)
Personal *cis* variant
1
Target gene
2
Self reported medical summary
Any relationship?
3
Is the group of affected target genes enriched for a particular function or phenotype?

Do individuals carry pathway specific cis-reg mutation load?
Which binding sites?

- Biochemically active ≠ functional
- Under purifying selection

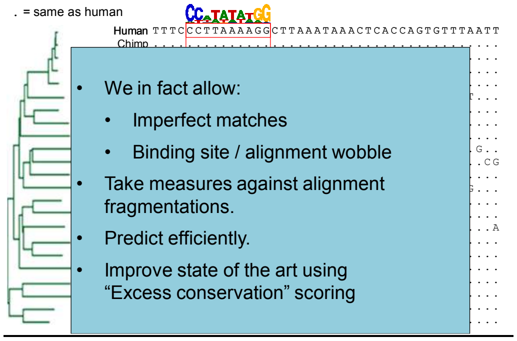
From which context?
The genome, which codes for *all* contexts.

TF Motif Library (PBM+ChIP+SELEX)



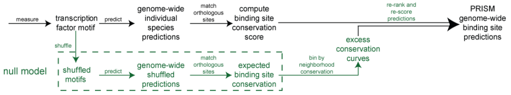
motifs for 657 TFs

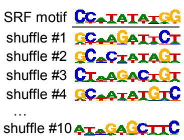
Predict conserved binding sites

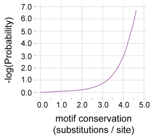


- We in fact allow:
  - Imperfect matches
  - Binding site / alignment wobble
- Take measures against alignment fragmentations.
- Predict efficiently.
- Improve state of the art using "Excess conservation" scoring

Compare *everything* to shuffled motifs & weed!


measure
transcription factor motif
predict
genome-wide individual species predictions
match orthologous sites
compute binding site conservation score
re-rank and re-score predictions
PRISM genome-wide binding site predictions
shuffle
null model
shuffled motifs
predict
genome-wide shuffled predictions
match orthologous sites
expected binding site conservation
bin by neighborhood conservation
excess conservation curves





[Wenger et al, Genome Research, 2013]

PRISM vs. ChIP-seq → GREAT

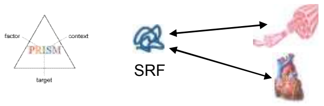
Actin cytoskeleton

Term	PRISM	T cell ChIP-seq	
actin cytoskeleton	✓	✓	Known

PRISM vs. ChIP-seq → GREAT

Actin cytoskeleton

Term	PRISM	T cell ChIP-seq	
actin cytoskeleton	✓	✓	Known
structural constituent of muscle	✓	✗	Known
dilated heart ventricles	✓	✗	Known
regulation of insulin secretion	✓	✗	Novel



PRISM vs. ChIP-seq → GREAT

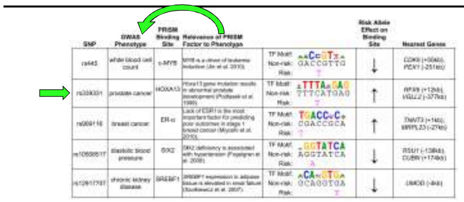
Actin cytoskeleton

Term	PRISM	T cell ChIP-seq	
actin cytoskeleton	✓	✓	Known
structural constituent of muscle	✓	✗	Known
dilated heart ventricles	✓	✗	Known
regulation of insulin secretion	✓	✗	Novel

Every known function is supported
by dozens or hundreds of *novel* binding sites.

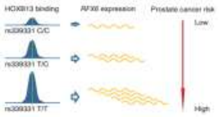
Note: Sensitivity vs Specificity

GWAS SNPs: Predict upstream regulator

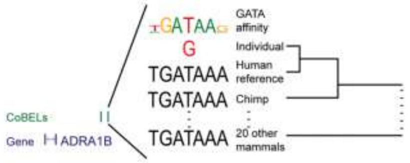


SNP	GWAS Phenotype	PRISM Binding Site	Relevance of PRISM Factor to Phenotype		Risk Allele Effect on Binding Site	Nearest Genes
rs445	white blood cell count	c-MYB	MYB is a driver of leukemia induction (Jin et al. 2010)	TF Motif: Non-risk: GACCGTTG Risk: T	↓	CDK6 (+50kb), PEX1 (-251kb)
rs339331	prostate cancer	HOXA13	Hoxa13 gene mutation results in abnormal prostate development (Podlasek et al. 1999)	TF Motif: Non-risk: TTTCATGAG Risk: T	↑	RFX6 (+12kb), VGLL2 (-377kb)
rs909116	breast cancer	ER-α	Lack of ESR1 is the most important factor for predicting poor outcomes in stage 1 breast cancer (Mycek et al. 2010)	TF Motif: Non-risk: CGACCGCA Risk: T	↑	TNNT3 (+1kb), MRPL23 (-27kb)
rs10503817	diastolic blood pressure	SIX2	SIX2 deficiency is associated with hypertension (Fogelgren et al. 2008)	TF Motif: Non-risk: AGGTATCA Risk: A	↓	RSU1 (-138kb), CUBN (+174kb)
rs12917707	chronic kidney disease	SREBF1	SREBP1 expression in adipose tissue is elevated in renal failure (Szolkiewicz et al. 2007)	TF Motif: Non-risk: GCAGGTGA Risk: T	↓	UMOD (-4kb)

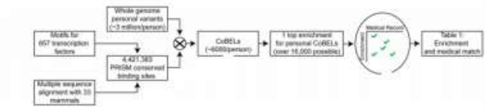
Huang et al. A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. Nat. Genet. 2014 Feb;46(2):126–135.



Personal Deleterious Binding Sites

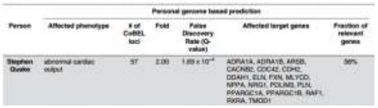


COBELs = Conserved Binding Site Eroding Loci


Motifs for 657 transcription factors
Whole genome personal variants (~3 million/person)
4,421,383 PRISM conserved binding sites
Multiple sequence alignment with 33 mammals
CoBELs (~6000/person)
1 top enrichment for personal CoBELs (over 16,000 possible)
Medical Record
Table 1: Enrichment and medical match

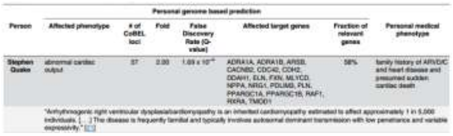
Individuals w public medical records

COBELs → GREAT



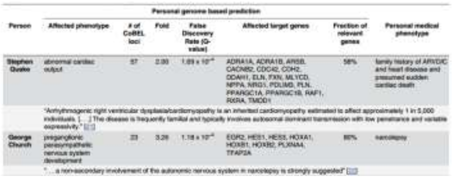
Person	Personal genome based prediction					
	Affected phenotype	# of CoBEL loci	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant genes
Stephen Quake	abnormal cardiac output	37	2.00	1.69×10^{-4}	ADRA1A, ADRA1B, ARSB, CACNB2, CDC42, CDH2, DDAH1, ELN, FXN, MLYCD, NPPA, NRG1, PDLIM3, PLN, PPARGC1A, PPARGC1B, RAF1, RXRA, TMOD1	58%

COBELs → GREAT



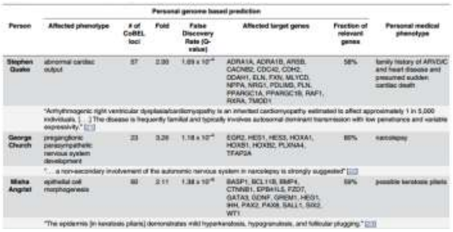
Person	Personal genome based prediction						
	Affected phenotype	# of CoBEL loci	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant genes	Personal medical phenotype
Stephen Quake	abnormal cardiac output	37	2.00	1.69×10^{-4}	ADRA1A, ADRA1B, ARSB, CACNB2, CDC42, CDH2, DDAH1, ELN, FXN, MLYCD, NPPA, NRG1, PDLIM3, PLN, PPARGC1A, PPARGC1B, RAF1, RXRA, TMOD1	58%	family history of ARVD/C and heart disease and presumed sudden cardiac death
	"Arrhythmogenic right ventricular dysplasia/cardiomyopathy is an inherited cardiomyopathy estimated to affect approximately 1 in 5,000 individuals. [...] The disease is frequently familial and typically involves autosomal dominant transmission with low penetrance and variable expressivity."						

COBELs → GREAT



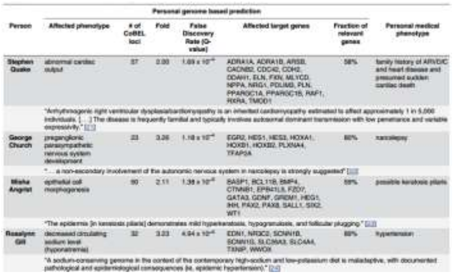
	Personal genome based prediction						
Person	Affected phenotype	# of CoBEL loci	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant genes	Personal medical phenotype
Stephen Quake	abnormal cardiac output	57	2.00	1.69×10^{-4}	ADRA1A, ADRA1B, ARSB, CACNB2, CDC42, CDH2, DDAH1, ELN, FXN, MLYCD, NPPA, NRG1, PDLIM3, PLN, PPARGC1A, PPARGC1B, RAF1, RXRA, TMOD1	58%	family history of ARVD/C and heart disease and presumed sudden cardiac death
	"Arrhythmogenic right ventricular dysplasia/cardiomyopathy is an inherited cardiomyopathy estimated to affect approximately 1 in 5,000 individuals. [...] The disease is frequently familial and typically involves autosomal dominant transmission with low penetrance and variable expressivity." [21]						
George Church	preganglionic parasympathetic nervous system development	23	3.26	1.18×10^{-4}	EGR2, HES1, HES3, HOXA1, HOXB1, HOXB2, PLXNA4, TFAP2A	80%	narcolepsy
	"... a non-secondary involvement of the autonomic nervous system in narcolepsy is strongly suggested" [22]						

COBELs → GREAT



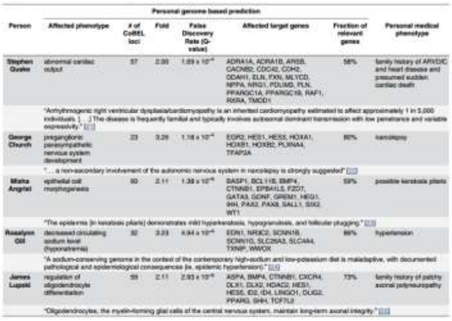
	Personal genome based prediction						
Person	Affected phenotype	# of CoBEL loci	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant genes	Personal medical phenotype
Stephen Quake	abnormal cardiac output	57	2.00	1.69×10^{-4}	ADRA1A, ADRA1B, ARSB, CACNB2, CDC42, CDH2, DDAH1, ELN, FXN, MLYCD, NPPA, NRG1, PDLIM3, PLN, PPARGC1A, PPARGC1B, RAF1, RXRA, TMOD1	58%	family history of ARVD/C and heart disease and presumed sudden cardiac death
	"Arrhythmogenic right ventricular dysplasia/cardiomyopathy is an inherited cardiomyopathy estimated to affect approximately 1 in 5,000 individuals. [...] The disease is frequently familial and typically involves autosomal dominant transmission with low penetrance and variable expressivity." [21]						
George Church	preganglionic parasympathetic nervous system development	23	3.26	1.18×10^{-4}	EGR2, HES1, HES3, HOXA1, HOXB1, HOXB2, PLXNA4, TFAP2A	80%	narcolepsy
	"... a non-secondary involvement of the autonomic nervous system in narcolepsy is strongly suggested" [22]						
Misha Angrist	epithelial cell morphogenesis	60	2.11	1.38×10^{-6}	BASP1, BCL11B, BMP4, CTNNB1, EPB41L5, FZD7, GATA3, GDNF, GREM1, HEG1, IHH, PAX2, PAX8, SALL1, SIX2, WT1	59%	possible keratosis pilaris
	"The epidermis [in keratosis pilaris] demonstrates mild hyperkeratosis, hypogranulosis, and follicular plugging." [23]						

COBELs → GREAT



	Personal genome based prediction						
Person	Affected phenotype	# of CoBEL loci	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant genes	Personal medical phenotype
Stephen Quake	abnormal cardiac output	57	2.00	1.69×10^{-4}	ADRA1A, ADRA1B, ARSB, CACNB2, CDC42, CDH2, DDAH1, ELN, FXN, MLYCD, NPPA, NRG1, PDLIM3, PLN, PPARGC1A, PPARGC1B, RAF1, RXRA, TMOD1	58%	family history of ARVD/C and heart disease and presumed sudden cardiac death
	"Arrhythmogenic right ventricular dysplasia/cardiomyopathy is an inherited cardiomyopathy estimated to affect approximately 1 in 5,000 individuals. [...] The disease is frequently familial and typically involves autosomal dominant transmission with low penetrance and variable expressivity." [21]						
George Church	preganglionic parasympathetic nervous system development	23	3.26	1.18×10^{-4}	EGR2, HES1, HES3, HOXA1, HOXB1, HOXB2, PLXNA4, TFAP2A	80%	narcolepsy
	"... a non-secondary involvement of the autonomic nervous system in narcolepsy is strongly suggested" [22]						
Misha Angrist	epithelial cell morphogenesis	60	2.11	1.38×10^{-6}	BASP1, BCL11B, BMP4, CTNNB1, EPB41L5, FZD7, GATA3, GDNF, GREM1, HEG1, IHH, PAX2, PAX8, SALL1, SIX2, WT1	59%	possible keratosis pilaris
	"The epidermis [in keratosis pilaris] demonstrates mild hyperkeratosis, hypogranulosis, and follicular plugging." [23]						
Rosalynn Gill	decreased circulating sodium level (hyponatremia)	32	3.23	4.94×10^{-4}	EDN1, NR3C2, SCNN1B, SCNN1G, SLC26A3, SLC4A4, TXNIP, WWOX	88%	hypertension
	"A sodium-conserving genome in the context of the contemporary high-sodium and low-potassium diet is maladaptive, with documented pathological and epidemiological consequences (ie, epidemic hypertension)." [24]						

COBELs → GREAT

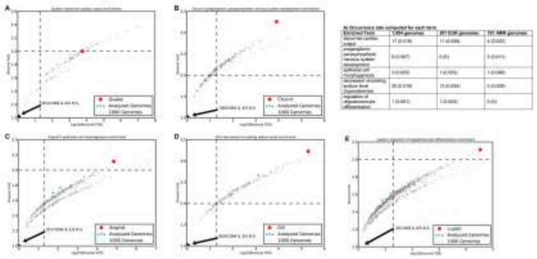


	Personal genome based prediction						
Person	Affected phenotype	# of CoBEL loci	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant genes	Personal medical phenotype
Stephen Quake	abnormal cardiac output	57	2.00	1.69×10^{-4}	ADRA1A, ADRA1B, ARSB, CACNB2, CDC42, CDH2, DDAH1, ELN, FXN, MLYCD, NPPA, NRG1, PDLIM3, PLN, PPARGC1A, PPARGC1B, RAF1, RXRA, TMOD1	58%	family history of ARVD/C and heart disease and presumed sudden cardiac death
	"Arrhythmogenic right ventricular dysplasia/cardiomyopathy is an inherited cardiomyopathy estimated to affect approximately 1 in 5,000 individuals. [...] The disease is frequently familial and typically involves autosomal dominant transmission with low penetrance and variable expressivity." [21]						
George Church	preganglionic parasympathetic nervous system development	23	3.26	1.18×10^{-4}	EGR2, HES1, HES3, HOXA1, HOXB1, HOXB2, PLXNA4, TFAP2A	80%	narcolepsy
	"... a non-secondary involvement of the autonomic nervous system in narcolepsy is strongly suggested" [22]						
Misha Angrist	epithelial cell morphogenesis	60	2.11	1.38×10^{-6}	BASP1, BCL11B, BMP4, CTNNB1, EPB41L5, FZD7, GATA3, GDNF, GREM1, HEG1, IHH, PAX2, PAX8, SALL1, SIX2, WT1	59%	possible keratosis pilaris
	"The epidermis [in keratosis pilaris] demonstrates mild hyperkeratosis, hypogranulosis, and follicular plugging." [23]						
Rosalynn Gill	decreased circulating sodium level (hyponatremia)	32	3.23	4.94×10^{-4}	EDN1, NR3C2, SCNN1B, SCNN1G, SLC26A3, SLC4A4, TXNIP, WWOX	88%	hypertension
	"A sodium-conserving genome in the context of the contemporary high-sodium and low-potassium diet is maladaptive, with documented pathological and epidemiological consequences (ie, epidemic hypertension)." [24]						
James Lupski	regulation of oligodendrocyte differentiation	59	2.11	2.93×10^{-6}	ASPA, BMP4, CTNNB1, CXCR4, DLX1, DLX2, HDAC2, HES1, HES5, ID2, ID4, LINGO1, OLIG2, PPARG, SHH, TCF7L2	73%	family history of patchy axonal polyneuropathy
	"Oligodendrocytes, the myelin-forming glial cells of the central nervous system, maintain long-term axonal integrity." [25]						

Randomize COBELs

- Replace every CoBEL with a random binding site prediction for the same transcription factor of same affinity and similar cross-species conservation.
- Using 10,000 random control sets, the likelihood of obtaining the functions reported in Table 1 as top prediction due to bias in the distribution of binding sites in the genome is low (Quake P = 3 x 10−4, Church P = 5.7 x 10−3, Angrist P = 4.8 x 10−3, Gill P = 1 x 10−4, Lupski P = 1.9 x 10−3, and combined P = 1.6 x 10−15).
- Significance remains high when we relax the requirement to recover each exact same term with matching any one of a broader group of 12–60 related functions as a top prediction (Quake P = 1.1 x 10−3, Church P = 1.3 x 10−2, Angrist P = 7.7 x 10−3, Gill P = 7.4 x 10−3, Lupski P = 6.5 x 10−3, and combined P = 5.2 x 10−12).

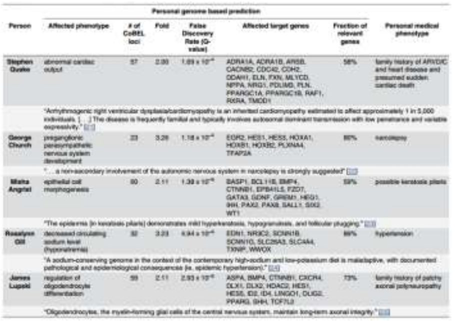
KGP as Controls



Randomize Medical Histories

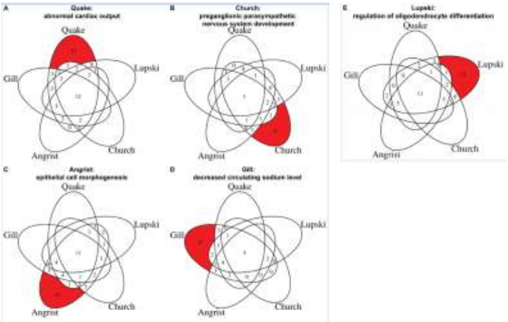
- Define an **association matrix linking enrichment and medical history**, with the phenotypes observed in the five individuals as rows, and top enriched terms in all as columns. A cell in the matrix would be marked "true" only where the enriched term (of any individual) is thought to be related to the etiology of the phenotype (of any individual).
- One instance of this matrix was filled by a **medical doctor** based on their medical knowledge and training and another instance was independently filled using a **literature survey**. The objective was to compute the chance of associating a set of five individuals with random medical histories with the observed enrichments using one of the two association matrices as the "gold" association.
- We generated 1,000 sets of five individuals with random medical histories composed of similar disease profiles and assessed the likelihood of being able to associate them with enrichments. Successfully linking five random individuals with enrichments was highly significant using the association matrix generated by the **medical doctor (P = 3.0 x 10−3)** and by the matrix generated by **literature survey (P = 3.0 x 10−2)** suggesting our links between enrichment and medical histories are not just a function of the listed histories.

COBELs ≠ GWAS SNPs or HGMD

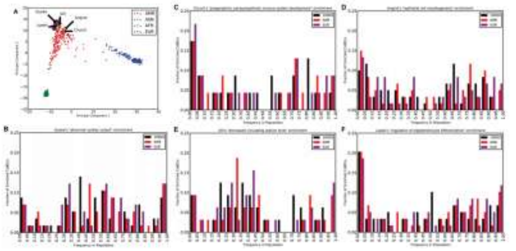


Person	Personal genome based prediction						Personal medical phenotype
	Affected phenotype	# of CoBEL loci	Fold	False Discovery Rate (Q-value)	Affected target genes	Fraction of relevant genes	
Stephen Quake	abnormal cardiac output	57	2.00	1.69×10^{-4}	ADRA1A, ADRA1B, ARSB, CACNB2, CDC42, CDH2, DDAH1, ELN, FXN, MLYCD, NPPA, NRG1, PDLIM3, PLN, PPARGC1A, PPARGC1B, RAF1, RXRA, TMOD1	58%	family history of ARVD/C and heart disease and presumed sudden cardiac death
	"Arrhythmogenic right ventricular dysplasia/cardiomyopathy is an inherited cardiomyopathy estimated to affect approximately 1 in 5,000 individuals. [. . .] The disease is frequently familial and typically involves autosomal dominant transmission with low penetrance and variable expressivity." [illegible]						
George Church	preganglionic parasympathetic nervous system development	23	3.26	1.18×10^{-4}	EGR2, HES1, HES5, HOXA1, HOXB1, HOXB2, PLXNA4, TFAP2A	80%	narcolepsy
	". . . a non-secondary involvement of the autonomic nervous system in narcolepsy is strongly suggested" [illegible]						
Misha Angrist	epithelial cell morphogenesis	80	2.11	1.38×10^{-6}	BASP1, BCL11B, BMP4, CTNNB1, EPB41L5, FZD7, GATA3, GDNF, GREM1, HEG1, IHH, PAX2, PAX8, SALL1, SIX2, WT1	59%	possible keratosis pilaris
	"The epidermis [in keratosis pilaris] demonstrates mild hyperkeratosis, hypogranulosis, and follicular plugging." [illegible]						
Rosalynn Gill	decreased circulating sodium level (hyponatremia)	32	3.23	4.94×10^{-4}	EDN1, NR3C2, SCNN1B, SCNN1G, SLC26A3, SLC4A4, TXNIP, WWOX	88%	hypertension
	"A sodium-conserving genome in the context of the contemporary high-sodium and low-potassium diet is maladaptive, with documented pathological and epidemiological consequences (ie, epidemic hypertension)." [illegible]						
James Lupski	regulation of oligodendrocyte differentiation	59	2.11	2.93×10^{-6}	ASPA, BMP4, CTNNB1, CXCR4, DLX1, DLX2, HDAC2, HES1, HES5, ID2, ID4, LINGO1, OLIG2, PPARG, SHH, TCF7L2	73%	family history of patchy axonal polyneuropathy
	"Oligodendrocytes, the myelin-forming glial cells of the central nervous system, maintain long-term axonal integrity." [illegible]						

Most Predictive COBELs are Private



Contributions from Common & Rare



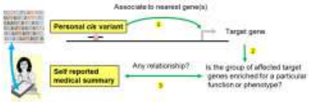
Subset to 1% freq in KGP → lose all enrichments

Summary

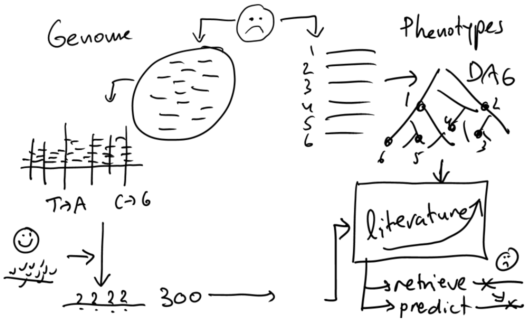
- We define likely deleterious events as personal variants that erode the affinity of human conserved binding sites.
- When the set of all such events is probed for lying next to gene sets of particular function or phenotype, we repeatedly get a solid match between top genomic prediction and self reported medical summary.
- Top genomic predictions are eroded at both gene and gene set level.
- The variants we highlight appear to be part of the mutational load pre-disposing individual lineages to different diseases.

ATAA
G
TGATAAA
TGATAAA
TGATAAA

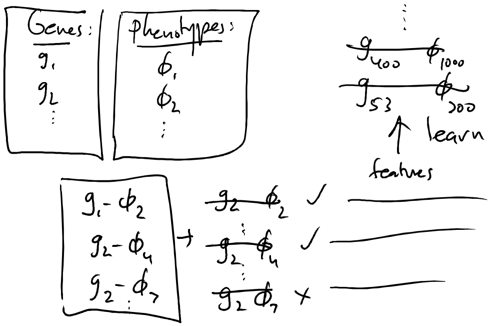
COBELs


Associate to nearest gene(s)
Personal cis variant
Target gene
1
2
Self reported medical summary
Any relationship?
3
Is the group of affected target genes enriched for a particular function or phenotype?

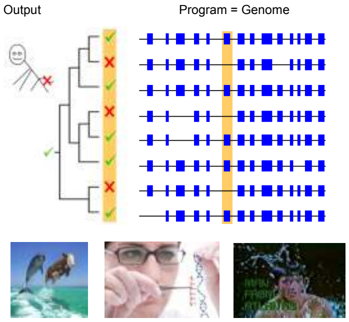
Other Lab Interests: 1) Solve Patient



2) Automate Patient Solving



3) Discover Mammalian Adaptations



Kudos

COBELs: (PLoS Comp Bio, 2016)
Harendra Guturu, Sandeep Chinchali, Shoa Clarke

PRISM: (Genome Research, 2013)
Aaron Wenger, Shoa Clarke, Harendra Guturu, Jenny Chen, Bruce Schaar, Cory McLean

GREAT: (Nature Biotechnology, 2010)
Cory McLean, Dave Bristor, Michael Hiller, Shoa Clarke, Bruce Schaar, Craig Lowe, Aaron Wenger

Bejerano Lab past & present The Organizers