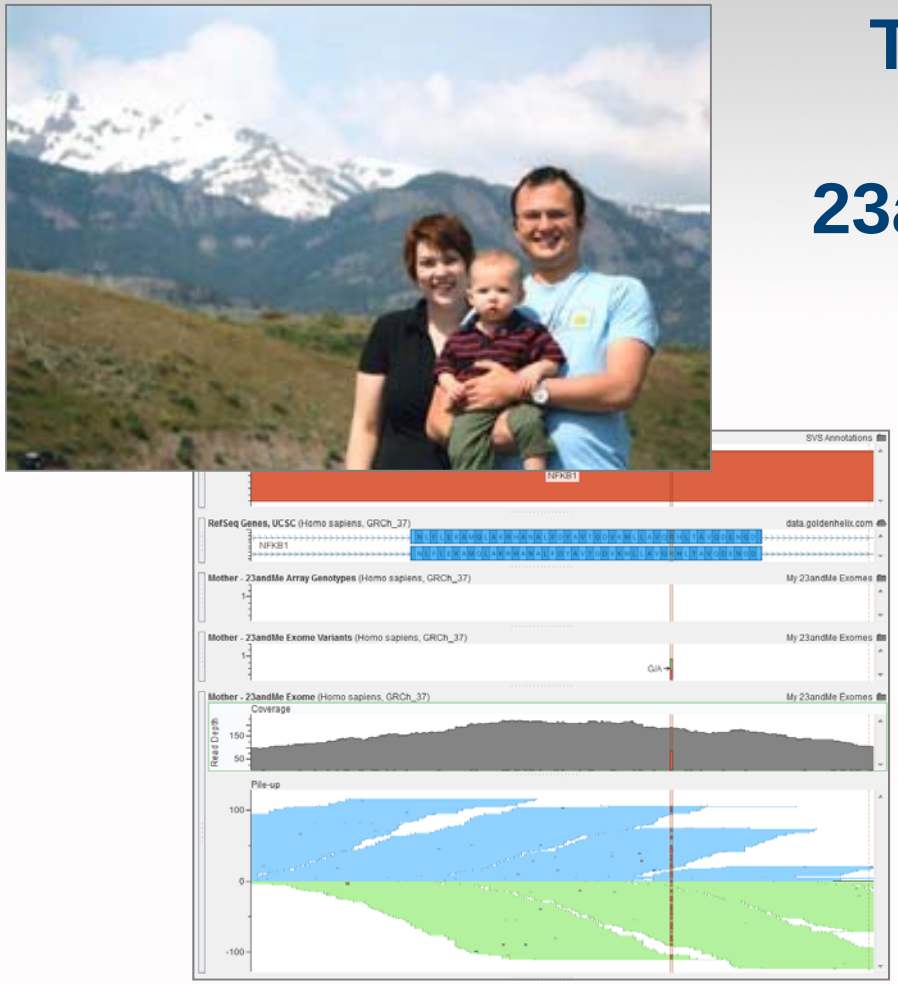
Home Brewed Personalized Genomics



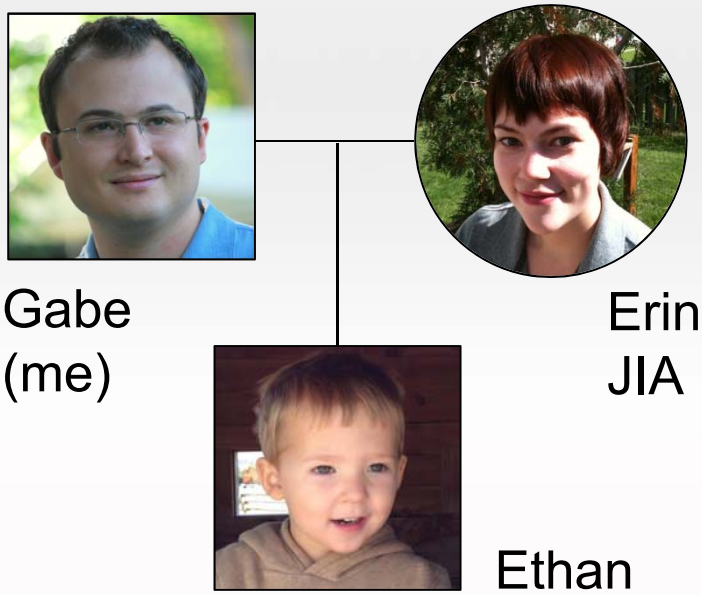
The Quest for Meaningful Analysis Results of a 23andMe Exome Pilot Trio of Myself, Wife, and Son

February 22, 2013

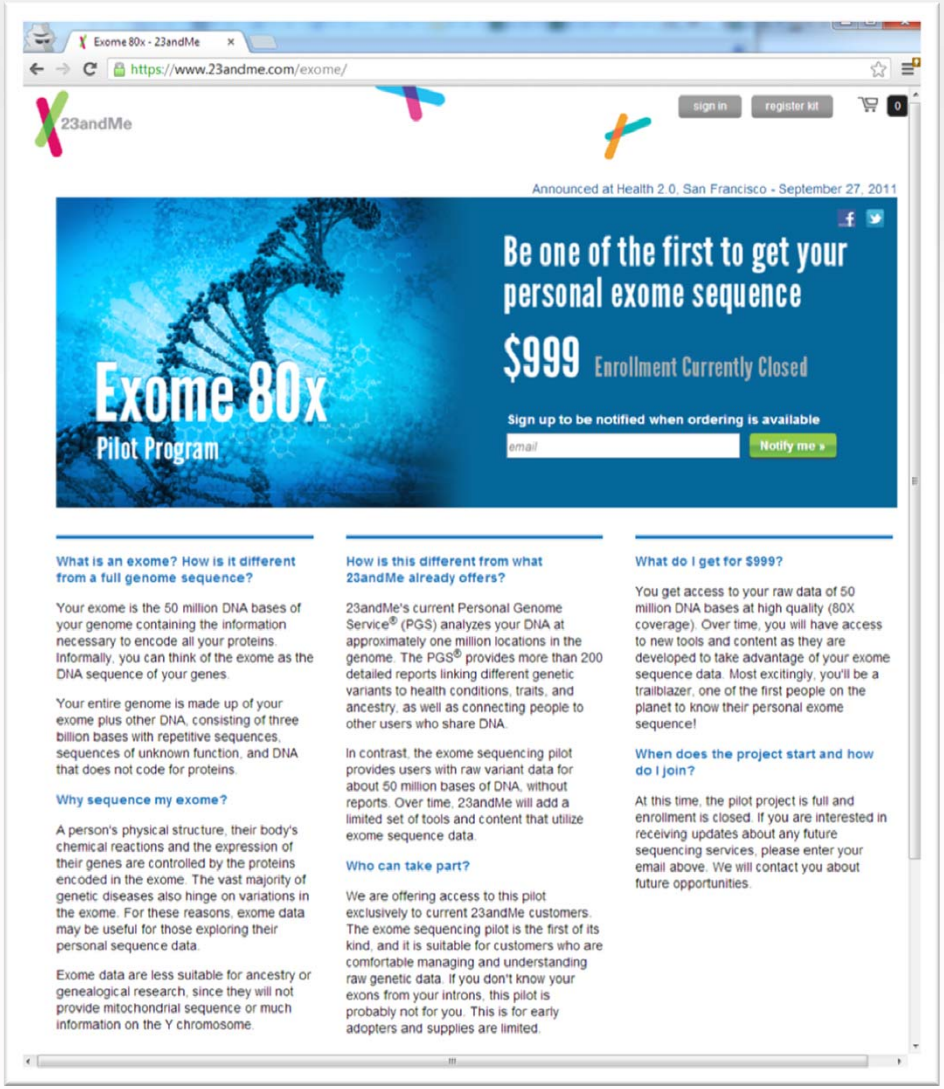
Gabe Rudy, Vice President of Product Development

GOLDEN HELIX
Accelerating the Quest for Significance™

Exome Sequencing in Consumer Genomics



- Exomes done as part of Pilot Program
- 80x coverage
- Raw data with no interpretation



Overview

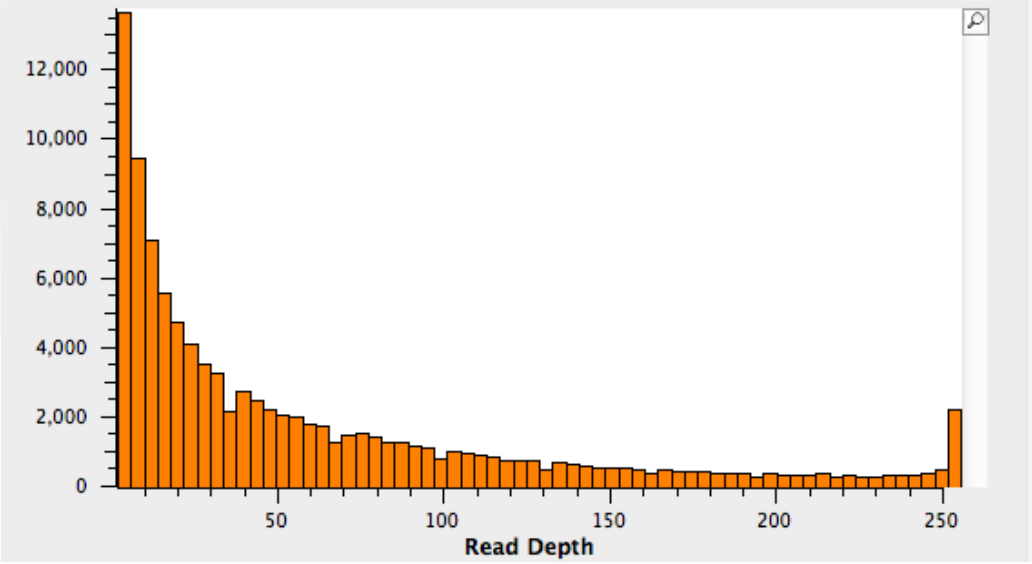
1. Consumer genetics data: research or clinical grade?
2. Treating my healthy self to a Mendelian disease analysis
3. Using exome data to explain a rare autoimmune disorder

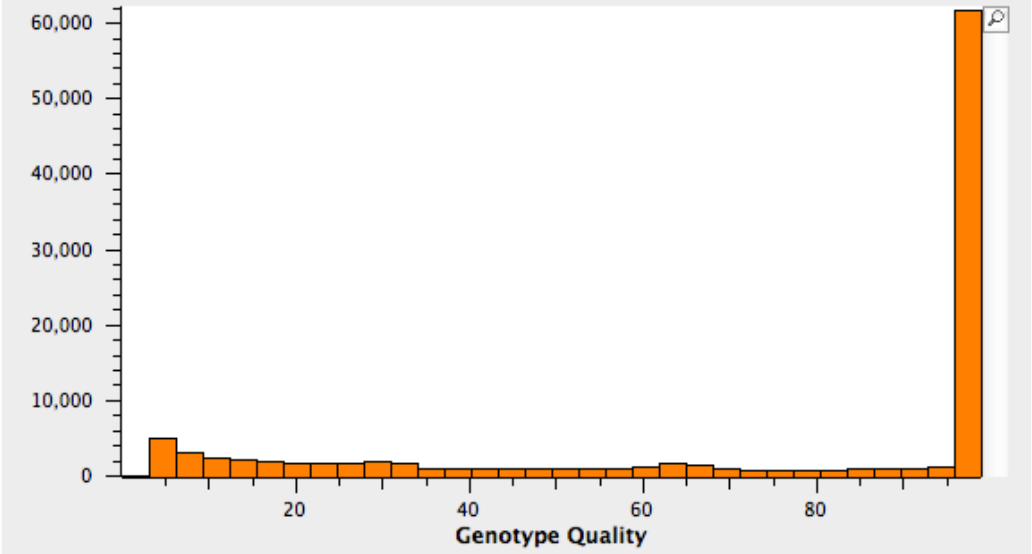
Consumer exomes: done using best practices

- **Sequencing done on HiSeq 2000**
  - 75bp PE
  - Agilent SureSelect exome capture
- **Aligned and called with BWA/GATK**
  - Broad's Best Practices with GATK Guide
  - Indel realignment
  - UnifiedGenotyper called samples concurrently
- **Deliverables**
  - BAM (minus indel realignments)
  - VCF (some filters applied)
  - PDF of Summary Report

Research or clinical grade?

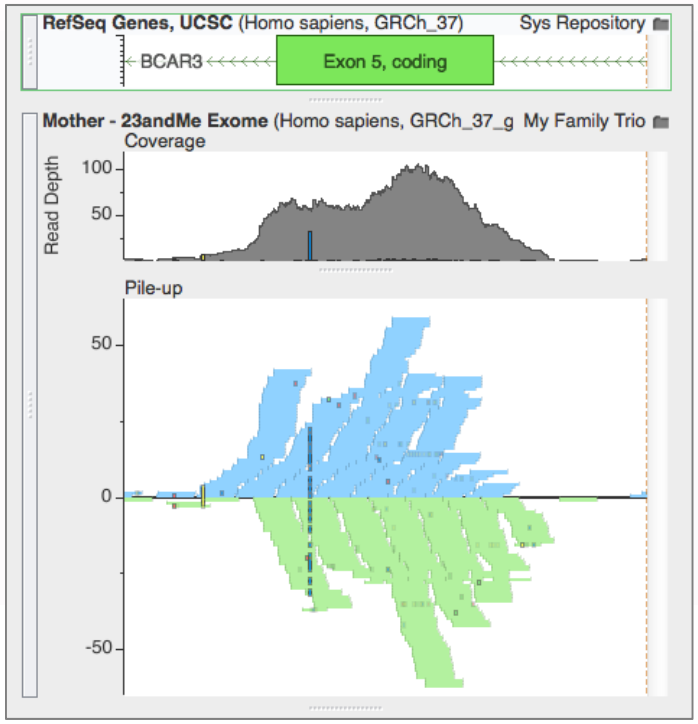
Total Reads	140M
Unique Align	87%
Mean Target	105x
% Target at 2x	97%
% Target at 10x	94%
% Target at 20x	89%
% Target at 30x	83%

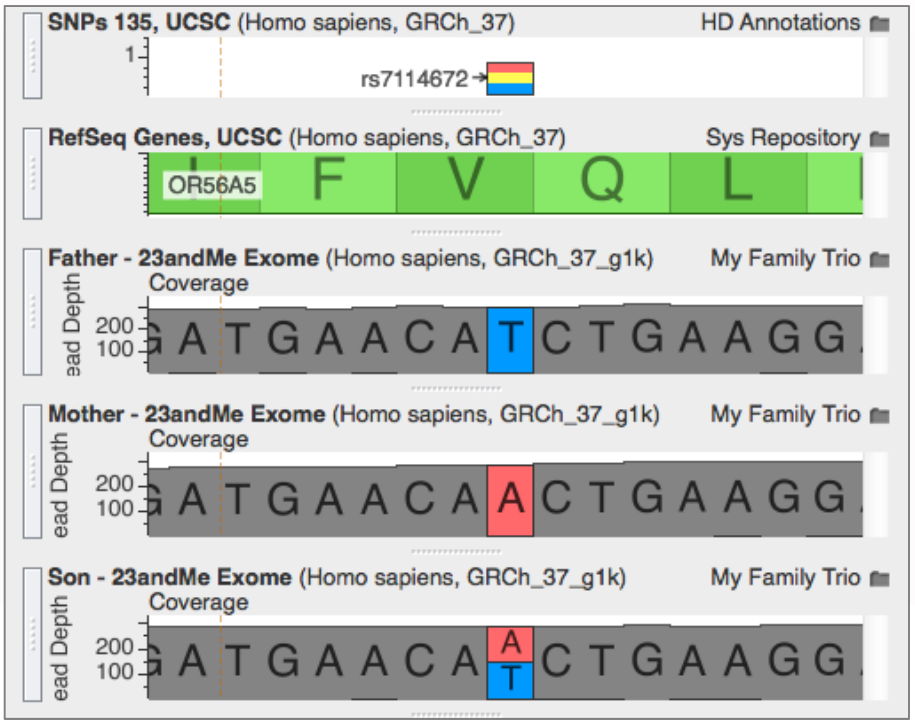




Clinical grade

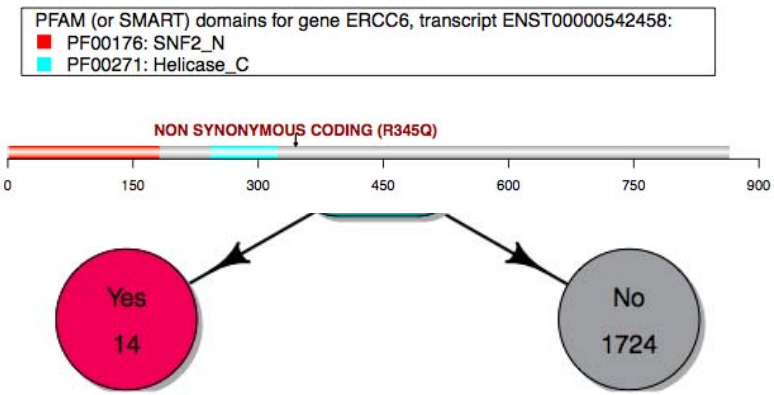
		Unfiltered	Provided	RD>10 & GQ>20	Exonic
Gabe	**SNPs**	98621	89132	65009	19365
	InDels	8141	7800	6503	428
	Ts/Tv	2.36	2.45	2.54	3.26
Trio	**Mendel Errors**	234	202	46	3



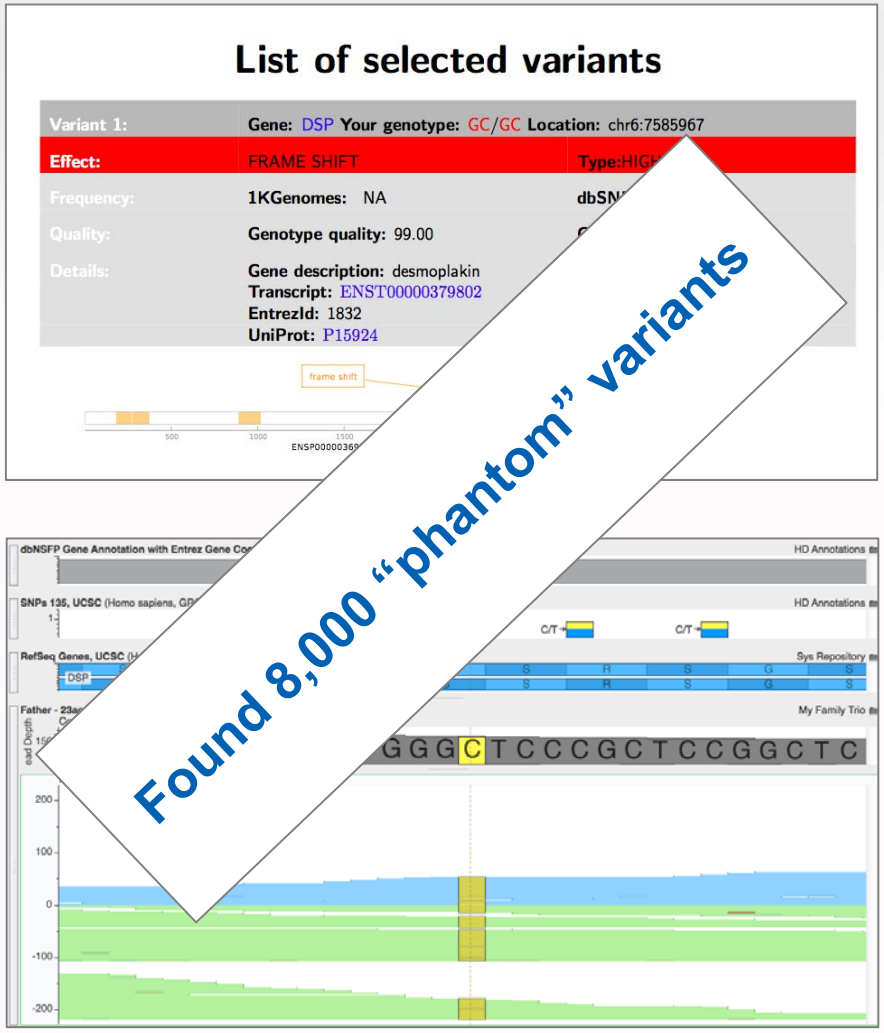


Summary Report

Variant 1:	**Gene:** ERCC6 **Your genotype:** C/T **Location:** chr10:50680422	
Effect:	**Impact:** NON SYNONYMOUS CODING	**Type:** MODERATE
Frequency:	**1KGenomes:** 0.00230	**dbSNP:** rs145720191
Quality:	**Genotype quality:** 99	**Coverage depth:** 142
Details:	**Gene description:** excision repair cross-complementing rodent repair deficiency, complementation group 6	
	Transcript: ENST00000542458	**AA change:** R345Q
	EntrezId: 2074	**EnsemblId:** ENSG00000225830
	UniProt: Q03468	**OMIM:** 609413



Updated VCF and report at end of October



GATK is a Research Tool. Clinics Beware.

GOLDEN HELIX
Accelerating the Quest for Significance™

Overview

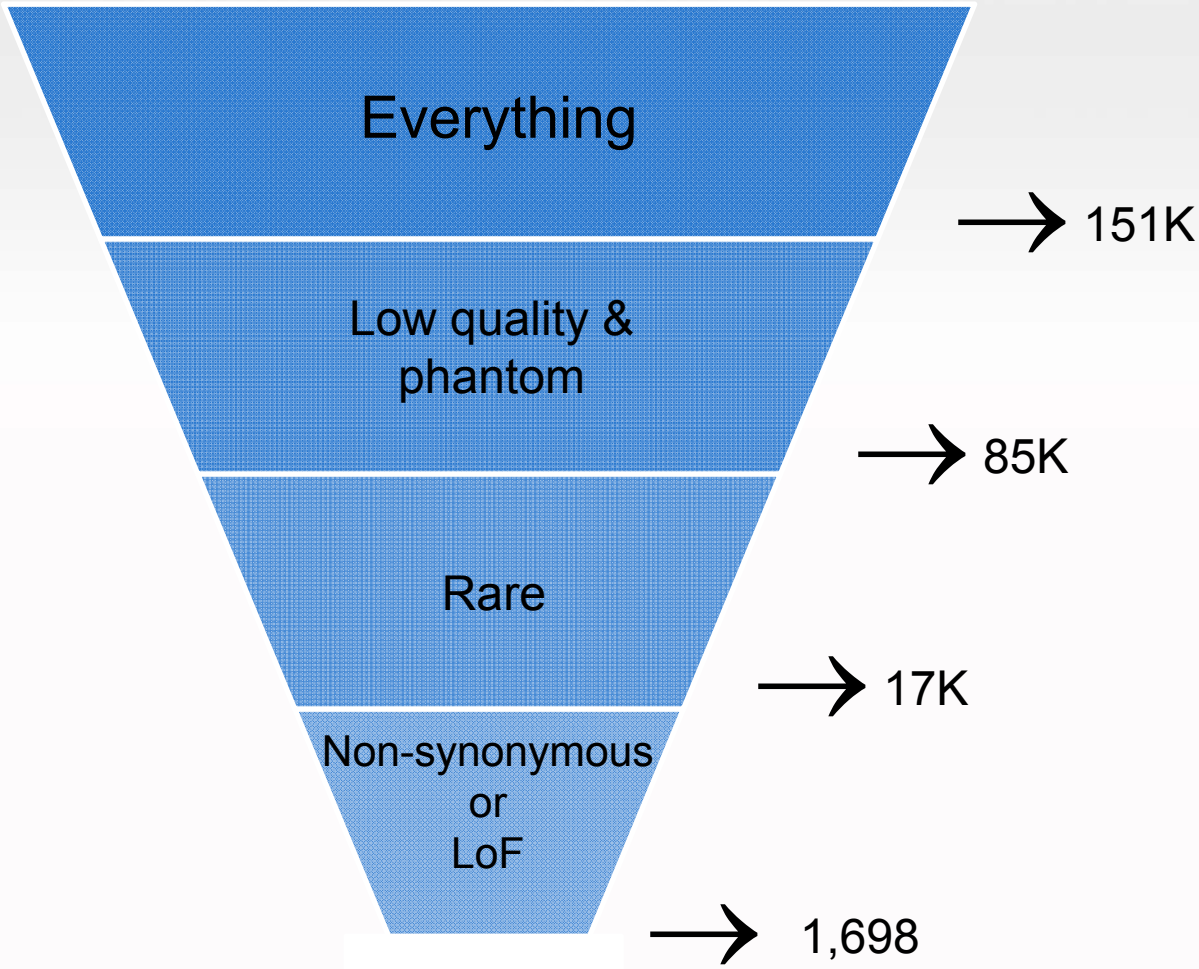
1. Consumer genetics data: research or clinical grade?
2. Treating my healthy self to a Mendelian disease analysis
3. Using exome data to ex

Rare Variant Analysis:
The Hammer

My Exome:
The Nail

Filtering and analysis strategy

- Follow best practices for high-impact variants
- Weed out false-positives
- Use functional prediction
- Interpretation more open-ended



Population catalog and variant classification

Non-coding

	Variant	Rare (Novel)
Intergenic	8,462	3,609 (1,130)
Intronic	48,826	7,516 (4,418)
UTR 3/5	4,128	669 (303)
Non-coding	1,643	648 (183)

Coding

	Variant	Rare (Novel)
Splicing	79	28 (17)
Frameshift Ins/Del	196	138 (118)
Stop gain/loss	113	31 (9)
Non-synonymous	10,252	1,501 (447)
In-frame Ins/Del	256	162 (89)
Synonymous	11,080	885 (215)
Unknown	589	176 (36)

Loss-of-function: 197
Nonsynonymous: 1,501

More filtering strategies

- Regions of Chromosomal Duplication (SuperDups)
- Look at genes in OMIM (most)
- Use predictions of genes as recessive/haploinsufficient to weed out low-priority genes
- For nonsynonymous missense variants can use functional prediction (SIFT/Polyphen2) to annotate

Genes of interest and homozygous variants

Loss of Function

	Rare	!Dups	OMIM	Rec Genes
Splicing	28	17	12 (2)	0 (0)
Frameshift Del	60	44	32 (4)	1 (0)
Frameshift Ins	78	66	46 (5)	3 (1)
Stop gain	31	8	7 (0)	0 (0)

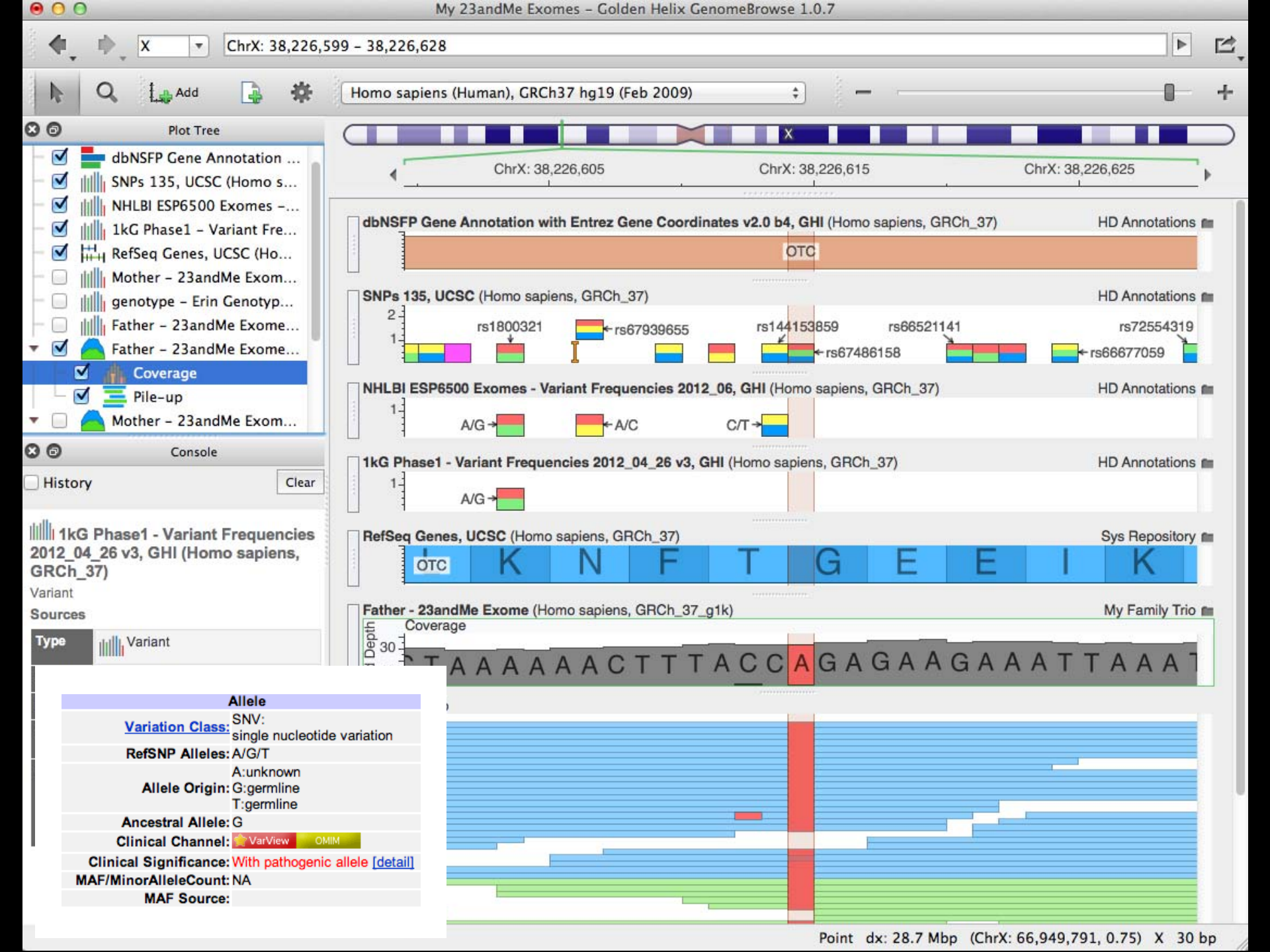
Non-Synonymous

	Rare	!Dups	OMIM	Rec Genes
Damaging (3/3)	337	108	85 (0)	9 (0)
Damaging (2/3)	205	139	97 (0)	7 (0)
Damaging (1/3)	136	194	129 (1)	12 (0)
Tolerated/Unk	781	339	204 (4)	10 (1)

Homozygous: in OMIM: 16
Heterozygous in Rec Genes: 40

Homozygous variants

Note	Variant	AD	DP	GQ	Gene(s)	Classification	HGVS Coding 1
common	1:54605319-Ins	50,26	76	99	CDCP2	Frameshift Ins	c.1224_1225insC
reference	2:71062833-Ins	106,1	107	99	CD207	Splicing	
in-wife	5:156721864-Ins	6,91	97	99	CYFIP2	Frameshift Ins	c.279_280insC
bad-call	6:44269193-Del	120,1	121	99	AARS2	Frameshift Del	c.2607delG
bad-call?	10:46999604-SNV	21,140	161	99	GPRIN2	Nonsyn SNV	c.724A>G
common	12:26834806-Ins	95,1	96	99	ITPR2	Splicing	
in-wife	14:63784408-Ins	3,141	144	99	GPHB5	Frameshift Ins	c.156_157insC
bad-call	17:7606722-Del	161,6	167	99	WRAP53	Frameshift Del	c.1565delC
bad-call	19:54649671-Del	142,1	143	99	CNOT3	Frameshift Del	c.729delT
in-wife	22:19189004-Ins	6,183	189	99	CLTCL1	Frameshift Ins	c.3601_3602insG
VUS	X:16657321-SNV	0,54	54	99	CTPS2	Nonsyn SNV	c.1342A>C
pathogenic	X:38226614-SNV	0,29	29	84.27	OTC	Nonsyn SNV	c.148G>A
VUS	X:100496711-SNV	0,65	66	99	DRP2	Nonsyn SNV	c.380C>T
VUS/in-5-M	X:105167411-SNV	0,16	16	48.13	NRK	Nonsyn SNV	c.2912A>G
wrong-geno	X:112022302-Ins	61,1	62	99	AMOT	Frameshift Ins	c.3080_3081insCC
VUS/common	X:150349559-Del	106,4	110	96.99	GPR50	Frameshift Del	c.1504_1514delACCACTGGCCA

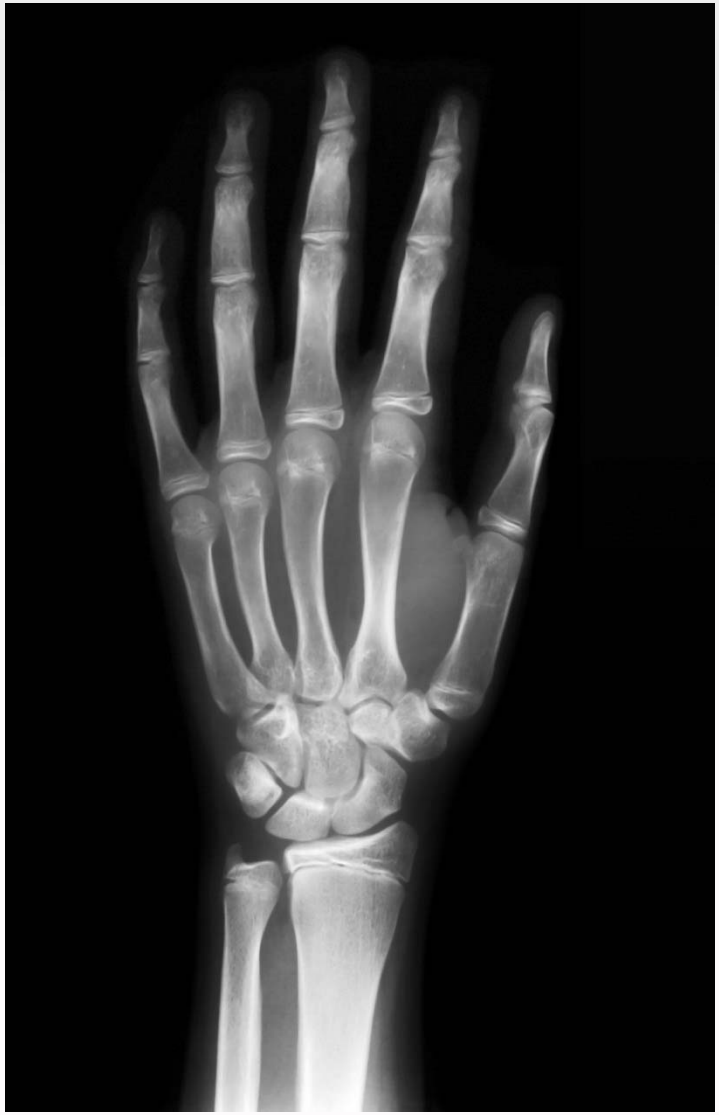

My 23andMe Exomes – Golden Helix GenomeBrowse 1.0.7
X
ChrX: 38,226,599 – 38,226,628
Add
Homo sapiens (Human), GRCh37 hg19 (Feb 2009)
Plot Tree
dbNSFP Gene Annotation ...
SNPs 135, UCSC (Homo s...
NHLBI ESP6500 Exomes -...
1kG Phase1 – Variant Fre...
RefSeq Genes, UCSC (Ho...
Mother – 23andMe Exom...
genotype – Erin Genotyp...
Father – 23andMe Exome...
Father – 23andMe Exome...
Coverage
Pile-up
Mother – 23andMe Exom...
Console
History
Clear
1kG Phase1 – Variant Frequencies 2012_04_26 v3, GHI (Homo sapiens, GRCh_37)
Variant
Sources
Type
Variant
ChrX: 38,226,605
ChrX: 38,226,615
ChrX: 38,226,625
dbNSFP Gene Annotation with Entrez Gene Coordinates v2.0 b4, GHI (Homo sapiens, GRCh_37)
HD Annotations
OTC
SNPs 135, UCSC (Homo sapiens, GRCh_37)
rs1800321
rs67939655
rs144153859
rs66521141
rs72554319
rs67486158
rs66677059
NHLBI ESP6500 Exomes - Variant Frequencies 2012_06, GHI (Homo sapiens, GRCh_37)
A/G
A/C
C/T
1kG Phase1 - Variant Frequencies 2012_04_26 v3, GHI (Homo sapiens, GRCh_37)
RefSeq Genes, UCSC (Homo sapiens, GRCh_37)
Sys Repository
Father - 23andMe Exome (Homo sapiens, GRCh_37_g1k)
My Family Trio
Coverage
Depth
Allele
Variation Class: SNV: single nucleotide variation
RefSNP Alleles: A/G/T
Allele Origin: A:unknown G:germline T:germline
Ancestral Allele: G
Clinical Channel: VarView OMIM
Clinical Significance: With pathogenic allele [detail]
MAF/MinorAlleleCount: NA
MAF Source:
Point dx: 28.7 Mbp (ChrX: 66,949,791, 0.75) X 30 bp

Overview

1. Consumer genetics data: research or clinical grade?
2. Treating my healthy self to a Mendelian disease analysis
3. Using exome data to explain a rare autoimmune disorder

Juvenile Idiopathic Arthritis (JIA)

- Unknown cause, onset before 16
- Between 8 and 150 of every 100,000 children
- 50% have pauciarticular JIA
  40% have polyarticular JIA
- Polyarticular RF negative sub-phenotype has heritability similar to Rheumatoid Arthritis (RA)
  - RA is expected to be 60% heritable
  - 51% explained through current genetic associations
  - 36% of heritability in HLA



IVA – Rare deleterious variance within 1 hop of JIA

Summary | **Variants** | Genes | Groups/Complexes | Pathways | Processes | Diseases | Overview

Edit Columns | Export | Create List | Search for gene name/symbol | 36 variants

Chr...	Position	Gene Region	Gene Symbol	Protein Variant	Case Samples	Sample ...	Translation Impact
1	29356974	Exonic	EPB41	Y187C, Y361C,		44; -	missense
1	94054600	Exonic	BCAR3	R197K, R288K		48; -	missense
3	184008959	Exonic	ECE2	G627S, G656S,		132; -	missense
4	3076672	Exonic	HTT	42_43insPP		10; -	in-frame
4	103518782	Exonic	NFKB1	R533H, R534H		161; -	missense
5	132158821	Intronic, Splice	SHROOM1			27; -	
5	137088945	Exonic	HNRNPA0	270_271insS		17; -	in-frame
6	32548556	Exonic	HLA-DRB1	A244T		250; -	missense
6	32548628	Exonic	HLA-DRB1	R220W		248; -	missense
6	32548632	Exonic	HLA-DRB1	R218S		246; -	missense
6	32549345	Exonic	HLA-DRB1	T214fs*		139; -	frameshift
6	32549361	Exonic	HLA-DRB1	V209M		250; -	missense
6	32549374	Exonic	HLA-DRB1	E198fs*		140; -	frameshift
6	32549402	Exonic	HLA-DRB1	R195fs*		147; -	frameshift
6	32549424	Exonic	HLA-DRB1	V188M		250; -	missense
6	32549531	Exonic	HLA-DRB1	Y152C		231; -	missense
6	32610487	Exonic	HLA-DQA1	F238L		217; -	missense
6	32632694	Exonic	HLA-DQB1	R87P		158; -	missense
6	32714164	Exonic	HLA-DQA2	L254P		146; -	missense
6	160952816	Exonic	LPA	Y2023C		109; -	missense
7	42007506	Exonic	GLI3	P707S		191; -	missense

Filter Cascade

- **Common Variants**: 19848 | 7010
- **Predicted Deleterious**: 1505 | 907
- **Confidence**: 1505 | 907
- **Genetic Analysis**: 527 | 419
- **Biological Context**: 36 | 25

Variant: chr4 | 103518782 | q24 | SNV | A

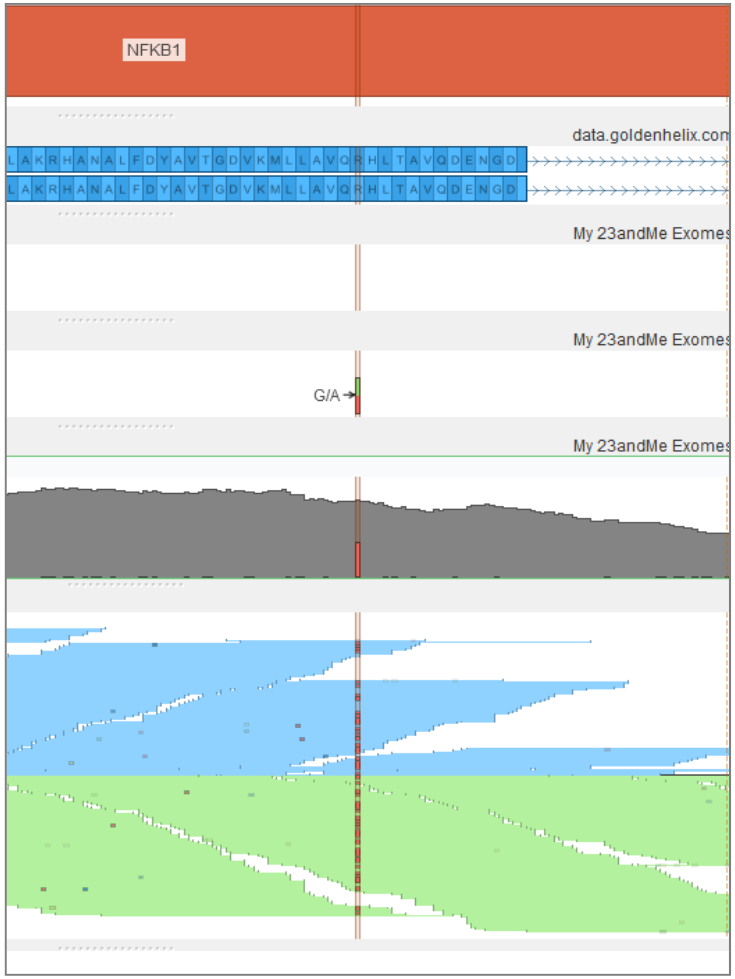
Details | Path to Phenotype | Haploinsufficiency | Diseases

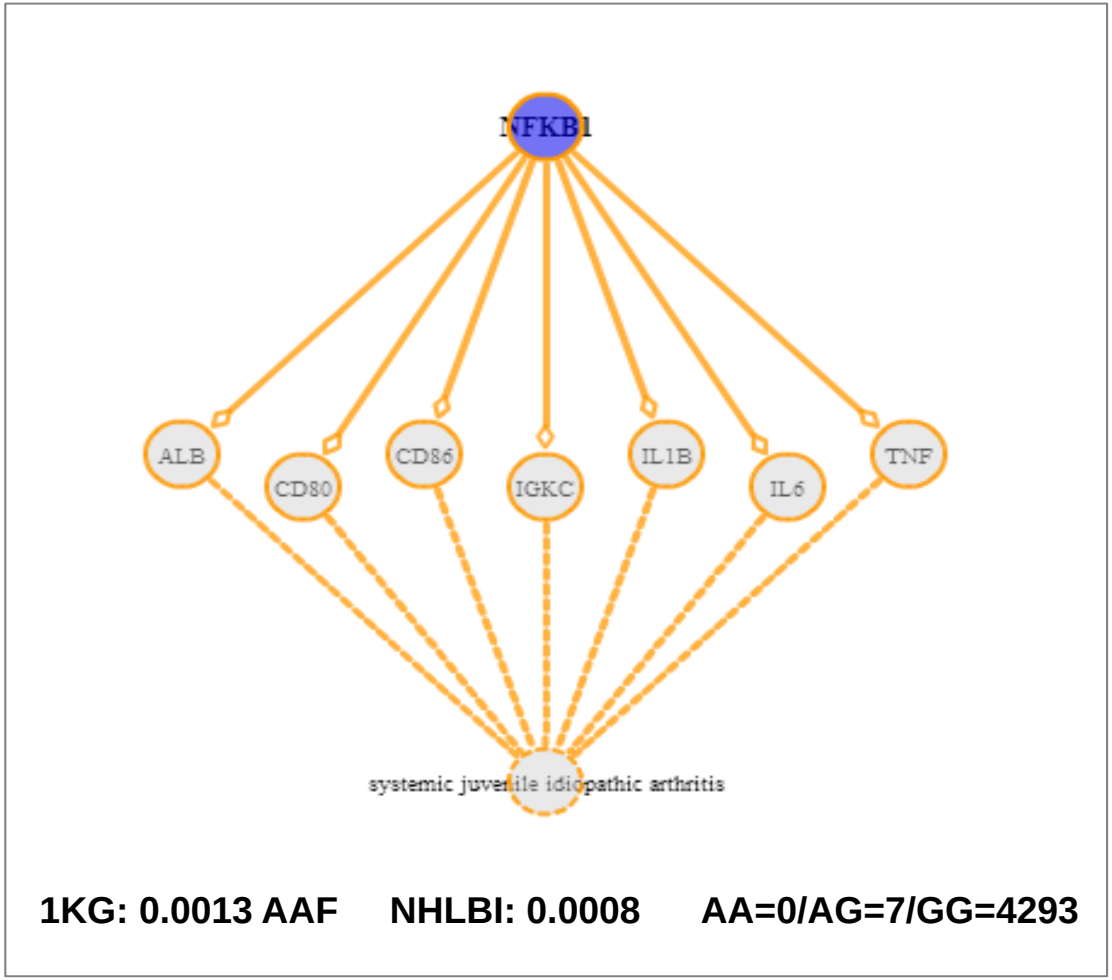
Sample Details (1 of 1 case, 0 of 1 control)

	Name	Subject	State	Genotype	Compound Hete	Call Quality	Copies	Read Depth	Inferred Activity
▬	LK8327 - Mother	Mother	case	Het	No	2134	-	161	loss

Coding Effects

Gene Symbol	Region	Transcript ID	Transcript Variant	Protein Variant	Translation Impact	SIFT Function Prediction
NFKB1	Exonic	NM_001165412.1	1598G>A	R533H	missense	Damaging
NFKB1	Exonic	NM_003998.3	1601G>A	R534H	missense	Damaging





1KG: 0.0013 AAF NHLBI: 0.0008 AA=0/AG=7/GG=4293

Altered T Cell and B Cell Signaling – Mostly HLA-DRB1

Chro...	Position	Gene Region	Gene Symbol	Protein Variant	Case Samples	Sample Read...	Translation Impact	SIFT Functio...	PolyPhen-2 Function Pr...	Conservation p...
4	103518782	Exonic	NFKB1	R533H, R534H		161; -	missense	Damaging	Probably Damaging	7.780E-4
6	32548556	Exonic	HLA-DRB1	A244T		250; -	missense	Damaging	Benign	
6	32548628	Exonic	HLA-DRB1	R220W		248; -	missense	Damaging	Benign	
6	32548632	Exonic	HLA-DRB1	R218S		246; -	missense	Damaging	Benign	
6	32549345	Exonic	HLA-DRB1	T214fs*		139; -	frameshift			
6	32549361	Exonic	HLA-DRB1	V209M		250; -	missense	Damaging	Benign	
6	32549374	Exonic	HLA-DRB1	E198fs*		140; -	frameshift			
6	32549402	Exonic	HLA-DRB1	R195fs*		147; -	frameshift			
6	32549424	Exonic	HLA-DRB1	V188M		250; -	missense	Damaging	Possibly Damaging	1.186E-3
6	32549531	Exonic	HLA-DRB1	Y152C		231; -	missense	Damaging	Probably Damaging	
6	32610487	Exonic	HLA-DQA1	F238L		217; -	missense	Activating	Benign	
6	32632694	Exonic	HLA-DQB1	R87P		158; -	missense	Damaging	Benign	

Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis

Eli A Stahl, Daniel Wegmann, Gosia Trynka, Javier Gutierrez-Achury, Ron Voight, Peter Kraft, Robert Chen, Henrik J Kallberg, Fina A S Kurreeman, Genetics Replication and Meta-analysis Consortium, Myocardial Infarctio Consortium, Sekar Kathiresan, Cisca Wijmenga, Peter K Gregersen, Lars Katherine A Siminovitch, Jane Worthington, Paul I W de Bakker, Soumya Robert M Plenge

Affiliations | Contributions | Corresponding authors

Nature Genetics **44**, 483–489 (2012) | doi:10.1038/ng.2232

Received 06 May 2011 | Accepted 01 March 2012 | Published online 25 March

Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis

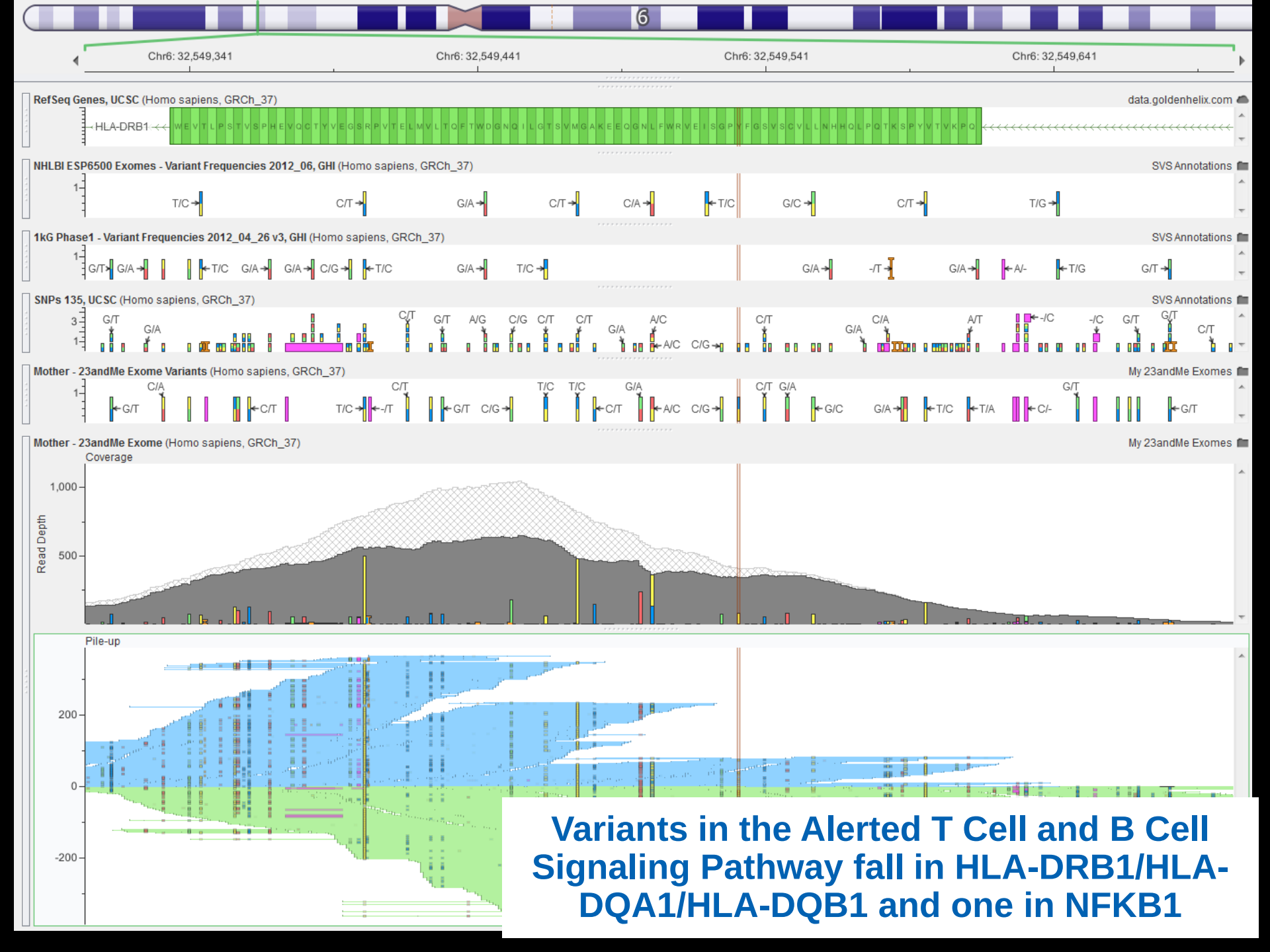
Soumya Raychaudhuri, Cynthia Sandor, Eli A Stahl, Jan Freudenberg, Hye-Soon Lee, Xiaoming Jia, Lars Alfredsson, Leonid Padyukov, Lars Klareskog, Jane Worthington, Katherine A Siminovitch, Sang-Cheol Bae, Robert M Plenge, Peter K Gregersen & Paul I W de Bakker

Affiliations | Contributions | Corresponding authors

Nature Genetics **44**, 291–296 (2012) | doi:10.1038/ng.1076

Received 30 September 2011 | Accepted 12 December 2011 | Published online 29 January 2012

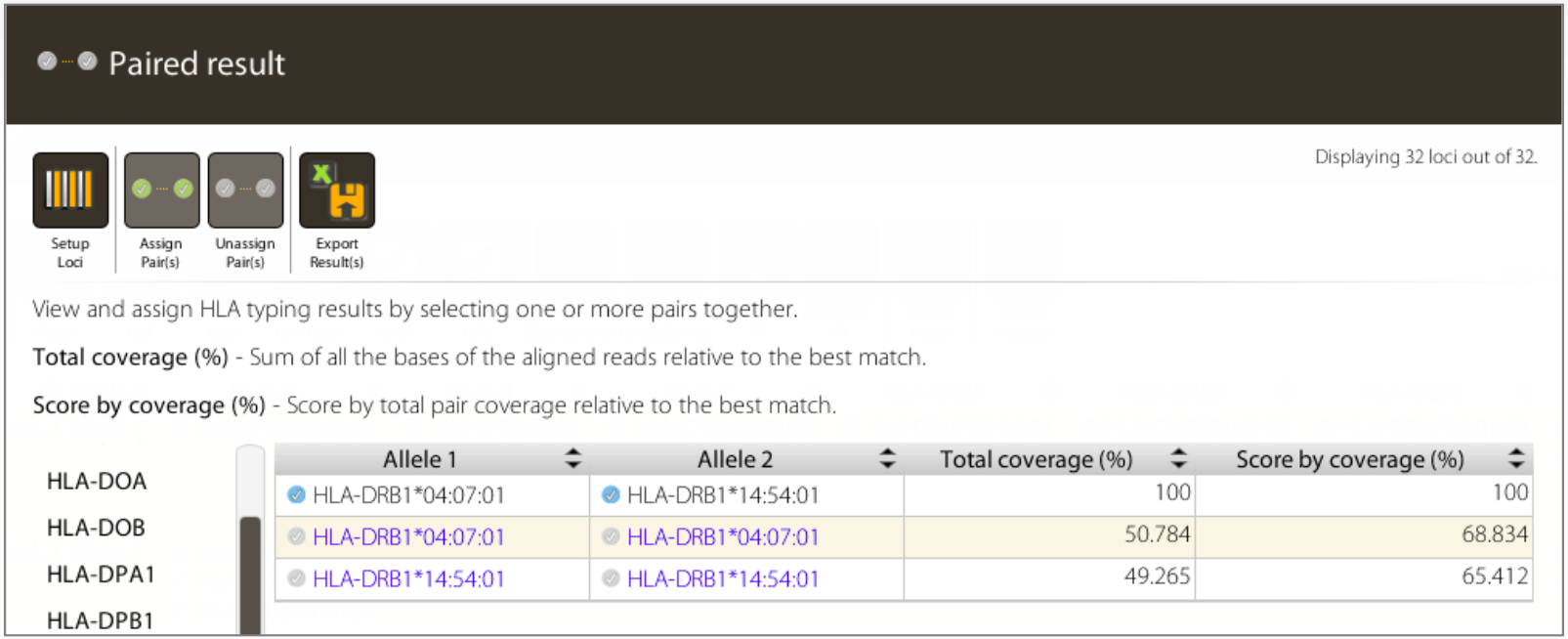
Accelerating the Quest for Significance™



Variants in the Alerted T Cell and B Cell Signaling Pathway fall in HLA-DRB1/HLA-DQA1/HLA-DQB1 and one in NFKB1

HLA Typing with Omixon Target HLA

- 32 HLA genes, “types” based on IMGT/HLA nomenclature
- HLA-A*02 predisposes to early-onset JIA
- HLA-DRB1*08 and HLA-DPB1*03 predispose to poly RF- JIA
- HLA-DRB1*04, HLA-DQA1*03 and HLA-DQB1*03 predispose to poly RF+JIA



HLA-A	HLA-DRB1	HLA-DQA1	HLA-DPB1
A*24:02:01:01	DRB1*04:07:01	DQA1*01:04:01	DPB1*04:01:01
A*03:01:01:01	DRB1*14:54:01	DQA1*03:03:01	DPB1*04:01:01

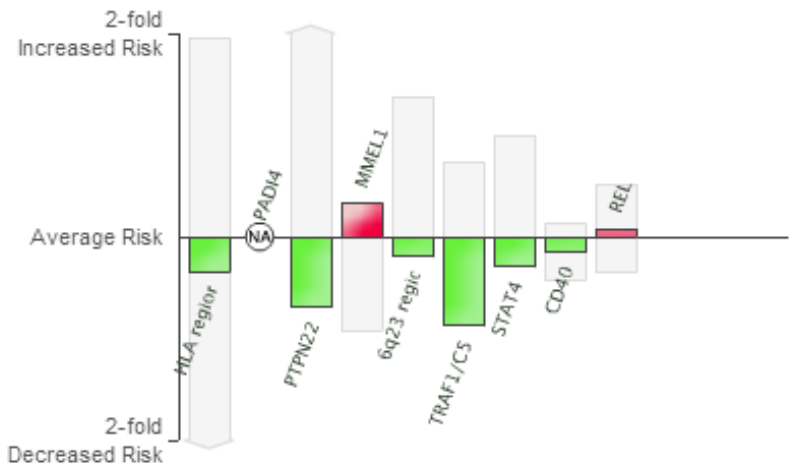
Predicting Lifetime Risk from Population Studies

disease risk

Next ▸
Sarcoidosis

Rheumatoid Arthritis
Like · 4 others like this

Share



What does this chart show?

The chart shows the approximate effects of the selected person's genotype at the 9 reported markers. Higher, red bars indicate increased risk from the average, while lower, green bars indicate decreased risk from the average. The light gray bars show the maximum possible effects for the possible genotypes at the marker.

Mouse over individual bars to view additional information about each marker. Click on a bar to view detailed information about that marker below. You can read more about all markers in the technical report.

HLA region

Marker: rs6457617

The "HLA region"—also known as the "major histocompatibility complex" or MHC—is a region of DNA that contains many genes involved in the immune system's recognition of invaders. Several SNPs have been found in a set of these genes known to be involved in triggering immune cells to attack. Different versions of these HLA genes might determine what kinds of proteins immune cells are presented with as foreign invaders.

Citations

Plenge et al. (2007) . "TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study." *N Engl J Med* 357(12):1199-209.

Wellcome Trust Case Control Consortium (2007) . "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." *Nature* 447(7145):661-78.

GWAS for JIA

Arthritis Rheum. 2010 Nov;62(11):3265-76. doi: 10.1002/art.27688.

The susceptibility loci juvenile idiopathic arthritis shares with other autoimmune diseases extend to PTPN2, COG6, and ANGPT1.

Thompson SD, Sudman M, Ramos PS, Marion MC, Ryan M, Tsoras M, Weiler T, Wagner M, Keddache M, Haas JP, Mueller C, Prahalad S, Bohnsack J, Wise CA, Punaro M, Zhang D, Rosé CD, Comeau ME, Divers J, Glass DN, Langefeld CD.

Cincinna

Arthritis Rheum. 2012 Aug;64(8):2781-91. doi: 10.1002/art.34429.

Genome-wide association analysis of juvenile idiopathic arthritis identifies a new susceptibility locus at chromosomal region 3q13.

Thompson SD, Marion MC, Sudman M, Ryan M, Tsoras M, Howard TD, Barnes MG, Ramos PS, Thomson W, Hinks A, Haas JP, Prahalad S, Bohnsack JF, Wise CA, Punaro M, Rosé CD, Pajewski NM, Spigarelli M, Keddache M, Wagner M, Langefeld CD, Glass DN.

Cincinnati Chi

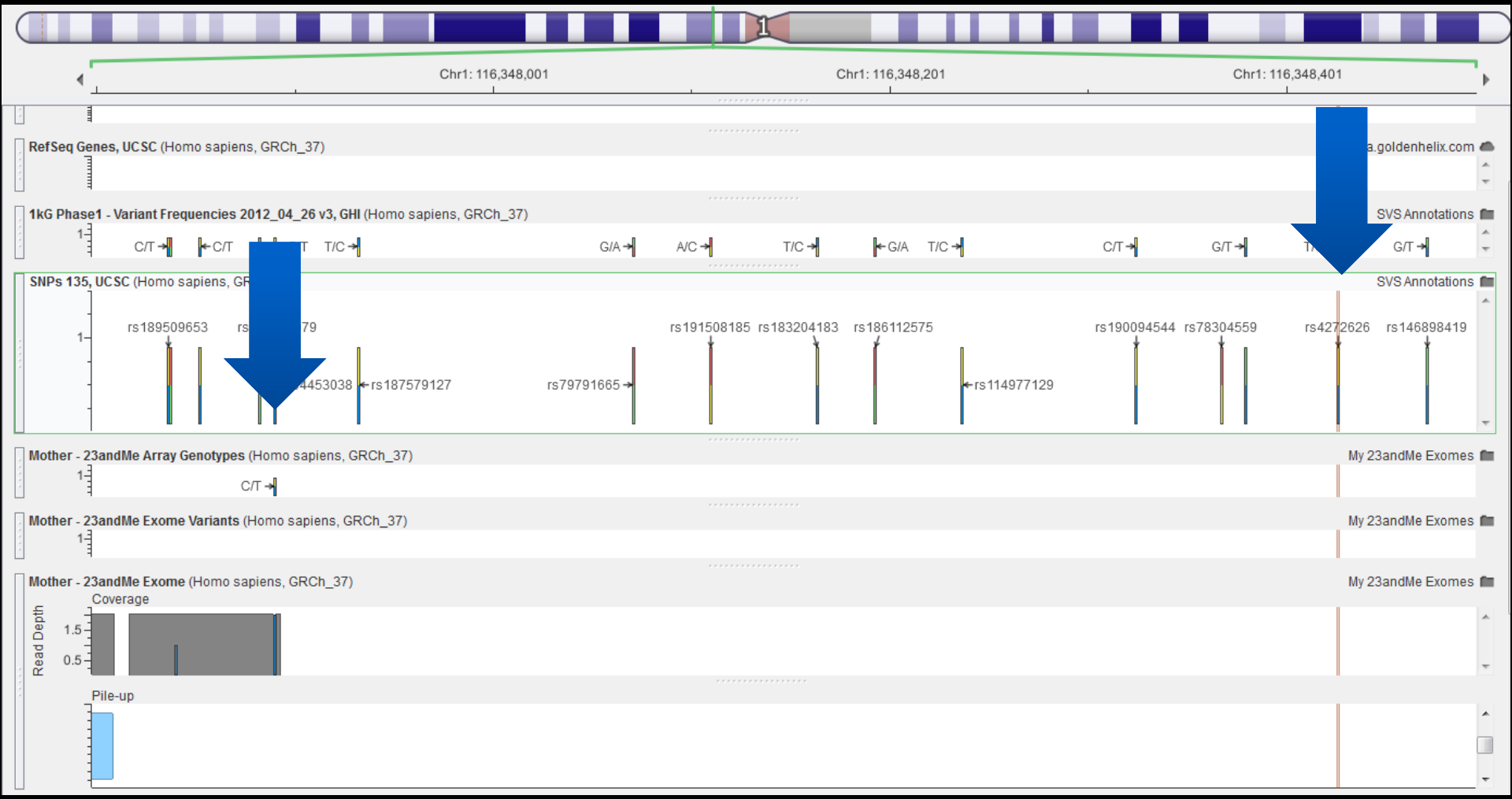
Ann Rheum Dis. 2012 Jul;71(7):1117-21. doi: 10.1136/annrheumdis-2011-200814. Epub 2012 Jan 31.

Investigation of rheumatoid arthritis susceptibility loci in juvenile idiopathic arthritis confirms high degree of overlap.

Hinks A, Cobb J, Sudman M, Eyre S, Martin P, Flynn E, Packham J; Childhood Arthritis Prospective Study (CAPS); UK RA Genetics (UKRAG) Consortium; British Society of Paediatric and Adolescent Rheumatology (BSPAR) Study Group, Barton A, Worthington J, Langefeld CD, Glass DN, Thompson SD, Thomson W.

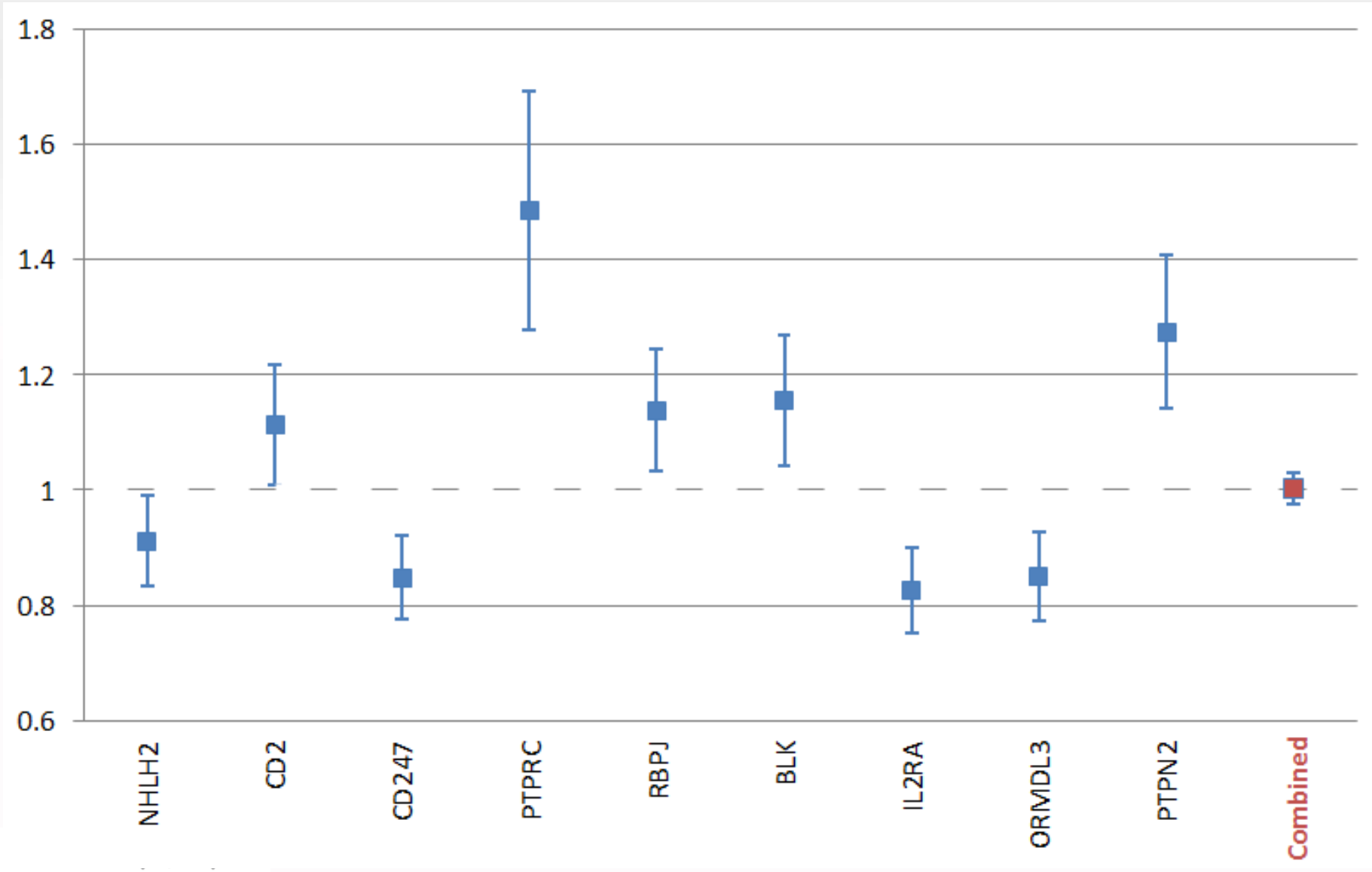
Collaborators (61)

Arthritis Research UK Epidemiology Unit, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK. anne.hinks@manchester.ac.uk

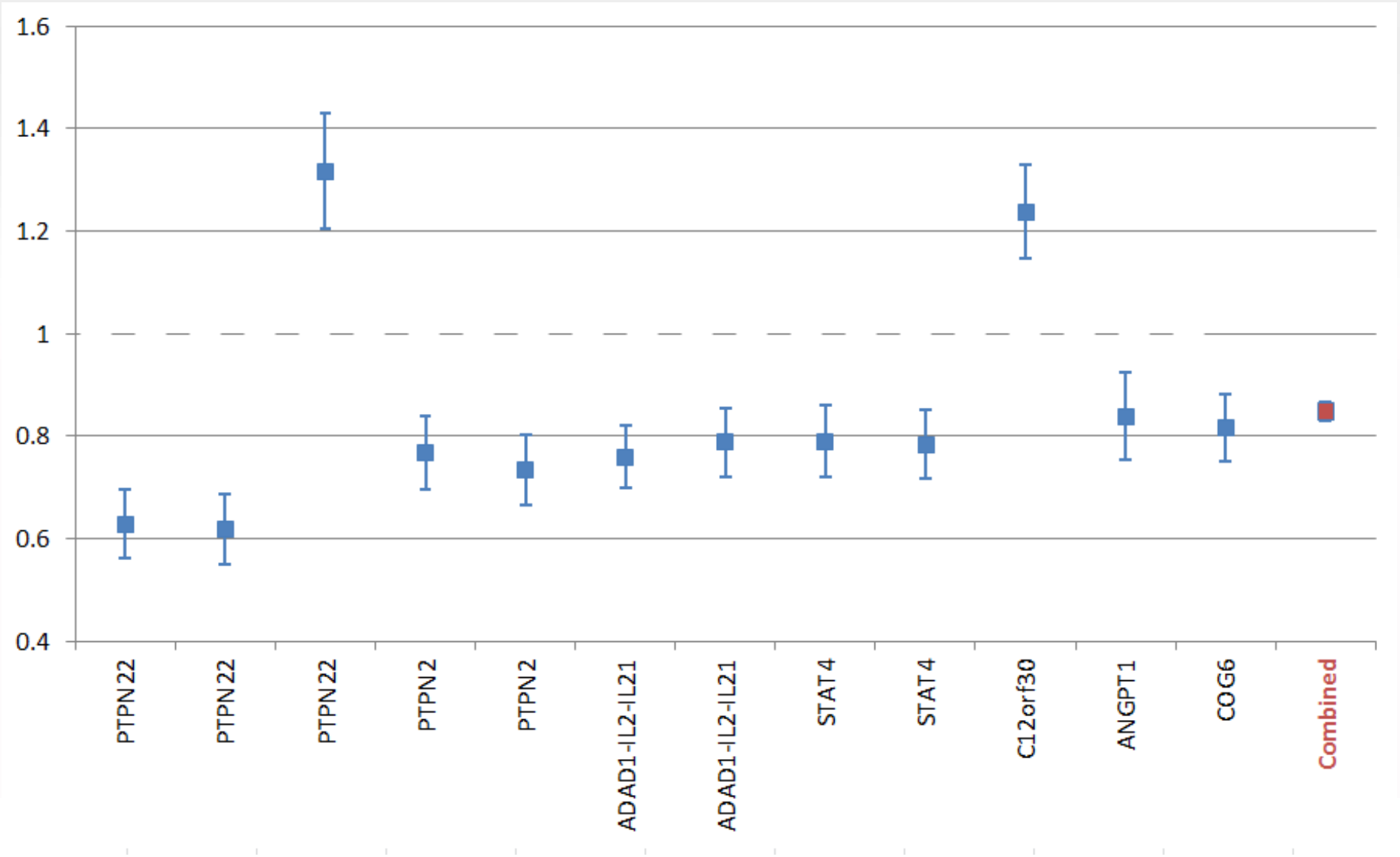


rs4272626 has R^2 of 1 with rs4453038 which is 537bp away

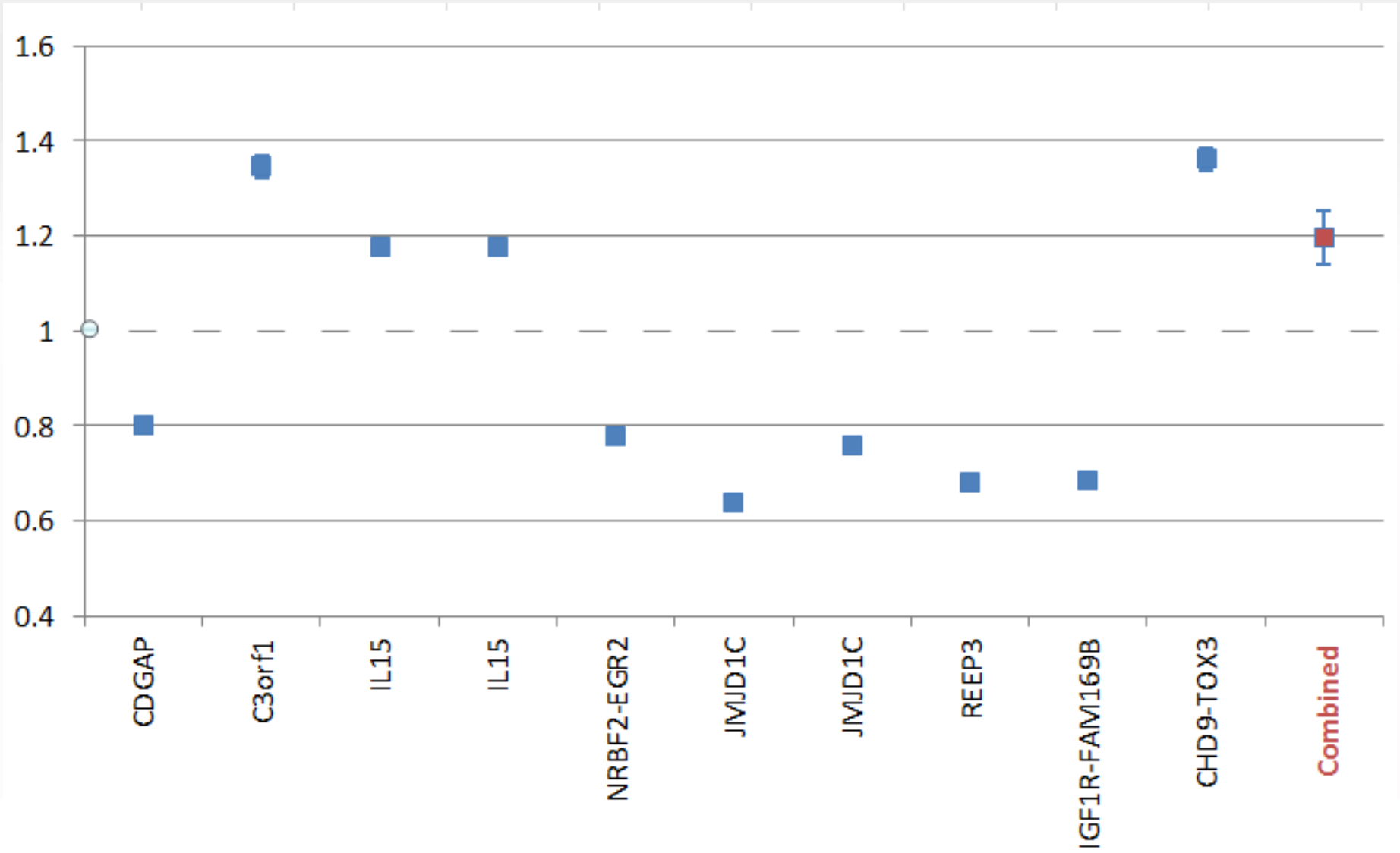
UK 2012


1.8
1.6
1.4
1.2
1
0.8
0.6
NHLH2
CD2
CD247
PTPRC
RBPJ
BLK
IL2RA
ORMDL3
PTPN2
Combined

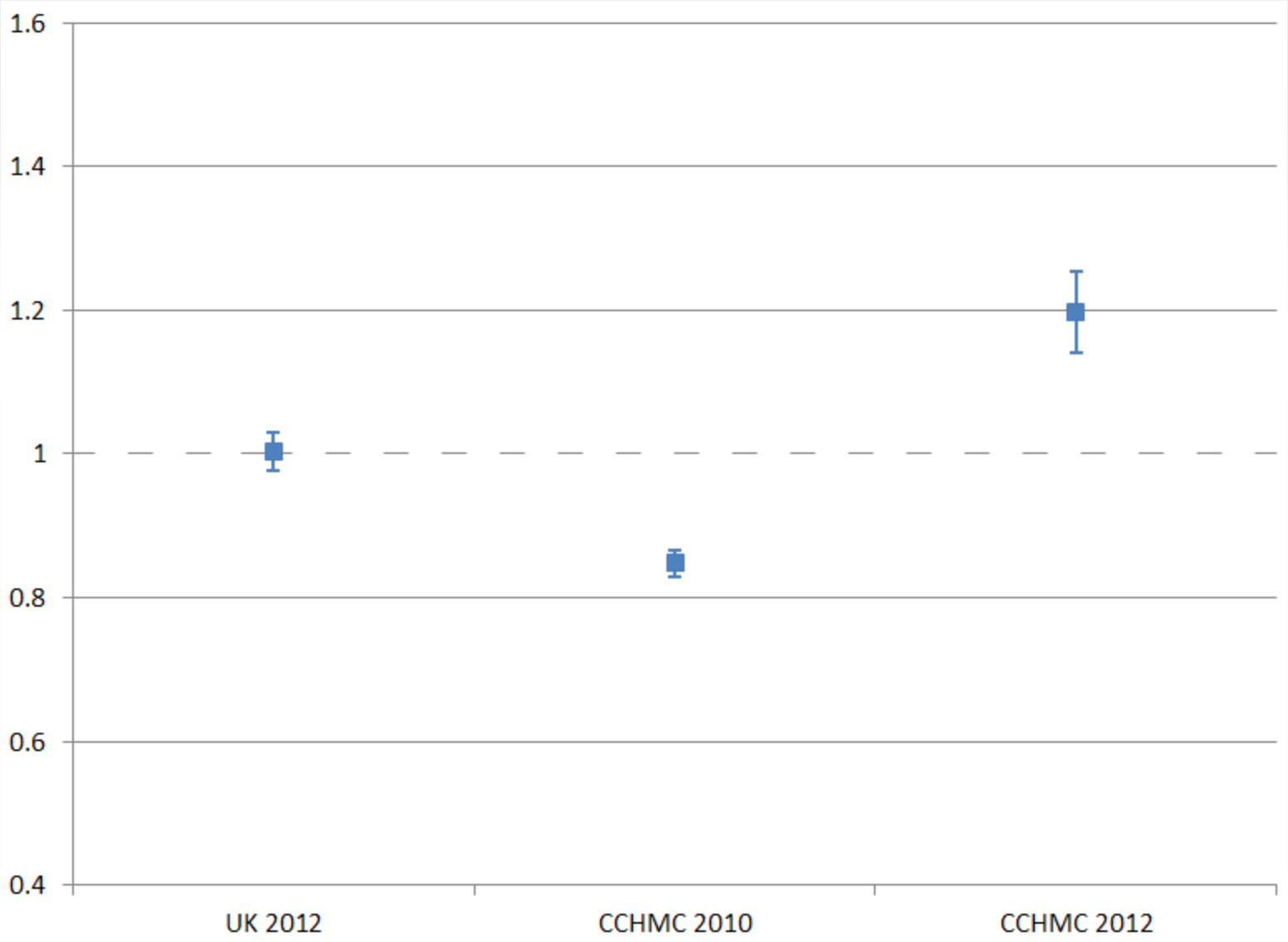
CCHMC 2010


1.6
1.4
1.2
1
0.8
0.6
0.4
PTPN22
PTPN22
PTPN22
PTPN2
PTPN2
ADAD1-IL2-IL21
ADAD1-IL2-IL21
STAT4
STAT4
C12orf30
ANGPT1
COG6
Combined

CCHMC 2012


1.6
1.4
1.2
1
0.8
0.6
0.4
CDGAP
C3orf1
IL15
IL15
NRBF2-EGR2
JMJD1C
JMJD1C
REEP3
IGF1R-FAM169B
CHD9-TOX3
Combined

Combined



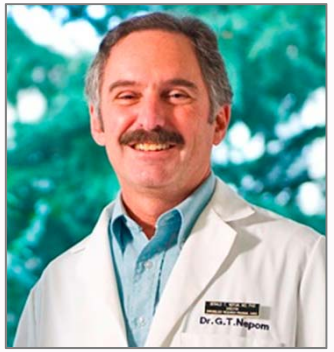
Final thoughts

- No smoking gun, but variants of interest
- Ongoing RA research with population level WGS and family NGS
- To be seen how much of autoimmune disorder heritability is explained by rare variants with higher effect sizes.
- Most promising signal is in genotype SNPs that might be tagging for functional mutations in regulatory regions.
- Rare sub-classifications like JIA polyarticular RF negative may be difficult to nail down with population studies
- Family studies looking at shared biomarkers along with symptoms may be better suited to find cause-effect relationships
- Wife's nuclear family has diagnosed cases of:
  - Pheochromocytoma: 2-8 per 1,000,000
  - Guillain-Barré: 0.6-4 per 100,000

Acknowledgements

- **Dr. Peter Gregersen**
  - Director, Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute
- **Dr. Gerald Nepom**
  - Director, Benaroya Research Institute, Director, Immune Tolerance Network
- **Golden Helix**
  - SNP & Variation Suite
  - GenomeBrowse
- **Sean Scott - Ingenuity**
  - Variant Analysis
- **Tim Hague - Omixon**
  - Target HLA Typing
- **Dr. Brian Naughton, 23andMe**
  - Trio exome sequencing

Dr. Peter Gregersen



Dr. Gerald Nepom