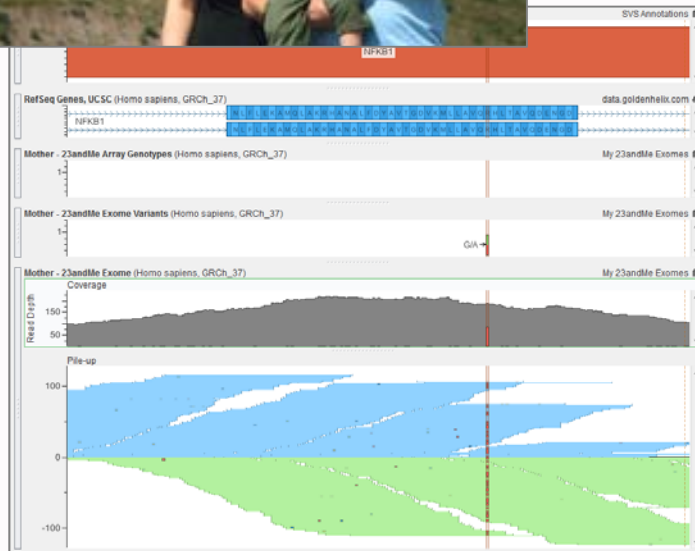




## The Quest for Meaningful Analysis Results of a 23andMe Exome Pilot Trio of Myself, Wife, and Son



February 22, 2013

Gabe Rudy, Vice President of Product Development

# Exome Sequencing in Consumer Genomics



Gabe  
(me)



Erin  
JIA



Ethan

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

Exome 80x - 23andMe

https://www.23andme.com/exome/

sign in register kit

Announced at Health 2.0, San Francisco - September 27, 2011

Be one of the first to get your personal exome sequence

**\$999** Enrollment Currently Closed

Sign up to be notified when ordering is available

email

**What is an exome? How is it different from a full genome sequence?**

Your exome is the 50 million DNA bases of your genome containing the information necessary to encode all your proteins. Informally, you can think of the exome as the DNA sequence of your genes.

Your entire genome is made up of your exome plus other DNA, consisting of three billion bases with repetitive sequences, sequences of unknown function, and DNA that does not code for proteins.

**Why sequence my exome?**

A person's physical structure, their body's chemical reactions and the expression of their genes are controlled by the proteins encoded in the exome. The vast majority of genetic diseases also hinge on variations in the exome. For these reasons, exome data may be useful for those exploring their personal sequence data.

Exome data are less suitable for ancestry or genealogical research, since they will not provide mitochondrial sequence or much information on the Y chromosome.

**How is this different from what 23andMe already offers?**

23andMe's current Personal Genome Service<sup>®</sup> (PGS) analyzes your DNA at approximately one million locations in the genome. The PGS<sup>®</sup> provides more than 200 detailed reports linking different genetic variants to health conditions, traits, and ancestry, as well as connecting people to other users who share DNA.

In contrast, the exome sequencing pilot provides users with raw variant data for about 50 million bases of DNA, without reports. Over time, 23andMe will add a limited set of tools and content that utilize exome sequence data.

**Who can take part?**

We are offering access to this pilot exclusively to current 23andMe customers. The exome sequencing pilot is the first of its kind, and it is suitable for customers who are comfortable managing and understanding raw genetic data. If you don't know your exons from your introns, this pilot is probably not for you. This is for early adopters and supplies are limited.

**What do I get for \$999?**

You get access to your raw data of 50 million DNA bases at high quality (80X coverage). Over time, you will have access to new tools and content as they are developed to take advantage of your exome sequence data. Most excitingly, you'll be a trailblazer, one of the first people on the planet to know their personal exome sequence!

**When does the project start and how do I join?**

At this time, the pilot project is full and enrollment is closed. If you are interested in receiving updates about any future sequencing services, please enter your email above. We will contact you about future opportunities.



**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



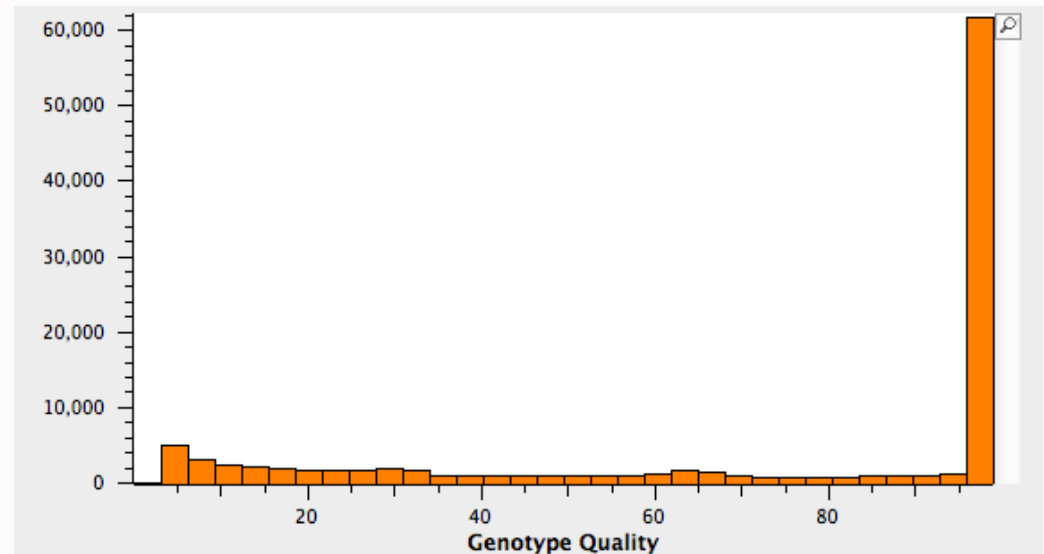
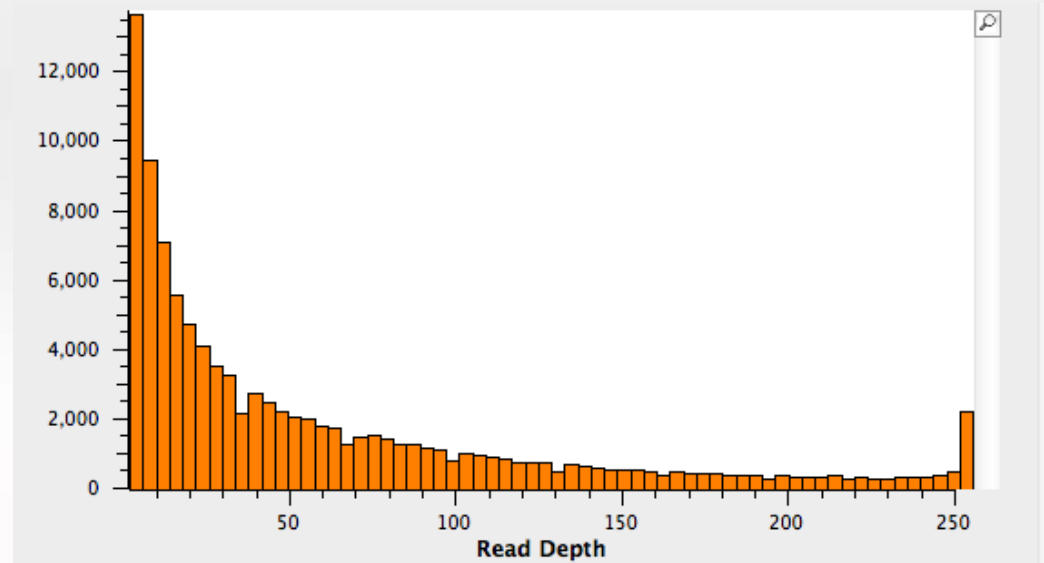
- **Sequencing done on HiSeq 2000**
  - 75bp PE
  - Agilent SureSelect exome capture
- **Aligned and called with BWA/GATK**
  - Broad's Best Practices with GATK Guide
  - Indel realignment
  - UnifiedGenotyper called samples concurrently
- **Deliverables**
  - BAM (minus indel realignments)
  - VCF (some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



# Research or clinical grade?



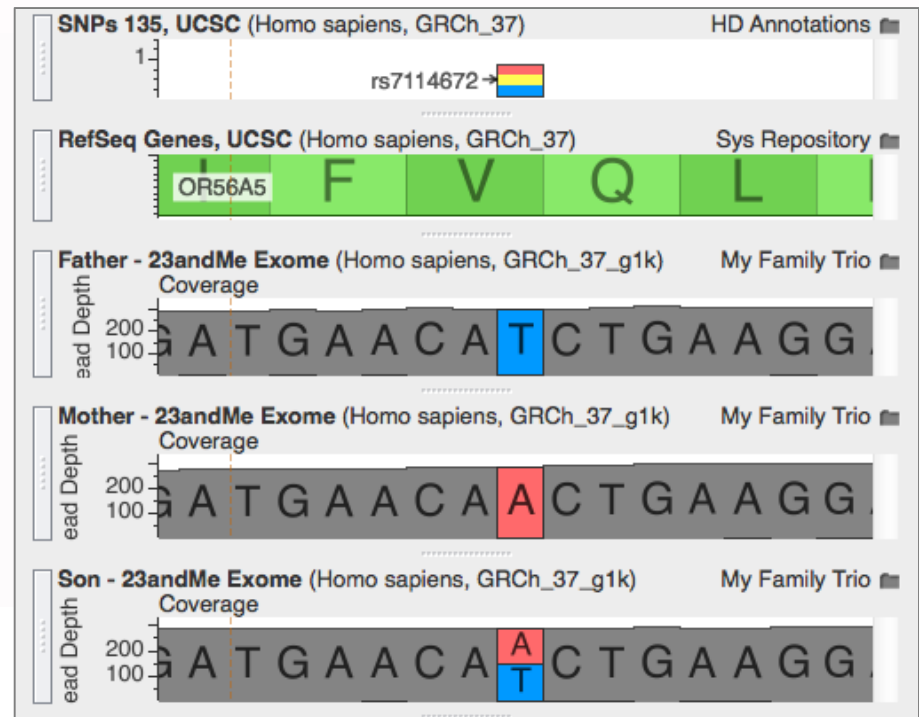
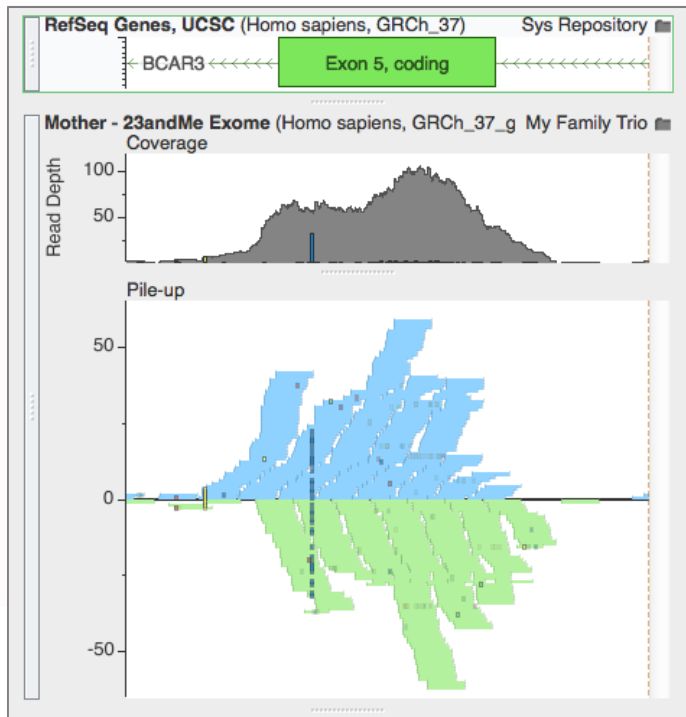
<b>Total Reads</b>	140M
<b>Unique Align</b>	87%
<b>Mean Target</b>	105x
<b>% Target at 2x</b>	97%
<b>% Target at 10x</b>	94%
<b>% Target at 20x</b>	89%
<b>% Target at 30x</b>	83%



# Clinical grade



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



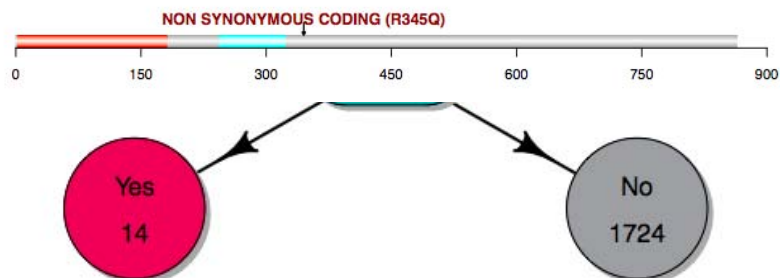
# Summary Report



Exome  
106762

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

PFAM (or SMART) domains for gene ERCC6, transcript ENST00000542458:  
■ PF00176: SNF2\_N  
■ PF00271: Helicase\_C

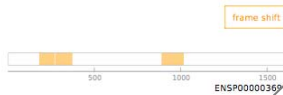


# Updated VCF and report at end of October

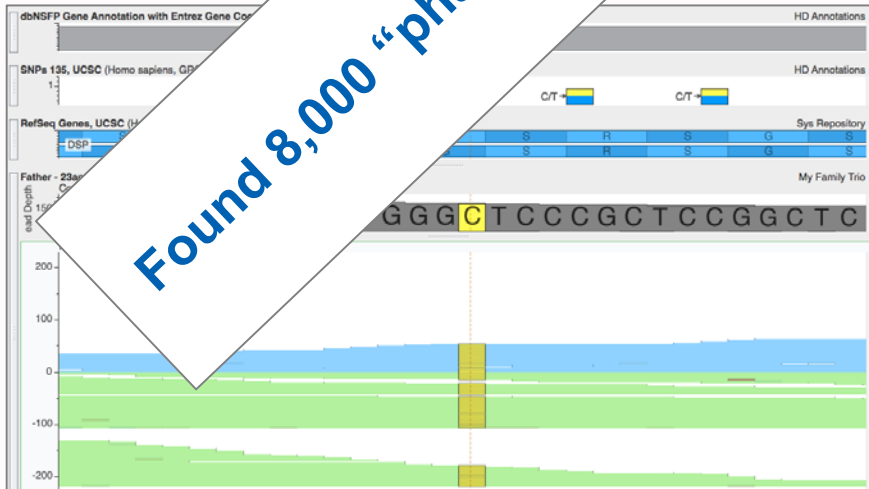


## List of selected variants

Variant 1:	Gene: <b>DSP</b> Your genotype: <b>GC/GC</b> Location: chr6:7585967
Effect:	<b>FRAME SHIFT</b> Type: HIGH
Frequency:	1KGenomes: NA dbSNP
Quality:	Genotype quality: 99.00
Details:	Gene description: desmoplakin Transcript: <a href="#">ENST00000379802</a> EntrezId: 1832 UniProt: <a href="#">P15924</a>



Found 8,000 "phantom" variants



## GATK is a Research Tool. Clinics Beware.

@gabeinformatics

AN "OUR 2 SNPs..." BLOG BY GOLDEN HELIX

Home Authors @gabeinformatics

— Upcoming Webcast: 23andMe Variant Analysis of My Personal Exome

### GATK is a Research Tool. Clinics Beware.

Posted on December 3, 2012 by Gabe Rudy

[t](#) [f](#) [s](#)

In preparation for a webcast I'll be giving on Wednesday on my own exome, I've been spending more time with variant callers and the myriad of false-positives one has to wade through to get to interesting, or potentially significant, variants.

So recently, I was happy to see a message in my inbox from the 23andMe exome team saying they had been continuing to work on improving their exome analysis and that a "final" analysis was now ready to download.

This meant I had both an updated "variants of interest" report as well as updated variant calls in a new VCF file. I'll get to the report in a second, which lists rare or novel variants in clinically associated genes, but first let's look at what changed in the variant calls.

**New... and Improved?**

At first blush, the variant files look quite different as you can see in the first Venn diagram below comparing variants (both unique and common) between the original and new files. But after I applied a conservative filter (Genotype Quality (GQ) > 10 and Read Depth (RD) > 10) on the variants, things start to look less dramatic. So what is with all the new variants? It looks like many are just more aggressive variant calls. In fact there were ten thousand variants with a read depth of just one (a single read) in the new file!

Category	Count
Original VCF File (unique)	106,760
Final VCF File (unique)	152,205
Overlap	105,786
Original only	974
Final only	46,419

**Follow...**

[Subscribe](#)

**Categories**

- About GHI (4)
- Add-on scripts & data repository (9)
- Assessment of new methods (4)
- Best practices in genetic analysis (9)
- Big picture (17)
- Bioinformatic support (7)
- Clinical Genetics (2)
- Customer success (15)
- General statistical genetics principles (5)
- How to's and advanced workflows (11)
- News, events, & announcements (25)
- Paper review (2)
- Personal genomics (5)
- Plant & Animal (4)
- Technology review (19)

**Recent Posts**





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*

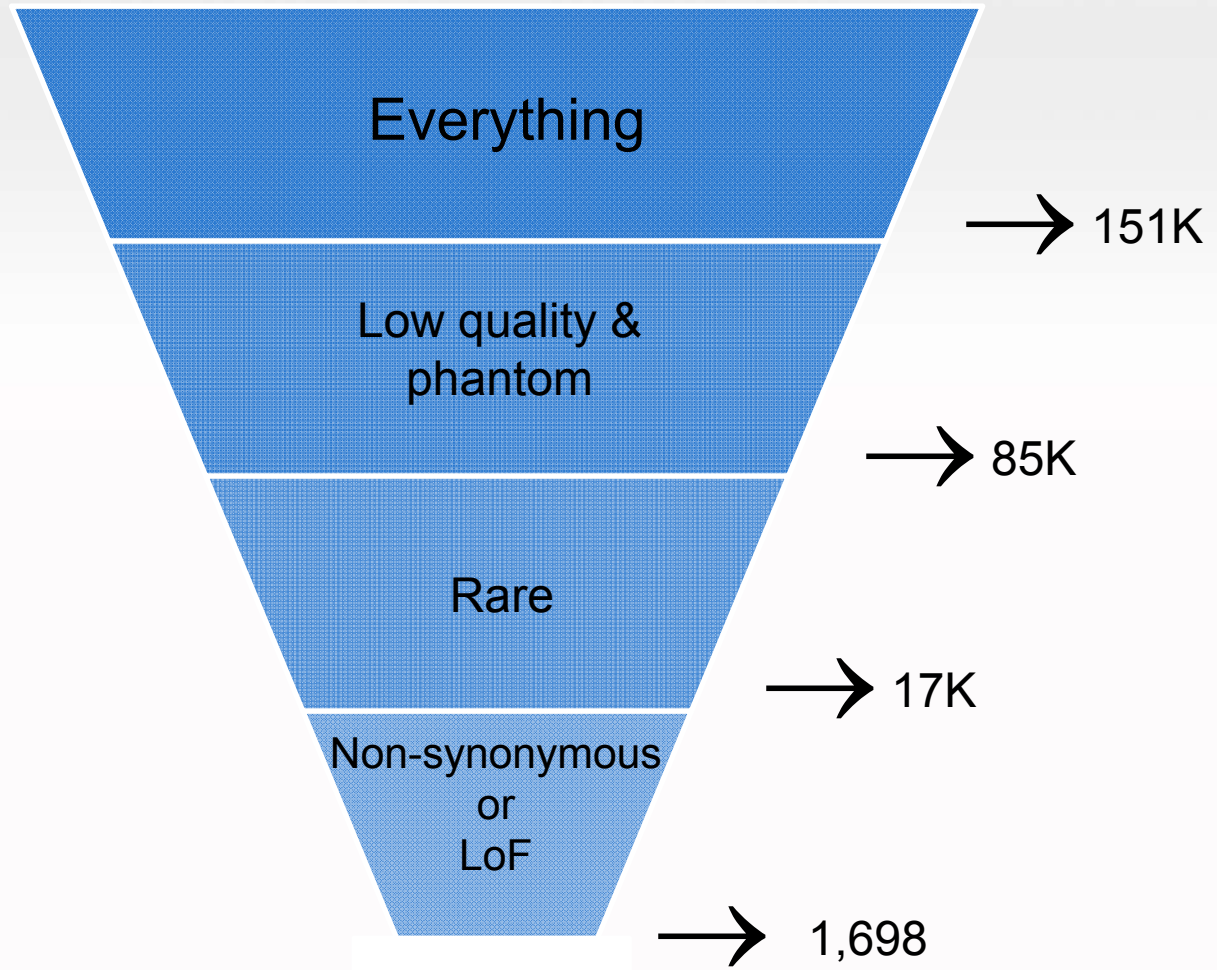
*My Exome:  
The Nail*



# 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**



- 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



		Rare	!Dups	OMIM	Rec Genes
Loss of Function	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)

		Rare	!Dups	OMIM	Rec Genes
Non-Synonymous	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_1514delACCACT GGCCA

ChrX: 38,226,599 - 38,226,628

Homo sapiens (Human), GRCh37 hg19 (Feb 2009)

Plot Tree

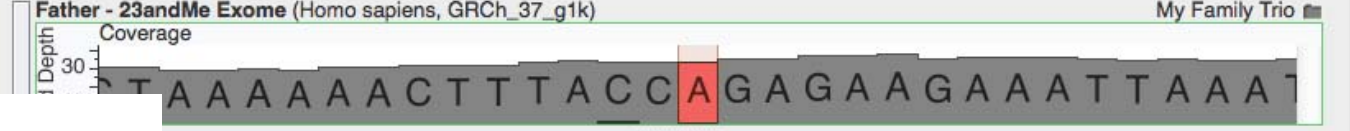
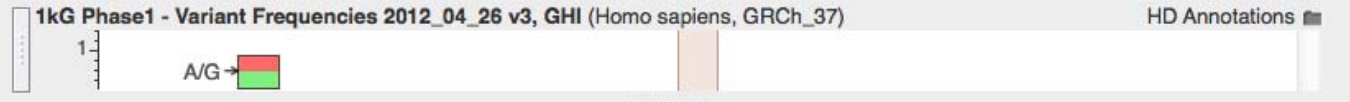
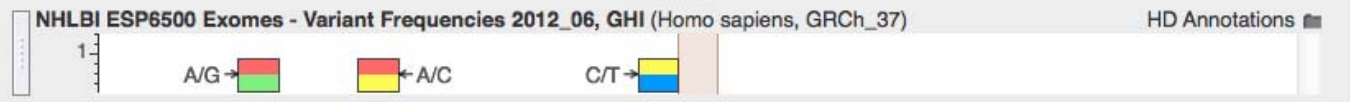
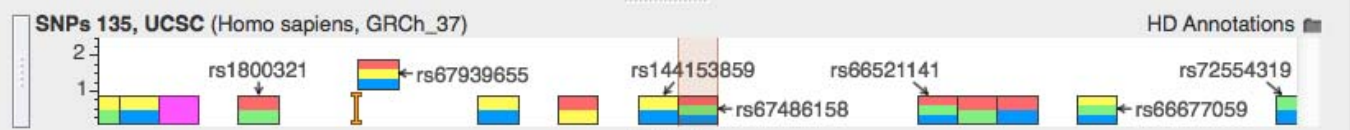
- dbNSFP Gene Annotation ...
- SNPs 135, UCSC (Homo s...
- NHLBI ESP6500 Exomes - ...
- 1kG Phase1 - Variant Fre...
- RefSeq Genes, UCSC (Ho...
- Mother - 23andMe Exom...
- genotype - Erin Genotyp...
- Father - 23andMe Exome...
- Coverage
- Pile-up
- Mother - 23andMe Exom...

Console

History Clear

1kG Phase1 - Variant Frequencies 2012\_04\_26 v3, GHI (Homo sapiens, GRCh\_37)

Variant Sources  
Type: Variant



**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:**



**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



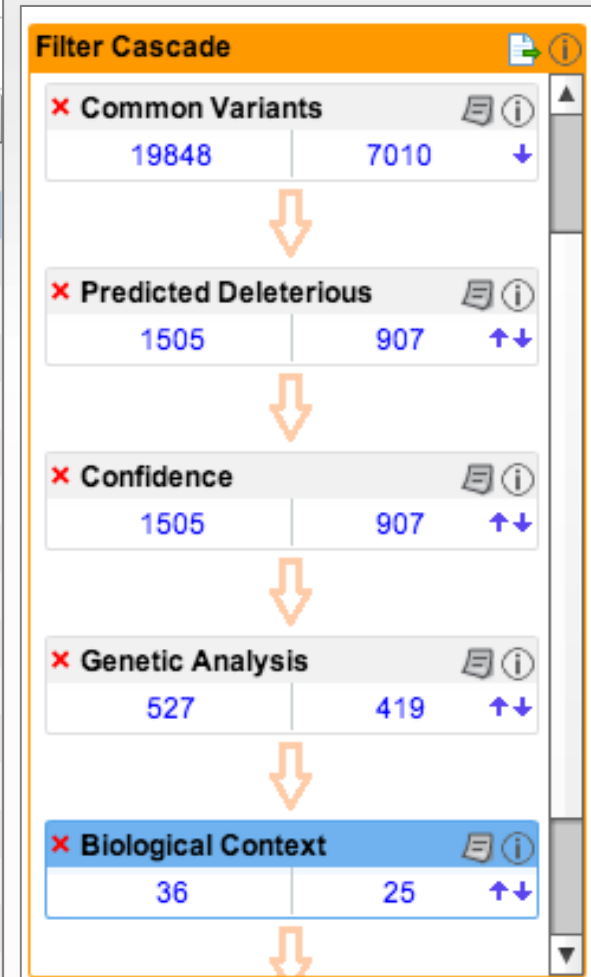
# IVA – Rare deleterious variance within 1 hop of JIA



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



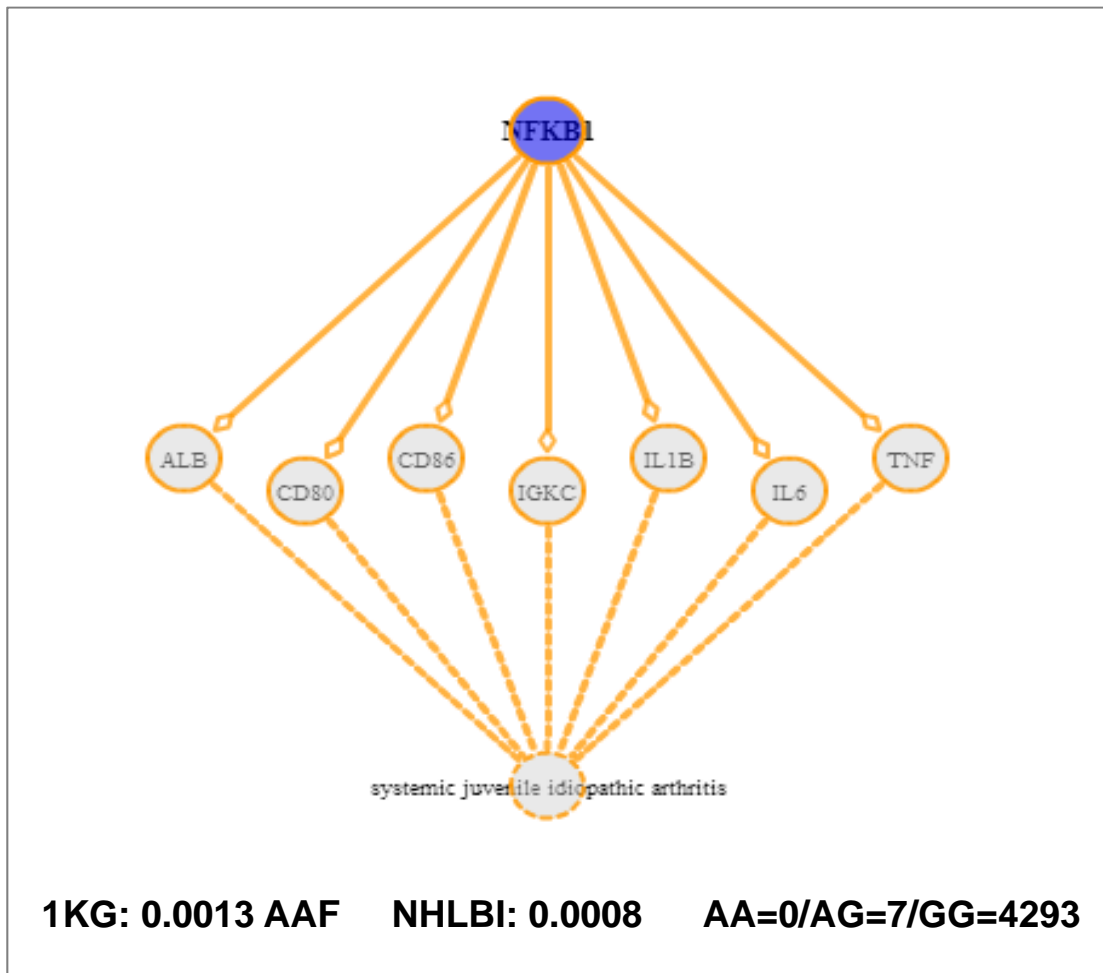
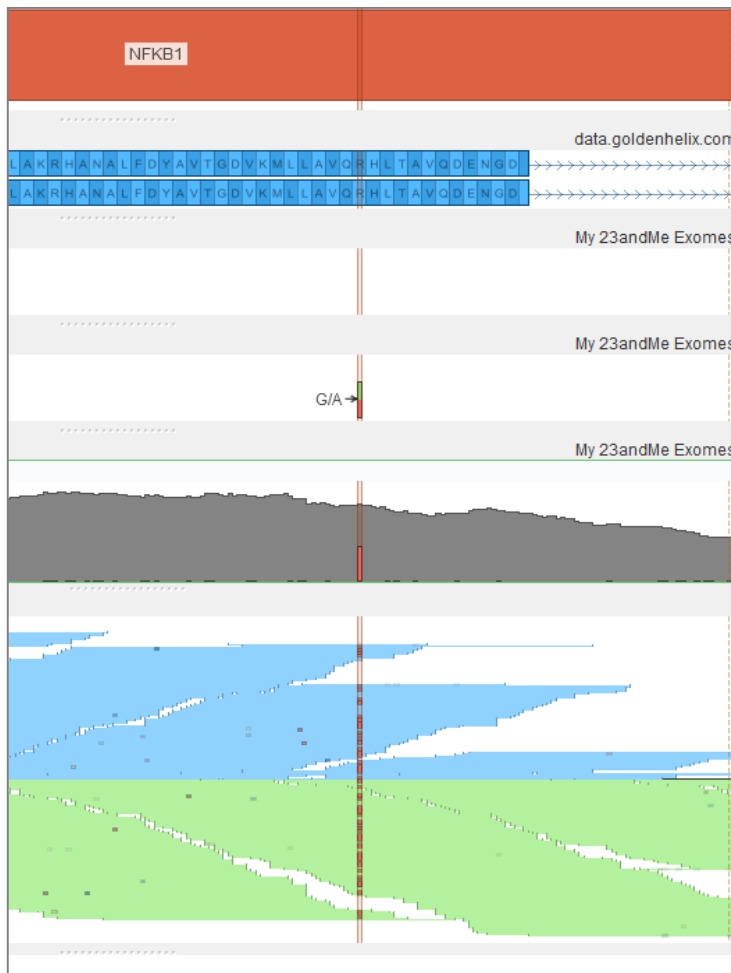
**Details** | [Path to Phenotype](#) | [Haploinsufficiency](#) | [Diseases](#)

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 Infarction 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 2012

## 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

Chr6: 32,549,341

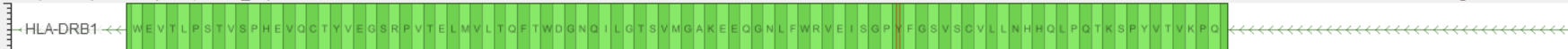
Chr6: 32,549,441

Chr6: 32,549,541

Chr6: 32,549,641

RefSeq Genes, UCSC (Homo sapiens, GRCh\_37)

data.goldenhelix.com



NHLBI ESP6500 Exomes - Variant Frequencies 2012\_06, GHI (Homo sapiens, GRCh\_37)

SVS Annotations



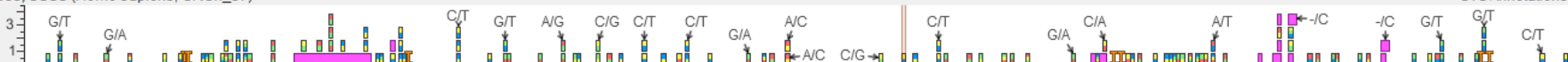
1kG Phase1 - Variant Frequencies 2012\_04\_26 v3, GHI (Homo sapiens, GRCh\_37)

SVS Annotations



SNPs 135, UCSC (Homo sapiens, GRCh\_37)

SVS Annotations



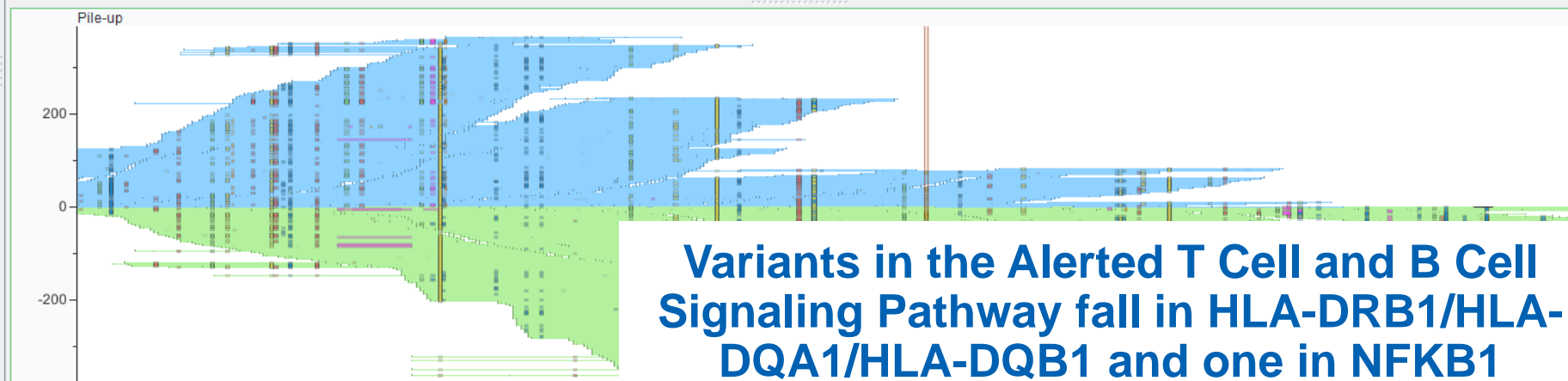
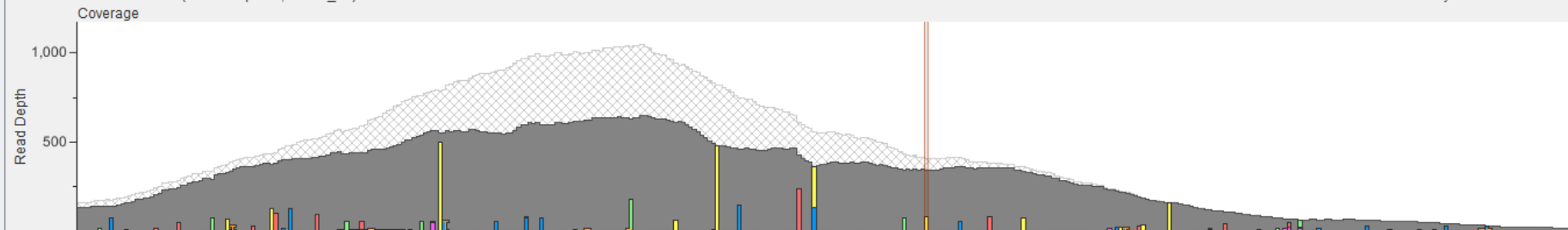
Mother - 23andMe Exome Variants (Homo sapiens, GRCh\_37)

My 23andMe Exomes



Mother - 23andMe Exome (Homo sapiens, GRCh\_37)

My 23andMe Exomes



**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

Paired result

Displaying 32 loci out of 32.

Setup Loci | Assign Pair(s) | Unassign Pair(s) | Export Result(s)

View and assign HLA typing results by selecting one or more pairs together.

Total coverage (%) - Sum of all the bases of the aligned reads relative to the best match.

Score by coverage (%) - Score by total pair coverage relative to the best match.

	Allele 1	Allele 2	Total coverage (%)	Score by coverage (%)
HLA-DOA	<input checked="" type="checkbox"/> HLA-DRB1*04:07:01	<input checked="" type="checkbox"/> HLA-DRB1*14:54:01	100	100
HLA-DOB	<input checked="" type="checkbox"/> HLA-DRB1*04:07:01	<input checked="" type="checkbox"/> HLA-DRB1*04:07:01	50.784	68.834
HLA-DPA1	<input checked="" type="checkbox"/> HLA-DRB1*14:54:01	<input checked="" type="checkbox"/> HLA-DRB1*14:54:01	49.265	65.412
HLA-DPB1				

## HLA-A

A\*24:02:01:01

A\*03:01:01:01

## HLA-DRB1

DRB1\*04:07:01

DRB1\*14:54:01

## HLA-DQA1

DQA1\*01:04:01

DQA1\*03:03:01

## HLA-DPB1

DPB1\*04:01:01

DPB1\*04:01:01

# Predicting Lifetime Risk from Population Studies



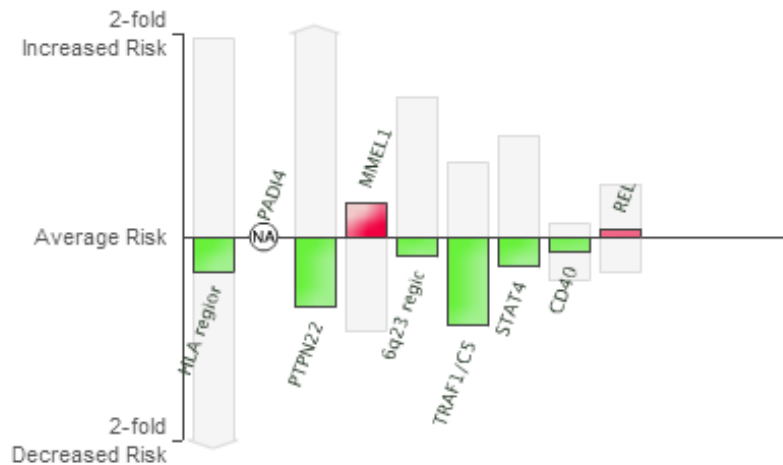
## disease risk

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.



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.

Cincinnati

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.

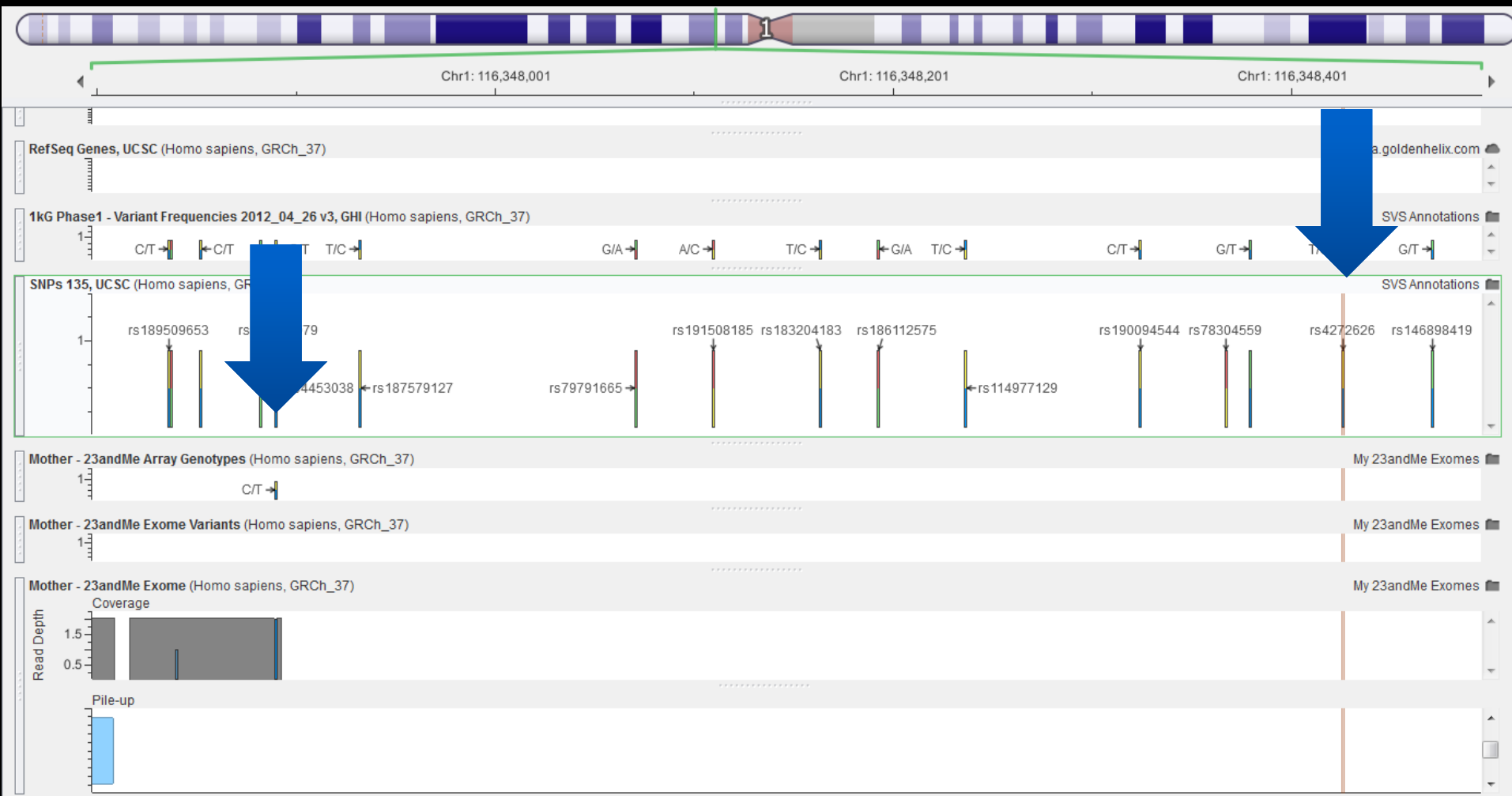
## **Investigation of rheumatoid arthritis susceptibility loci in juvenile idiopathic arthritis confirms high degree of overlap.**

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.

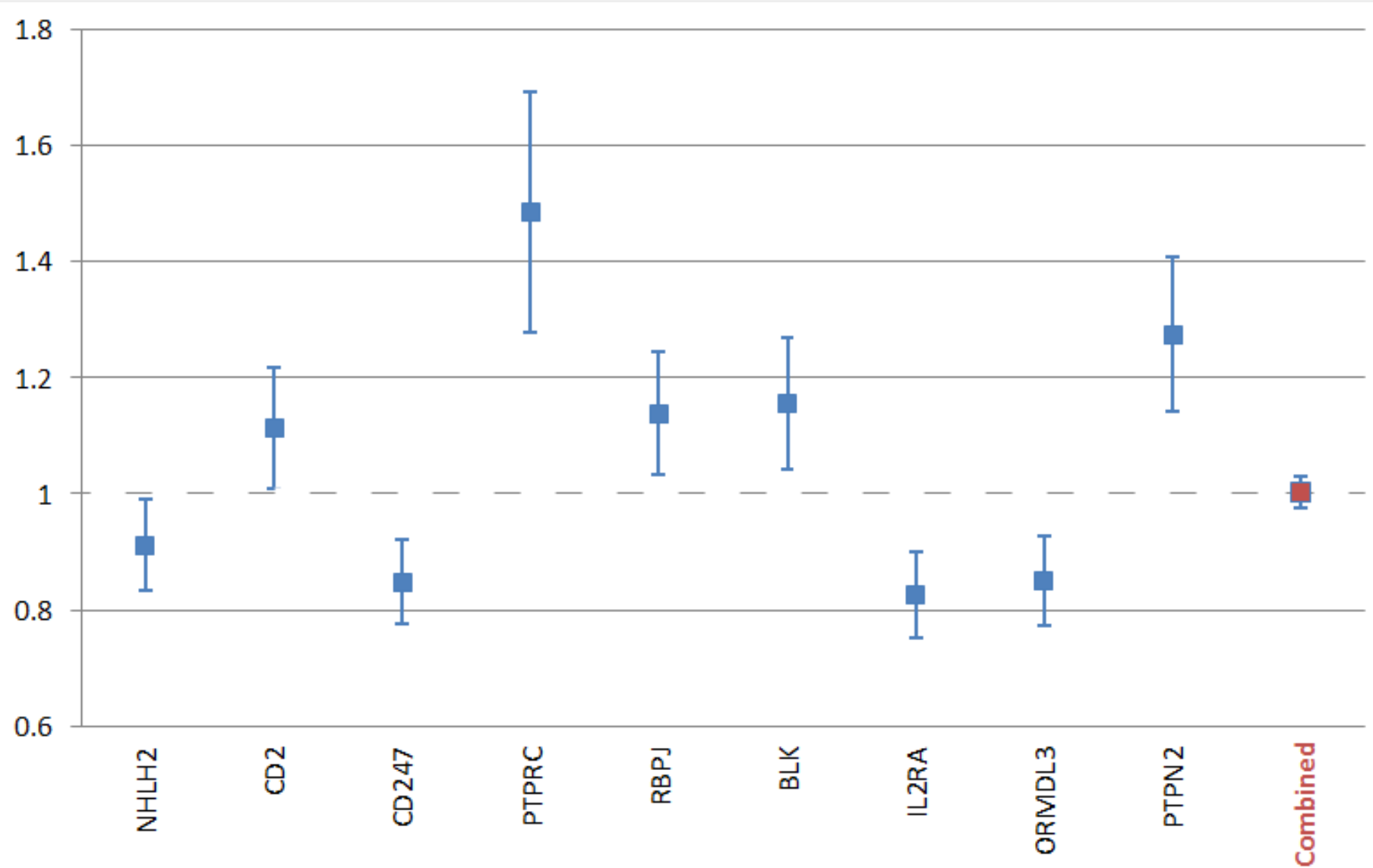
### **+ Collaborators (61)**

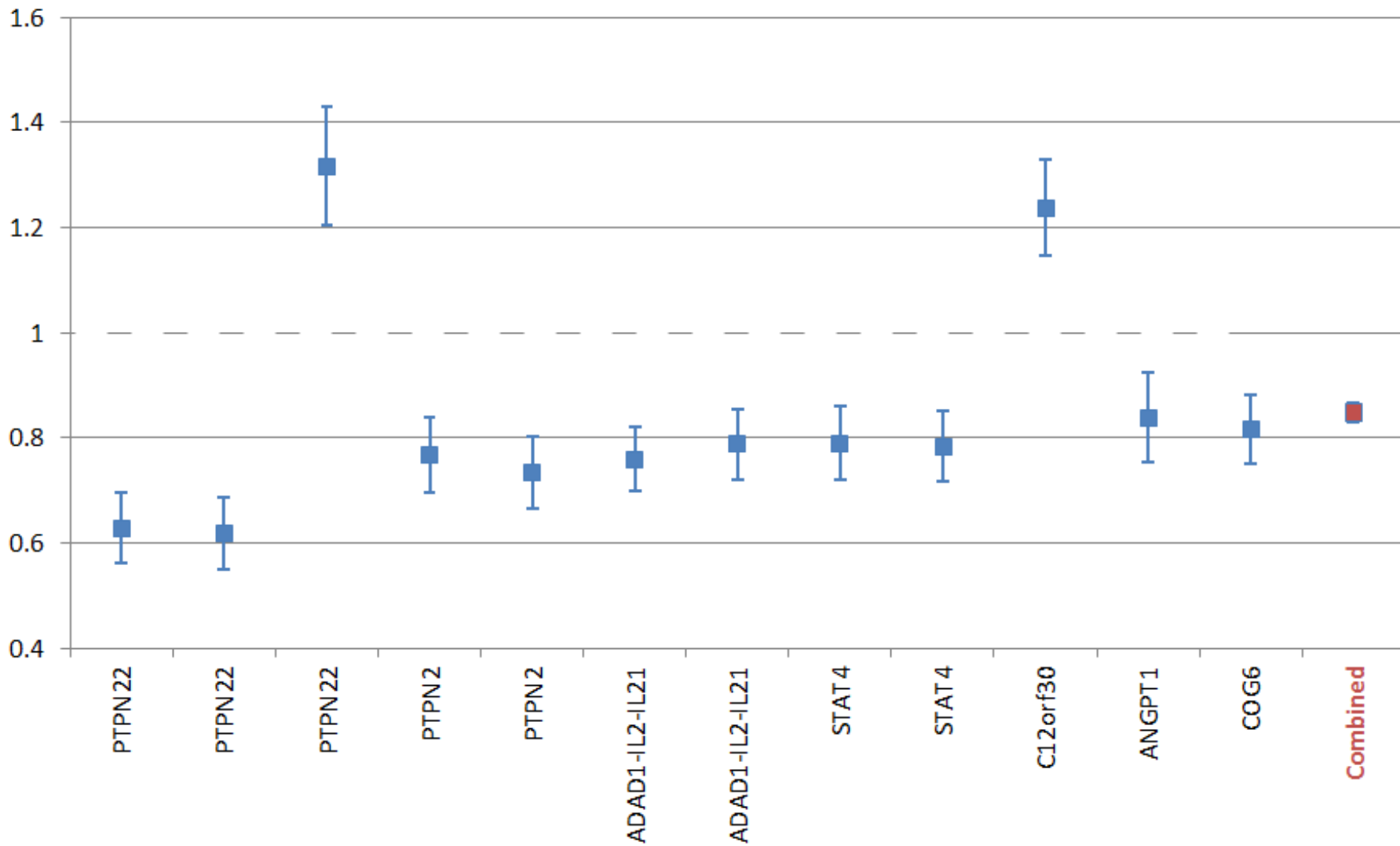
Arthritis Research UK Epidemiology Unit, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK. [anne.hinks@manchester.ac.uk](mailto:anne.hinks@manchester.ac.uk)

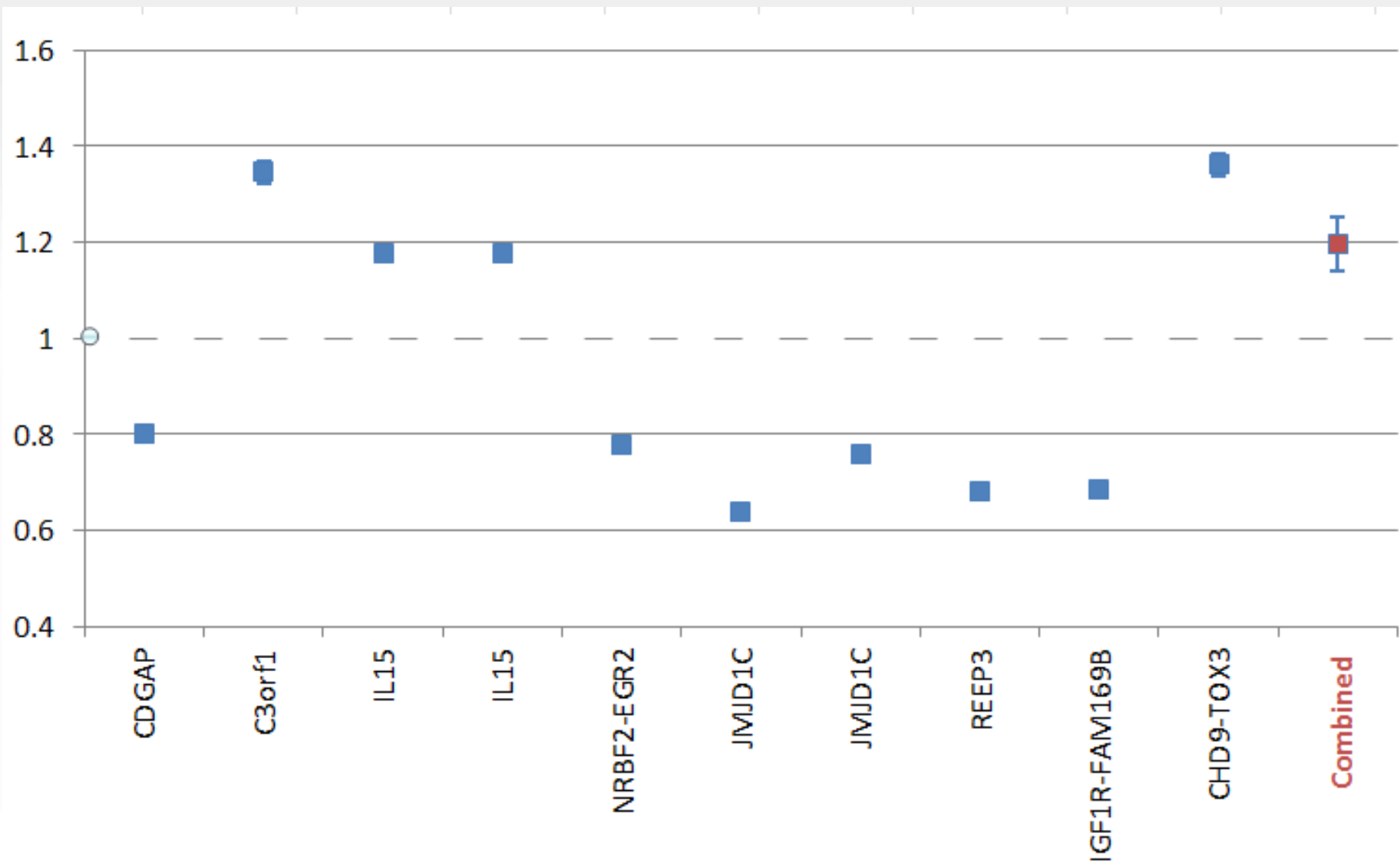




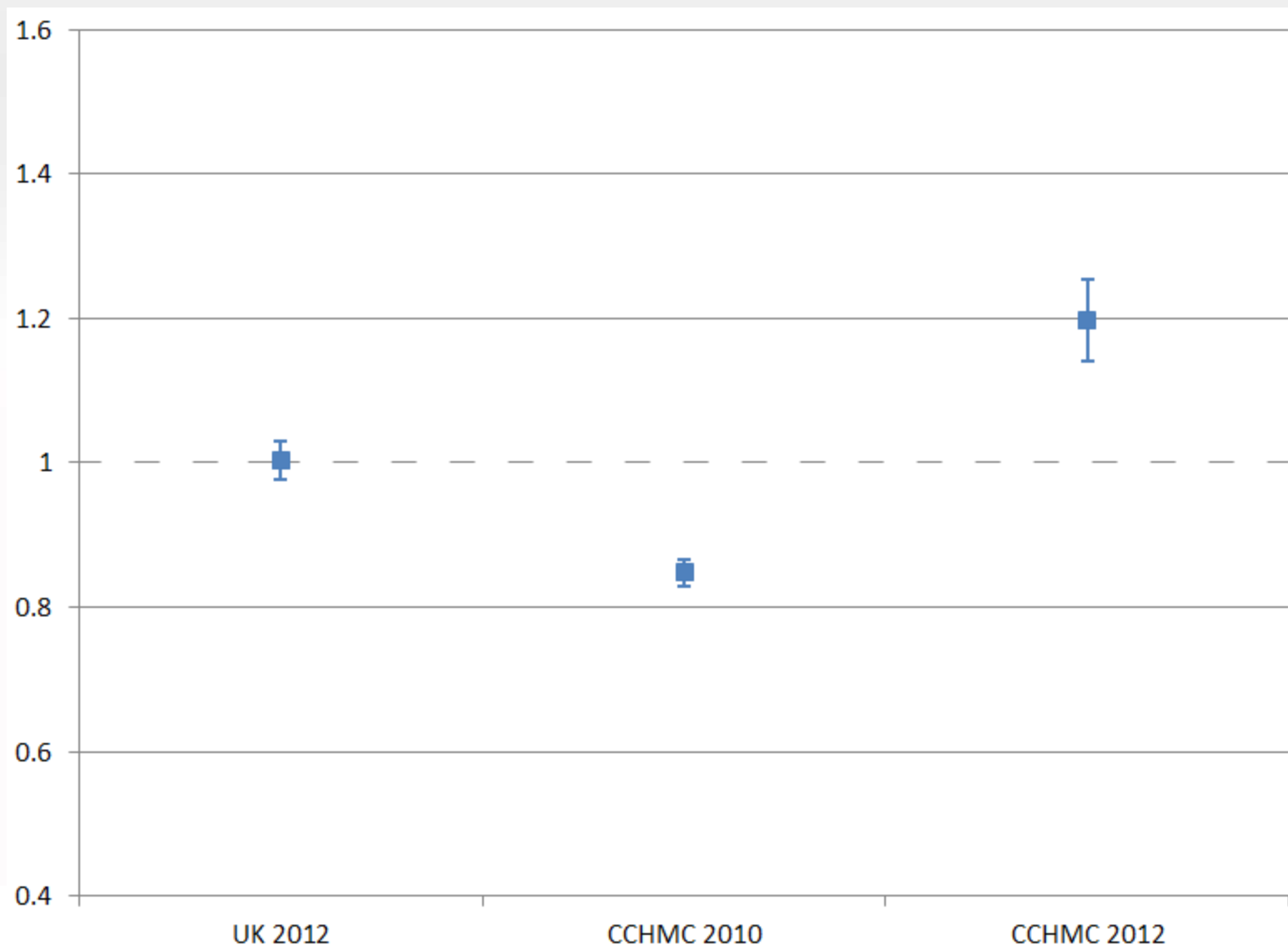
**rs4272626 has  $R^2$  of 1 with  
rs4453038 which is 537bp away**







# 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

SNP & VARIATION SUITE 



- **Sean Scott - Ingenuity**

- Variant Analysis



- **Tim Hague - Omixon**

- Target HLA Typing

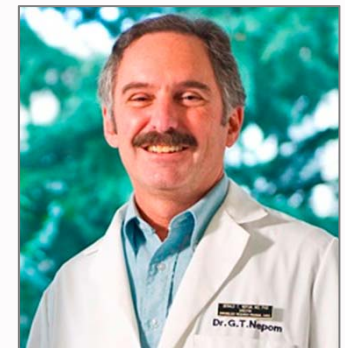


- **Dr. Brian Naughton, 23andMe**

- Trio exome sequencing



*Dr. Peter Gregersen*



*Dr. Gerald Nepom*