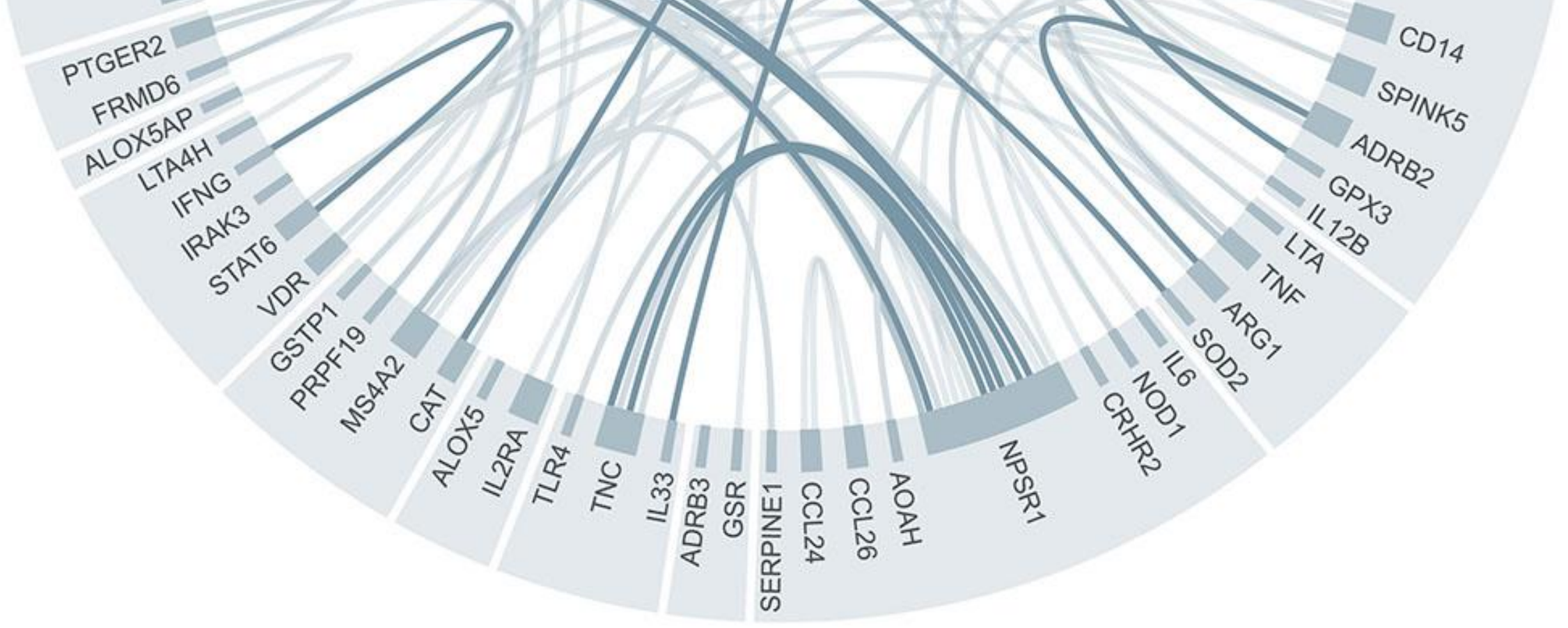


Confronting the Missing Epistasis Problem: On the Reproducibility of Gene-Gene Interactions

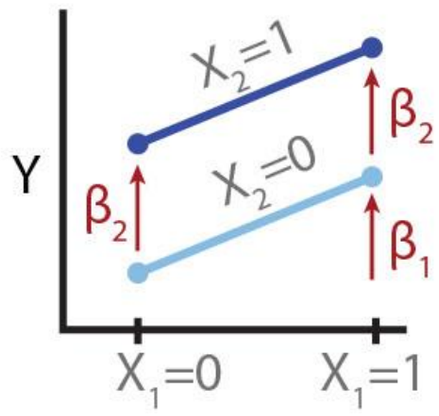
William Murk

Ph.D. Candidate

Yale School of Public Health

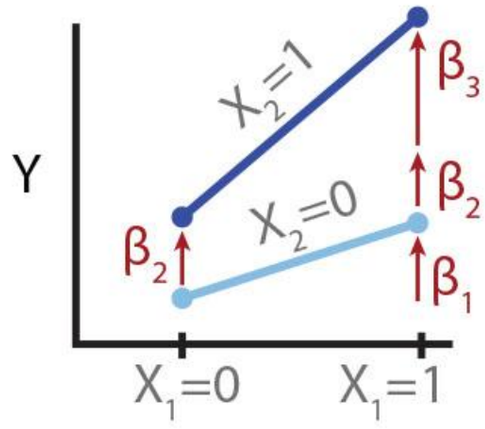
Statistical interaction

Additive

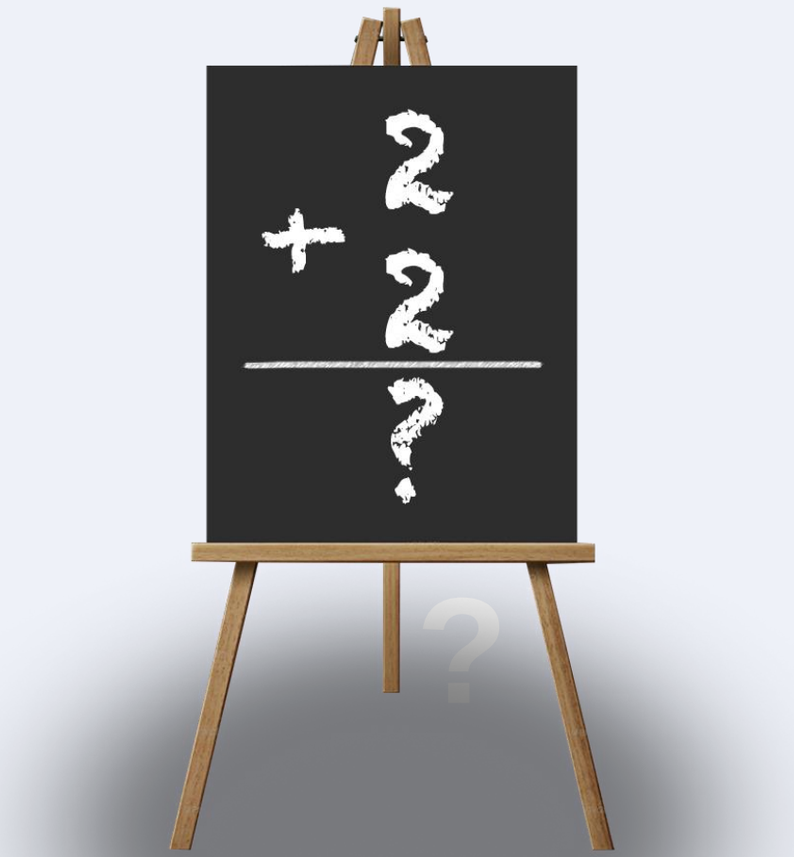


$$Y(0,0) = 0$$
$$Y(1,0) = \beta_1$$
$$Y(0,1) = \beta_2$$
$$Y(1,1) = \beta_1 + \beta_2$$
$$= Y(1,0) + Y(0,1)$$

Non-additive



$$Y(0,0) = 0$$
$$Y(1,0) = \beta_1$$
$$Y(0,1) = \beta_2$$
$$Y(1,1) = \beta_1+\beta_2+\beta_3$$
$$\neq Y(1,0) + Y(0,1)$$



$$Y = \text{Intercept} + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

$\beta_3 = 0 \longrightarrow$ No interaction

$\beta_3 \neq 0 \longrightarrow$ Interaction

Why study interactions?

PNAS

Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis

Haifeng Shao[a,b,1], Lindsay C. Burrage[a,b,1], David S. Sinasac[a,1], Annie E. Hill[a], Sheila R. Ernest[a], William O'Brien[c], Hayden-William Courtland[d], Karl J. Jepsen[d], Andrew Kirby[e], E. J. Kulbokas[e], Mark J. Daly[e,f], Karl W. Broman[g], Eric S. Lander[f,h,i,2,3], and Joseph H. Nadeau[a,b,j,k,2,3]

[a]Department of Genetics and [i]Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH 44106; [b]Center for Proteomics and Bioinformatics and [k]Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106; [c]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; [d]Leni and Peter W. May Department of Orthopaedics, Mount Sinai School of Medicine, New York, NY 10029; [e]Center for Human Genetics Research, Massachusetts General Hospital Simches Research Center, Boston, MA 02114; [f]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142; [g]Department of Biostatistics and

LETTER

doi:10.1038/nature11510

Epistasis as the primary factor in molecular evolution

Michael S. Breen[1], Carsten Kemena[1], Peter K. Vlasov[1], Cedric Notredame[1] & Fyodor A. Kondrashov[1,2]

25 OCTOBER 2012 | VOL 490 | NATURE | 535

The main forces directing long-term molecular evolution remain obscure. A sizable fraction of amino-acid substitutions seem to be fixed by positive selection[1–4], but it is unclear to what degree long-

adaptation, such as a housekeeping protein, then the same amino-acid state should be acceptable in an orthologous site in a different species. However, if epistasis is common then amino-acid substitutions that

PNAS

The mystery of missing heritability: Genetic interactions create phantom heritability

Or Zuk[a], Eliana Hechter[a], Shamil R. Sunyaev[a,b], and Eric S. Lander[a,1]

Broad Institute of MIT and Harvard, Cambridge, MA 02142; and [b]Genetics Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

Contributed by Eric S. Lander, December 5, 2011 (sent for review October 9, 2011)

The curse of dimensionality

Order:

k = 2

No. SNPs:	10^4	10^5	10^6
No. Interactions:	10^7	10^9	10^{11}
Time Required:	83 min	5.8 d	1.6 yr
Bonferroni threshold	10^{-9}	10^{-11}	10^{-13}

k = 3

No. SNPs:	10^4	10^5	10^6
No. Interactions:	10^{11}	10^{14}	10^{17}
Time Required:	192.8 d	528 yr	5.3×10^5 yr
Bonferroni threshold	10^{-13}	10^{-16}	10^{-19}

Assuming 10,000 interactions analyzed per second

Aim:

To evaluate the robustness and reproducibility of reported gene-gene interactions in asthma

Study design

Literature search

Identify all gene-gene interactions in asthma

Assess “robustness” of original reports
- Replication?
- Direction of effect described?
- Explicit test for interaction?

Replication in independent data

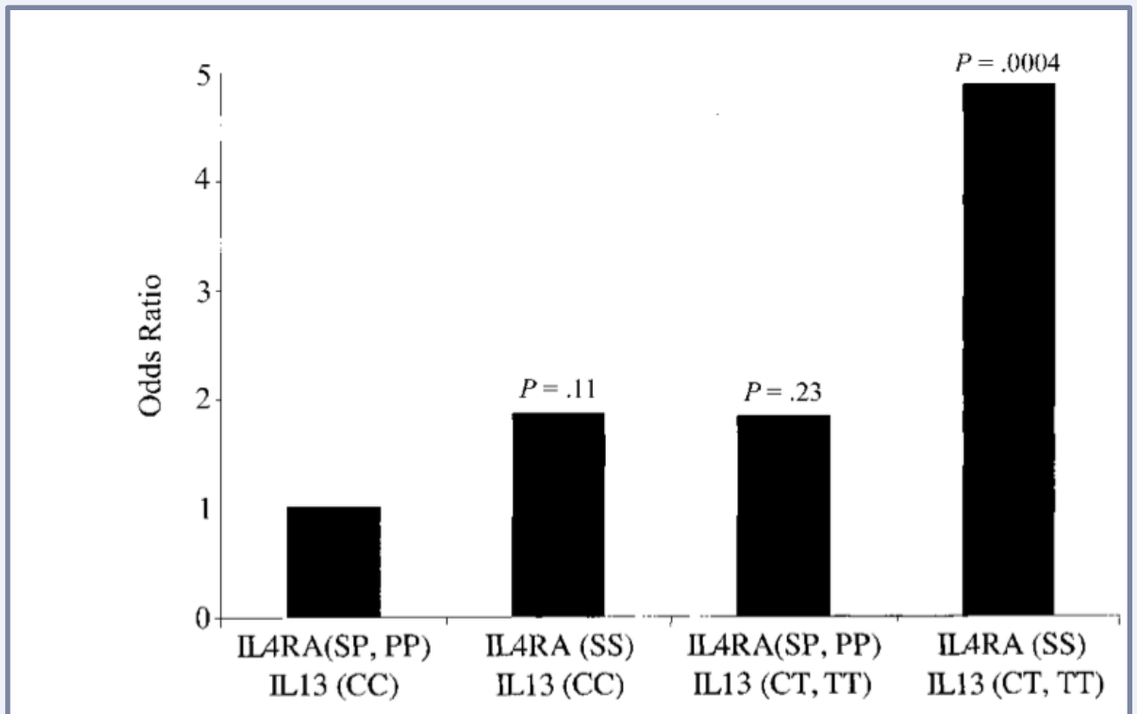
Explicit vs. non-explicit tests for interaction

Explicit :

$$\ln(p/(1-p)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

Non-explicit:

$$\ln(p/(1-p)) = \beta_0 + \beta_3 X_1 X_2$$



No. of cases	16	62	16	46
No. of controls	22	46	12	13

Figure 1 Interaction of *IL4RA* and *IL13* genotypes. Bars indicate the ORs between the different combinations of genotypes for *IL4RA* (S478P) and *IL13* (−1111 C/T). The nonrisk genotype for each gene was used as the reference OR.

Howard et al. Am J Hum Genet 2002. 70: 230-236.

Literature Search

(asthma OR asthmatic)

AND

(gene OR genes OR genetic OR locus OR loci OR genome OR genomic OR chromosome OR chromosomes OR genomewide OR "genome wide" OR GWAS OR polymorphism OR polymorphisms OR SNP OR SNPs)

AND

(epistasis OR epistatic OR interact OR interacts OR interaction OR interactions OR multilocus OR "multi locus" OR additive OR multiplicative OR MDR OR multifactor OR "multi factor" OR forest OR forests OR combinatorial OR partition OR partitions OR stratified OR strata OR stratum OR stratification OR tree OR trees OR cart OR polygenic OR multigenic OR multimarker)

From Agostino Ramelli, Le Diverse et Artifiose Machine (1588). Obtained from:
https://commons.wikimedia.org/wiki/File:Le_diverse_et_artificiose_machine_del_Capitano_Agostino_Ramelli_Figure_CLXXXVIII.jpg

N = 46 studies

Table 1 Original characteristics of included interactions

Characteristic	*N*	%
Total number of reported interactions	191	100
Interaction order		
2	139	72.8
3	35	18.3
4	8	4.2
5	5	2.6
6+	4	2.1
Analytical method		
Logistic regression	59	30.9
MDR (any variation)	45	23.6
MDR alone	(33)	(17.3)
MDR with logistic regression	(6)	(3.1)
GMDR	(3)	(1.6)
MB-MDR with logistic regression	(3)	(1.6)
Chi-square test (contingency table analysis)	32	16.8
ROSETTA	17	8.9
ALPS	9	4.7
BN-BMLA with logistic regression	8	4.2
Recursive partitioning	4	2.1
Set association analysis	4	2.1
FAMHAP	3	1.6
Generalized estimating equation	3	1.6
Log-linear modeling	2	1.0
Cox proportional hazards regression	1	0.5
PEAK	1	0.5
Unclear	3	1.6

Characteristic	*N*	%
Interaction explicitly identified		
Yes	110	57.6
No	76	39.8
Unclear	5	2.6
Internal replication		
Internal replication attempted	29	15.2
Attempted and replicated	14	7.3
Attempted and not replicated	15	7.9
Internal replication not attempted	162	84.8
Direction inferable		
Yes—multiple association measures reported or inferable	64	33.5
Yes—only one association measure reported or inferable	39	20.4
No—no association measure reported or inferable	88	46.1

Murk W et al. Hum Genet 2015. 134: 837-49.

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Large-Scale, Consortium-Based Genomewide Association Study of Asthma

Miriam F. Moffatt, D.Phil., Ivo G. Gut, Ph.D., Florence Demenais, M.D., David P. Strachan, M.D., Emmanuelle Bouzigon, M.D., Ph.D., Simon Heath, Ph.D., Erika von Mutius, M.D., Martin Farrall, F.R.C.Path., Mark Lathrop, Ph.D., and William O.C.M. Cookson, M.D., D.Phil., for the GABRIEL Consortium*

N Engl J Med 2010;363:1211-21.

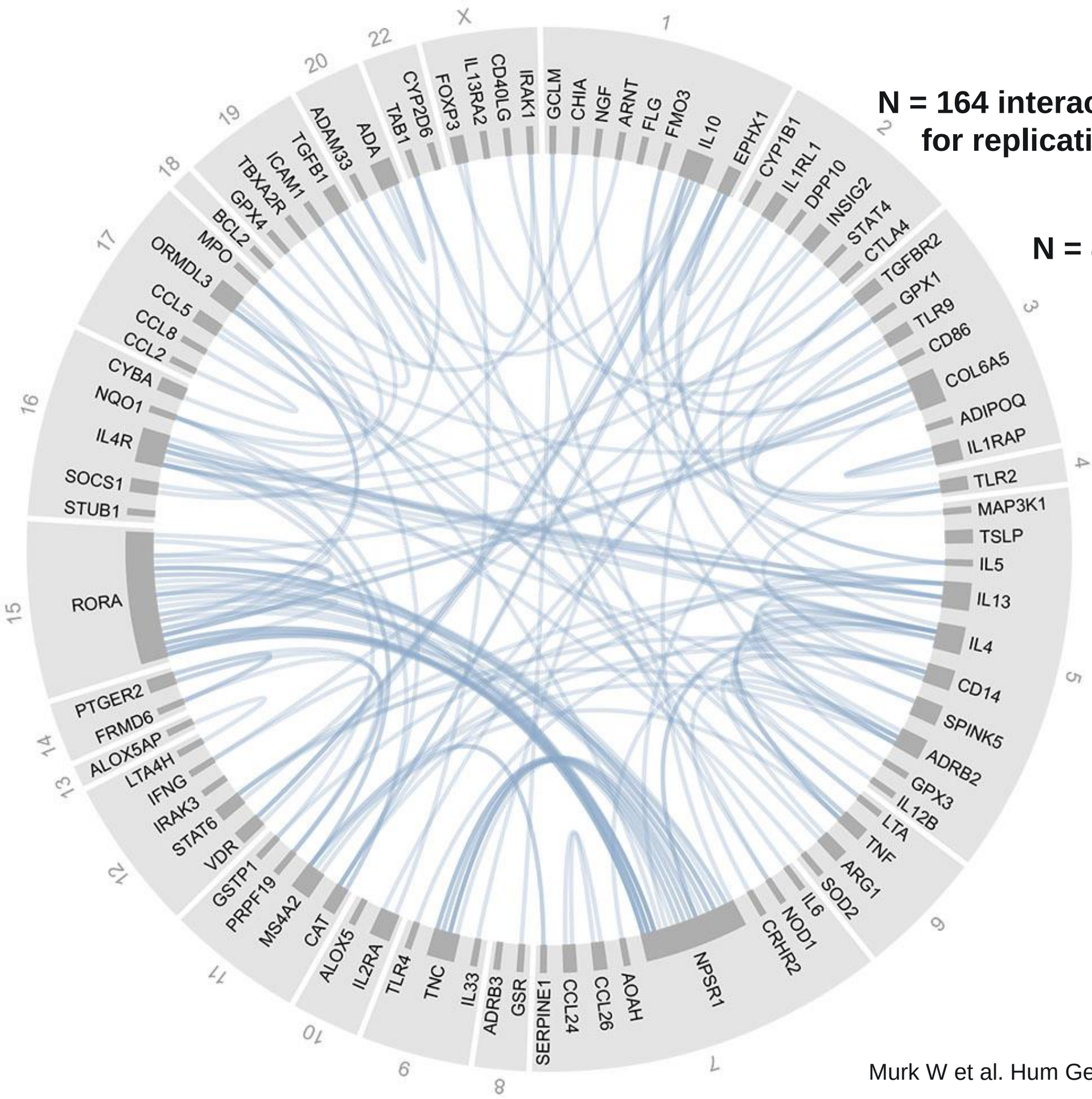
Study definitions and subject counts

Study	Country	Case definition	Control definition	N, cases	N, controls
BAMSE	Sweden	Ever had doctor-diagnosed asthma (self-report).	Did not have a history of asthma or other allergic diseases (self-report).	226	235
BUSSELTON	Australia	Ever had doctor-diagnosed asthma (self-report).	Did not ever have doctor-diagnosed asthma (self-report).	520	685
EGEA	France	Ever had asthma attacks (self-report).	Did not ever have asthma attacks (self-report).	120	444
GABRIEL-AS	Austria, Germany, Switzerland	Ever had doctor-diagnosed asthma or had asthmatic bronchitis at least twice (self-report).	Did not ever have doctor-diagnosed asthma and no asthmatic bronchitis diagnosed at least twice (self-report).	802	823
KARELIA	Finland, Russia	Ever had doctor-diagnosed asthma (self-report).	Did not ever have doctor-diagnosed asthma (self-report).	57	68
KMSU	Russia	Asthma diagnosed on the basis of symptoms (recurrent cough, wheezing, or dyspnea), airway obstruction reversibility, or airway methacholine hyper-responsiveness.	No symptoms or history of allergic disease, normal total serum IgE, normal pulmonary function.	285	261
PIAMA	The Netherlands	Ever had doctor-diagnosed asthma (self-report).	Did not have a history of asthma or other allergic diseases (self-report).	174	187
SAPALDIA	Switzerland	Ever had asthma (self-report).	Did not ever have asthma (self-report).	581	880
UFA	Russia	Asthma diagnosed on the basis of clinical examination, family and medication history, and lung function tests.	No symptoms or history of asthma or other pulmonary disease, no symptoms or history of atopy, no first-degree relatives with a history of asthma or atopy.	333	333
Total				**3,098**	**3,916**

Analytical approach

Interaction order	Analytical method
2	Logistic regression, MDR
3	Logistic regression, MDR
> 3	MDR

Also evaluated pairwise interactions nested within interactions of order > 2.



N = 164 interactions assessed for replication

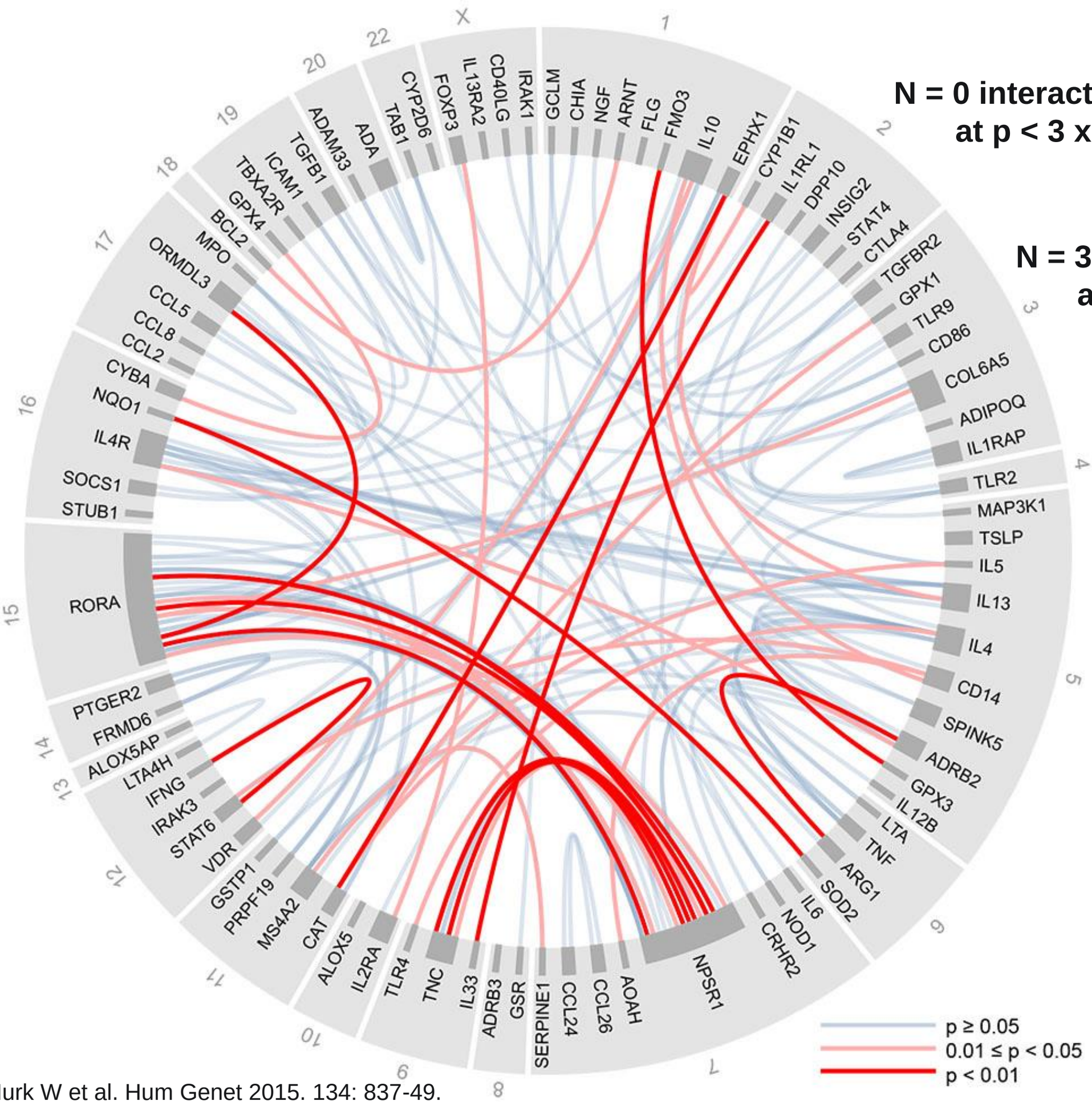
N = 87 genes

N = 179 SNPs

Murk W et al. Hum Genet 2015. 134: 837-49.

N = 0 interactions
at p < 3 x 10-4

N = 39 interactions
at p < 0.05



Murk W et al. Hum Genet 2015. 134: 837-49.

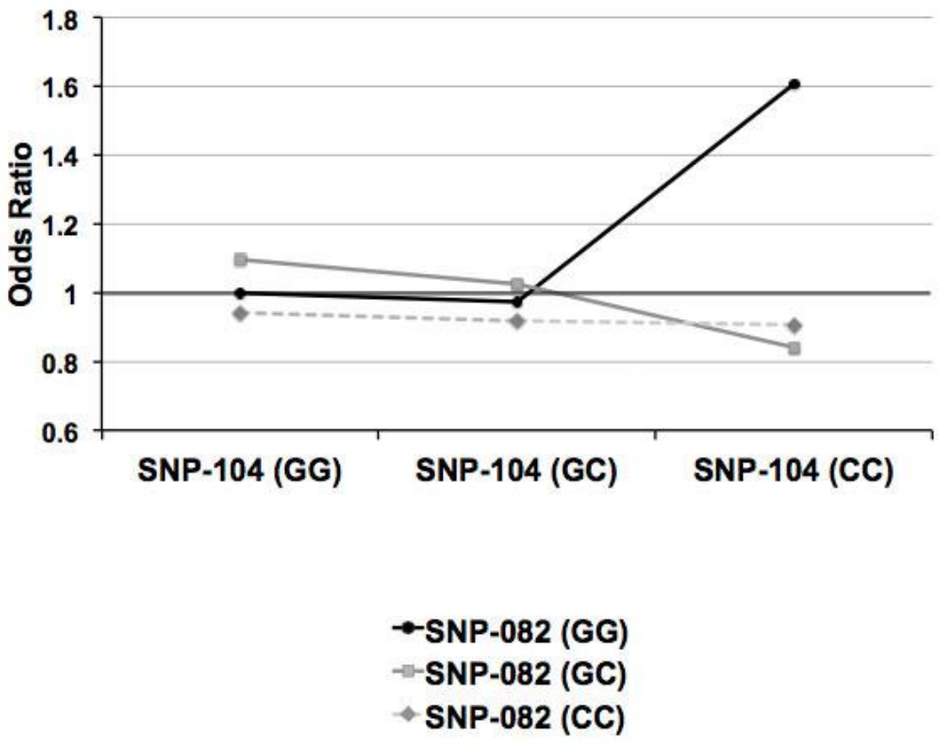
Interaction between rs3789873 (TNC) and rs323922 (NPSR1)

ORIGINAL STUDY

Associations:		SNP-104	
		GG	GC/CC
SNP-082	GG	1 [Ref.]	1.79 (1.13, 2.85)
	GC	1 [Ref.]	0.72 (0.48, 1.08)
	CC	1 [Ref.]	0.97 (0.55, 1.73)

CURRENT STUDY

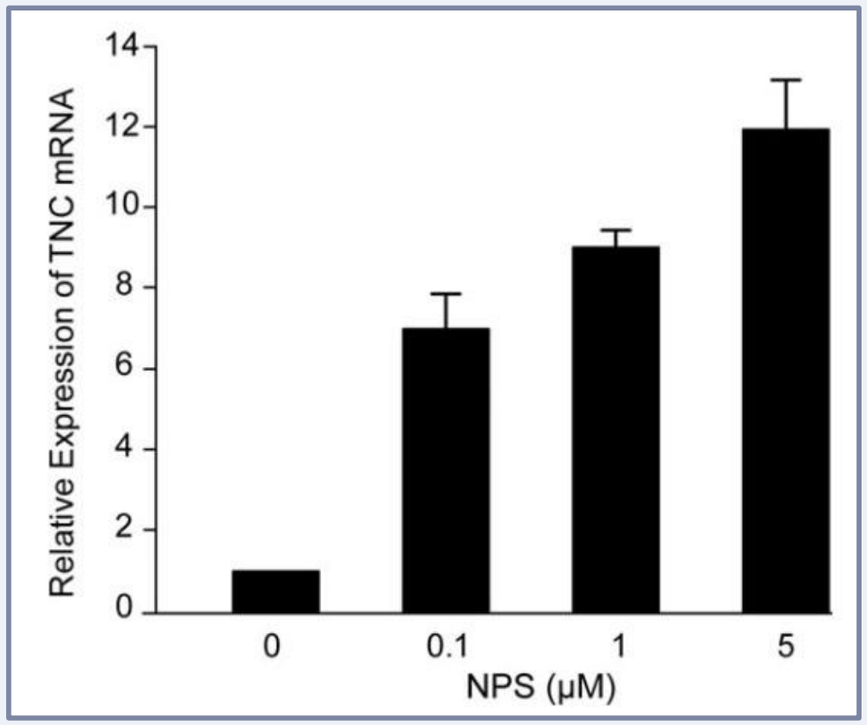
Associations:		SNP-104	
		GG	GC/CC
SNP-082	GG	1 [Ref.]	1.07 (0.90, 1.27) .429
	GC	1 [Ref.]	0.92 (0.79, 1.06) .224
	CC	1 [Ref.]	0.99 (0.77, 1.27) .948

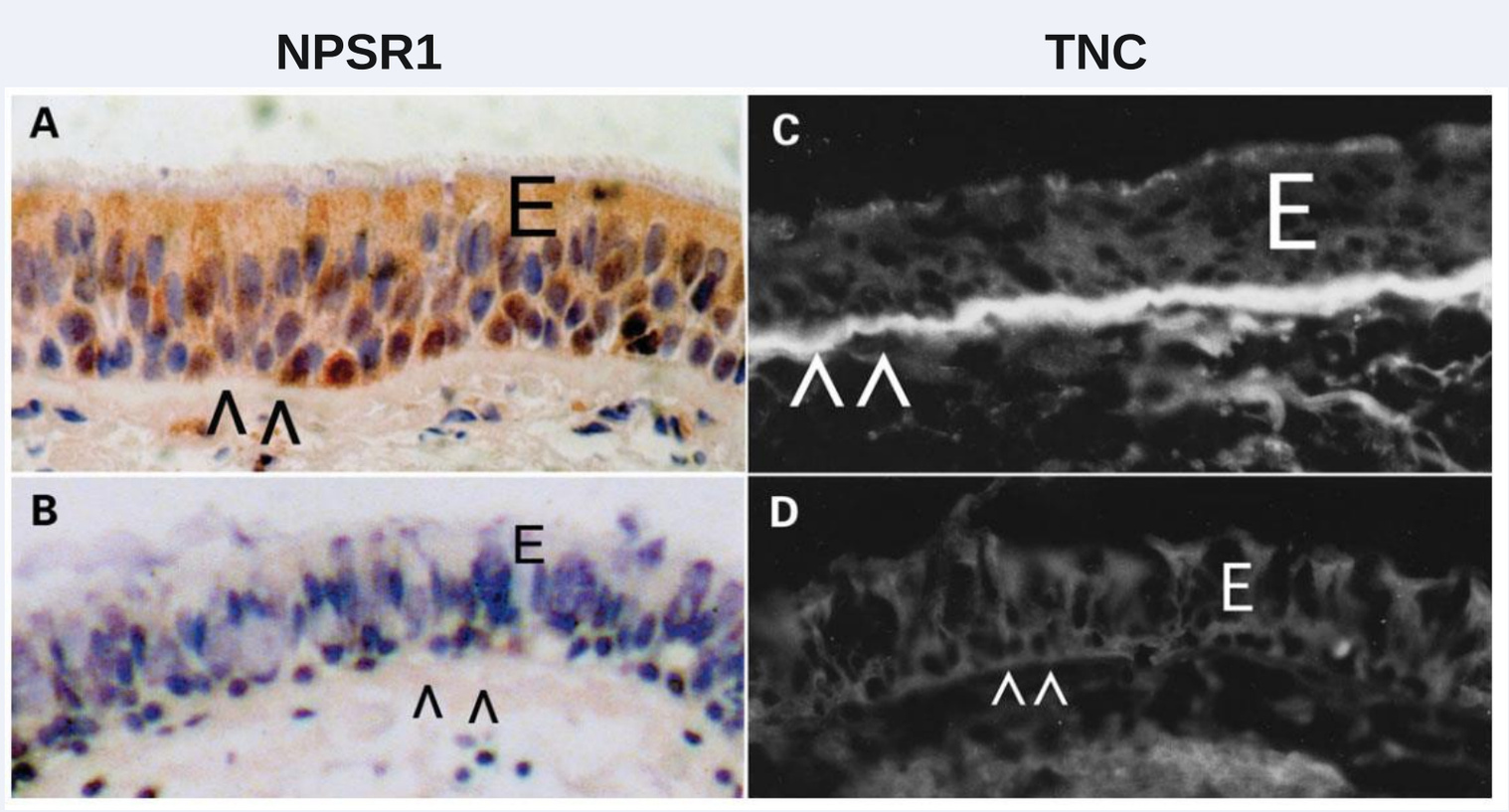


Sub-study:	Interaction Term: SNP-104 (Rec-C) × SNP-082 (Dom-C)	
	OR (95% CI)	P
BAMSE	0.55 (0.14, 2.20)	.401
BUSSELTON	0.18 (0.06, 0.54)	2.0 x 10-3
EGEA	0.82 (0.14, 4.81)	.829
GABRIELAS	0.84 (0.34, 2.06)	.705
KARELIA	N/A*	--
KMSU	0.93 (0.16, 5.41)	.939
PIAMA	1.30 (0.24, 6.91)	.760
SAPALDIA	0.48 (0.20, 1.19)	.115
UFA	0.15 (0.03, 0.71)	.016
Meta-analysis:		
Pooled estimate (FE)	0.50 (0.32, 0.77)	1.7 x 10-3
Heterogeneity: I^2	22.4 %	
Heterogeneity: Q (P)	.251	
Collapsed:		
Genotype probabilities	0.51 (0.34, 0.77)	1.3 x 10-3
Best-guess genotypes	0.49 (0.32, 0.74)	7.3 x 10-4

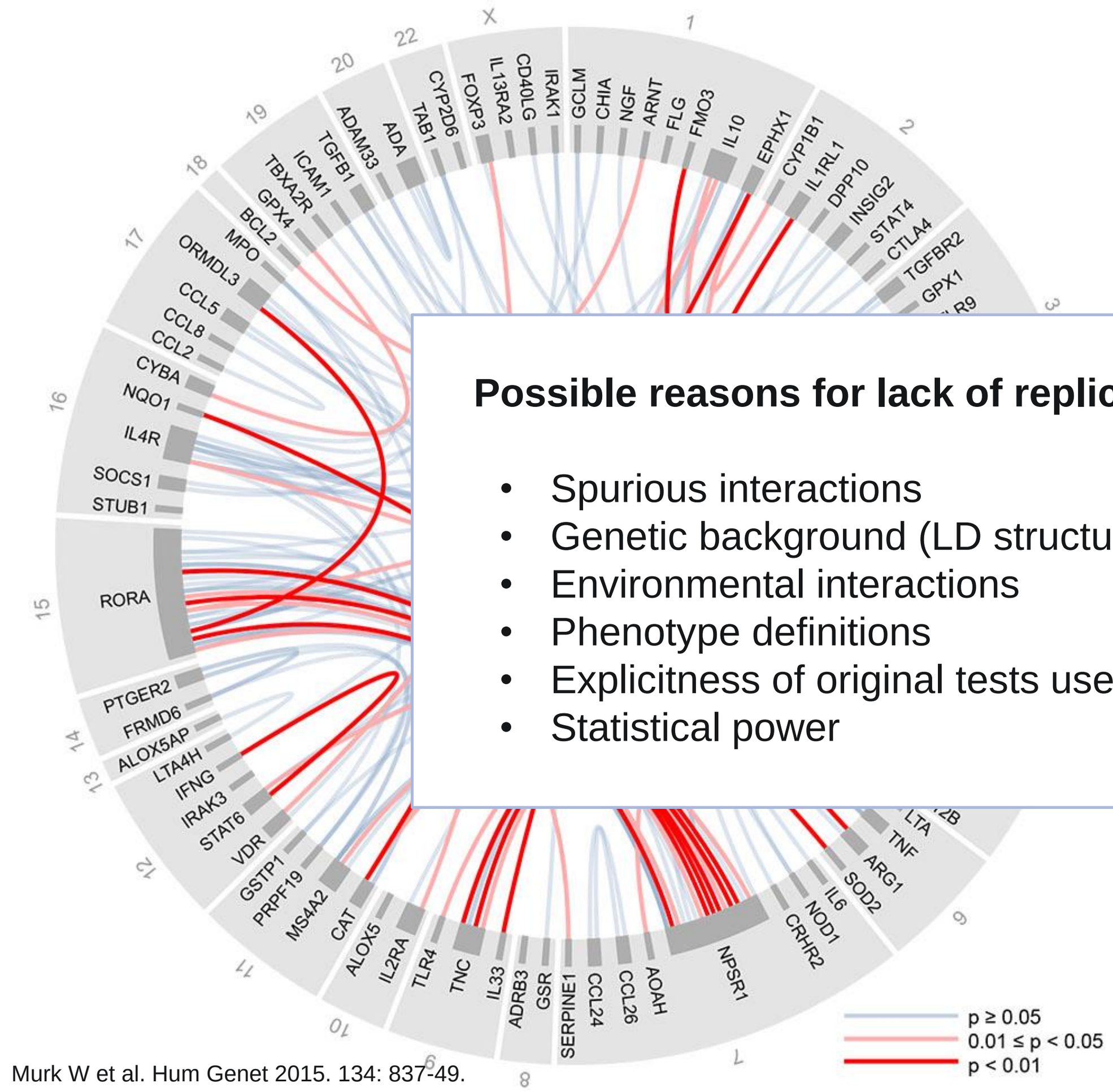
Murk W et al. Hum Genet 2015. 134: 837-49.

Biological relationship between TNC and NPSR1





Orsmark-Pietras et al. Hum Mol Genet 2008. 17: 1673-1682

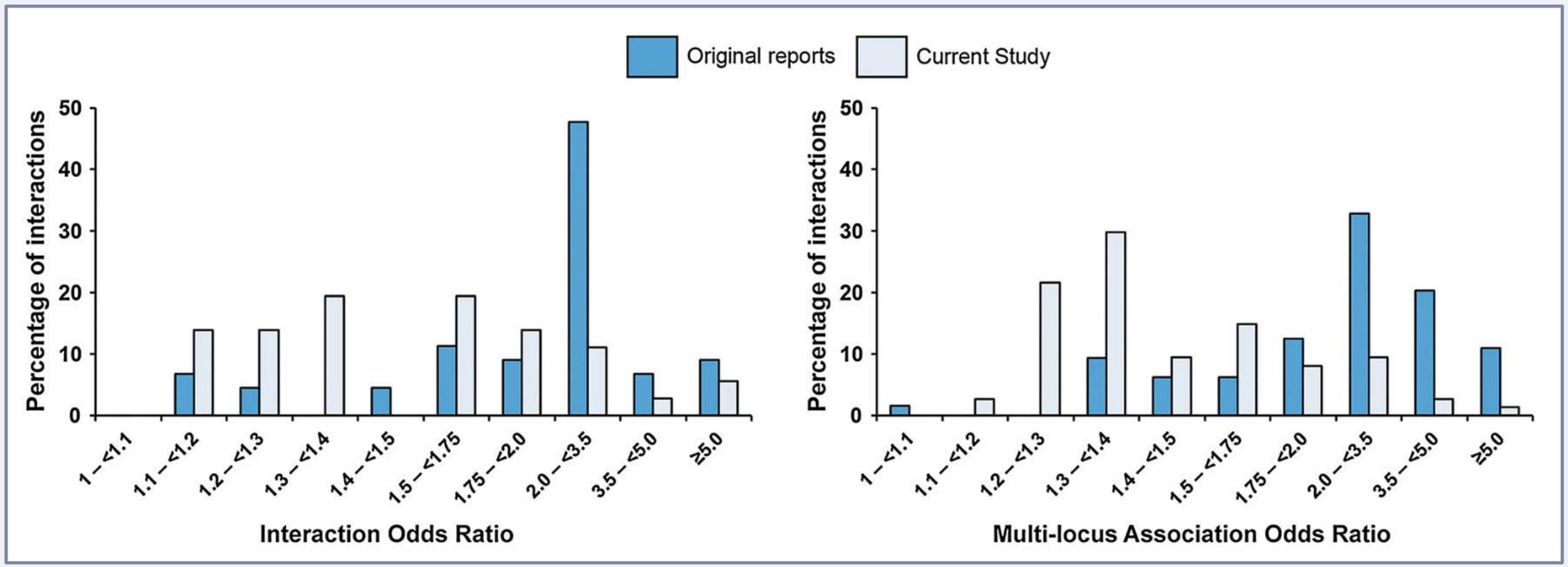


Possible reasons for lack of replication:

- Spurious interactions
- Genetic background (LD structure; MAF)
- Environmental interactions
- Phenotype definitions
- Explicitness of original tests used
- Statistical power

Murk W et al. Hum Genet 2015. 134: 837-49.

Winner's Curse



Murk W et al. Hum Genet 2015. 134: 837-49.

Conclusions

- Most published gene-gene interactions are not well supported, and many are not well described
- No interaction could be strictly replicated
- Suggestive evidence in support of a minority of the interactions

Acknowledgements

Andrew DeWan
Michael B. Bracken
Hongyu Zhao

Funding:
•Canadian Institutes of Health Research
Doctoral Foreign Study Award; Award No. DFS – 129311
•Yale School of Public Health

Data:
- GABRIEL Consortium
- European Genome-phenome Archive