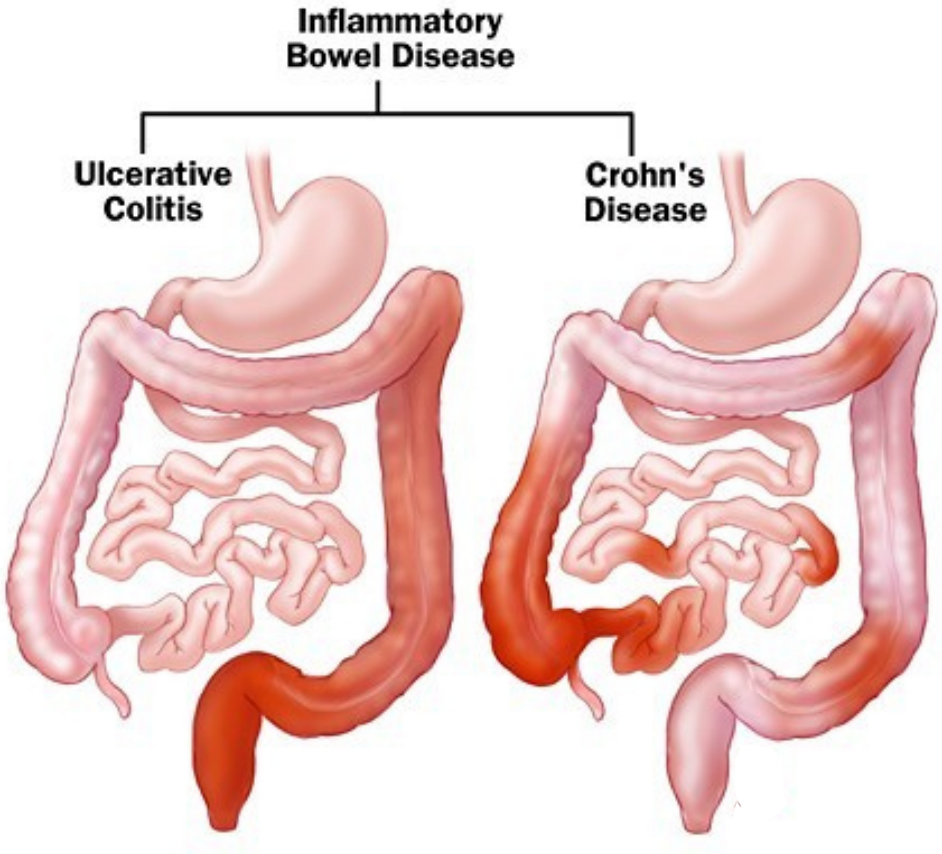
Genome-Wide Association Studies in Inflammatory Bowel Disease

Richard H. Duerr, M.D.
Associate Professor of Medicine and Human Genetics
University of Pittsburgh

Inflammatory bowel disease: ulcerative colitis and Crohn's disease

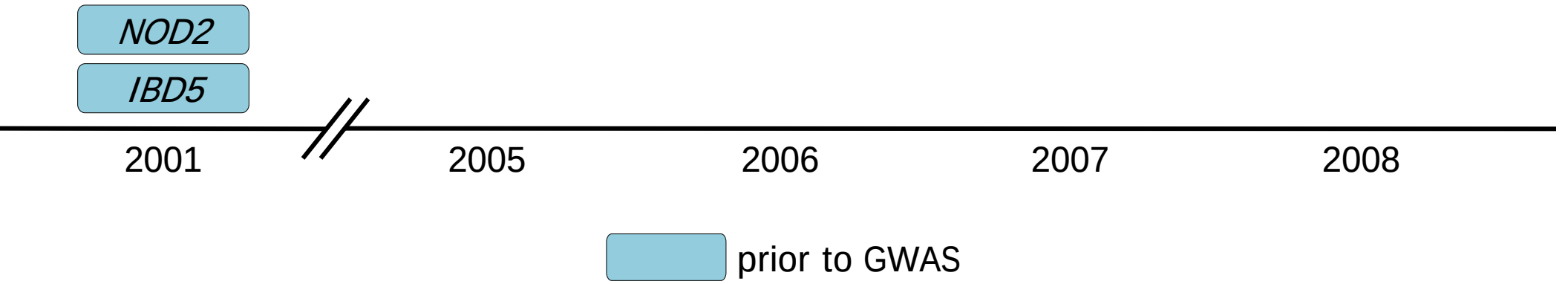


- Symptoms: diarrhea, abdominal pain, intestinal bleeding, growth retardation
- Pathophysiology: ubiquitous, commensal intestinal bacteria trigger an overactive and ongoing mucosal immune response in genetically predisposed hosts
- Prevalence: 100-200/100,000 for each disorder
- Peak age of onset: 2nd to 4th decades of life

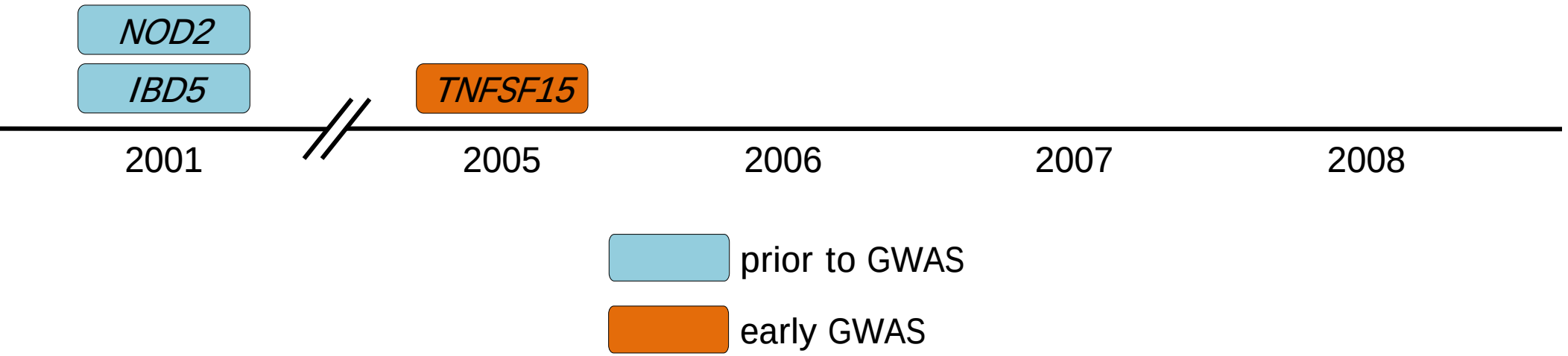
Lines of evidence for heritability of IBD

- Racial and ethnic differences in rates of IBD
  - Whites > Blacks > Hispanics and Asians in U.S.
  - Jews > non-Jews
- Familial aggregation of IBD
  - 5-20% of patients have a family hx of IBD
  - λ_s as high as 30-40 for CD and 10-20 for UC
  - 75% of multiplex families have only CD- or only UC-affected members; 25% of families "mixed" with CD and UC within the same family
- MZ:DZ twin concordance
  - CD: MZ=20-50% : DZ=0-7%
  - UC: MZ=14-19% : DZ=0-5%
- Associations between IBD and specific genetic variants

Confirmed Crohn's disease loci



Confirmed Crohn's disease loci



NIDDK IBD Genetics Consortium

Genetic Research Centers

Cedars – Sinai Medical Center
- Kent Taylor (PI)

Johns Hopkins University
- Steven Brant (PI)

Universite de Montreal
- John Rioux (PI)

University of Pittsburgh
- Richard Duerr (PI)

University of Toronto
- Mark Silverberg (PI)

Yale University
- Judy Cho (PI)

Data Coordinating Center

Yale University
-Judy Cho (PI), Yashoda Sharma

University of Chicago
- Dan Nicolae, Philip Schumm, Emily Kistner

Committees

Recruitment	Phenotyping	Genotyping	Analytic	Publications/IP
-Mark Silverberg	- Hillary Steinhart	- Richard Duerr	- Mark Daly	- Steven Brant, John Rioux

A Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene

Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH. Science 2006;314:1461-3.

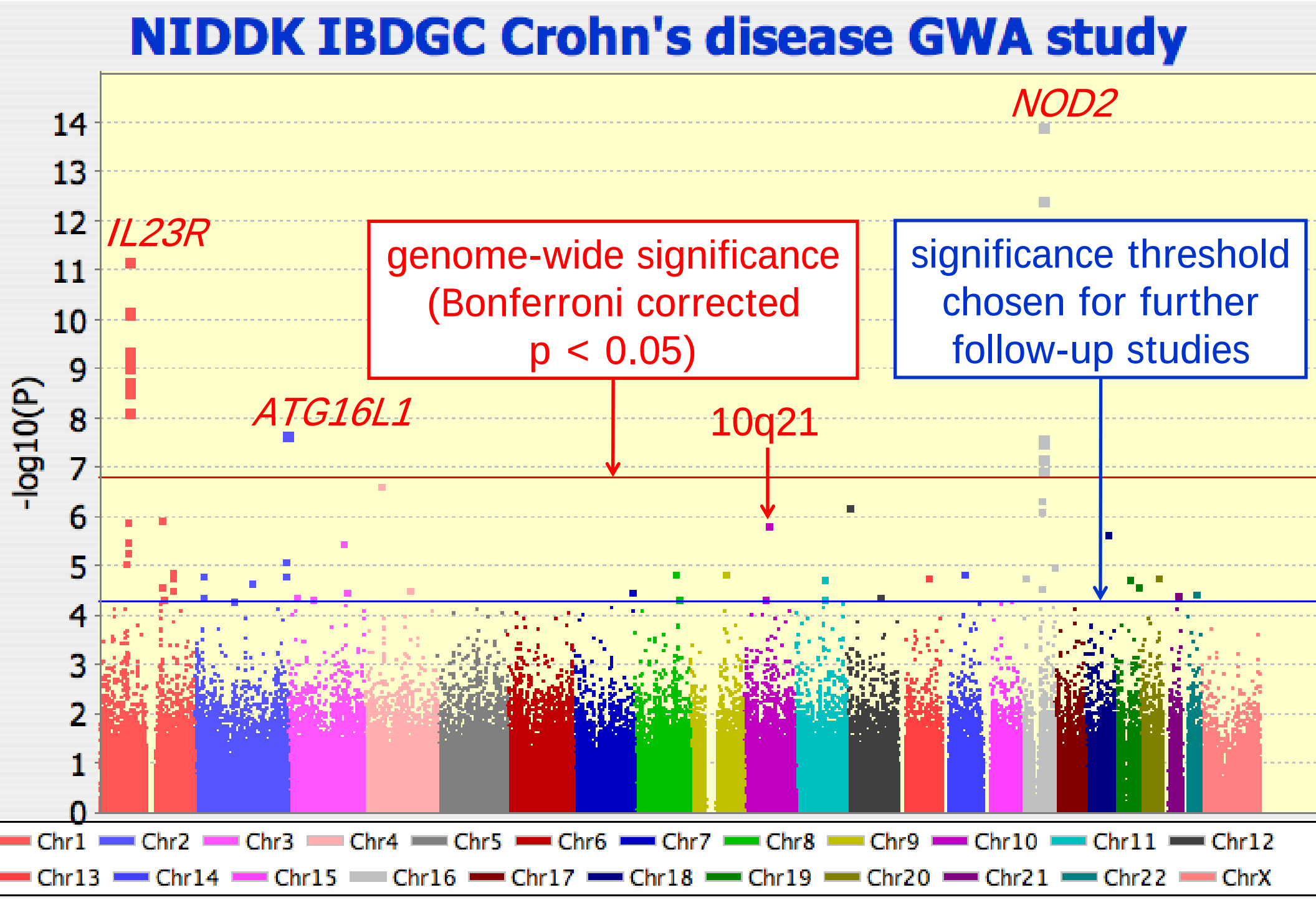
Genome-wide association study identifies new susceptibility loci for Crohn's disease and implicates autophagy in disease pathogenesis

Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR. Nat Genet. 2007;39:596-604.

GWAS Crohn's cases: N=946

GWAS controls: N=977

Platform: Illumina HumanHap300


NIDDK IBDGC Crohn's disease GWA study
NOD2
IL23R
genome-wide significance
(Bonferroni corrected
p < 0.05)
significance threshold
chosen for further
follow-up studies
ATG16L1
10q21
-log10(P)
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12
Chr13 Chr14 Chr15 Chr16 Chr17 Chr18 Chr19 Chr20 Chr21 Chr22 ChrX

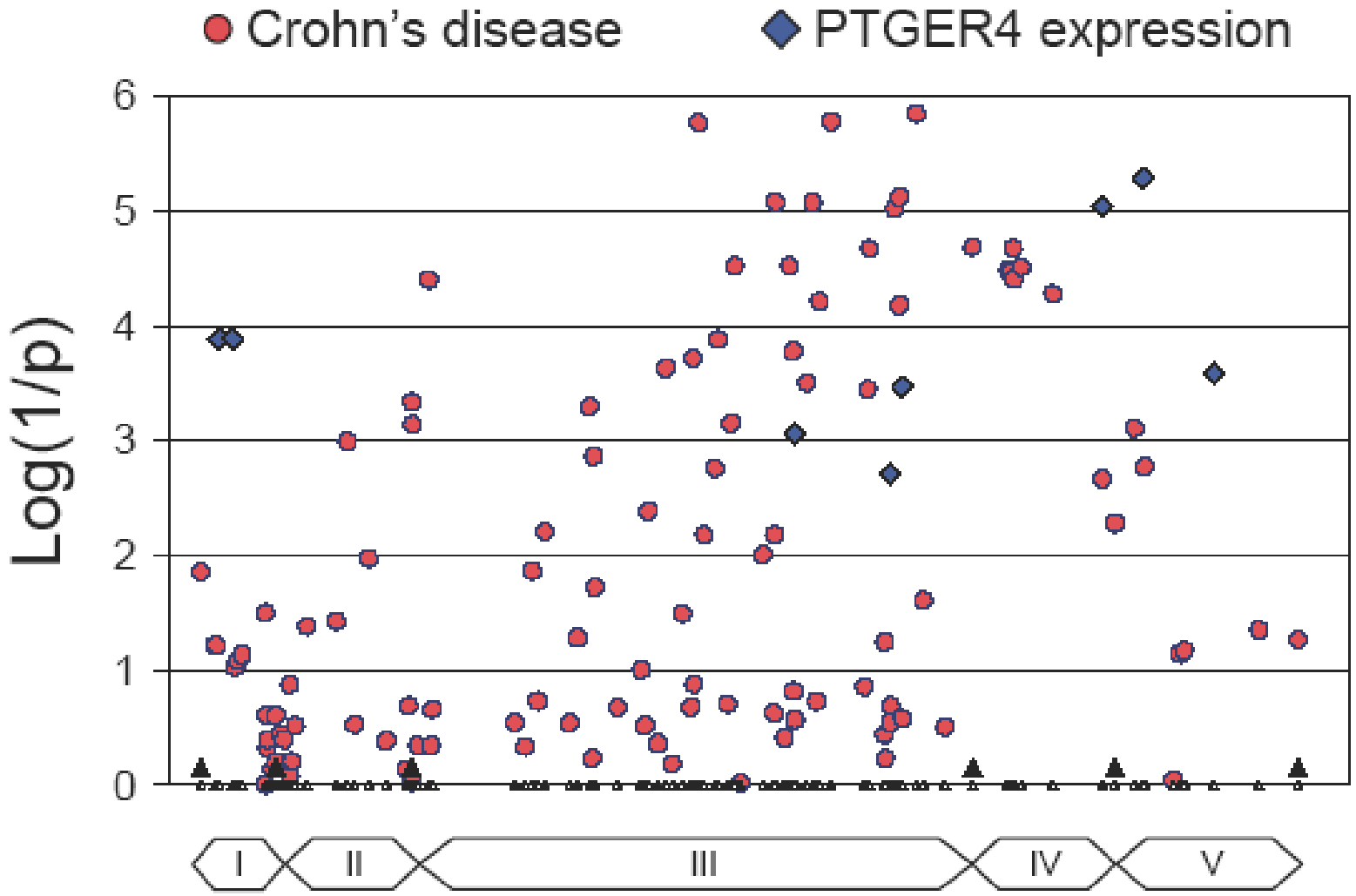
Belgian-French IBD consortium

- Teams:
  - University of Liège: *C. Libioulle, S. Hansoul, C. Sandor, M. Mni, J. Belaiche, E. Louis & M. Georges*
  - CNG: *I. Gut, S. Heath, D. Zelenika, M. Froglio, M. Lathrop*
  - Catholic University of Leuven: *S. Vermeire, P. Rutgeerts*
  - Free University of Brussels: *D. Franchimont, A. Van Gossum*
  - Ghent University: *M. De Vos, D. Laukens*
  - University Paris Diderot: *J.P. Hugot*
- Samples:
  - GWAS Crohn's cases: N=547
  - GWAS controls: N=928
  - Platform: Illumina HumanHap300

Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4

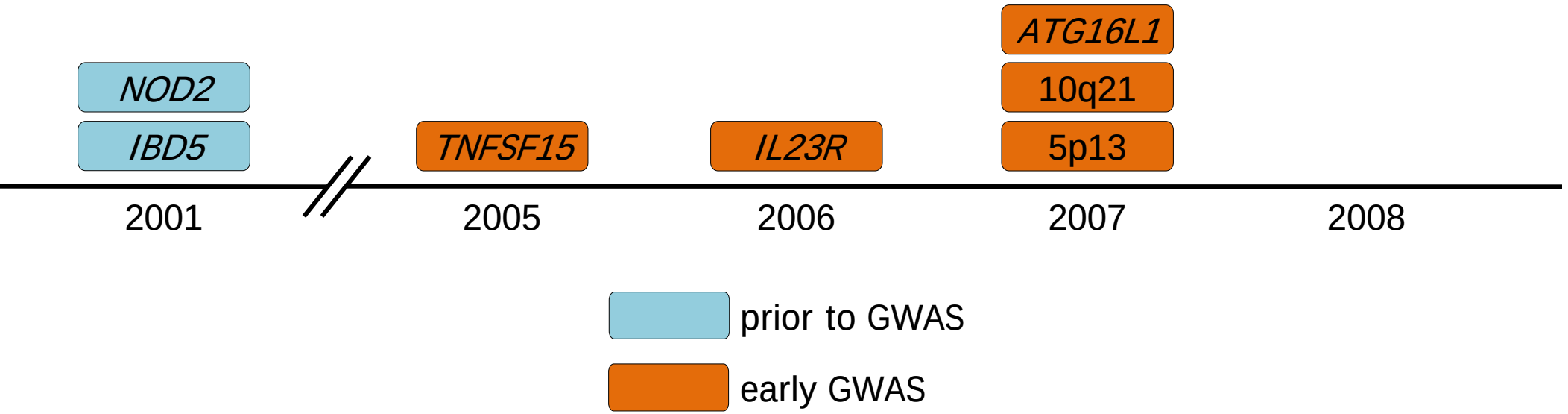
Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A, Demarche B, Gut I, Heath S, Foglio M, Liang L, Laukens D, Mni M, Zelenika D, Van Gossum A, Rutgeerts P, Belaiche J, Lathrop M, Georges M. PLoS Genet. 2007;3:e58.

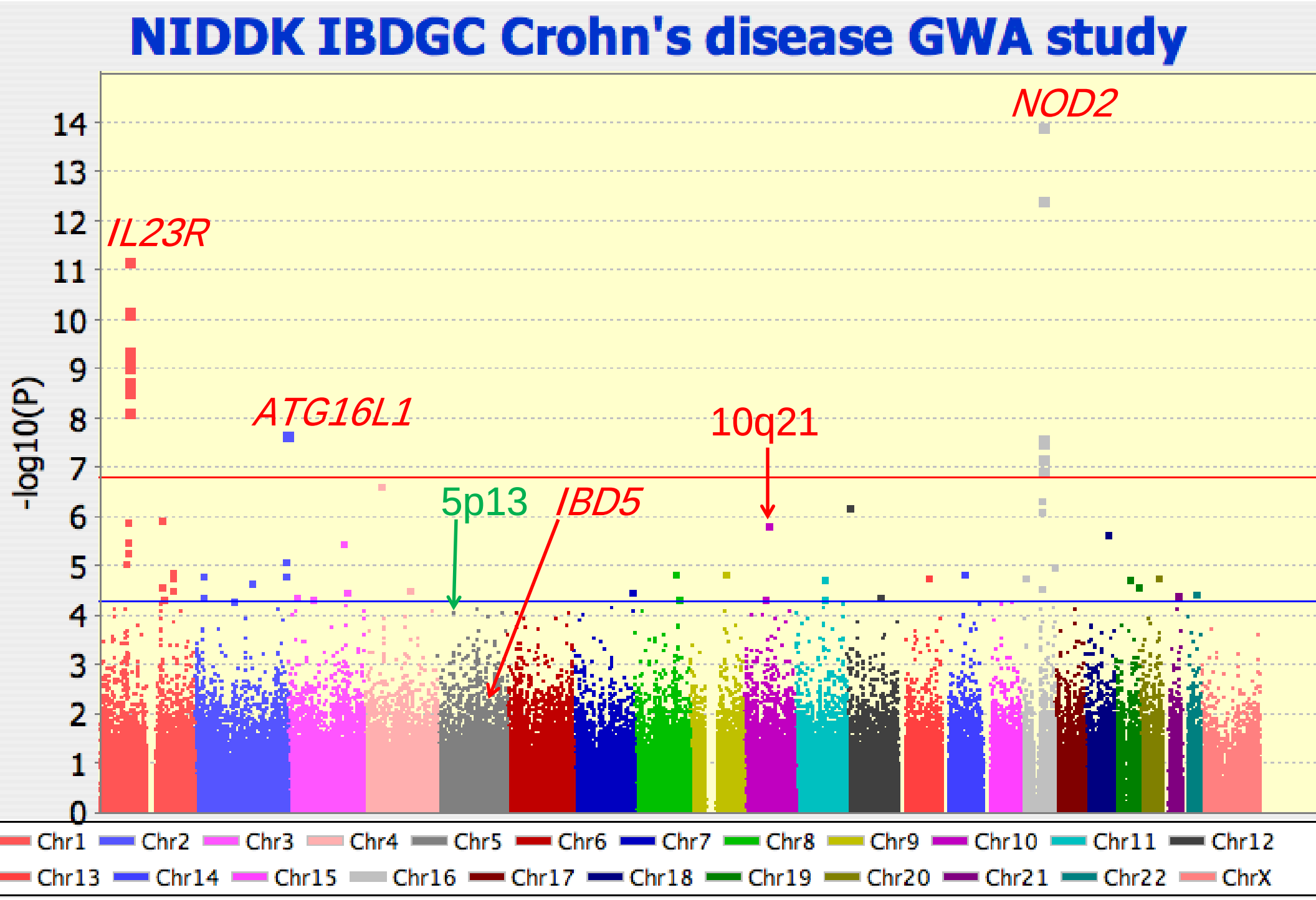
CD-associated chromosome 5p13 variants and *PTGER4* eQTL analysis in EBV-transformed lymphoblastoid cell lines



Libioulle et al. *PLoS Genet*. 2007;3:e58.

Confirmed Crohn's disease loci




NIDDK IBDGC Crohn's disease GWA study
NOD2
IL23R
ATG16L1
10q21
5p13
IBD5
-log10(P)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
ChrX

Wellcome Trust Case Control Consortium

UK IBD Genetics Group
Derek Jewell, Oxford
John Mansfield, Newcastle
Chris Mathew, London
Miles Parkes, Cambridge
Jack Satsangi, Edinburgh

WTCHG, Oxford
Jeff Barrett
Carl Anderson
Lon Cardon

King's College London:
Natalie Prescott
British 1958 Birth Cohort

nature genetics

Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci contribute to Crohn's disease susceptibility

Miles Parkes[1,13], Jeffrey C Barrett[2,13], Natalie J Prescott[3,13], Mark Tremelling[1], Carl A Anderson[2], Sheila A Fisher[3], Roland G Roberts[3], Elaine R Nimmo[4], Fraser R Cummings[5], Dianne Soars[3], Hazel Drummond[4], Charlie W Lees[4], Saud A Khawaja[3], Richard Bagnall[3], Denis A Burke[6], Catherine E Todhunter[6], Tariq Ahmad[5], Clive M Onnie[3], Wendy McArdle[7], David Strachan[8], Graeme Bethel[9], Claire Bryan[9], Cathryn M Lewis[3], Panos Deloukas[9], Alastair Forbes[10], Jeremy Sanderson[11], Derek P Jewell[5], Jack Satsangi[4], John C Mansfield[6], the Wellcome Trust Case Control Consortium[12], Lon Cardon[2] & Christopher G Mathew[3]

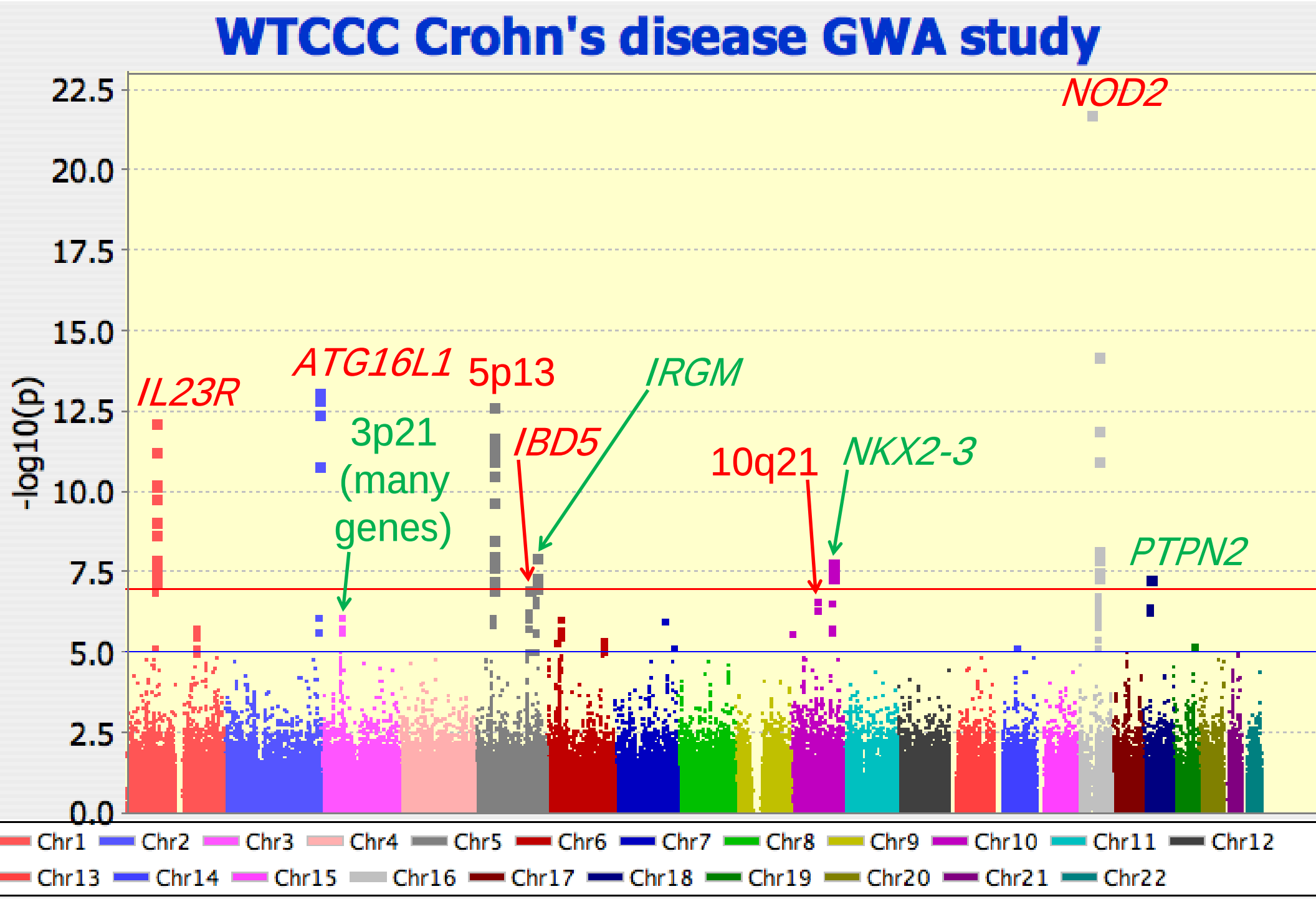
GWAS Crohn's cases: N=1,748
GWAS controls: N=2,938 ('58 BBC/NBS donors)
Platform: Affymetrix 500K

wellcome trust

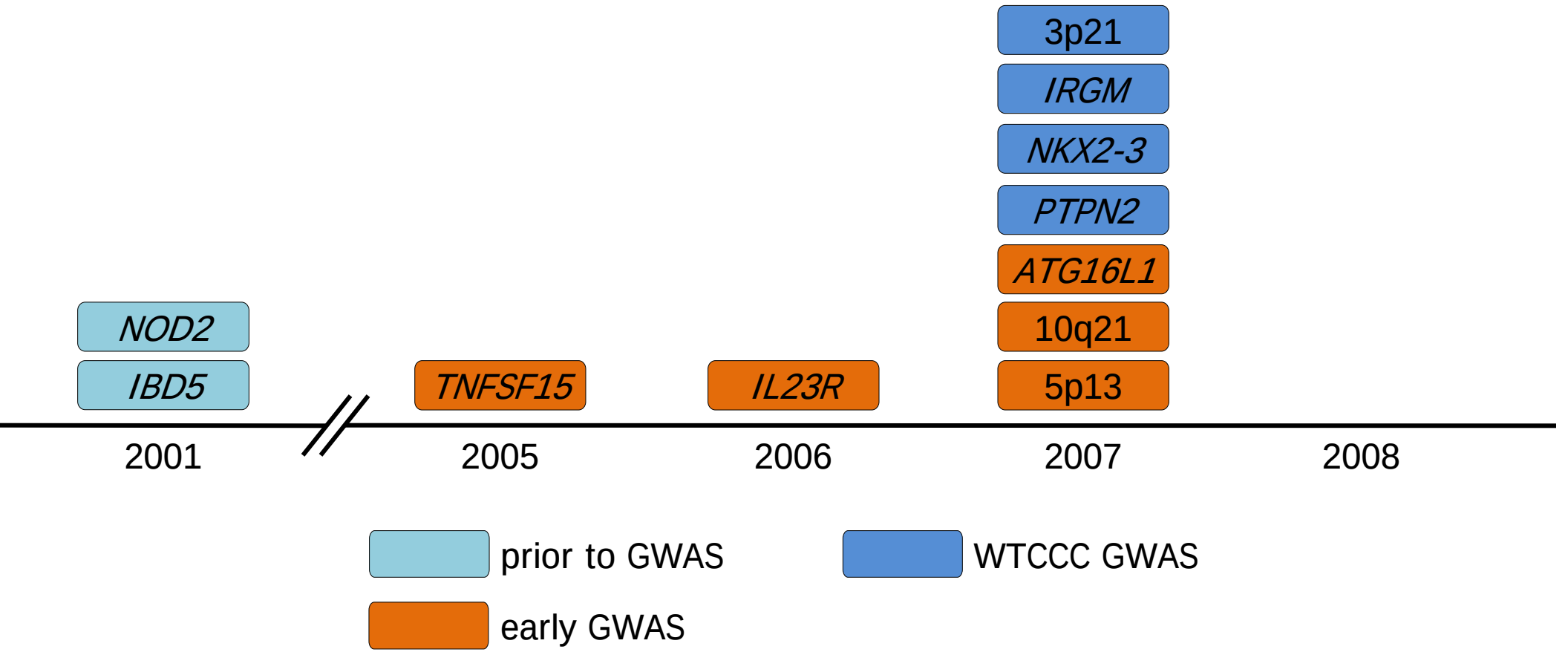
Colitis and Crohn's NACC
UK Registered Charity Number 282732

core FIGHTING GUT AND LIVER DISEASE

ucb


WTCCC Crohn's disease GWA study
-log10(p)
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
NOD2
IL23R
ATG16L1
3p21
(many
genes)
5p13
IBD5
IRGM
10q21
NKX2-3
PTPN2
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22

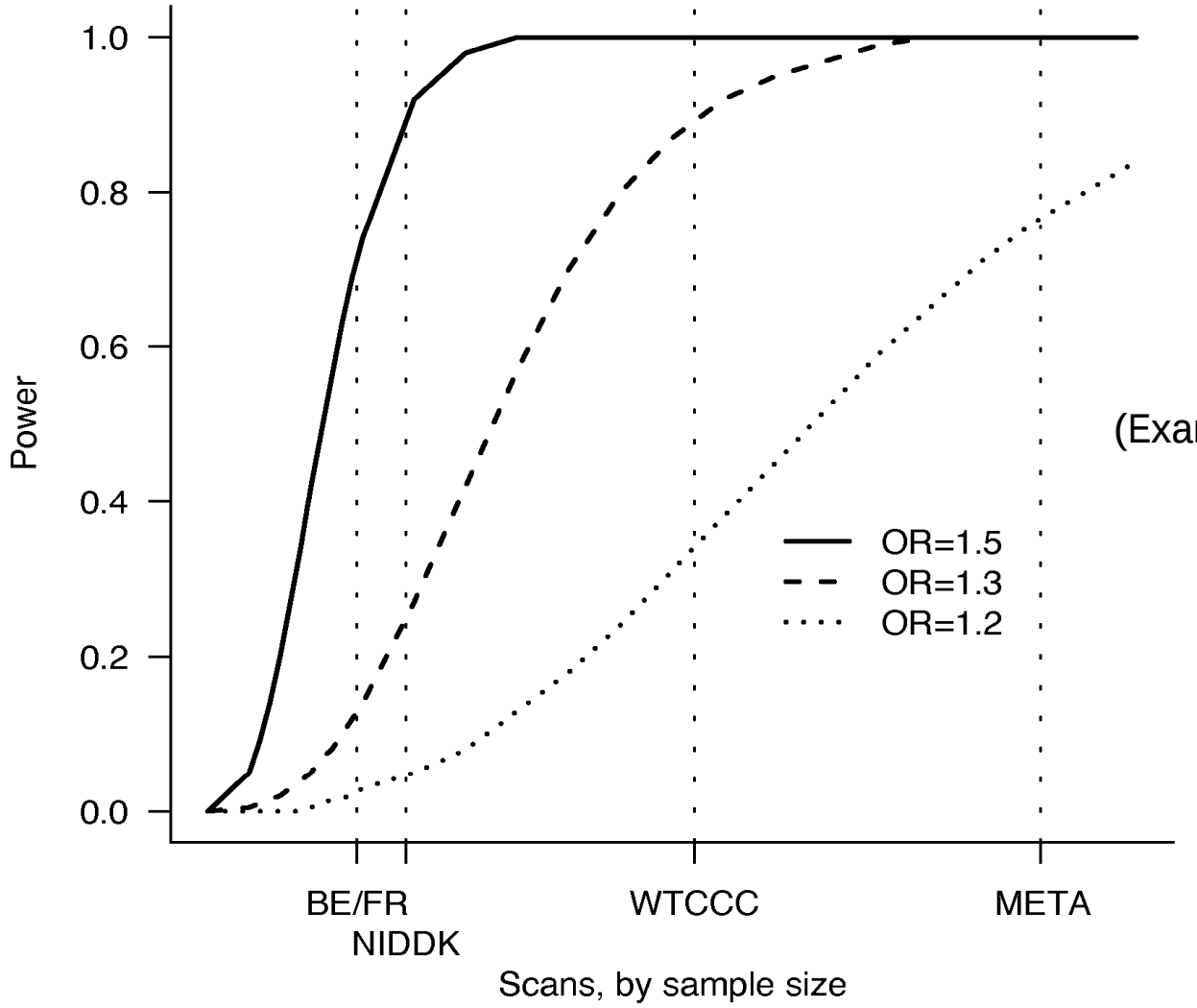
Confirmed Crohn's disease loci



Opportunity for additional disease locus discovery: Crohn's disease GWA meta-analysis

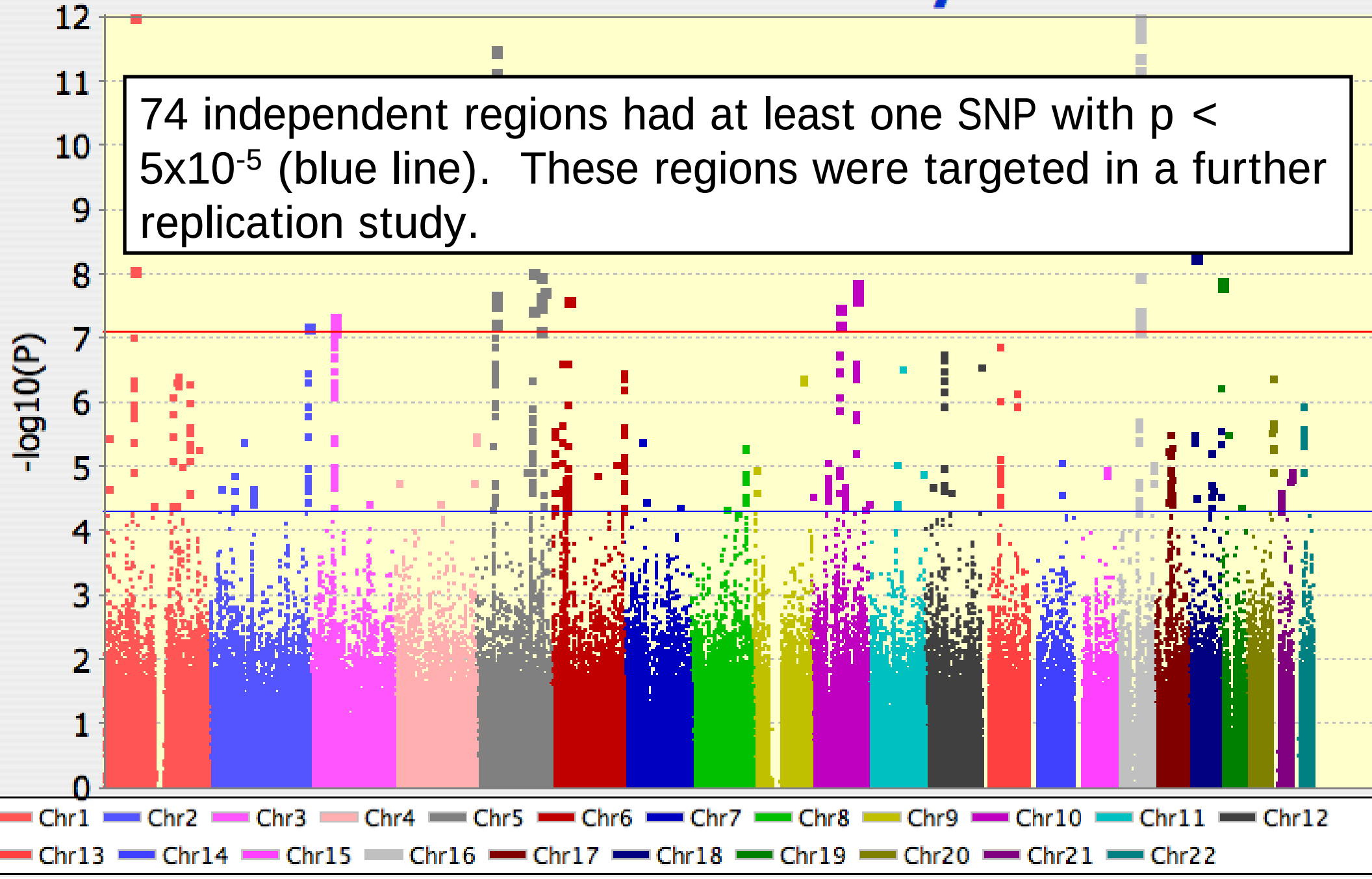
- meta-analysis of NIDDK + Belgian/French + WTCCC Crohn's disease GWA studies
- combined total of 3,230 cases and 4,829 controls with genome-wide SNP data post quality control
- 633,548 SNPs (directly genotyped SNPs or SNPs with genotypes imputed using genotyped SNPs and HapMap data)

Crohn's disease GWA meta-analysis provides substantial increase in discovery power



(Example – disease-associated SNP with MAF = 0.20)

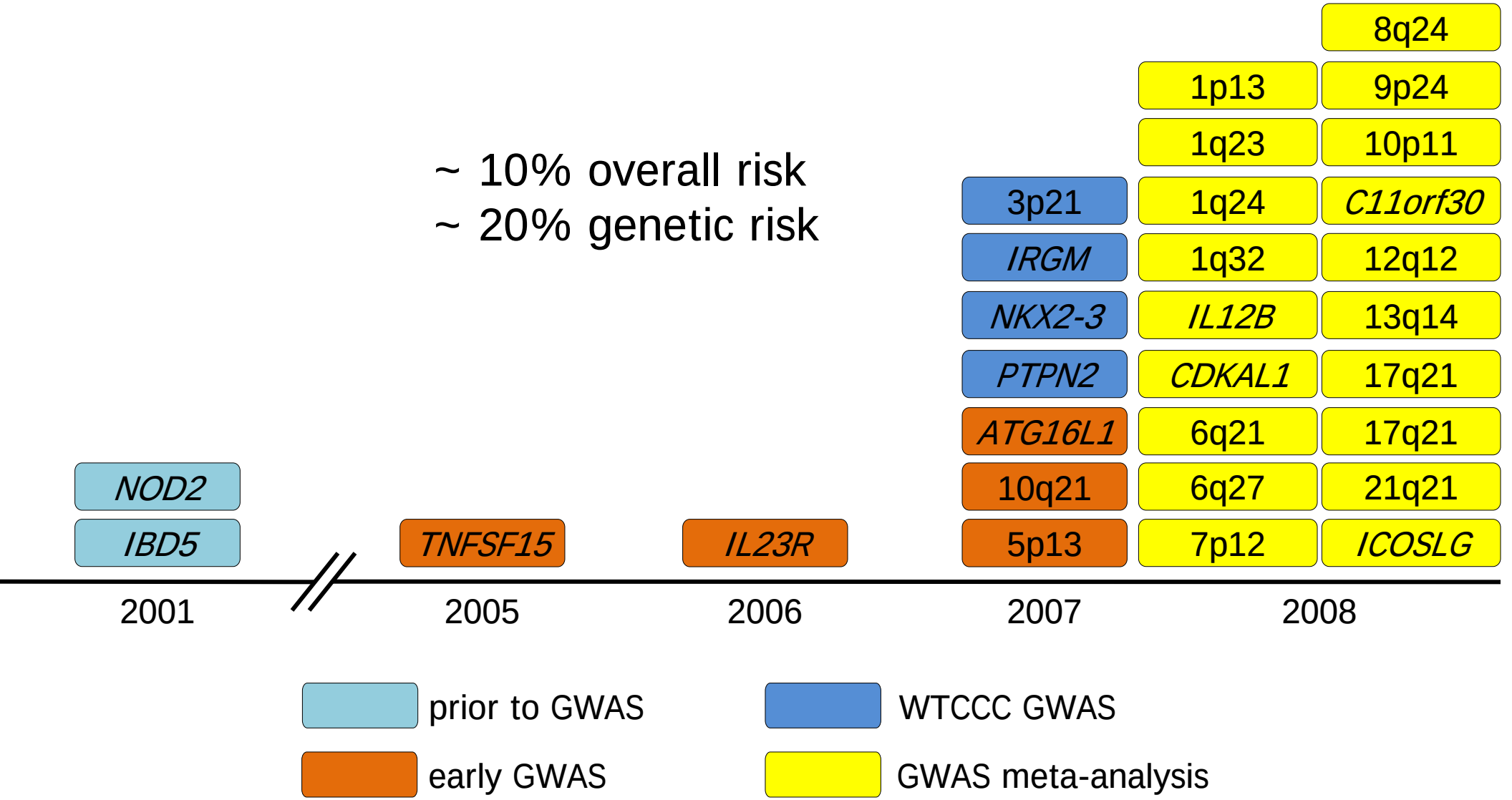
Crohns's disease GWA meta-analysis zoomed in



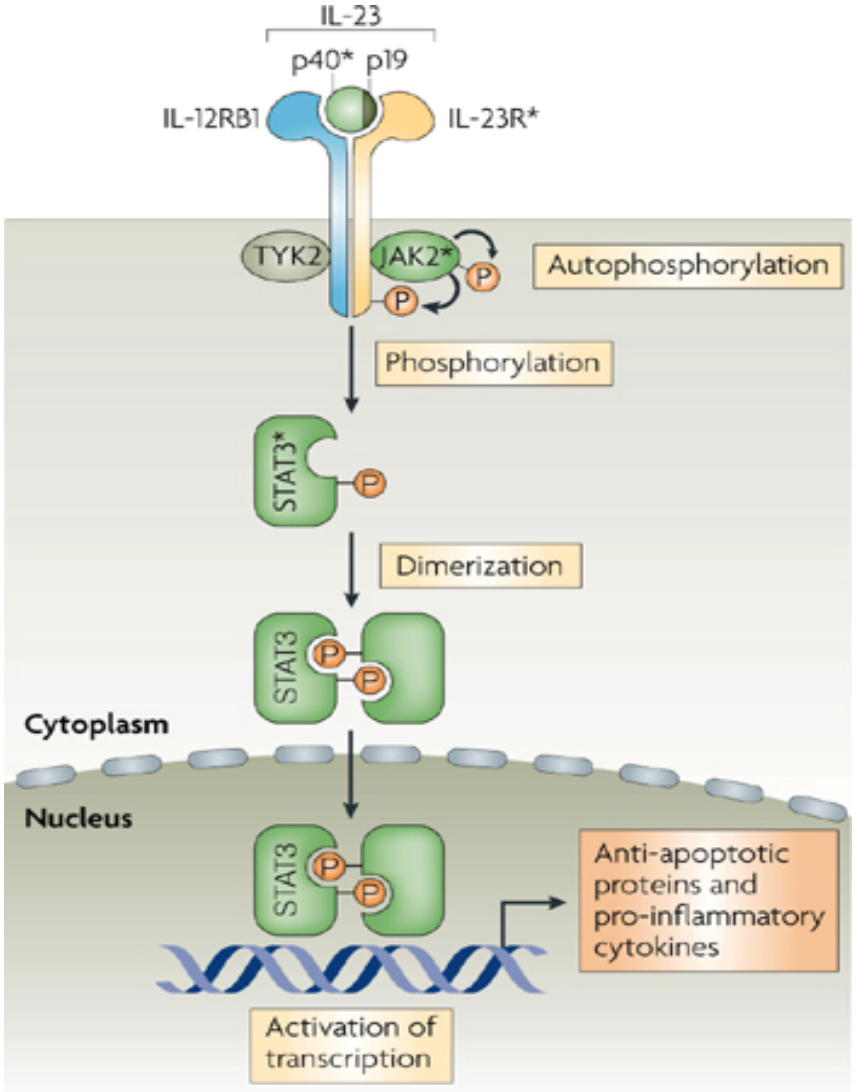
Crohn's disease GWA meta-analysis replication

- Replication study subjects, post quality control
  - 2,325 cases
  - 1,809 controls
  - 1,339 parent-parent-affected offspring trios
- 30 regions had significant replication evidence and had combined (GWA meta-analysis and replication) genome-wide significant $p < 5x10^{-8}$
- Another 10 regions were nominally replicated with $p < 0.05$ (would expect only 2-3 by chance); the MHC and chr 19p13 had combined (GWA meta-analysis and replication) genome-wide significant $p < 5x10^{-8}$

Confirmed Crohn's disease loci



IL23/TH17 pathway variants associated with risk for CD



Cho JH. Nat Rev Immunol. 2008;8:458-66

IL23R Arg381Gln

Gln allele is protective
Frequency: ~2% in cases
~7% in controls

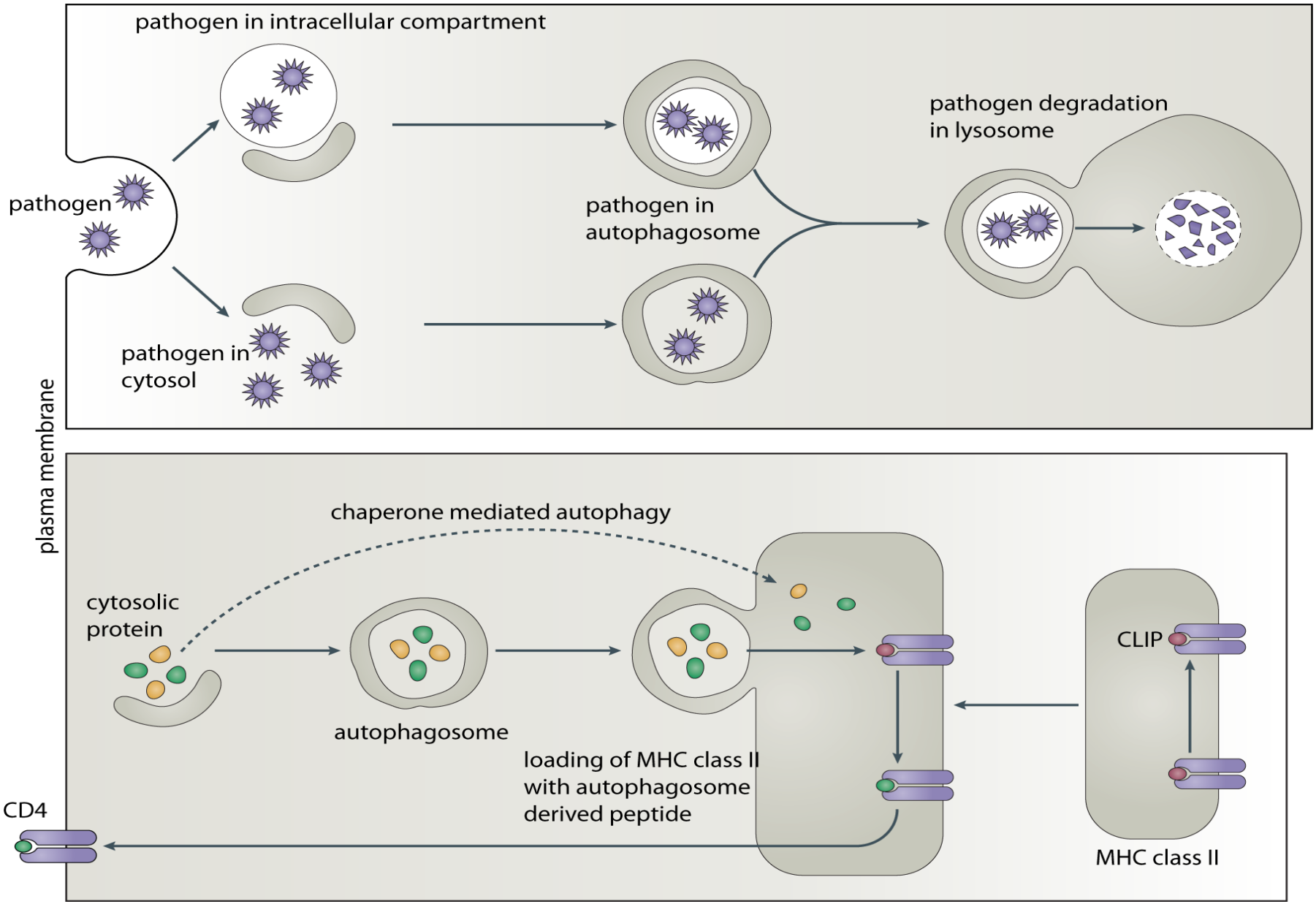
*** Also associated with UC and psoriasis ***

IL23R risk haplotype

Causal variant not known
Frequency: ~40% in cases
~30% in controls

Duerr et. al. *Science* 2006;314:1461-3

Autophagy in pathogen degradation



Autophagy in antigen presentation

Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease

Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, Lees CW, Onnie CM, Hanson CE, Blaszczyk K, Inouye M, Ewels P, Ravindrarajah R, Keniry A, Hunt S, Carter M, Watkins N, Ouwehand W, Lewis CM, Cardon L; Wellcome Trust Case Control Consortium, Lobo A, Forbes A, Sanderson J, Jewell DP, Mansfield JC, Deloukas P, Mathew CG, Parkes M, Satsangi J. *Nat Genet*. 2008;40:710-2.

- scan of 10,886 nsSNPs in 905 UC cases and 1,465 controls (panel 1)
- 33 SNPs from 21 distinct loci with $P < 0.001$ were followed up in 936 UC cases and 1,470 controls (panel 2)
- 2 SNPs with promising association evidence in the *ECM1* (extracellular matrix protein 1) locus, which had not previously been associated with IBD, were followed up in an additional 1,146 cases and 1,559 controls (panel 3)
- 16 SNPs tagging 13 Crohn's disease–associated loci identified by GWAS were tested in 1,841 ulcerative colitis cases (panels 1 and 2) and 1,470 controls (panel 2)
- association analyses using Cochran-Armitage trend tests

Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease

Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, Lees CW, Onnie CM, Hanson CE, Blaszczyk K, Inouye M, Ewels P, Ravindrarajah R, Keniry A, Hunt S, Carter M, Watkins N, Ouwehand W, Lewis CM, Cardon L; Wellcome Trust Case Control Consortium, Lobo A, Forbes A, Sanderson J, Jewell DP, Mansfield JC, Deloukas P, Mathew CG, Parkes M, Satsangi J. *Nat Genet*. 2008;40:710-2.

SNP	Chr	Pos	Gene	P_{GC} panel 1	P panel 2	P panel 3	P combined
rs3737240	1	148749979	*ECM1*	4.5×10^{-4}	0.013	0.038	2.3×10^{-6}
rs13294	1	148751611	*ECM1*	7.1×10^{-4}	0.020	0.060	7.9×10^{-6}
rs3197999	3	49696536	*MST1*	0.0010	0.0061		8.0×10^{-6}
rs9268480	6	32471822	*BTNL2*	0.0015	0.0064		7.2×10^{-6}
rs660895	6	32685358	*HLA-DRB1*	1.5×10^{-5}	0.0035		2.8×10^{-8}

Genetic determinants of ulcerative colitis include the *ECM1* locus and five loci implicated in Crohn's disease

Fisher SA, Tremelling M, Anderson CA, Gwilliam R, Bumpstead S, Prescott NJ, Nimmo ER, Massey D, Berzuini C, Johnson C, Barrett JC, Cummings FR, Drummond H, Lees CW, Onnie CM, Hanson CE, Blaszczyk K, Inouye M, Ewels P, Ravindrarajah R, Keniry A, Hunt S, Carter M, Watkins N, Ouwehand W, Lewis CM, Cardon L; Wellcome Trust Case Control Consortium, Lobo A, Forbes A, Sanderson J, Jewell DP, Mansfield JC, Deloukas P, Mathew CG, Parkes M, Satsangi J. *Nat Genet*. 2008;40:710-2.

SNP	Chr	Pos	Gene	P combined
rs11805303	1	67448104	*IL23R*	2.2×10^{-4}
rs9858542	3	49676987	*MST1*	1.3×10^{-6}
rs9292777	5	40473705	gene desert	0.014
rs6556416	5	158751323	*IL12B*	6.8×10^{-4}
rs6887695	5	158755223	*IL12B*	0.0016
rs10761659	10	64115570	gene desert	0.0012
rs10883365	10		*NKX2-3*	2.4×10^{-6}

Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility

Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, Domingues FS, Albrecht M, Nothnagel M, Ellinghaus D, Sina C, Onnie CM, Weersma RK, Stokkers PC, Wijmenga C, Gazouli M, Strachan D, McArdle WL, Vermeire S, Rutgeerts P, Rosenstiel P, Krawczak M, Vatn MH; IBSEN study group, Mathew CG, Schreiber S. *Nat Genet.* 2008;40:1319-23.

- scan of 440,794 SNPs in 1,167 UC cases and 777 controls (panel A)
- top 20 SNPs were followed up in three case-control samples totalling 1,855 UC cases and 3,091 controls in stage 2 (panels B-D)

Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility

Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, Domingues FS, Albrecht M, Nothnagel M, Ellinghaus D, Sina C, Onnie CM, Weersma RK, Stokkers PC, Wijmenga C, Gazouli M, Strachan D, McArdle WL, Vermeire S, Rutgeerts P, Rosenstiel P, Krawczak M, Vatn MH; IBSEN study group, Mathew CG, Schreiber S. *Nat Genet.* 2008;40:1319-23.

SNP	Chr	Pos	Gene	Panel A	Panels B-D
rs11805303	1	67387537	*IL23R*	5.39×10^{-6}	1.09×10^{-5}
rs3024505	1	203328299	*IL10*	1.43×10^{-5}	1.35×10^{-12}
rs12612347	2	158755223	*ARPC2*	8.42×10^{-6}	2.00×10^{-4}
rs9268480	6	32471822	*BTNL2*	2.21×10^{-6}	3.15×10^{-9}
rs9268858	6	32537736	*HLA-DRA*	5.41×10^{-7}	2.58×10^{-12}
rs9268877	6	32539125	*HLA-DRA*	5.23×10^{-7}	6.48×10^{-18}

NIDDK IBD Genetics Consortium

Genetic Research Centers

Cedars – Sinai Medical Center
- Kent Taylor (PI)

Johns Hopkins University
- Steven Brant (PI)

Universite de Montreal
- John Rioux (PI)

University of Pittsburgh
- Richard Duerr (PI)

University of Toronto
- Mark Silverberg (PI)

Yale University
- Judy Cho (PI)

Data Coordinating Center

Yale University
-Judy Cho (PI), Yashoda Sharma

University of Chicago
- Dan Nicolae, Philip Schumm, Emily Kistner

Committees

Recruitment
-Mark Silverberg

Phenotyping
- Hillary Steinhart

Genotyping
- Richard Duerr

Analytic
- Mark Daly

Publications/IP
- Steven Brant, John Rioux

Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study

Silverberg MS, Cho JH, Rioux JD, McGovern DPB, Wu J, Annese V, Achkar J-P, Goyette P, Scott R, Xu W, Barmada MM, Klei L, Daly MJ, Abraham C, Bayless TM, Bossa F, Griffiths AM, Ippoliti AF, Lahaie RG, Latiano A, Paré P, Proctor DD, Regueiro MD, Steinhart AH, Targan SR, Schumm LP, Kistner EO, Lee AT, Gregersen PK, Rotter JI, Brant SR, Taylor KD, Roeder K, Duerr RH. *Nat Genet.* 2009;41:216-20.

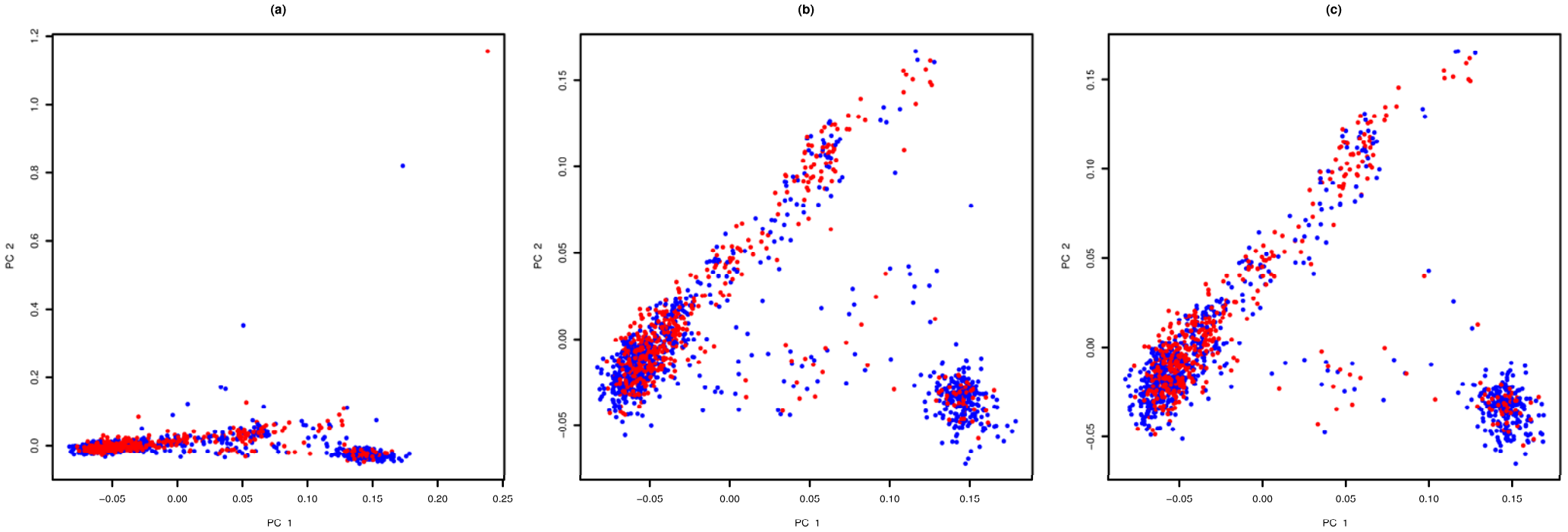
NIDDK IBDGC UC GWAS stage I

Study subjects	Genotyping array	Raw data (number of samples)	Post quality control data (number of samples)
Ulcerative colitis cases	Illumina HumanHap300v2	552	537
Ulcerative colitis cases	Illumina HumanHap550v3	500	485
Total ulcerative colitis cases		1,052	1,022
NIDDK IBDGC Crohn's disease GWAS controls	Illumina HumanHap300v1	994	951
Illumina iControlDB study 65 controls	Illumina HumanHap300v1	125	123
Illumina iControlDB study 65 controls	Illumina HumanHap550v1	619	609
Illumina iControlDB studies 64 and 65 controls	Illumina HumanHap550v3	833	820
Total controls		2,571	2,503

Genetic matching (GEM) to control for population structure in case-control association analysis

- create an L SNPs and N individuals matrix of allele counts from genome-wide tag SNP data
- compute eigenvector decomposition to describe the ancestry of individuals on a multidimensional “principal components of ancestry” map
- remove outliers for any significant principal component of ancestry and recompute eigenvector decomposition
- cluster individuals who appear to have common ancestry; remove individuals who are too distant from their counterparts (controls for cases, and vice versa)
- perform association analysis using cluster number as the conditional variable in conditional logistic regression analysis

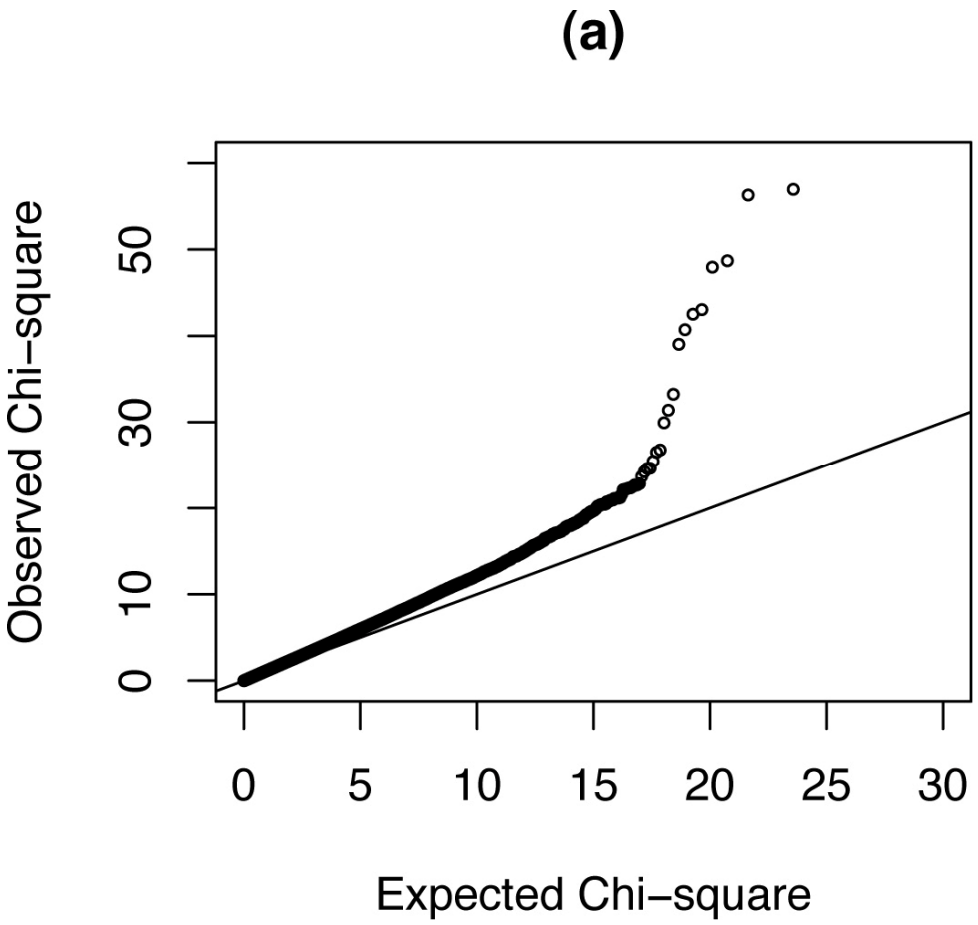
Principal components of ancestry in NIDDK IBDGC UC GWAS stage I female study subject tag SNP data



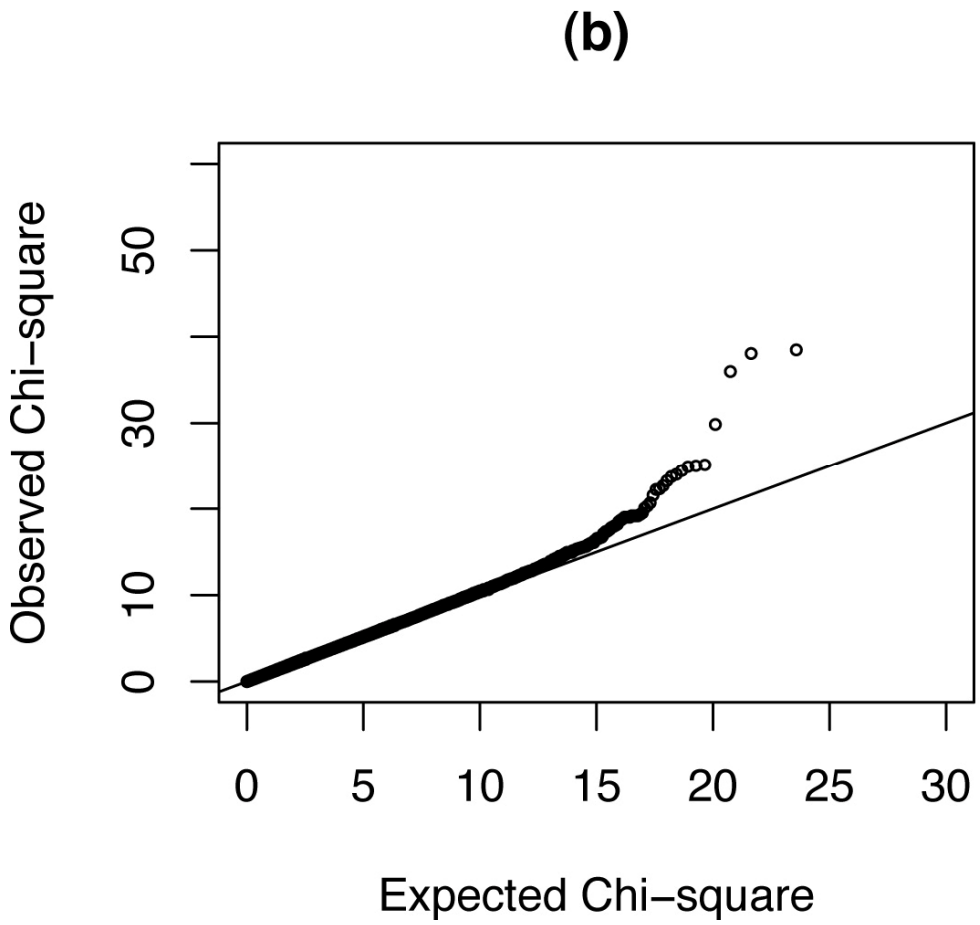
NIDDK IBDGC UC GWAS stage I

Study subjects	Genotyping method	All arrays shared SNPs gender- and ancestry-matched case-control strata (number of samples, strata or SNPs)	HumanHap550 only SNPs gender- and ancestry-matched case-control strata (number of samples, strata or SNPs)
Ulcerative colitis cases	Illumina HumanHap300v2	512	
Ulcerative colitis cases	Illumina HumanHap550v3	465	476
Total ulcerative colitis cases		977	476
NIDDK IBDGC Crohn's disease GWAS controls	Illumina HumanHap300v1	837	
Illumina iControlDB study 65 controls	Illumina HumanHap300v1	102	
Illumina iControlDB study 65 controls	Illumina HumanHap550v1	523	565
Illumina iControlDB studies 64 and 65 controls	Illumina HumanHap550v3	660	729
Total controls		2,122	1,294
Gender- and ancestry-matched case-control strata		909	447
SNPs		280,748	207,332

NIDDK IBDGC UC GWAS stage I quantile-quantile plots pre- and post-GEM



pre-GEM genomic control inflation factor = 1.17

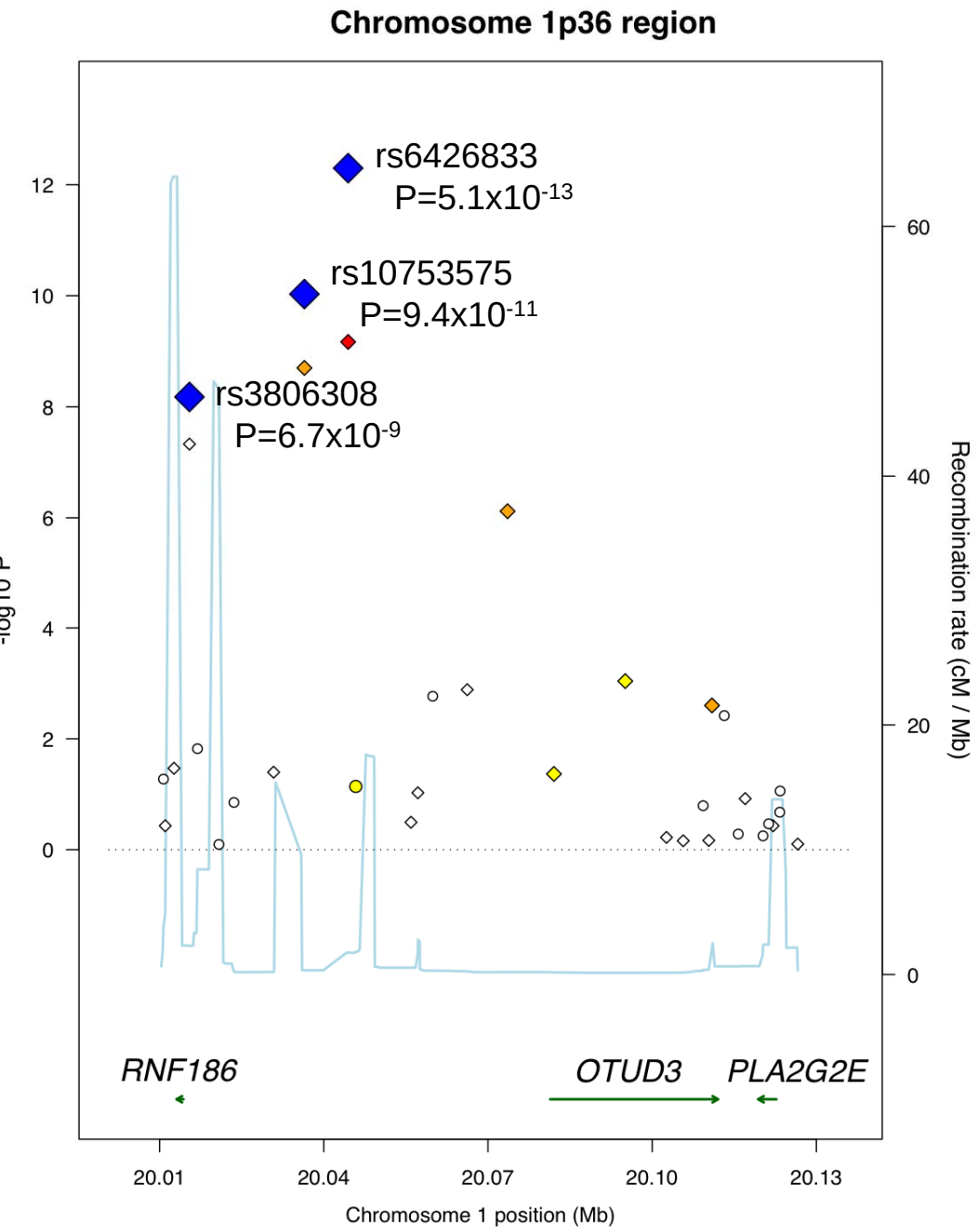


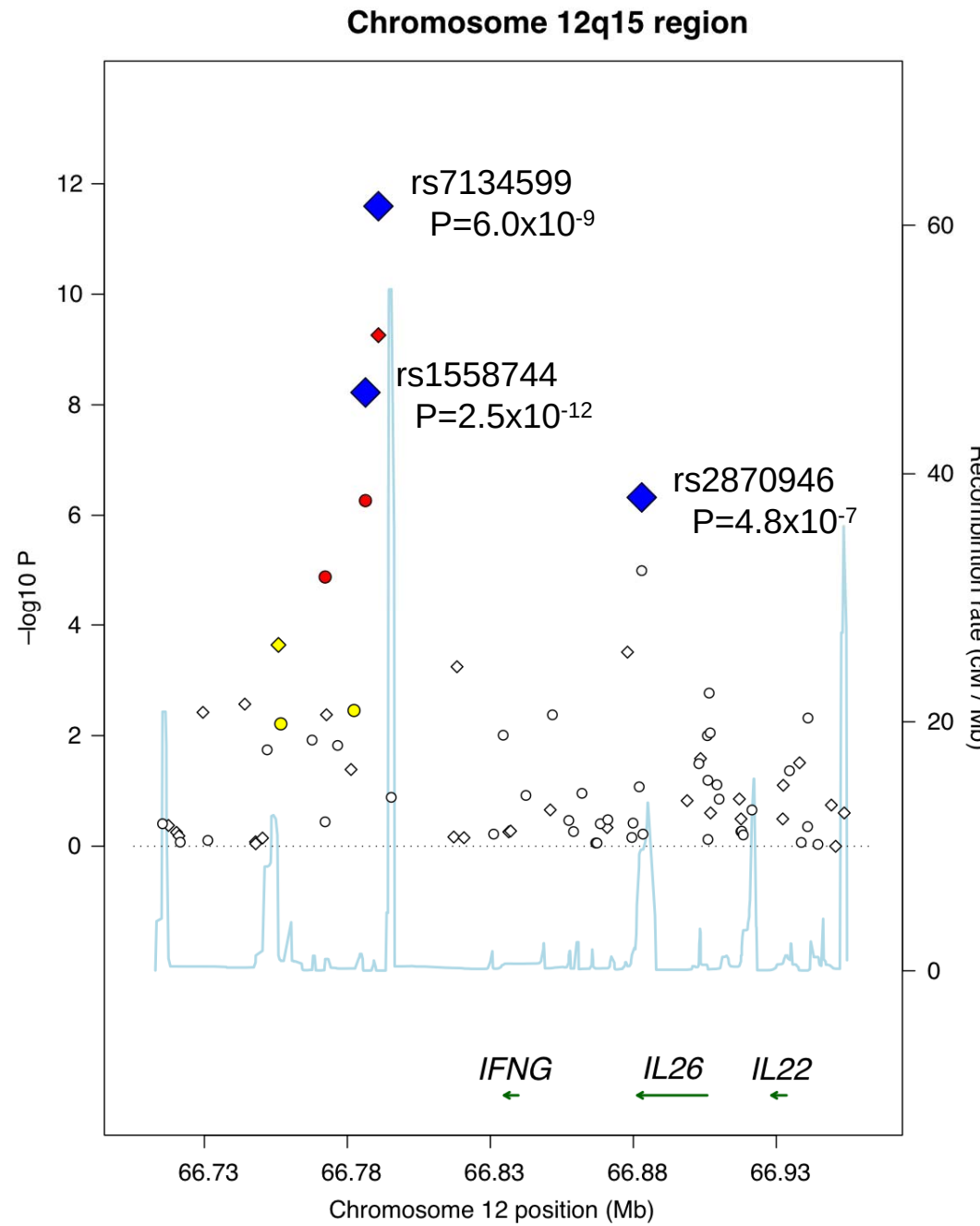
post-GEM genomic control inflation factor = 1.04

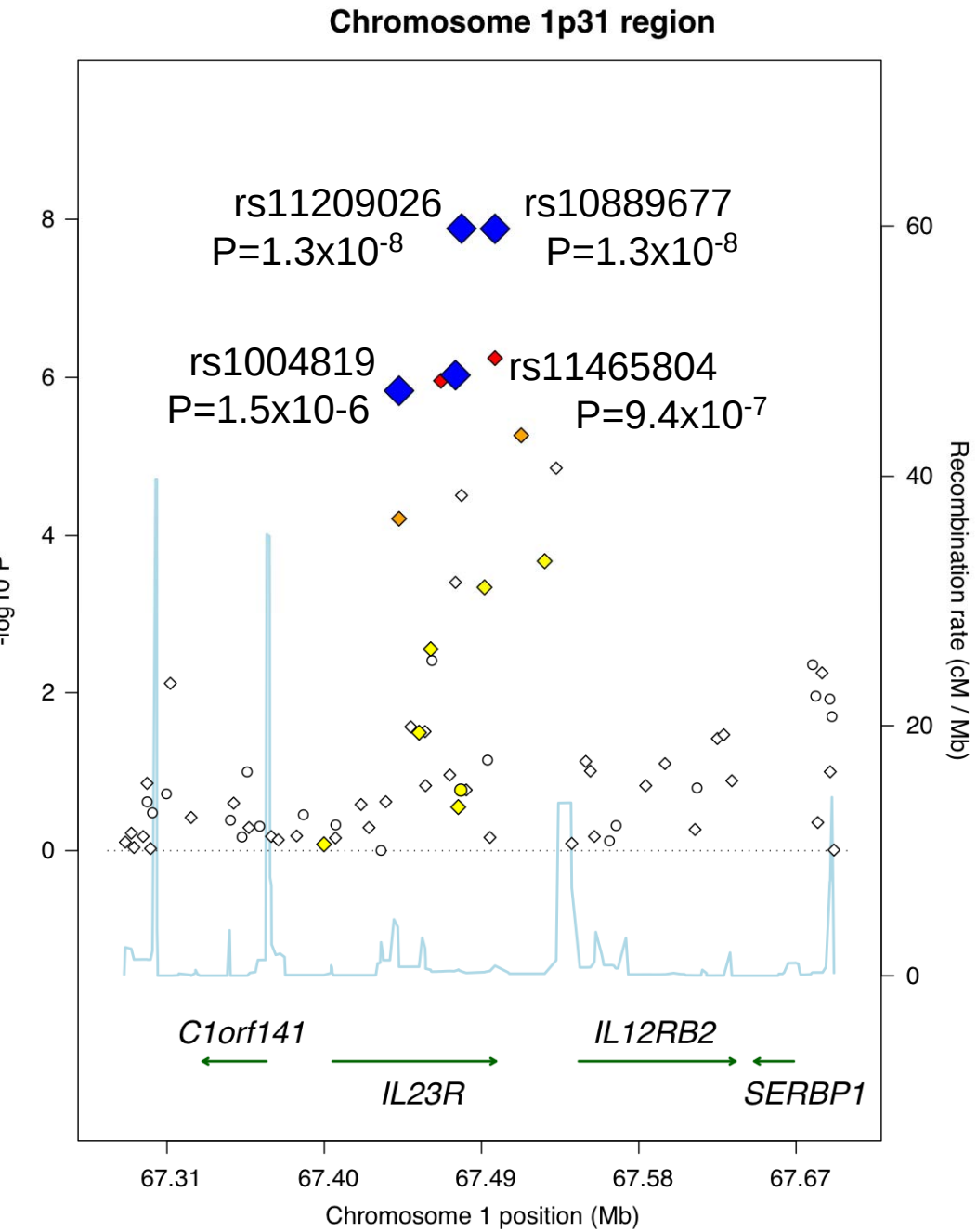
NIDDK IBDGC UC GWAS stage II

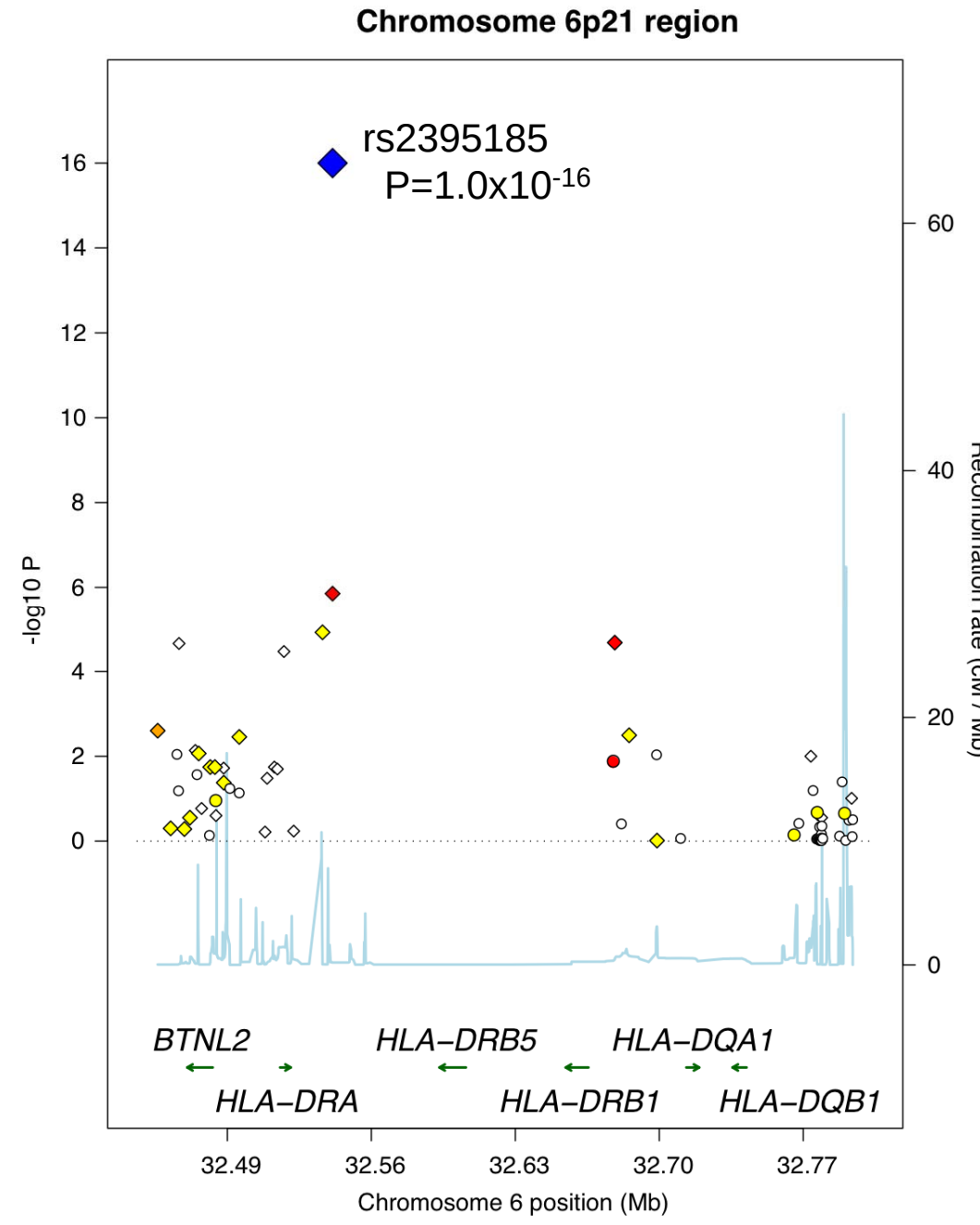
Study subjects	Raw data (number of samples)	Post quality control data (number of samples)
North American cases	769	768
North American controls	727	721
Italian cases	633	619
Italian controls	415	394

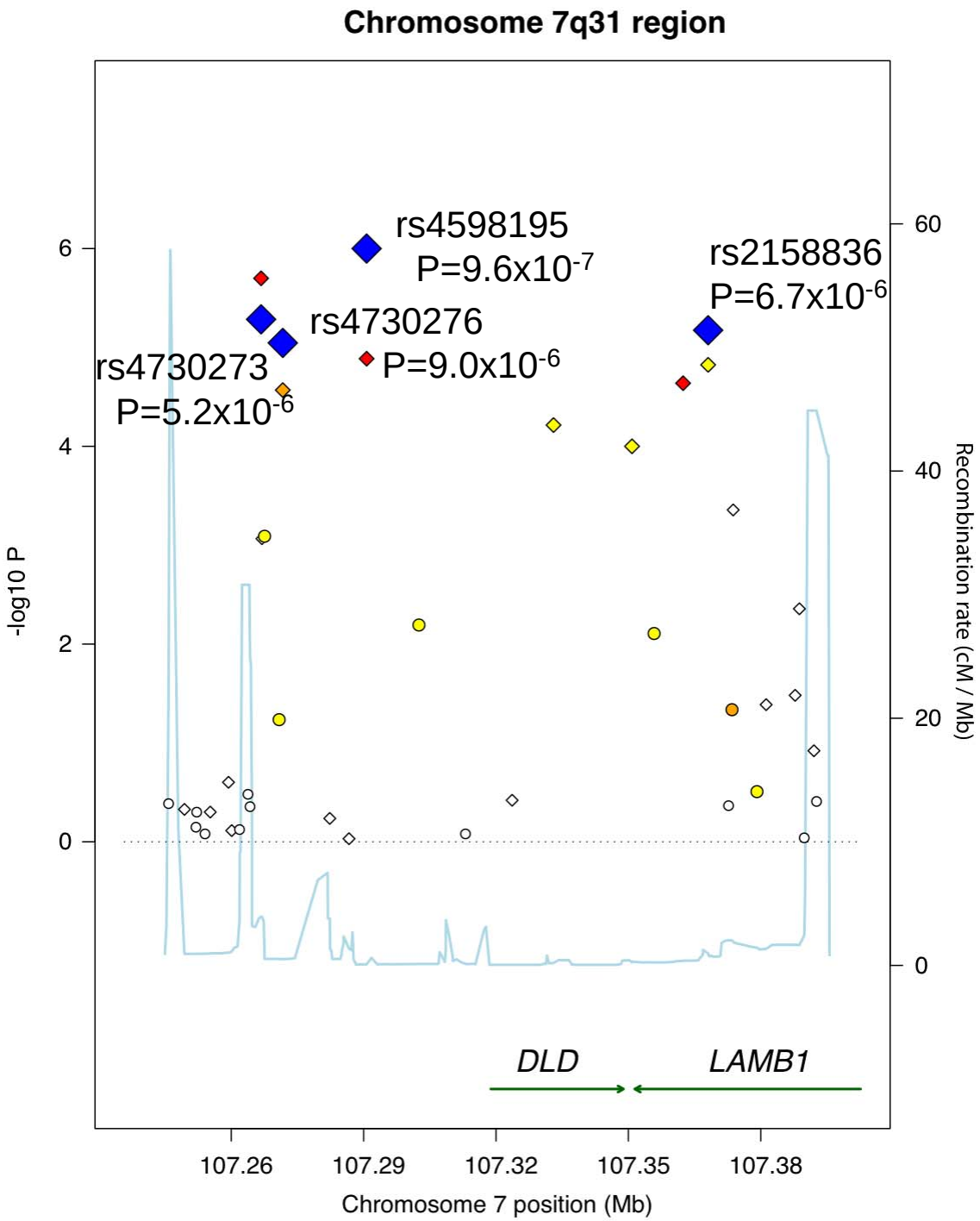
- successfully genotyped 54 of 65 independent ($r^2 < 0.5$) SNPs with $P < 1x10^{-4}$ in stage I
- Cochran-Mantel-Haenszel analysis of North American and Italian stage II case-control data
- stage I and stage II P values combined using Fisher method


Chromosome 1p36 region
rs6426833
P=5.1x10-13
rs10753575
P=9.4x10-11
rs3806308
P=6.7x10-9
RNF186
OTUD3
PLA2G2E
20.01
20.04
20.07
20.10
20.13
Chromosome 1 position (Mb)
Recombination rate (cM / Mb)


Chromosome 12q15 region
rs7134599
P=6.0x10-9
rs1558744
P=2.5x10-12
rs2870946
P=4.8x10-7
IFNG
IL26
IL22
66.73
66.78
66.83
66.88
66.93
Chromosome 12 position (Mb)


Chromosome 1p31 region
rs11209026
P=1.3x10-8
rs10889677
P=1.3x10-8
rs1004819
P=1.5x10-6
rs11465804
P=9.4x10-7
C1orf141
IL23R
IL12RB2
SERBP1
Recombination rate (cM / Mb)
67.31
67.40
67.49
67.58
67.67
Chromosome 1 position (Mb)


Chromosome 6p21 region
rs2395185
P=1.0x10-16
BTNL2
HLA–DRB5
HLA–DQA1
HLA–DRA
HLA–DRB1
HLA–DQB1
32.49
32.56
32.63
32.70
32.77
Chromosome 6 position (Mb)


Chromosome 7q31 region
rs4598195
P=9.6x10-7
rs2158836
P=6.7x10-6
rs4730276
P=9.0x10-6
rs4730273
P=5.2x10-6
-log10 P
Recombination rate (cM / Mb)
DLD
LAMB1
107.26
107.29
107.32
107.35
107.38
Chromosome 7 position (Mb)

UC GWAS Conclusions

- Loci identified in UC GWAS and confirmed in independent samples
  - chromosome 1p36 near *OTUD3* (OTU domain containing 3) *PLA2G2E* (phospholipase A2, group IIE), and *RNF186* (ring finger protein 186)
  - *IL23R* on chromosome 1p31
  - chromosome 1q32 near *IL10*
  - MHC on chromosome 6p21
  - chromosome 12q15 near *IFNG, IL26, IL22*
- Similar to the experience in CD, more UC loci will likely be found as additional UC GWAS and a UC GWAS meta-analysis are done

What next?

- Resequencing to identify low frequency variants
- “Fine mapping” including trans-racial/trans-ethnic mapping
- Copy number variants
- Multimarker model correlation with subphenotypes/disease outcomes
- Functional studies—GWAS findings have stimulated a resurgence of biological investigation!