Crohn's Disease Whole-genome Studies

Dan L. Nicolae

Departments of Medicine and Statistics
The University of Chicago

GAIN Analysis Workshop II
Bethesda, October 18, 2007

Crohn's disease

- type of inflammatory bowel disease (IBD); the other main form is ulcerative colitis (UC);
- causes inflammation of the digestive tract;
- can affect any area of the GI tract, from the mouth to the anus; most common - ileum.
- prevalence: 100-150 per 100,000 (European ancestry);
- Ashkenazi Jews - increased risk of developing Crohn's;
- African Americans, Hispanics and Asians - lower rates;
- forefront of genetic studies.

A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease

Yasunori Ogura*†, Denise K. Bonen‡†, Naohiro Inohara*, Dan L. Nicolae§, Felicia F. Chen*, Richard Ramos‡, Heidi Britton‡, Thomas Moran‡, Reda Karaliuskas‡, Richard H. Duerr‖, Jean-Paul Achkar¶, Steven R. Brant#, Theodore M. Bayless#, Barbara S. Kirschner☆, Stephen B. Hanauer‡, Gabriel Nuñez*†† & Judy H. Cho‡††

Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease

Jean-Pierre Hugot*†‡, Mathias Chamaillard*†, Habib Zouali*, Suzanne Lesage*, Jean-Pierre Cézard‡, Jacques Belaiche§, Sven Almer‖, Curt Tysk¶, Colm A. O'Morain#, Miquel Gassull☆, Vibeke Binder**, Yigael Finkel††, Antoine Cortot‡‡, Robert Modigliani§§, Pierre Laurent-Puig†, Corine Gower-Rousseau‡‡, Jeanne Macry‖‖, Jean-Frédéric Colombel‡‡, Mourad Sahbatou* & Gilles Thomas*†¶¶

Other loci

Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease

John D. Rioux[1], Mark J. Daly[1], Mark S. Silverberg[2,3], Kerstin Lindblad[1], Hillary Steinhart[2], Zane Cohen[4], Terrye Delmonte[1], Kerry Kocher[1], Katie Miller[1], Sheila Guschwan[1], Edward J. Kulbokas[1], Sinead O'Leary[1], Ellen Winchester[1], Ken Dewar[1], Todd Green[1], Valerie Stone[1], Christine Chow[1], Albert Cohen[7], Diane Langelier[8], Gilles Lapointe[9], Daniel Gaudet[9], Janet Faith[7], Nancy Branco[7], Shelley B. Bull[6], Robin S. McLeod[4], Anne M. Griffiths[5], Alain Bitton[7], Gordon R. Greenberg[2], Eric S. Lander[1,10,*], Katherine A. Siminovitch[2,3,*] & Thomas J. Hudson[1,7,*]

These authors co-directed the project.

Genetic variation in *DLG5* is associated with inflammatory bowel disease

Monika Stoll[1,7], Brit Corneliussen[2], Christine M Costello[1], Georg H Waetzig[3], Bjorn Mellgard[2], W Andreas Koch[1], Philip Rosenstiel[1], Mario Albrecht[4], Peter J P Croucher[1], Dirk Seegert[3], Susanna Nikolaus[1], Jochen Hampe[1,5], Thomas Lengauer[4], Stefan Pierrou[2], Ulrich R Foelsch[1], Christopher G Mathew[6], Maria Lagerstrom-Fermer[2] & Stefan Schreiber[1,5]

Genome-wide association scans

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

Richard H. Duerr,[1,2] Kent D. Taylor,[3,4] Steven R. Brant,[5,6] John D. Rioux,[7,8] Mark S. Silverberg,[9] Mark J. Daly,[8,10] A. Hillary Steinhart,[9] Clara Abraham,[11] Miguel Regueiro,[1] Anne Griffiths,[12] Themistocles Dassopoulos,[5] Alain Bitton,[13] Huiying Yang,[3,4] Stephan Targan,[4,14] Lisa Wu Datta,[5] Emily O. Kistner,[15] L. Philip Schumm,[15] Annette T. Lee,[16] Peter K. Gregersen,[16] M. Michael Barmada,[2] Jerome I. Rotter,[3,4] Dan L. Nicolae,[11,17] Judy H. Cho[18]*

A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in *ATG16L1*

Jochen Hampe[1,2,10], Andre Franke[1,10], Philip Rosenstiel[1,9], Andreas Till[1], Markus Teuber[1], Klaus Huse[3], Mario Albrecht[4], Gabriele Mayr[4], Francisco M De La Vega[5], Jason Briggs[5], Simone Günther[5], Natalie J Prescott[6], Clive M Onnie[6], Robert Häsler[1], Bence Sipos[7], Ulrich R Fölsch[2], Thomas Lengauer[4], Matthias Platzer[3], Christopher G Mathew[6], Michael Krawczak[8] & Stefan Schreiber[1,2]

OPEN ACCESS Freely available online PLoS GENETICS

Novel Crohn Disease Locus Identified by Genome-Wide Association Maps to a Gene Desert on 5p13.1 and Modulates Expression of *PTGER4*

Cécile Libioulle[1], Edouard Louis[2], Sarah Hansoul[1], Cynthia Sandor[1], Frédéric Farnir[1], Denis Franchimont[3], Séverine Vermeire[4], Olivier Dewit[5], Martine de Vos[6], Anna Dixon[7], Bruno Demarche[8], Ivo Gut[9], Simon Heath[9], Mario Foglio[9], Liming Liang[10], Debby Laukens[6], Myriam Mni[1], Diana Zelenika[9], André Van Gossum[3], Paul Rutgeerts[4], Jacques Belaiche[2], Mark Lathrop[9], Michel Georges[1]*

Genome-wide association scans (more...)

http://www.nature.com/naturegenetics

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

John D Rioux[1,2], Ramnik J Xavier[3], Kent D Taylor[4], Mark S Silverberg[5], Philippe Goyette[1], Alan Huett[3], Todd Green[2], Petric Kuballa[3], M Michael Barmada[6], Lisa Wu Datta[7], Yin Yao Shugart[8], Anne M Griffiths[9], Stephan R Targan[4], Andrew F Ippoliti[4], Edmond-Jean Bernard[10], Ling Mei[4], Dan L Nicolae[11], Miguel Regueiro[12], L Philip Schumm[13], A Hillary Steinhart[5], Jerome I Rotter[4], Richard H Duerr[6,12], Judy H Cho[14,16], Mark J Daly[2,15,16] & Steven R Brant[7,8,16]

Publishing Group http://www.nature.com/naturegenetics

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]

PNAS

Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci

John V. Raelson*†, Randall D. Little*, Andreas Ruether‡, Hélène Fournier*, Bruno Paquin*, Paul Van Eerdewegh*, W. E. C. Bradley§, Pascal Croteau*, Quynh Nguyen-Huu*, Jonathan Segal*, Sophie Debrus*, René Allard*, Philip Rosenstiel‡, Andre Franke‡, Gunnar Jacobs‡, Susanna Nikolaus¶‖, Jean-Michel Vidal*, Peter Szego‖, Nathalie Laplante*, Hilary F. Clark**, René J. Paulussen*, John W. Hooper*, Tim P. Keith*, Abdelmajid Belouchi*, and Stefan Schreiber‡¶‖

OPEN ACCESS Freely available online PLoS one

Systematic Association Mapping Identifies *NELL1* as a Novel IBD Disease Gene

Andre Franke[1], Jochen Hampe[2], Philip Rosenstiel[1], Christian Becker[3,4], Florian Wagner[4], Robert Häsler[1], Randall D. Little[5], Klaus Huse[6], Andreas Ruether[7], Tobias Balschun[1], Michael Wittig[7], Abdou ElSharawy[1], Gabriele Mayr[8], Mario Albrecht[8], Natalie J. Prescott[9], Clive M. Onnie[9], Hélène Fournier[5], Tim Keith[5], Uwe Radelof[4], Matthias Platzer[6], Christopher G. Mathew[9], Monika Stoll[10], Michael Krawczak[7,11], Peter Nürnberg[3,12], Stefan Schreiber[1,2,7]*

NIDDK IBD Genetics Consortium GWAS

- centers: University of Chicago (Yale), Cedars-Sinai, Johns Hopkins, University of Montreal, University of Pittsburgh, University of Toronto
- study design:
  - ileal Crohn's
  - non-Jewish: 547 cases and 548 controls
  - Jewish: 401 cases and 433 controls
- Illumina HumanHap300 BeadChip: 317,503 SNPs (308,332 autosomal)
- family-based cohort for replication (883 nuclear families); both CD and UC

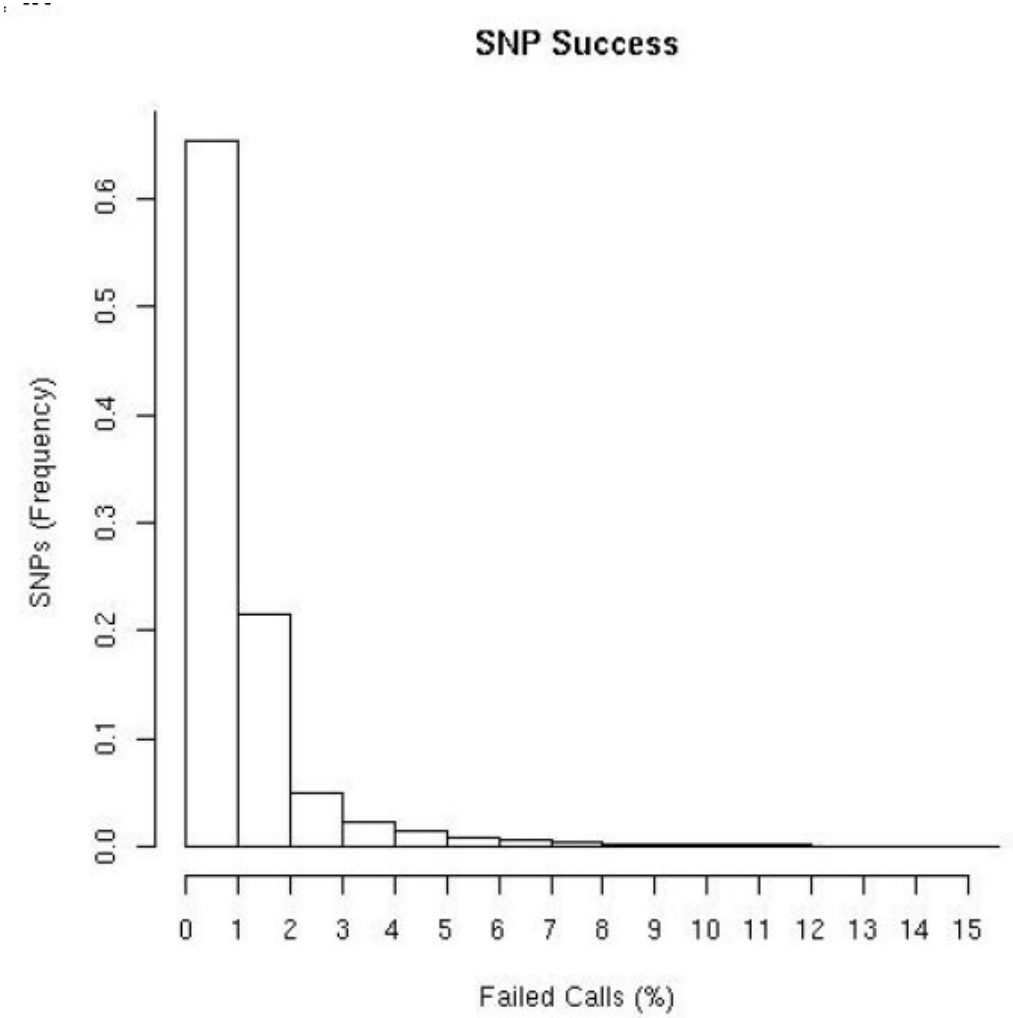
QC checks

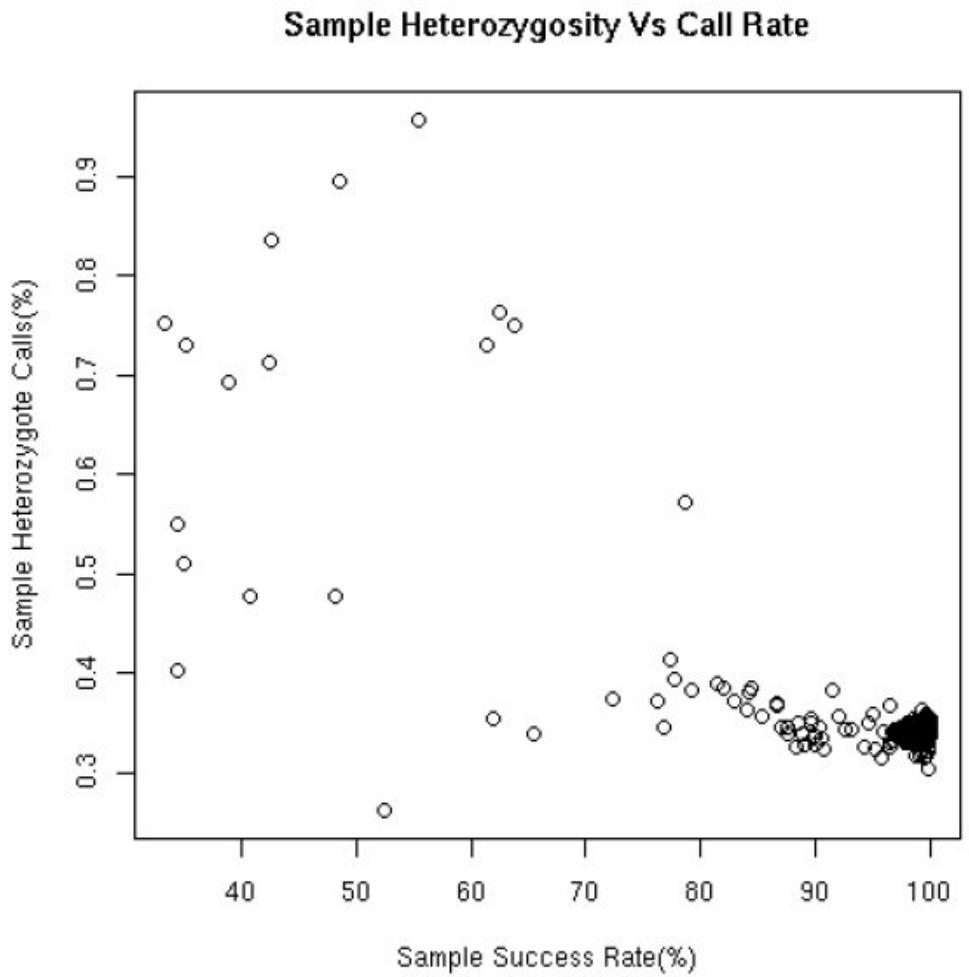
Sample quality filtering

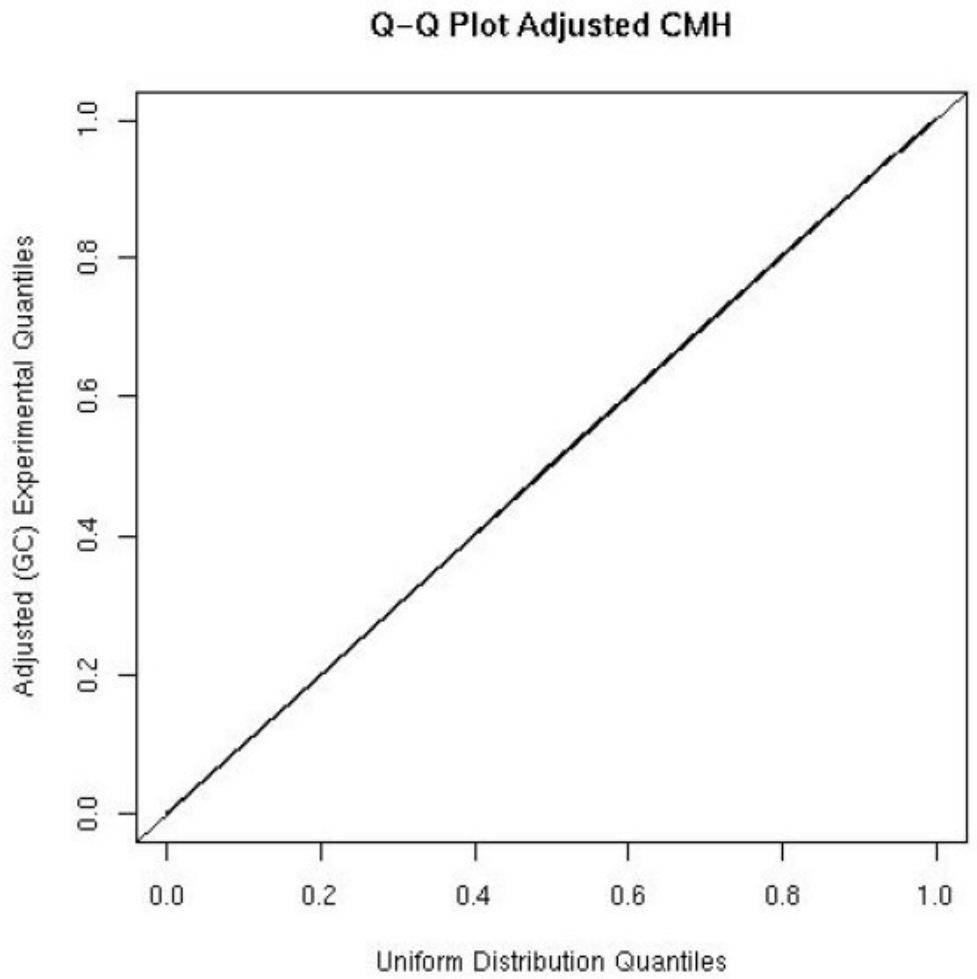
- relatedness check using genome-wide data (eight duplicate samples, ten related pairs);
- call rate threshold 93% (determined from heterozygosity)

SNP quality filtering (304,413 SNPs left; average call rate 99.35%)

- call rate threshold 95% (determined from HWE tests, genomic control)
- Hardy-Weinberg equilibrium test
- genomic control correction of 1.16


SNP Success
SNPs (Frequency)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Failed Calls (%)


Sample Heterozygosity Vs Call Rate
Sample Heterozygote Calls(%)
0.3
0.4
0.5
0.6
0.7
0.8
0.9
40
50
60
70
80
90
100
Sample Success Rate(%)


Q−Q Plot Adjusted CMH
Adjusted (GC) Experimental Quantiles
Uniform Distribution Quantiles
0.0
0.2
0.4
0.6
0.8
1.0

Interim Results (Duerr et al, 2006, Science)

- three SNPs genome-wide significant in the NJ scan
  - rs2066843 (p=2.86×10^{-9}) in NOD2
  - rs2076756 (p=5.12×10^{-10}) in NOD2
  - rs1120902 (p=5.05×10^{-9}) in IL23R
- IL23R
  - rs1120902 is a non-synonymous SNP (Arg381Gln)
  - multiple signals
  - IL23R protein - extracellular domain, a single transmembrane domain, and acytoplasmic domain
  - mouse models involve IL-23 in murine colitis, experimental autoimmune encephalitis, collagen-induced arthritis
  - blockade of the IL-23 signaling pathway - possible therapeutic strategy for IBD

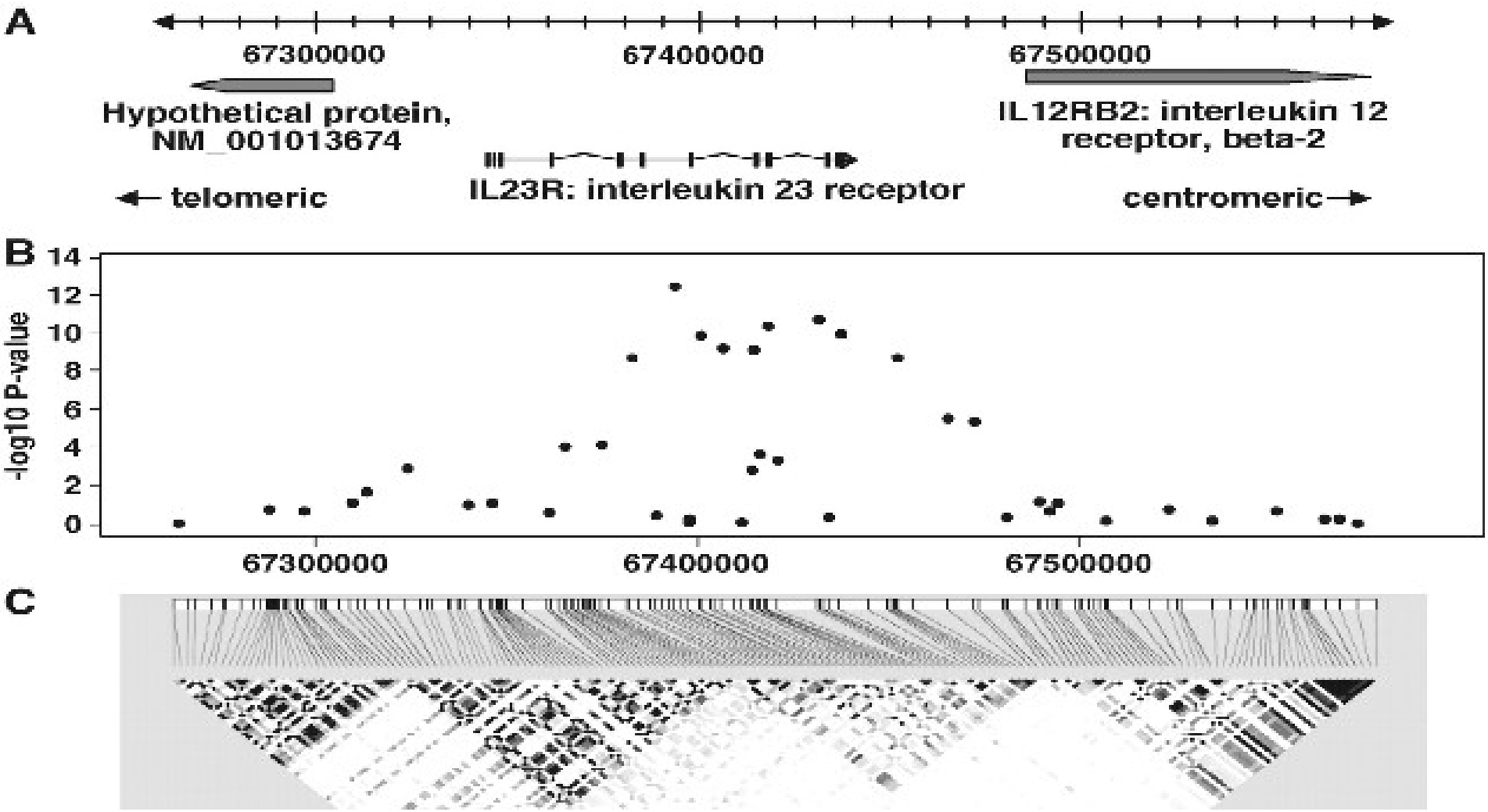

A
67300000
67400000
67500000
Hypothetical protein, NM_001013674
IL12RB2: interleukin 12 receptor, beta-2
IL23R: interleukin 23 receptor
telomeric
centromeric
B
-log10 P-value
14
12
10
8
6
4
2
0
67300000
67400000
67500000
C

Table 1. Non-Jewish and Jewish ileal Crohn's disease (CD) case-control association study results for IL23R region markers with *P*-values < 0.0001 in the non-Jewish cohort. Minor allele frequencies (MAF), allelic test *P*-values, and odds ratios (OR) with 95% confidence intervals (CI) are shown for each case-control cohort (*8*). The ORs shown are for the minor allele. Combined Cochran-Mantel-Haenszel *P*-values are also shown (*8*). UTR, untranslated region.

Marker	Location	Non-Jewish case-control cohort				Jewish case-control cohort				Combined *P*-value
		CD (n = 547) MAF	Control (n = 548) MAF	*P*-value	OR [95% CI]	CD (n = 401) MAF	Control (n = 433) MAF	*P*-value	OR [95% CI]	
rs1004819	Intron	0.374	0.280	3.79×10^{-6}	1.53 [1.27,1.84]	0.426	0.334	1.00×10^{-4}	1.48 [1.21,1.82]	1.54×10^{-9}
rs7517847	Intron	0.331	0.443	1.09×10^{-7}	0.62 [0.52,0.74]	0.240	0.352	5.84×10^{-7}	0.58 [0.47,0.72]	3.36×10^{-13}
rs10489629	Intron	0.378	0.475	4.27×10^{-6}	0.67 [0.56,0.80]	0.355	0.465	5.79×10^{-6}	0.63 [0.52,0.77]	1.14×10^{-10}
rs2201841	Intron	0.385	0.291	4.57×10^{-6}	1.52 [1.27,1.83]	0.414	0.315	2.92×10^{-5}	1.53 [1.25,1.89]	5.46×10^{-10}
rs11465804	Intron	0.020	0.063	7.52×10^{-7}	0.30 [0.18,0.51]	0.048	0.096	1.39×10^{-4}	0.47 [0.31,0.71]	5.97×10^{-10}
rs11209026	Arg381Gln	0.019	0.070	5.05×10^{-9}	0.26 [0.15,0.43]	0.033	0.070	7.95×10^{-4}	0.45 [0.27,0.73]	3.55×10^{-11}
rs1343151	Intron	0.275	0.370	2.26×10^{-6}	0.65 [0.54,0.78]	0.229	0.336	1.69×10^{-6}	0.59 [0.47,0.73]	1.64×10^{-11}
rs10889677	Exon-3'UTR	0.385	0.288	1.82×10^{-6}	1.55 [1.29,1.86]	0.419	0.316	1.51×10^{-5}	1.56 [1.27,1.91]	9.58×10^{-11}
rs11209032	Intergenic	0.393	0.293	1.03×10^{-6}	1.56 [1.30,1.87]	0.382	0.298	3.49×10^{-4}	1.45 [1.18,1.79]	1.60×10^{-9}
rs1495965	Intergenic	0.498	0.412	2.93×10^{-5}	1.44 [1.21,1.71]	0.469	0.412	2.04×10^{-2}	1.26 [1.03,1.53]	2.55×10^{-6}

Table 2. Family-based and combined (case-control and family-based) association results. Family-based association *P*-values were computed using the empirical variance estimator implemented in the FBAT software package (*8*). Combined Fisher *P*-values for all case-control (Table 1) and nuclear family cohorts are also shown (*8*). UTR, untranslated region.

Marker	Location	Non-Jewish CD (518 families, 651 affected offspring)	Non-Jewish UC (215 families, 251 affected offspring)	Jewish CD (77 families, 99 affected offspring)	Jewish UC (80 families, 91 affected offspring)	All IBD (883 families, 1,119 affected offspring)	Combined (family-based and case-control) *P*-value
		P-value	*P*-value	*P*-value	*P*-value	*P*-value	
rs1004819	Intron	3.60×10^{-5}	1.20×10^{-3}	1.24×10^{-2}	5.47×10^{-1}	6.06×10^{-8}	1.78×10^{-14}
rs7517847	Intron	2.30×10^{-5}	2.71×10^{-1}	3.50×10^{-2}	5.00×10^{-1}	1.80×10^{-5}	9.99×10^{-16}
rs10489629	Intron	1.87×10^{-3}	2.70×10^{-1}	4.33×10^{-1}	8.21×10^{-1}	1.27×10^{-3}	1.62×10^{-11}
rs2201841	Intron	5.80×10^{-4}	3.21×10^{-4}	3.50×10^{-2}	5.69×10^{-1}	1.04×10^{-7}	1.10×10^{-14}
rs11465804	Intron	1.32×10^{-4}	2.70×10^{-3}	8.90×10^{-5}	3.71×10^{-1}	3.46×10^{-9}	3.33×10^{-16}
rs11209026	Arg381Gln	8.00×10^{-6}	2.97×10^{-4}	9.41×10^{-4}	4.91×10^{-1}	1.32×10^{-10}	6.62×10^{-19}
rs1343151	Intron	9.63×10^{-2}	8.51×10^{-2}	3.30×10^{-2}	1.89×10^{-1}	1.24×10^{-3}	2.74×10^{-12}
rs10889677	Exon-3'UTR	2.60×10^{-3}	3.35×10^{-4}	5.88×10^{-2}	7.32×10^{-1}	1.65×10^{-6}	3.40×10^{-14}
rs11209032	Intergenic	2.68×10^{-3}	3.57×10^{-4}	3.48×10^{-2}	7.50×10^{-1}	2.41×10^{-6}	5.50×10^{-13}
rs1495965	Intergenic	4.07×10^{-4}	1.74×10^{-2}	3.93×10^{-2}	9.21×10^{-1}	1.72×10^{-5}	3.55×10^{-9}

Entire Dataset (Rioux et al., 2007, NG)

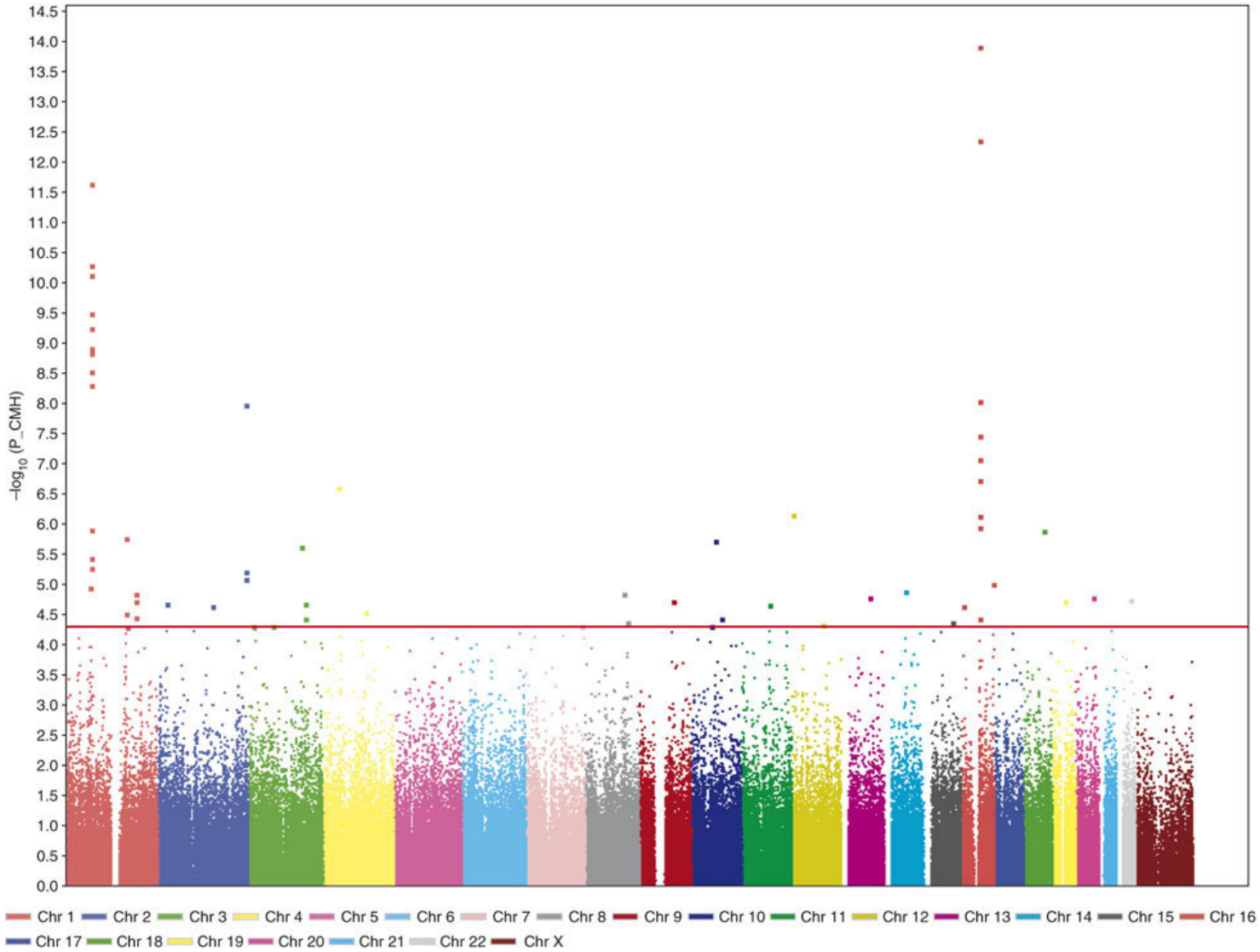


Table 1 Summary of the GWA study and replication studies

				GWA			Replication cohort 1		Replication cohort 2		Combined replication		
Rank	Number of SNPs	Chr	RS number	MAF iCD	MAF CTL	*P* value	T	U	MAF iCD	MAF CTL	OR	*P* value	Gene
1	8	16	rs2076756	0.358	0.244	7.01×10^{-14}	P.T.	P.T.	P.T.	P.T.	P.T.	P.T.	*CARD15*
2	13	1	rs7517847	0.295	0.403	3.06×10^{-12}	P.T.	P.T.	P.T.	P.T.	P.T.	P.T.	*IL23R*
3	3	2	**rs2241880**	0.364	0.453	6.38×10^{-8}	**220**	**306**	0.353	0.478	0.68	4.1×10^{-8}	*ATG16L1*
4	1	4	**rs16853571**	0.038	0.077	7.68×10^{-7}	**39**	**75**	0.057	0.047	0.69	0.0084	*PHOX2B*
5	1	12	rs886898	0.156	0.102	1.93×10^{-6}	121	136	N.D.	N.D.	N.D.	N.D.	–
6	2	1	rs2343331	0.279	0.212	2.49×10^{-6}	Failed	Failed	N.D.	N.D.	N.D.	N.D.	–
7	1	18	rs937815	0.054	0.094	3.25×10^{-6}	96	99	N.D.	N.D.	N.D.	N.D.	–
8	1	3	rs6439924	0.218	0.160	6.00×10^{-6}	166	140	N.D.	N.D.	N.D.	N.D.	–
9	1	10	**rs224136**	0.134	0.191	7.90×10^{-6}	**94**	**149**	0.140	0.230	0.60	2.9×10^{-7}	Intergenic
10	1	9	rs10821091	0.399	0.332	1.44×10^{-5}	274	252	N.D.	N.D.	N.D.	N.D.	–
11	1	14	rs1188157	0.487	0.417	1.58×10^{-5}	254	240	N.D.	N.D.	N.D.	N.D.	–
12	1	1	rs2819130	0.177	0.126	2.10×10^{-5}	130	144	N.D.	N.D.	N.D.	N.D.	–
13	1	11	rs2712800	0.373	0.441	2.38×10^{-5}	242	222	N.D.	N.D.	N.D.	N.D.	–
14	1	22	**rs4821544**	0.397	0.333	2.89×10^{-5}	**267**	**221**	0.374	0.339	1.19	0.0090	*NCF4*
15	1	2	rs6733000	0.081	0.124	3.03×10^{-5}	81	77	N.D.	N.D.	N.D.	N.D.	–
16	1	2	rs7603516	0.064	0.102	3.10×10^{-5}	73	62	N.D.	N.D.	N.D.	N.D.	–
17	1	16	**rs8050910**	0.388	0.458	3.28×10^{-5}	**221**	**271**	0.400	0.430	0.84	0.0085	*FAM92B*
18	2	1	rs2490271	0.206	0.152	3.44×10^{-5}	175	166	N.D.	N.D.	N.D.	N.D.	–
19	1	20	rs4810663	0.236	0.180	3.45×10^{-5}	182	178	N.D.	N.D.	N.D.	N.D.	–
20	1	8	rs10505007	0.400	0.332	3.78×10^{-5}	221	248	N.D.	N.D.	N.D.	N.D.	–
21	1	8	rs2044999	0.330	0.395	3.84×10^{-5}	NT	NT	N.D.	N.D.	N.D.	N.D.	–
22	1	9	rs4878061	0.418	0.485	4.64×10^{-5}	NT	NT	N.D.	N.D.	N.D.	N.D.	–
23	1	13	**rs11617463**	0.044	0.077	4.85×10^{-5}	**59**	**80**	Failed	Failed	N.D.	N.D.	–

Where next?

- several replicated genes/regions;
- more studies planned; e.g. joint analysis of the NIDDK, WTCCC, Belgium/France Consortia datasets (Mark Daly presentation at ASHG);
- better understanding of risk variation in identified genes; e.g. for IL23R, genotype additional variation for complete coverage; sequencing projects;
- better understanding of phenotype-genotype association (age of onset, disease location, GxE);
- interactions (e.g. no obvious interaction between NOD2 and IL23R)

Acknowledgments

NIDDK Consortium

Yale University	Judy H. Cho
University of Toronto	Mark S. Silverberg
Johns Hopkins University	Steven R. Brant
University of Pittsburgh	Richard H. Duerr
University of Montreal	John D. Rioux Mark J. Daly (Broad)
Cedars-Sinai	Kent D. Taylor
University of Chicago	Philip Schumm Emily O. Kistner