What constitutes evidence 2012

Mark Daly
Chief, Analytic and Translational Genetics Unit
Massachusetts General Hospital
&
Co-director, Medical and Population Genetics
Broad Institute

The story of dysbindin: 2002-2012

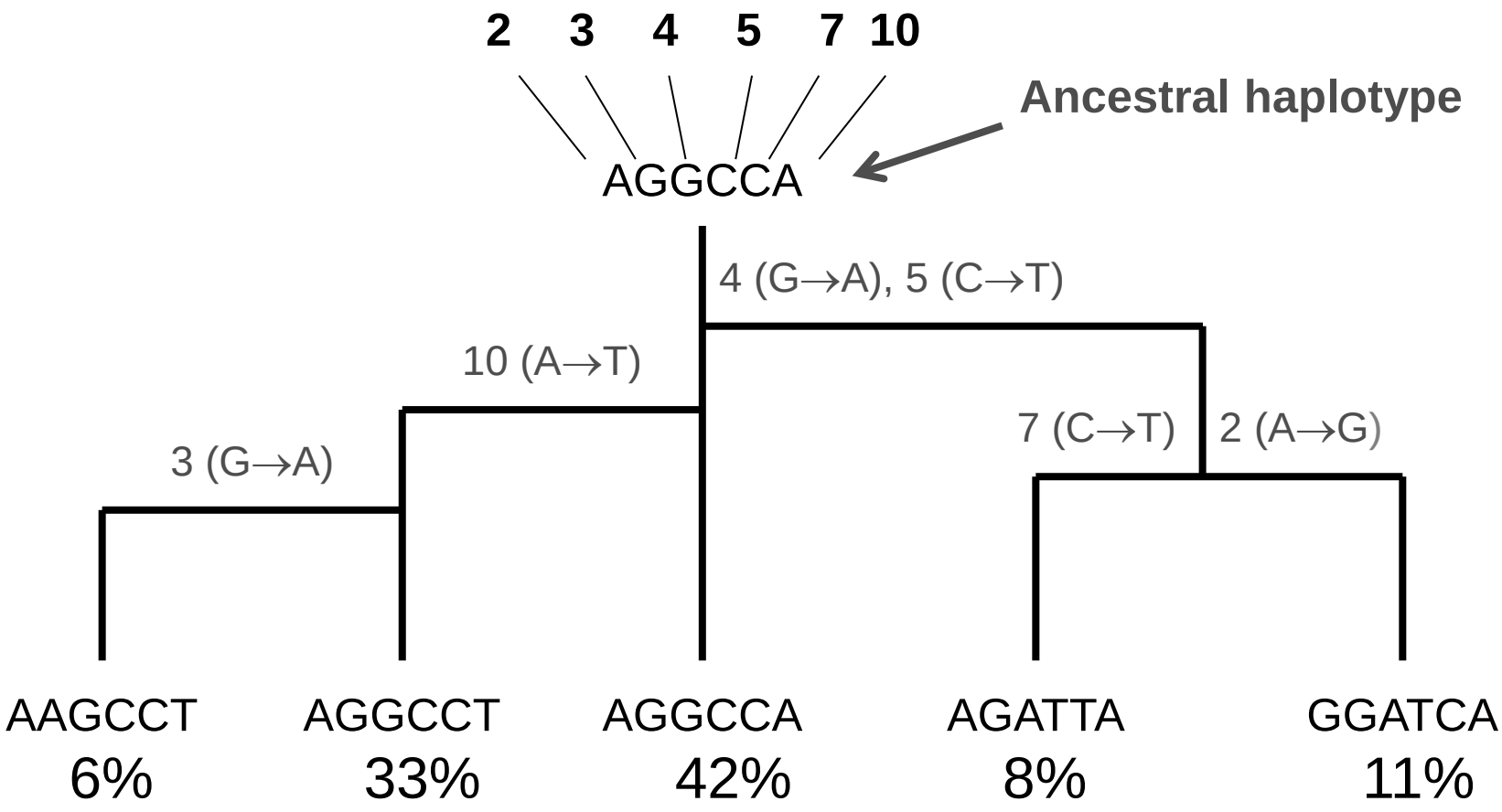
Am. J. Hum. Genet. 71:337–348, 2002

Genetic Variation in the 6p22.3 Gene *DTNBP1*, the Human Ortholog of the Mouse Dysbindin Gene, Is Associated with Schizophrenia

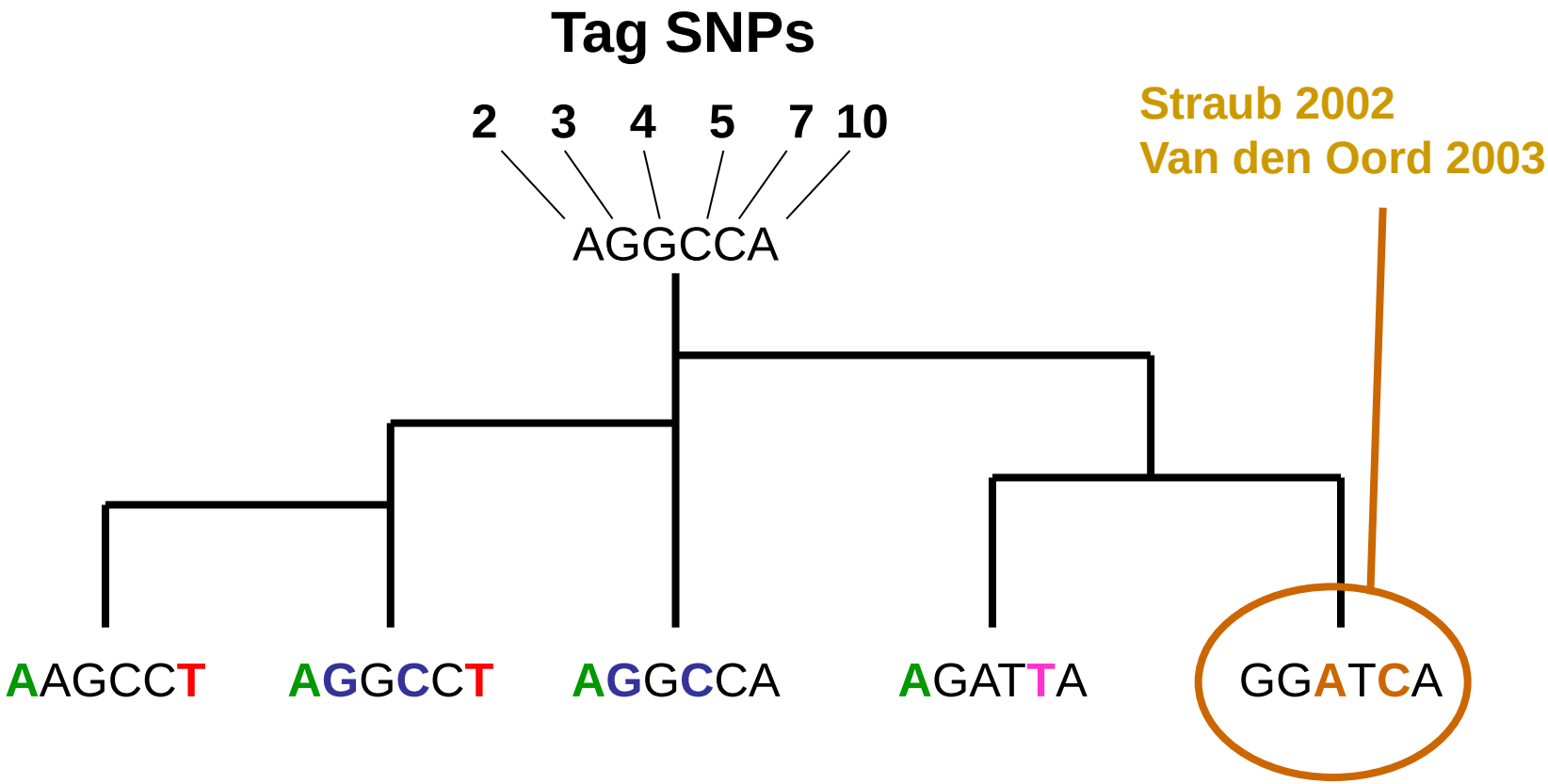
Uncorrected, empirical *P* values produced by the program TRANSMIT were significant ($P < .01$) for a number of individual SNP markers, and most remained significant when the data were restricted to include only one affected offspring per nuclear family per extended pedigree; multiple three-marker haplotypes were highly significant ($P = .008–.0001$) under the restricted conditions.

Extensive LD across DTNBP1

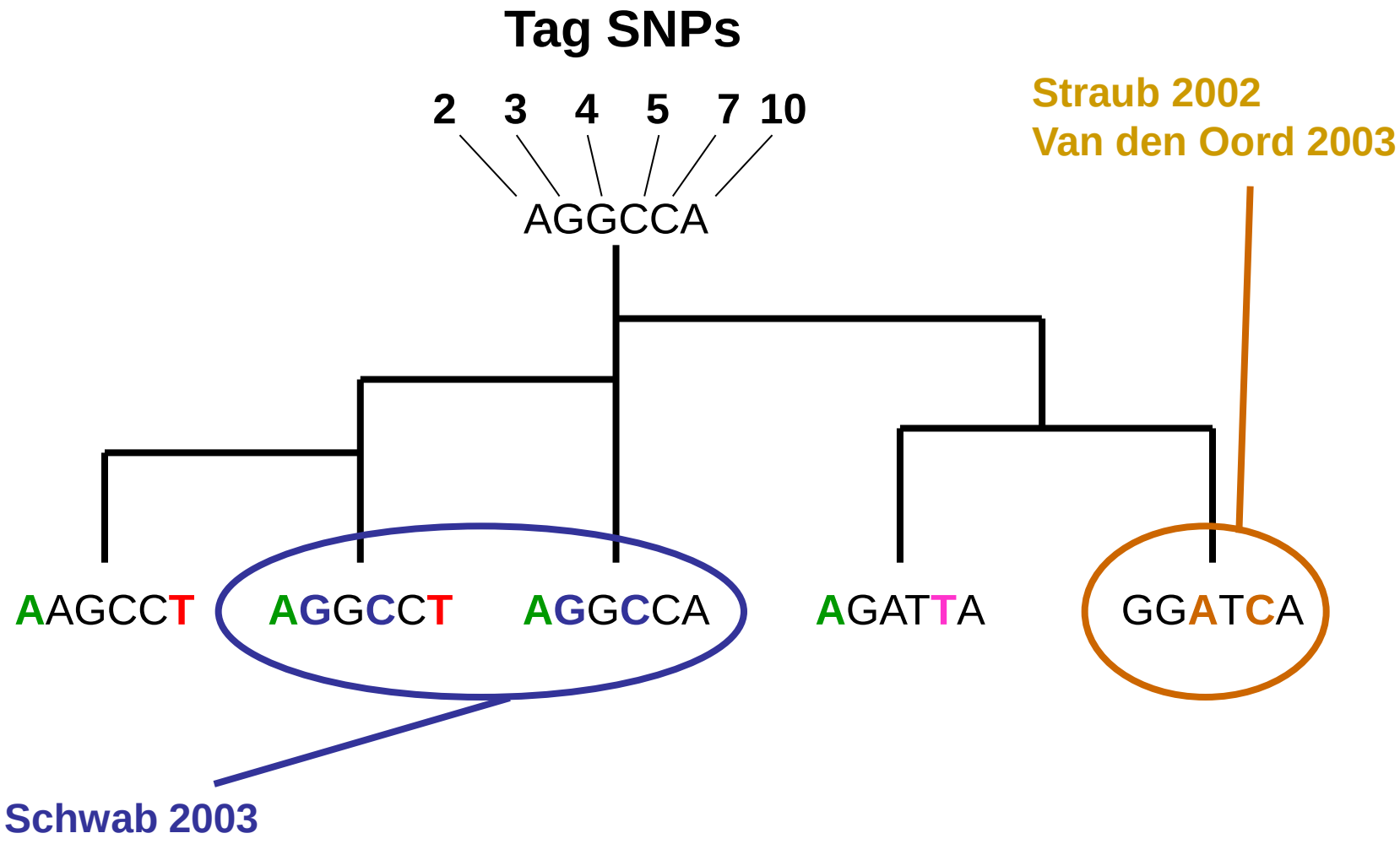
Phylogeny of DTNBP1 tag SNPs



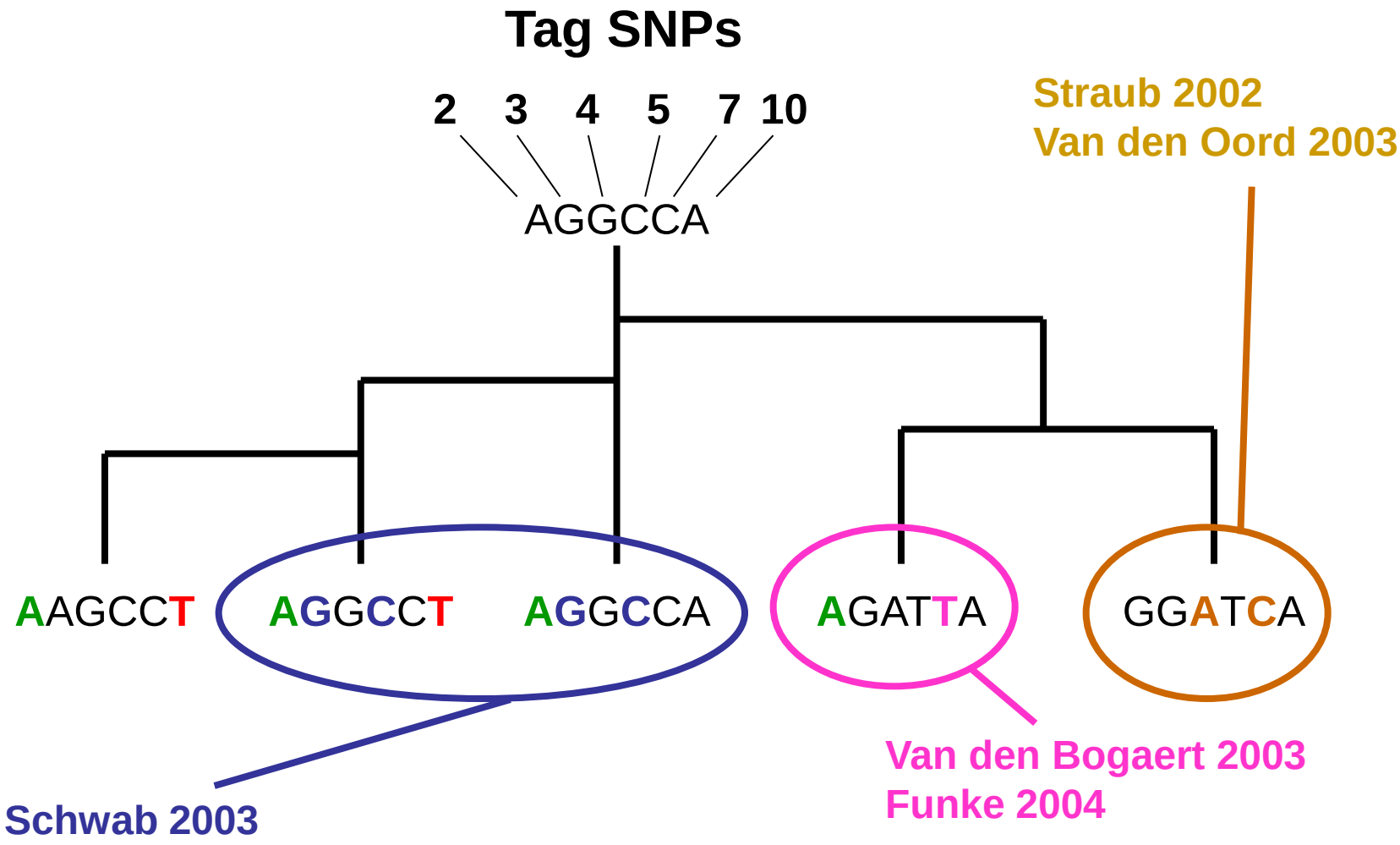
Associated alleles reported



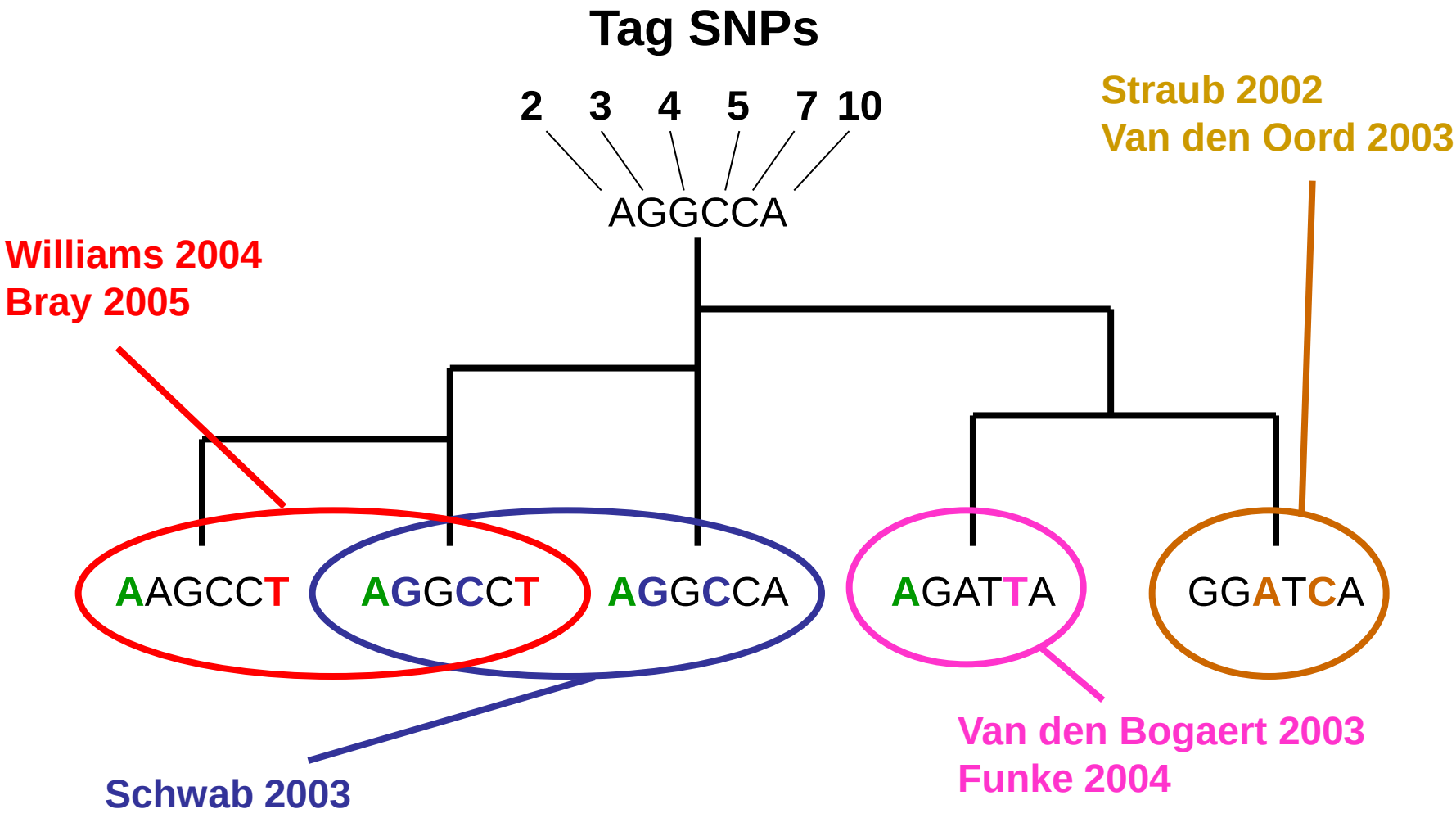
Associated alleles reported



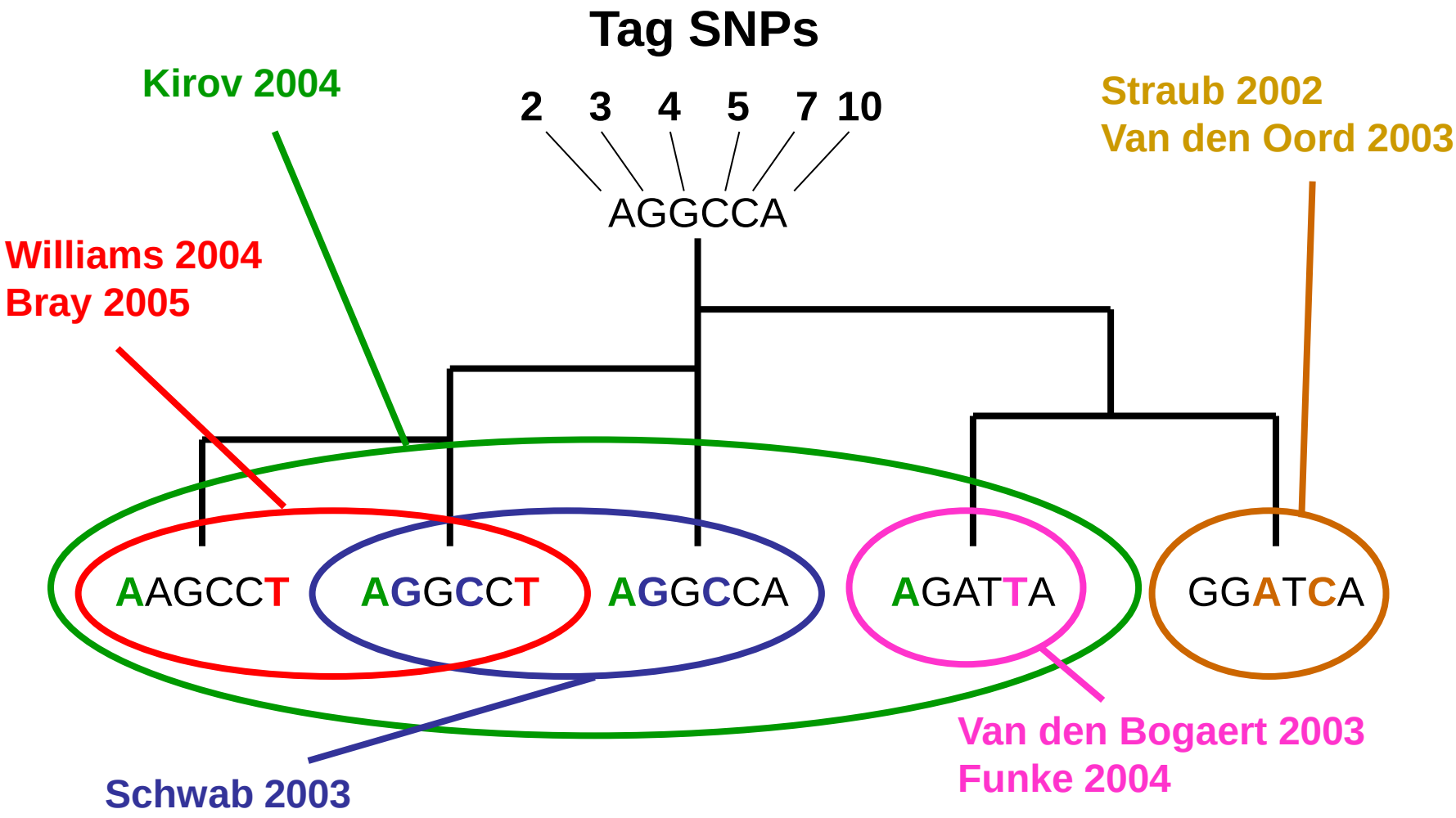
Associated alleles reported



Associated alleles reported



Associated alleles reported



November 2006 - AJHG

ARTICLE

Analysis of High-Resolution HapMap of *DTNBP1* (Dysbindin) Suggests No Consistency between Reported Common Variant Associations and Schizophrenia

Mousumi Mutsuddi,* Derek W. Morris,* Skye G. Waggoner, Mark J. Daly, Edward M. Scolnick, and Pamela Sklar

"***Evidence of association is, at present, equivocal and unsatisfactory***."

* No consistently associated SNP/haplotype pattern across studies

* All studies (European-derived populations) had allele/haplotype frequencies compatible with HapMap-CEU sample

* Reference variation sets such as HapMap can successfully relate associations from diverse marker sets

What's happened since then?

- There are 238 PubMed entries when searching "DTNBP1 and schizophrenia" – 168 of these in 2007-2012 !!

- September 2011 – World Congress of Psychiatric Genetics – an entire session was devoted to dysbindin

- No evidence supporting an association to DTNBP1 in current GWAS meta-analysis (Sept 2011, Nat Gen, n>9000 cases)

  all SNPs within 100 kb the gene have p > .001

Reversing the curse: the story of GWAS

1996: Risch and Merikangas propose that a **p**-value of **5×10^{-8}** (equivalent to a **p**-value of 0.05 after a Bonferroni CORRECTION for 1 million independent tests) is a conservative threshold for declaring significant association in a genome-wide study.

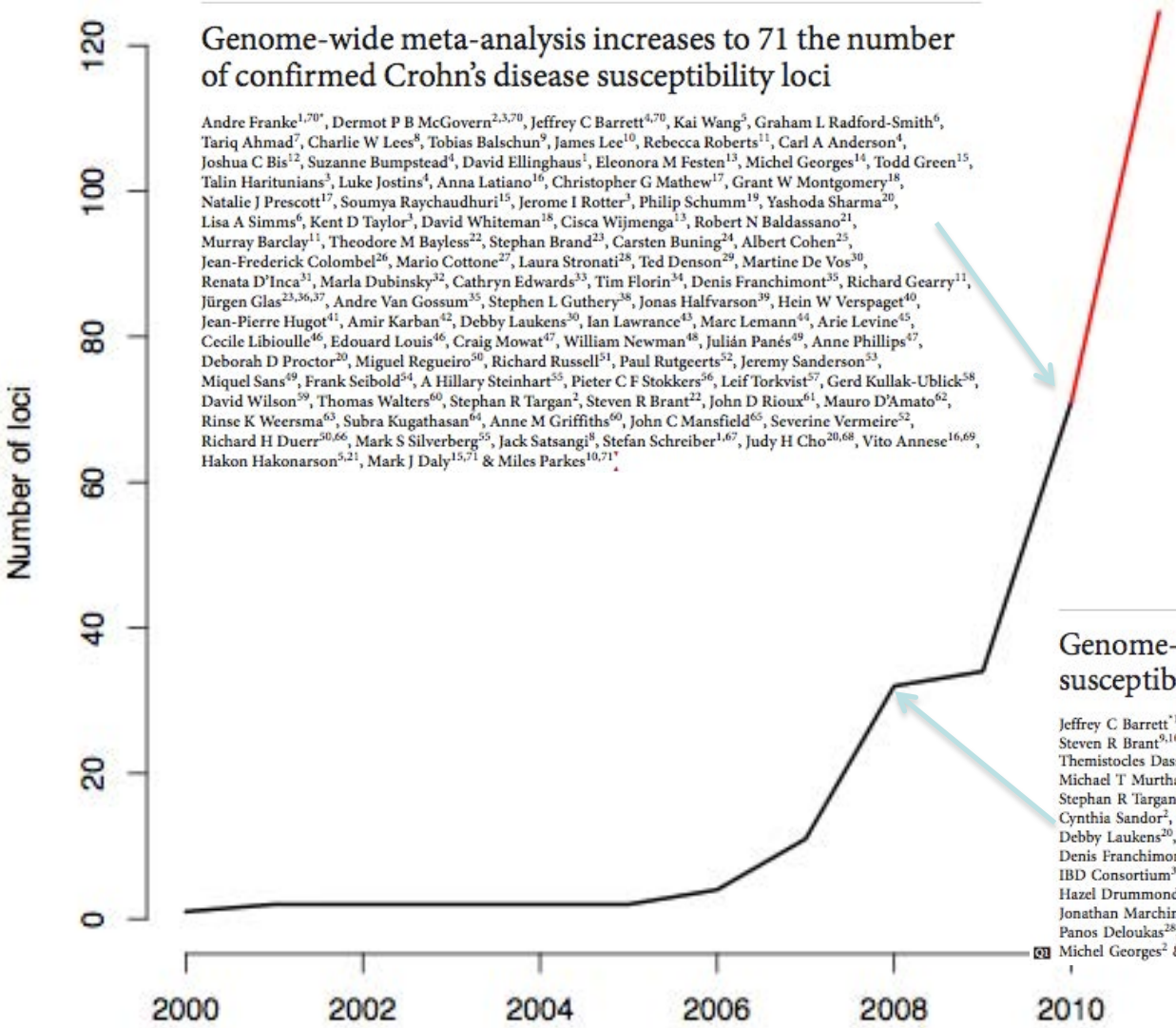
2008: 3 groups publish empirically derived estimates using quite different approaches and data sets that estimate appropriate dense-map numbers to be in the range of **2.5 to 7.2×10^{-8}**

Crohn's Disease: Barrett et al 2008

Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease

32 genome-wide significant hits defined – $p<5x10^{-8}$

Crohn's Disease: Jostins Ripke et al 2012



Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci

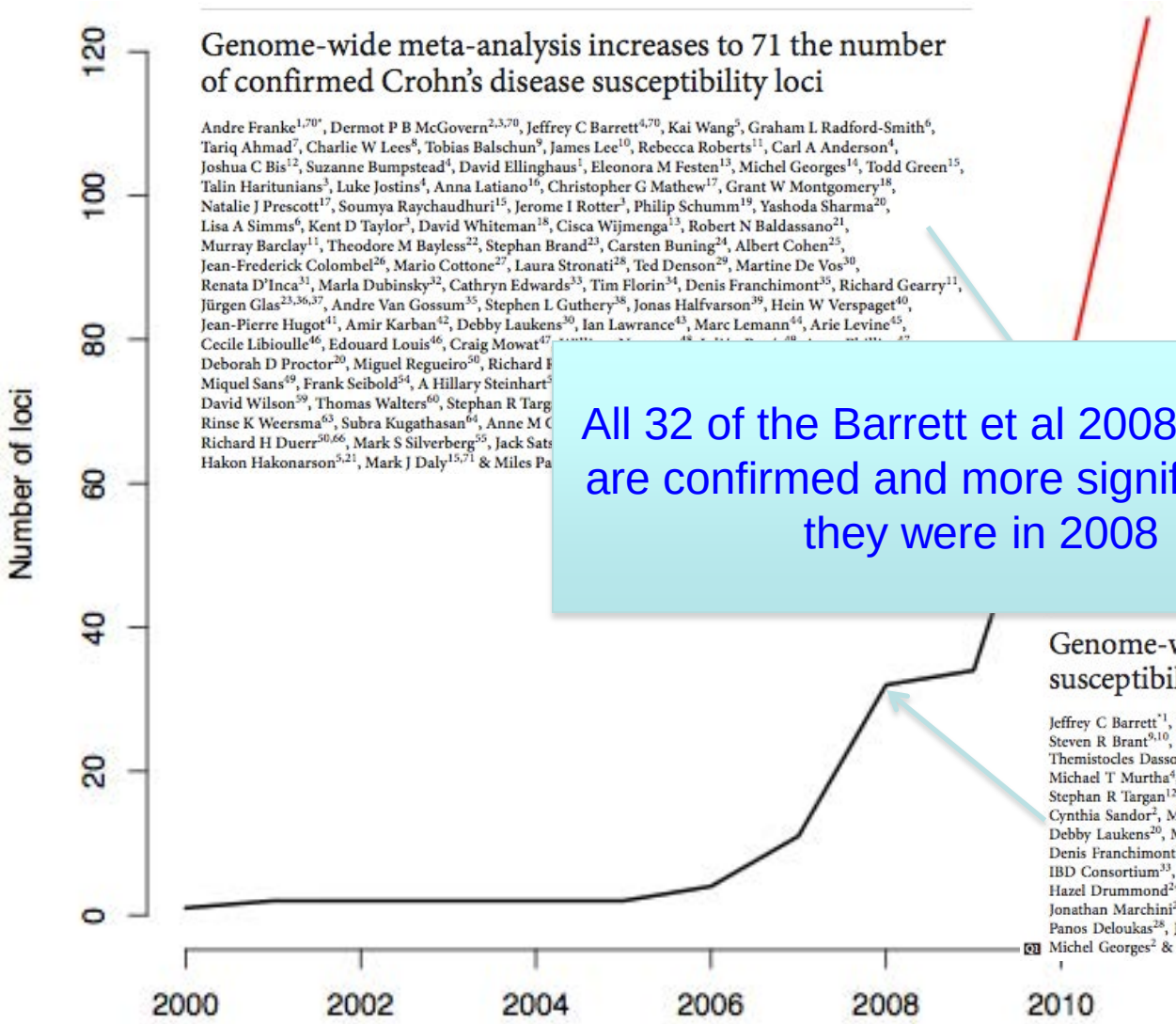
Andre Franke[1,70*], Dermot P B McGovern[2,3,70], Jeffrey C Barrett[4,70], Kai Wang[5], Graham L Radford-Smith[6], Tariq Ahmad[7], Charlie W Lees[8], Tobias Balschun[9], James Lee[10], Rebecca Roberts[11], Carl A Anderson[4], Joshua C Bis[12], Suzanne Bumpstead[4], David Ellinghaus[1], Eleonora M Festen[13], Michel Georges[14], Todd Green[15], Talin Haritunians[3], Luke Jostins[4], Anna Latiano[16], Christopher G Mathew[17], Grant W Montgomery[18], Natalie J Prescott[17], Soumya Raychaudhuri[15], Jerome I Rotter[3], Philip Schumm[19], Yashoda Sharma[20], Lisa A Simms[6], Kent D Taylor[3], David Whiteman[18], Cisca Wijmenga[13], Robert N Baldassano[21], Murray Barclay[11], Theodore M Bayless[22], Stephan Brand[23], Carsten Buning[24], Albert Cohen[25], Jean-Frederick Colombel[26], Mario Cottone[27], Laura Stronati[28], Ted Denson[29], Martine De Vos[30], Renata D'Inca[31], Marla Dubinsky[32], Cathryn Edwards[33], Tim Florin[34], Denis Franchimont[35], Richard Gearry[11], Jürgen Glas[23,36,37], Andre Van Gossum[35], Stephen L Guthery[38], Jonas Halfvarson[39], Hein W Verspaget[40], Jean-Pierre Hugot[41], Amir Karban[42], Debby Laukens[30], Ian Lawrance[43], Marc Lemann[44], Arie Levine[45], Cecile Libioulle[46], Edouard Louis[46], Craig Mowat[47], William Newman[48], Julián Panés[49], Anne Phillips[47], Deborah D Proctor[20], Miguel Regueiro[50], Richard Russell[51], Paul Rutgeerts[52], Jeremy Sanderson[53], Miquel Sans[49], Frank Seibold[54], A Hillary Steinhart[55], Pieter C F Stokkers[56], Leif Torkvist[57], Gerd Kullak-Ublick[58], David Wilson[59], Thomas Walters[60], Stephan R Targan[2], Steven R Brant[22], John D Rioux[61], Mauro D'Amato[62], Rinse K Weersma[63], Subra Kugathasan[64], Anne M Griffiths[60], John C Mansfield[65], Severine Vermeire[52], Richard H Duerr[50,66], Mark S Silverberg[55], Jack Satsangi[8], Stefan Schreiber[1,67], Judy H Cho[20,68], Vito Annese[16,69], Hakon Hakonarson[5,21], Mark J Daly[15,71] & Miles Parkes[10,71*]

Jostins, Ripke et al
Nature, in press
163 distinct confirmed associated loci in IBD (CD and/or UC)
N=75,000 samples

Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease

Jeffrey C Barrett[*1], Sarah Hansoul[2], Dan L Nicolae[3], Judy H Cho[4], Richard H Duerr[5,6], John D Rioux[7,8], Steven R Brant[9,10], Mark S Silverberg[11], Kent D Taylor[12], M Michael Barmada[5], Alain Bitton[13], Themistocles Dassopoulos[9], Lisa Wu Datta[9], Todd Green[8], Anne M Griffiths[14], Emily O Kistner[15], Michael T Murtha[4], Miguel D Regueiro[6], Jerome I Rotter[12], L Philip Schumm[15], A Hillary Steinhart[11], Stephan R Targan[12], Ramnik J Xavier[16], the NIDDK IBD Genetics Consortium[33], Cécile Libioulle[2], Cynthia Sandor[2], Mark Lathrop[17], Jacques Belaiche[18], Olivier Dewit[19], Ivo Gut[17], Simon Heath[17], Debby Laukens[20], Myriam Mni[2], Paul Rutgeerts[21], André Van Gossum[22], Diana Zelenika[17], Denis Franchimont[22], JP Hugot[23], Martine de Vos[20], Severine Vermeire[21], Edouard Louis[18], the Belgian-Frenc IBD Consortium[33], the Wellcome Trust Case Control Consortium[33], Lon R Cardon[1], Carl A Anderson[1], Hazel Drummond[24], Elaine Nimmo[24], Tariq Ahmad[25], Natalie J Prescott[26], Clive M Onnie[26], Sheila A Fisher[2 Jonathan Marchini[27], Jilur Ghori[28], Suzannah Bumpstead[28], Rhian Gwillam[28], Mark Tremelling[29], Panos Deloukas[28], John Mansfield[30], Derek Jewell[31], Jack Satsangi[24], Christopher G Mathew[26], Miles Parkes[29], Michel Georges[2] & Mark J Daly[8,32]

Crohn's Disease: Jostins Ripke et al 2012



Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci

Andre Franke[1,70*], Dermot P B McGovern[2,3,70], Jeffrey C Barrett[4,70], Kai Wang[5], Graham L Radford-Smith[6], Tariq Ahmad[7], Charlie W Lees[8], Tobias Balschun[9], James Lee[10], Rebecca Roberts[11], Carl A Anderson[4], Joshua C Bis[12], Suzanne Bumpstead[4], David Ellinghaus[1], Eleonora M Festen[13], Michel Georges[14], Todd Green[15], Talin Haritunians[3], Luke Jostins[4], Anna Latiano[16], Christopher G Mathew[17], Grant W Montgomery[18], Natalie J Prescott[17], Soumya Raychaudhuri[15], Jerome I Rotter[3], Philip Schumm[19], Yashoda Sharma[20], Lisa A Simms[6], Kent D Taylor[3], David Whiteman[18], Cisca Wijmenga[13], Robert N Baldassano[21], Murray Barclay[11], Theodore M Bayless[22], Stephan Brand[23], Carsten Buning[24], Albert Cohen[25], Jean-Frederick Colombel[26], Mario Cottone[27], Laura Stronati[28], Ted Denson[29], Martine De Vos[30], Renata D'Inca[31], Marla Dubinsky[32], Cathryn Edwards[33], Tim Florin[34], Denis Franchimont[35], Richard Gearry[11], Jürgen Glas[23,36,37], Andre Van Gossum[35], Stephen L Guthery[38], Jonas Halfvarson[39], Hein W Verspaget[40], Jean-Pierre Hugot[41], Amir Karban[42], Debby Laukens[30], Ian Lawrance[43], Marc Lemann[44], Arie Levine[45], Cecile Libioulle[46], Edouard Louis[46], Craig Mowat[47], Deborah D Proctor[20], Miguel Regueiro[50], Richard H Miquel Sans[49], Frank Seibold[54], A Hillary Steinhart[5] David Wilson[59], Thomas Walters[60], Stephan R Targ Rinse K Weersma[63], Subra Kugathasan[64], Anne M C Richard H Duerr[50,66], Mark S Silverberg[55], Jack Sats Hakon Hakonarson[5,21], Mark J Daly[15,71] & Miles Pa

Jostins, Ripke et al
Nature, in press
163 distinct confirmed associated loci in IBD (CD and/or UC)
N=75,000 samples

All 32 of the Barrett et al 2008 GWS hits are confirmed and more significant than they were in 2008

Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease

Jeffrey C Barrett*[1], Sarah Hansoul[2], Dan L Nicolae[3], Judy H Cho[4], Richard H Duerr[5,6], John D Rioux[7,8], Steven R Brant[9,10], Mark S Silverberg[11], Kent D Taylor[12], M Michael Barmada[5], Alain Bitton[13], Themistocles Dassopoulos[9], Lisa Wu Datta[9], Todd Green[8], Anne M Griffiths[14], Emily O Kistner[15], Michael T Murtha[4], Miguel D Regueiro[6], Jerome I Rotter[12], L Philip Schumm[15], A Hillary Steinhart[11], Stephan R Targan[12], Ramnik J Xavier[16], the NIDDK IBD Genetics Consortium[33], Cécile Libioulle[2], Cynthia Sandor[2], Mark Lathrop[17], Jacques Belaiche[18], Olivier Dewit[19], Ivo Gut[17], Simon Heath[17], Debby Laukens[20], Myriam Mni[2], Paul Rutgeerts[21], André Van Gossum[22], Diana Zelenika[17], Denis Franchimont[22], JP Hugot[23], Martine de Vos[20], Severine Vermeire[21], Edouard Louis[18], the Belgian-Frenc IBD Consortium[33], the Wellcome Trust Case Control Consortium[33], Lon R Cardon[1], Carl A Anderson[1], Hazel Drummond[24], Elaine Nimmo[24], Tariq Ahmad[25], Natalie J Prescott[26], Clive M Onnie[26], Sheila A Fisher[2 Jonathan Marchini[27], Jilur Ghori[28], Suzannah Bumpstead[28], Rhian Gwillam[28], Mark Tremelling[29], Panos Deloukas[28], John Mansfield[30], Derek Jewell[31], Jack Satsangi[24], Christopher G Mathew[26], Miles Parkes[29], Michel Georges[2] & Mark J Daly[8,32]

Moreover…

2008

suggesting that many additional loci remain to be identified. This is reinforced by the additional eight nominal replications (**Table 3**), where only two or three would be expected by chance,

2012

Of those 8 additional loci, 4 are now genome-wide significant, 2 have been sufficiently rejected, 2 are yet to be confirmed or rejected convincingly

Isn’t rare variation different?

I’m just looking at coding variation in the exome – why worry?

SIAE: original report

LETTERS

Functionally defective germline variants of sialic acid acetylesterase in autoimmunity

Ira Surolia[1]*, Stephan P. Pirnie[1]*, Vasant Chellappa[1]*, Kendra N. Taylor[1]*, Annaiah Cariappa[1]*, Jesse Moya[1], Haoyuan Liu[1], Daphne W. Bell[1]†, David R. Driscoll[1], Sven Diederichs[1]†, Khaleda Haider[1], Ilka Netravali[1], Sheila Le[1], Roberto Elia[1], Ethan Dow[1], Annette Lee[2], Jan Freudenberg[2], Philip L. De Jager[3,4], Yves Chretien[5], Ajit Varki[6], Marcy E. MacDonald[7], Tammy Gillis[7], Timothy W. Behrens[8], Donald Bloch[9], Deborah Collier[9], Joshua Korzenik[10], Daniel K. Podolsky[10]†, David Hafler[3,4], Mandakolathur Murali[11], Bruce Sands[10], John H. Stone[9], Peter K. Gregersen[2] & Shiv Pillai[1]

- "Loss of function" or "defective" alleles of *SIAE* defined biochemically
- One common "defective" allele (M89V) treated as deleterious only in homozygotes; rarer "LoF" alleles treated as deleterious in heterozygotes
- Variants (thus defined) were present in 24/923 cases of autoimmune disease, 2/648 controls
- Odds ratio 8.62, *P* value 0.0002

SIAE: independent follow-up

CORRESPONDENCE

Rare and functional *SIAE* variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry

- Targeted genotyping of reported *SIAE* variants in much larger numbers of cases and controls
- M89V homozygosity ~0.3% in cases, ~0.3% in controls (N = 66,924, *P* = 0.45)
- C196F – confirmed defective variant – 0.199% in cases, 0.196% controls (N=43,378)
- Rarer variant heterozygosity in 0.12% of cases, 0.08% of controls (N = 43,378, *P* = 0.44)

Surely if I already know the gene or a closely related gene is involved in the phenotype there’s nothing to worry about?

CETP and high HDL

Loss-of-function mutations in CETP reliably associated with significant increase in serum HDL

Published in Volume 92, Issue 4 (October 1993)
J Clin Invest. 1993;92(4):2060–2064. doi:10.1172/JCI116802.

Research Article

A missense mutation in the cholesteryl ester transfer protein gene with possible dominant effects on plasma high density lipoproteins.

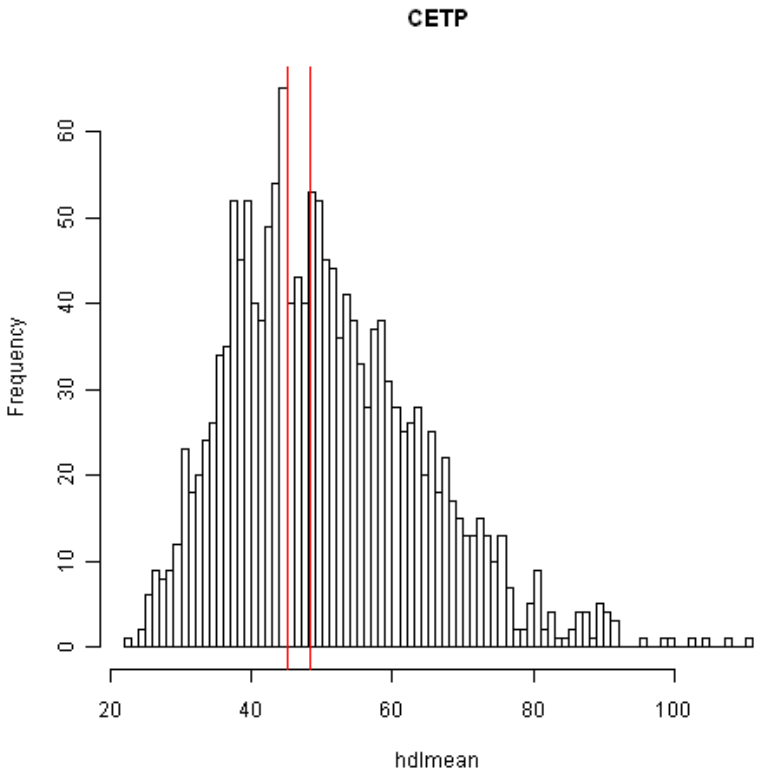
K Takahashi, X C Jiang, N Sakai, S Yamashita, K Hirano, H Bujo, H Yamazaki, J Kusunoki, T Miura and P Kussie

Second Department of Internal Medicine, Chiba University, Japan.

Published October 1993

Plasma HDL are a negative risk factor for atherosclerosis. Cholesteryl ester transfer protein (CETP; 476 amino acids) transfers cholesteryl ester from HDL to other lipoproteins. Subjects with homozygous CETP deficiency caused by a gene splicing defect have markedly elevated HDL; however, heterozygotes have only mild increases in HDL. We describe two probands with a CETP missense mutation (442 D:G). Although heterozygous, they have threefold increases in HDL concentration and markedly decreased plasma CETP mass and activity, suggesting that the mutation has dominant effects on CETP and HDL in vivo. Cellular expression of mutant cDNA results in secretion of only 30% of wild type CETP activity. Moreover, coexpression of wild type and mutant cDNAs leads to inhibition of wild type secretion and activity. The dominant effects of the CETP missense mutation during cellular expression probably explains why the probands have markedly increased HDL in the heterozygous state, and suggests that the active molecular species of CETP may be multimeric.

CETP and high HDL



HDL value of two individuals with D442G mutation in CETP

Published in Volume 92, Issue 4 (October 1993)
J Clin Invest. 1993;92(4):2060–2064. doi:10.1172/JCI116802.

Research Article

A missense mutation in the cholesteryl ester transfer protein gene with possible dominant effects on plasma high density lipoproteins.

K Takahashi, X C Jiang, N Sakai, S Yamashita, K Hirano, H Bujo, H Yamazaki, J Kusunoki, T Miura and P Kussie

Second Department of Internal Medicine, Chiba University, Japan.

Published October 1993

Plasma HDL are a negative risk factor for atherosclerosis. Cholesteryl ester transfer protein (CETP; 476 amino acids) transfers cholesteryl ester from HDL to other lipoproteins. Subjects with homozygous CETP deficiency caused by a gene splicing defect have markedly elevated HDL; however, heterozygotes have only mild increases in HDL. We describe two probands with a CETP missense mutation (442 D:G). Although heterozygous, they have threefold increases in HDL concentration and markedly decreased plasma CETP mass and activity, suggesting that the mutation has dominant effects on CETP and HDL in vivo. Cellular expression of mutant cDNA results in secretion of only 30% of wild type CETP activity. Moreover, coexpression of wild type and mutant cDNAs leads to inhibition of wild type secretion and activity. The dominant effects of the CETP missense mutation during cellular expression probably explains why the probands have markedly increased HDL in the heterozygous state, and suggests that the active molecular species of CETP may be multimeric.

MSUD mutations

Molecular basis of intermittent maple syrup urine disease: novel mutat
in the *E2* gene of the branched-chain α-keto acid dehydrogenase comp

Tsuruta et al, J. Hum Genet (1998)

- ‘Mutations’ identified in 3 MSUD patients
- One of the three, G323S, has a 90% population frequency

Both the CETP and E2 variants are still in HGMD

Many reported disease mutations are not actually likely disease-causing

- Bell et al. (2011) *Sci Transl Med* 3:65ra4
  - "In total, 27% (122 of 460) of literature-cited disease mutations were omitted, because they were adjudged to be common polymorphisms or sequencing errors or because of a lack of evidence of pathogenicity."
- Lupski et al. (2010) *New Engl J Med* 362:1181-1191
  - Lupski is homozygous for five reported recessive mutations and carries one reported dominant mutation linked to diseases he does not have
- 1000 Genomes pilot project found 71 reported severe disease mutations in HGMD with population frequency >10%

What's in an exome?

- ~20,000 DNA variants in/near protein coding DNA
- ~100 rare missense variants
- ~100 loss-of-function variants (~20 rare or private)

A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes

Daniel G. MacArthur,[1,2*] Suganthi Balasubramanian,[3,4] Adam Frankish,[1] Ni Huang,[1] James Morris,[1] Klaudia Walter,[1] Luke Jostins,[1] Lukas Habegger,[3,4] Joseph K. Pickrell,[5] Stephen B. Montgomery,[6,7] Cornelis A. Albers,[1,8] Zhengdong D. Zhang,[9] Donald F. Conrad,[10] Gerton Lunter,[11] Hancheng Zheng,[12] Qasim Ayub,[1] Mark A. DePristo,[13] Eric Banks,[13]

A few things are a bit more interpretable…but not absolute slam dunks…

- ~1 *de novo* variant per exome (only 5% LoF)
- <5% chance that an individual has a complete knockout of even a single gene where LoF mutations are rare in general

But, of course those *de novo* mutations in autism must be causal, right?

945 trios and counting…

		OBSERVED	EXPECTED	
		de novo per trio	*de novo* per trio	p-value
Affected	**Synonymous**	0.269	0.270	ns
	Missense	0.645	0.611	ns
	Loss of Function	**0.156**	**0.086**	**1.08E-08**
Unaffected	**Synonymous**	0.213	0.268	ns
	Missense	0.612	0.612	ns
	Loss of Function	0.087	0.087	ns

- Comparisons to well-calibrated mutation model and to observed rates in unaffecteds
- Vast majority of de novo missense variants and half the de novo LoF are unrelated to autism risk

Most compelling results still not clearly significant…

Gene	# *de novo* LoF	# *de novo* missense	p-value	# LoF cases	# LoF control	# LoF ESP
KATNAL2	2	0	9.96E-06	3	0	4
DYRK1A	2	0	2.16E-05	0	0	0
POGZ	2	0	7.62E-05	0	0	0
SCN2A	2	1	1.33E-04	0	0	0
CHD8	2	0	2.17E-04	3	0	0

Appropriate threshold for genome-wide significance roughly $1x10^{-6}$

$.05 / (N_{genes} * N_{tests})$

N_{tests} here set at 2 (all LoF, all LoF+missense)

This is important!!!

Thanks

Daniel MacArthur

Monkol Lek

Jeff Barrett