National Human Genome Research Institute

National Institutes of Health

U.S. Department of Health and Human Services

NCI/NHGRI Working Group on Criteria for Replication of Genotype-Phenotype Associations

U.S. Department of Health and Human Services
National Institutes of Health
National Cancer Institute
National Human Genome Research Institute
Michael Boehnke
Eric Boerwinkle
Stephen Chanock
David Hunter
Teri Manolio

Need for Consensus on What Constitutes Replication

- Avalanche of GWA and candidate gene studies now and in near future
- Replication held as *sine qua non*
- Likelihood of single study establishing an association is low until sample sizes increase sufficiently and analytical methods improve substantially
- Common problem of how to interpret confusing and spurious findings

Goal: To Develop Consensus on…

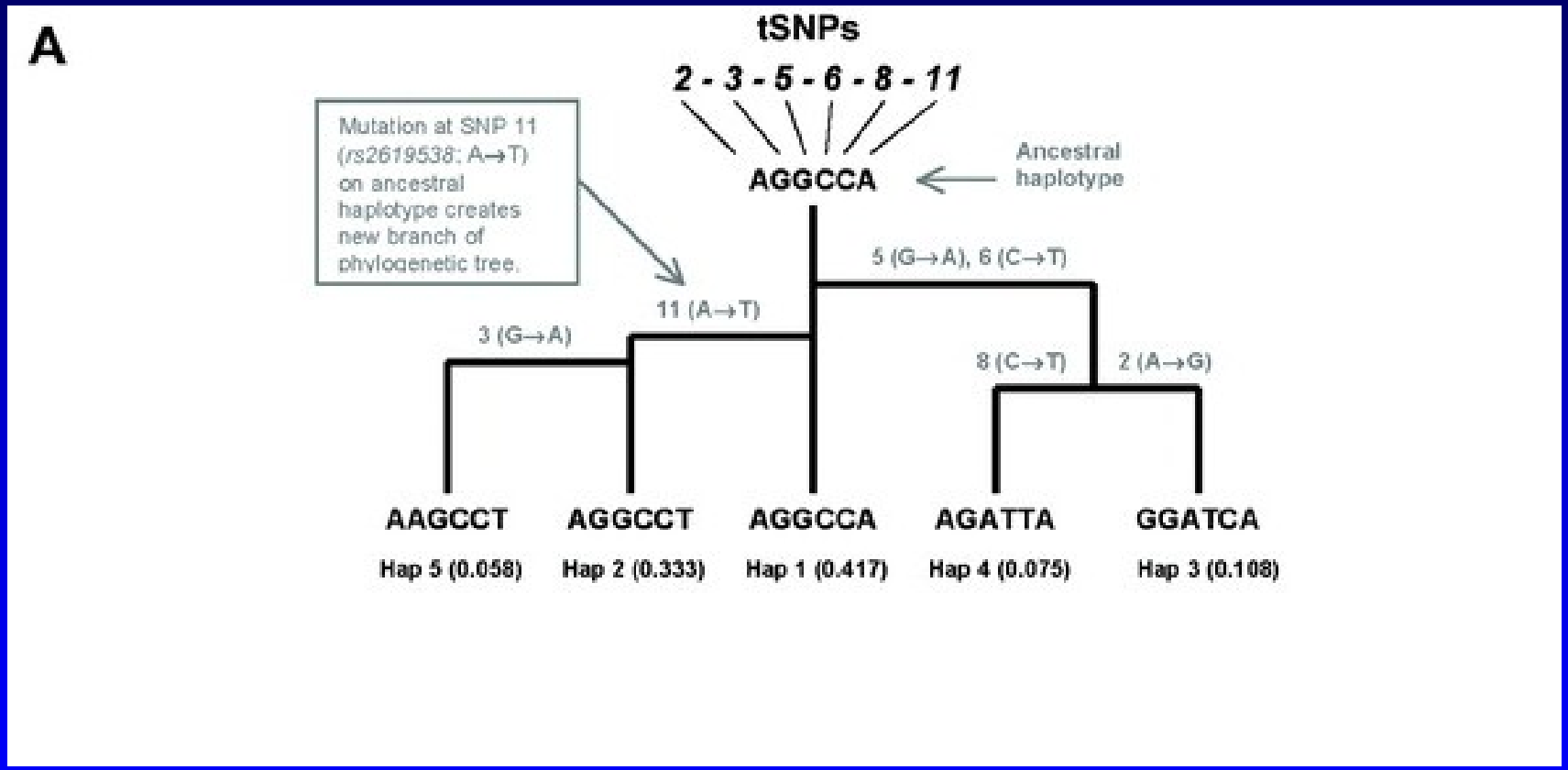
- Validity of single study
- Criteria for confirmation, replication, validation
- Priority for follow-up studies of GWA
- Guidelines for publication

Case in Point: *DTNBP1* and Schizophrenia

- Dystrobrevin binding protein 1 first identified as schizophrenia-susceptibility gene in Irish pedigrees
- Reported confirmation in several replication studies in independent European samples but reported risk alleles and haplotypes appeared to differ between studies
- Comparison among studies difficult because different marker sets used by each group
- HapMap data and all identified polymorphisms typed in CEPH samples to produce high density reference map

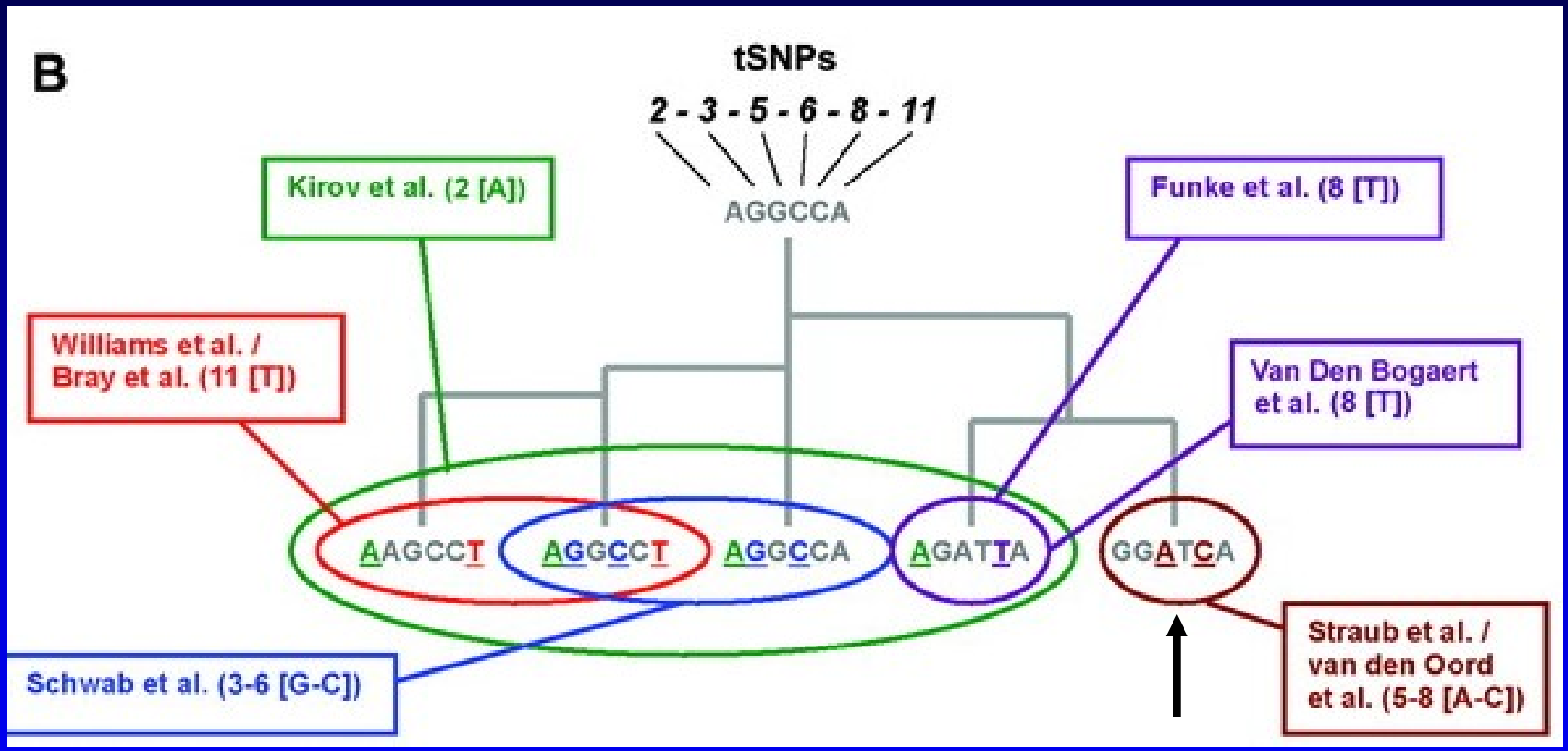
Mutsuddi et al, *Am J Hum Genet* 2006; 79:903-909.

Likely Phylogenetic Tree of Five Common Haplotypes of *DTNBP1*



Mutsuddi et al, *Am J Hum Genet* 2006; 79:903-909.

Positively Associated Haplotypes Differ in All Six Studies



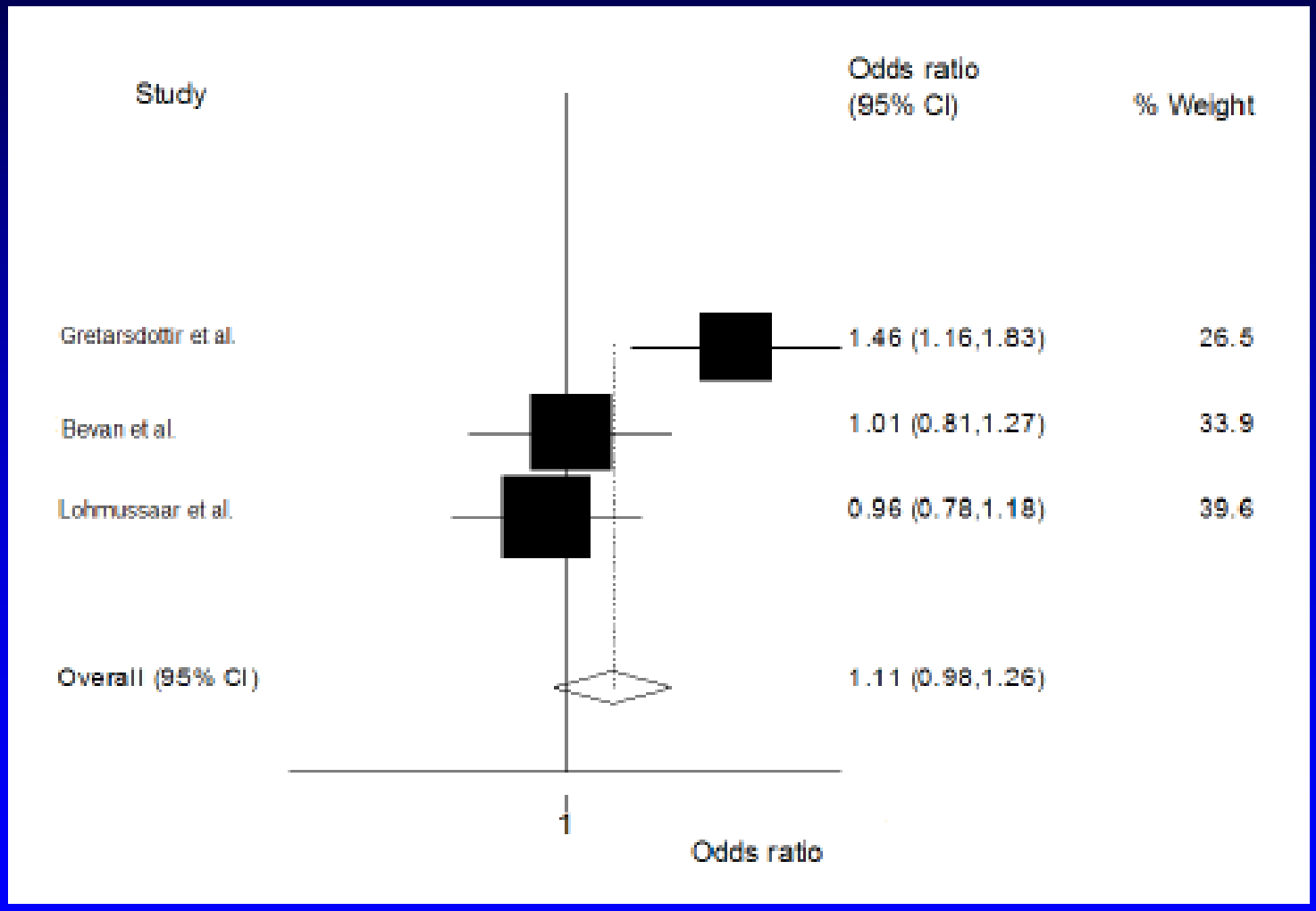
Each common DTNBP1 haplotype was tagged by association signal of at least one study, implying there is not one common variant contributing to schizophrenia risk at DTNBP1 locus

Mutsuddi et al, *Am J Hum Genet* 2006; 79:903-909.

How NOT To Do A Replication Study

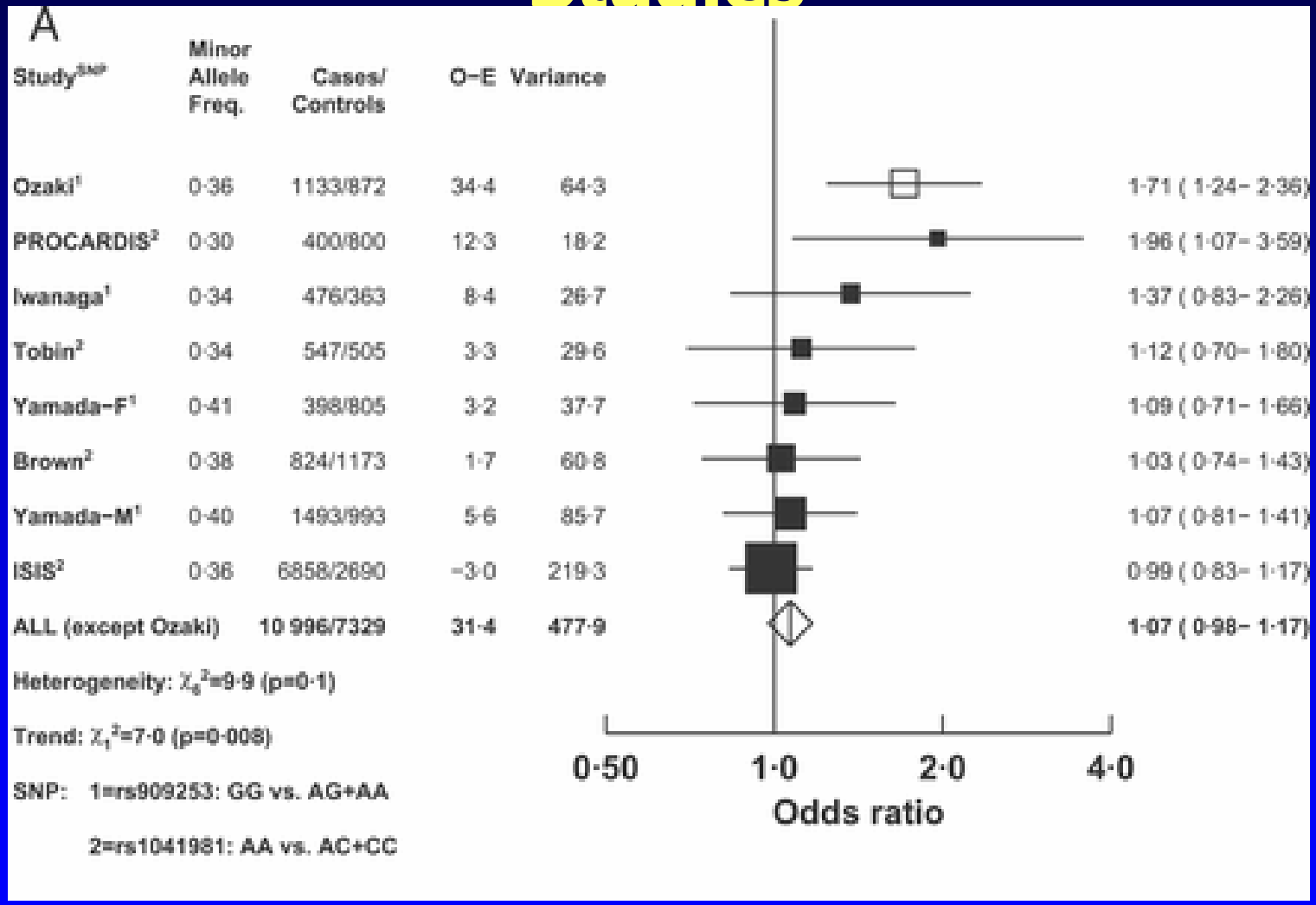
- Use a different phenotype
- Use different markers
- Mix fine-mapping and replication
- Use different analytic methods (haplotype vs. single marker)
- Use different populations

Odds Ratio for Stroke Associated with *PDE4D* in Three Studies

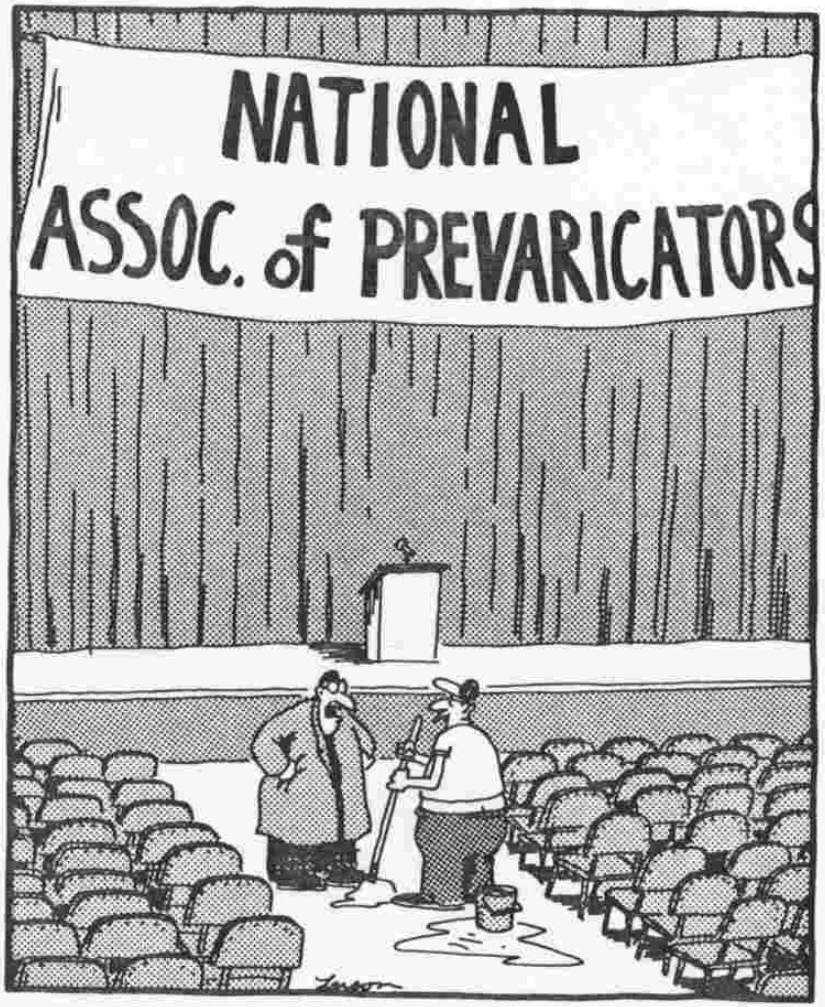


Rosand et al, *Nat Genet* 2006; 38:1091-1092.

Odds ratio (CI) for CHD Associated with *LTA* Genotypes in ISIS and Other Studies



Clarke et al, *PLoS Genet* 2006; 2:e107.



"Yesterday! I was told the meeting was *today*!"

The Far Side

November

10

Thursday

CGEMS

Cancer Genetic Markers of Susceptibility

Core Genotyping Facility

Office of Cancer Genomics

Division of Cancer Epidemiology and Genetics

Characteristics of SNPs with Strongest Signal of Association in 8q24

rsnumber	Position (b.35)	MAF (%)	HWE controls	Completion rate
rs4242382	128586755	0.14	0.76	1
rs7017300	128594450	0.18	0.16	1
rs7837688	128608542	0.14	0.87	0.999
rs1447295*	128554220	0.14	0.60	1

*** Amundadottir et al, Nat Genet 2006; 38:652-658.**

CGEMS

Cancer Genetic Markers of Susceptibility

Core Genotyping Facility

Office of Cancer Genomics

Division of Cancer Epidemiology and Genetics

LD of Top SNPs with rs1447295 and Associated RR of Aggressive Disease

rsnumber	Risk Allele (MAF %)	r^2	P–Assoc	Genotype RR	
				Heteroz	Homoz
rs4242382	A (0.10)	0.94	0.00007	1.24	1.46
rs7017300	A (0.10)	0.71	0.00009	1.27	1.39
rs7837688	C (0.13)	0.84	0.00003	1.17	1.37
rs1447295*	T (0.10)	--	0.0003	1.26	1.54

*** Amundadottir et al, Nat Genet 2006; 38:652-658.**

Replication of *TCF7L2* Association in 4,632 U.K. Cases and Controls[1]

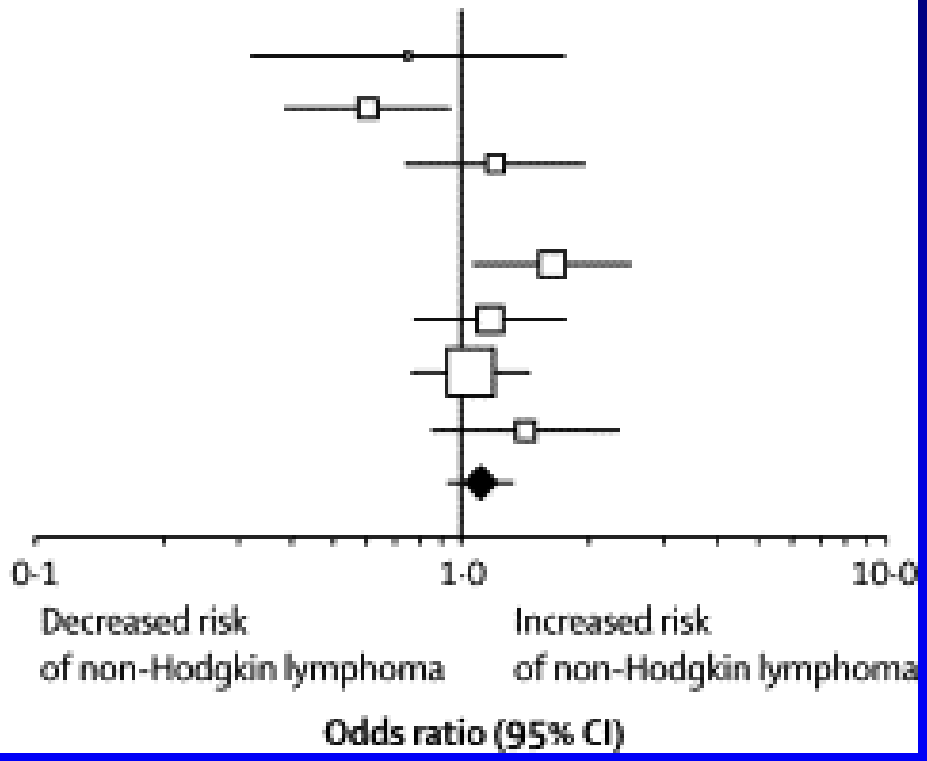
SNP	Allele RR (95% CI)	*P*
rs4506565	1.35 (1.23–1.47)	1.6×10^{-11}
rs7903146[2]	1.36 (1.24–1.48)	1.3×10^{-11}
rs12243326[2]	1.31 (1.19–1.43)	4.3×10^{-9}
rs12255372	1.29 (1.18–1.41)	2.2×10^{-8}

[1] **Groves et al *Diabetes* 2006; 55:2640-2644.**

[2] **Grant et al *Nat Genet* 2006; 38:320–323.**

Study-Specific and Pooled Risk Estimates of *IL1B –511C→T* for Non-Hodgkins Lymphoma

	Odds ratio (95% CI)
EPILYMPH—Italy	0·75 (0·32–1·75)
EPILYMPH—Spain[15]	0·60 (0·39–0·95)
University of California, San Francisco[16]	1·21 (0·74–1·96)
EPILYMPH—Germany[17]	1·64 (1·07–2·52)
Connecticut[18]	1·17 (0·78–1·77)
NCI-SEER[12]	1·05 (0·76–1·45)
British Columbia	1·42 (0·85–2·36)
Pooled	1·11 (0·92–1·35)



Rothman et al, *Lancet Oncol* 2006; 7: 27-38.

Definition of Robust Initial Finding

- Sufficient statistical power to observe reported effect, which will vary by magnitude of observed effect
- Highly significant analysis using stable method
- Consistent findings using straightforward, simple analytic approach
- Consistent findings in epidemiologically sound study
- Consistent findings overall and within key subgroups of initial study
- Consistent findings in same or highly similar phenotypes

Value of Single/First Study

- Initial study rarely definitive by itself but often represents important discovery tool
  - If consortium of multiple studies, stronger
- What to do with studies not having option for replication?
  - Don't change standards for definitiveness
- Don't just rely on GWA-- have multiple tools for identifying and understanding associations
- May need different standards for findings of major clinical significance, particularly

Importance of Significance Level

- Should we promulgate a specific number– NO, but in general, smaller is better
- General agreement: range is very broad, higher threshold for difficult to measure phenotype
- Beware of the very smallest
- If significance depends on analytic method or multiple comparison correction, BEWARE
- If significance or association depends on phenotype definition, BEWARE
- Randomize the phenotypes and report number of random associations significant at that level
- Biologic information may be useful *A PRIORI* but *a posteriori* can come up with almost anything

Importance of Genotyping Quality

- Report results of known study sample duplicates, HapMap or other standard duplicates
- Replicate small number of "significant" SNPs with second technology at some late stage
- May not be needed if nearby SNPs in strong LD show same results
- Strong caveats are needed regarding fallibility of genotyping
  - Results can change based on genotype calling algorithm
  - QC filters and consistency of results after applying them must be described

Proposed Criteria for Positive Replication

- Sufficient sample size to distinguish proposed effect from no effect convincingly
- Same or very similar trait (extension to related trait may increase confidence in finding, such as consistent finding for both dichotomized obesity and continuous BMI)
- Same or very similar population (extension to other populations may also increase confidence in finding, such as consistent association in populations of European, Asian, or even recent African ancestry

Proposed Criteria for Positive Replication

- Same inheritance model (dominant, co-dominant, recessive), though not necessarily same analytic method)
- Same gene, same SNP (or SNP in complete LD with prior SNP, $r^2 = 1$), same direction as original finding
- Highly significant association
- N.B.: Initial study must adequately describe these parameters

Proposed Criteria for True Non-Replication or "Meaningful Negativity"

- Same as for positive replication (same trait, same gene, same SNP, same direction, same genetic model)
- Must be identical trait and population to claim non-replication
- Powered to appropriate effect size (account for "winner's curse")

Information to be Included in Initial Report

- Study design and collection, including success rate for sample acquisition, extraction and analysis
- Standard “Table 1” of salient characteristics of cases and controls, including rates of missing data
- Analysis methods in sufficient detail to understand and reproduce what was done
- Availability of results for others to analyze with their methods
- Revelation of replication and analysis attempts by authors
- Attestation of no known non-replication by others (can this be meaningfully described,

Inclusion of Standard Analyses

- Average value of chi-square and full distribution
- Q-Q plots of chi square and p-values
- Genotyping cluster plots for SNPs of interest
- Signal at nearby or correlated SNPs
- Genotype QC filters applied, including Hardy-Weinberg

Points to Consider for Reviewers and Authors

- Strength of observation
  - Suitably large sample size
  - Suitably small p-value
  - Suitably high quality of study design, including selection of study population, reliability of phenotypes, measurement of potential confounders
- Conclusions commensurate with sample size, power, p-value, quality
- Usefulness of observations to others for subsequent research
- Attempts to consider alternative explanations

Importance of Negative Studies

- Journals should be urged to publish valid non-replication of original reports, similar to recent Parkinson's example
- Original journal should be encouraged to take responsibility for this, particularly for adequately powered and conducted non-replication findings
- Set timeframe, probably 3-9 months from initial publication, may need to be journal-specific
- Need public archive of negative studies (new PLOS publication, CEBP's "Null Results in Brief")

How to Make Data Available

- No journal has defined policy for genotype deposition, confidentiality issues
- Genotype summaries (not individual data) can and should be published on-line unless good reasons not to, such as special populations
- Value of placing genotypes and phenotypes in data archive—NCBI dbGaP is possibility
- What are journals' responsibilities for adhering to (or even being aware of) publication restrictions on widely available datasets such as GAIN?

Outcome of Meeting

- Validity of single study?
  - rare but possible
- Criteria for confirmation, replication, validation?
  - use only “replication,” don’t define it
- Priority for follow-up studies of GWA?
  - didn’t really tackle
- Guidelines for publication?
  - publish compendium of follow-up studies in same journal as initial report
  - develop “points to consider” and append to working group summary manuscript
  - need archive for negative results

Outcome of Meeting

- Validity of single study?
  - rare but possible
- Criteria for confirmation, replication, validation?
  - use only "replication," don't define it
- Priority for follow-up studies of GWA?
  - didn't really tackle
- Guidelines for publication?
  - publish compendium of follow-up studies in same journal as initial report
  - develop "points to consider" and append to working group summary manuscript
  - need archive for negative results

NCI/NHGRI Replication Working Group

Goncalo Abecasis
David Altshuler
Joan E. Bailey-Wilson
Lisa D. Brooks
Lon R. Cardon
Francis S. Collins
Mark Daly
Peter Donnelly
Joseph F. Fraumeni
Nelson B. Freimer
Daniela S. Gerhard
Chris Gunter
Alan E. Guttmacher
Mark S. Guyer
Joel Hirschhorn
Josephine Hoh
Robert Hoover
C. Augustine Kong
Kathleen R. Merikangas
Cynthia C. Morton
Lyle J. Palmer
John P. Rice
Jerry Roberts
Charles Rotimi
Gilles Thomas
Peggy Tucker
Kyle Vogan
Sholom Wacholder
Ellen M. Wijsman
Debbie Winn