Using the 1000 Genomes Data for Disease Studies
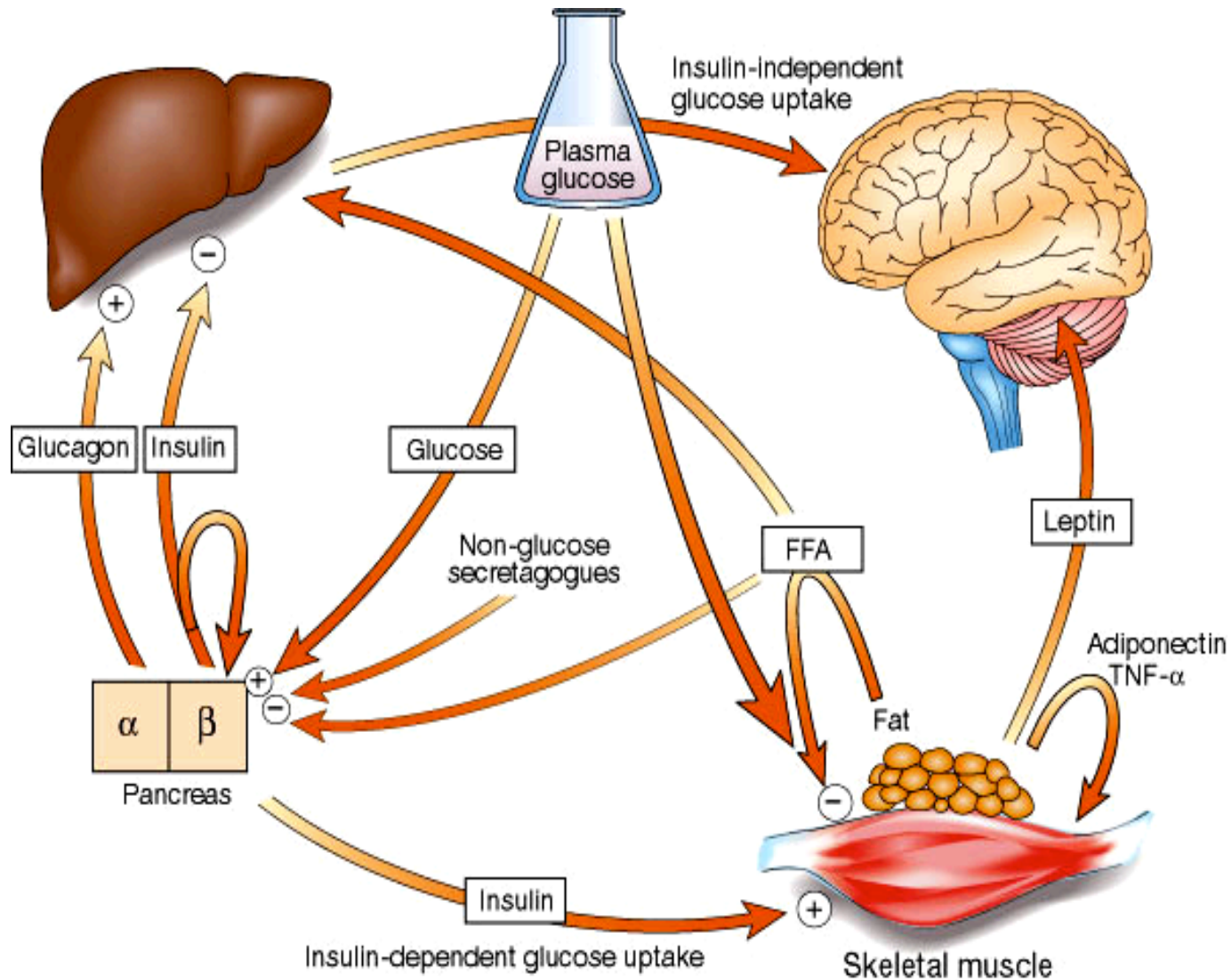
1000 Genomes Data Tutorial

American Society of Human Genetics

Karen L. Mohlke, Ph.D.

University of North Carolina

October 19, 2014

Genetics of complex metabolic traits
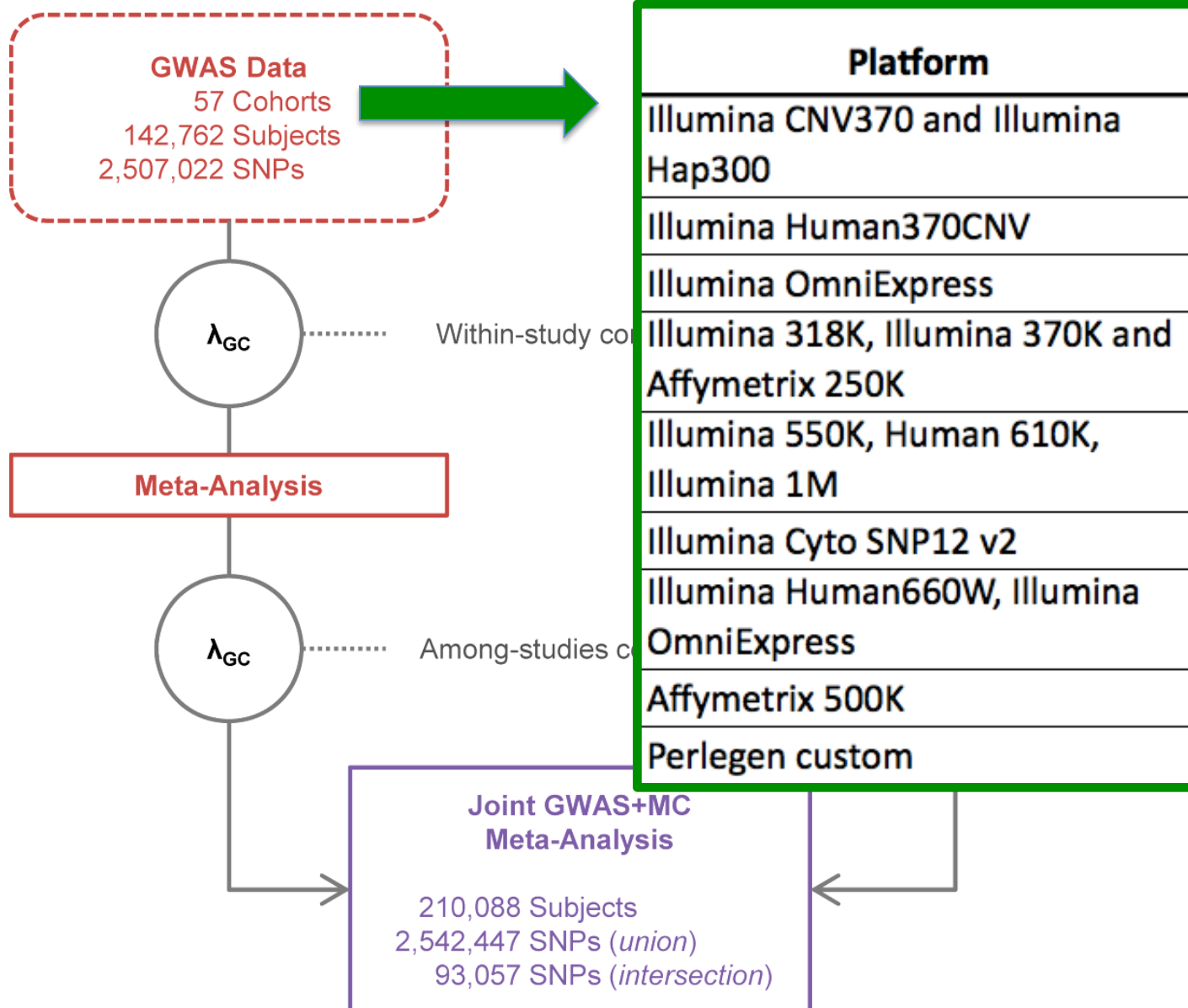

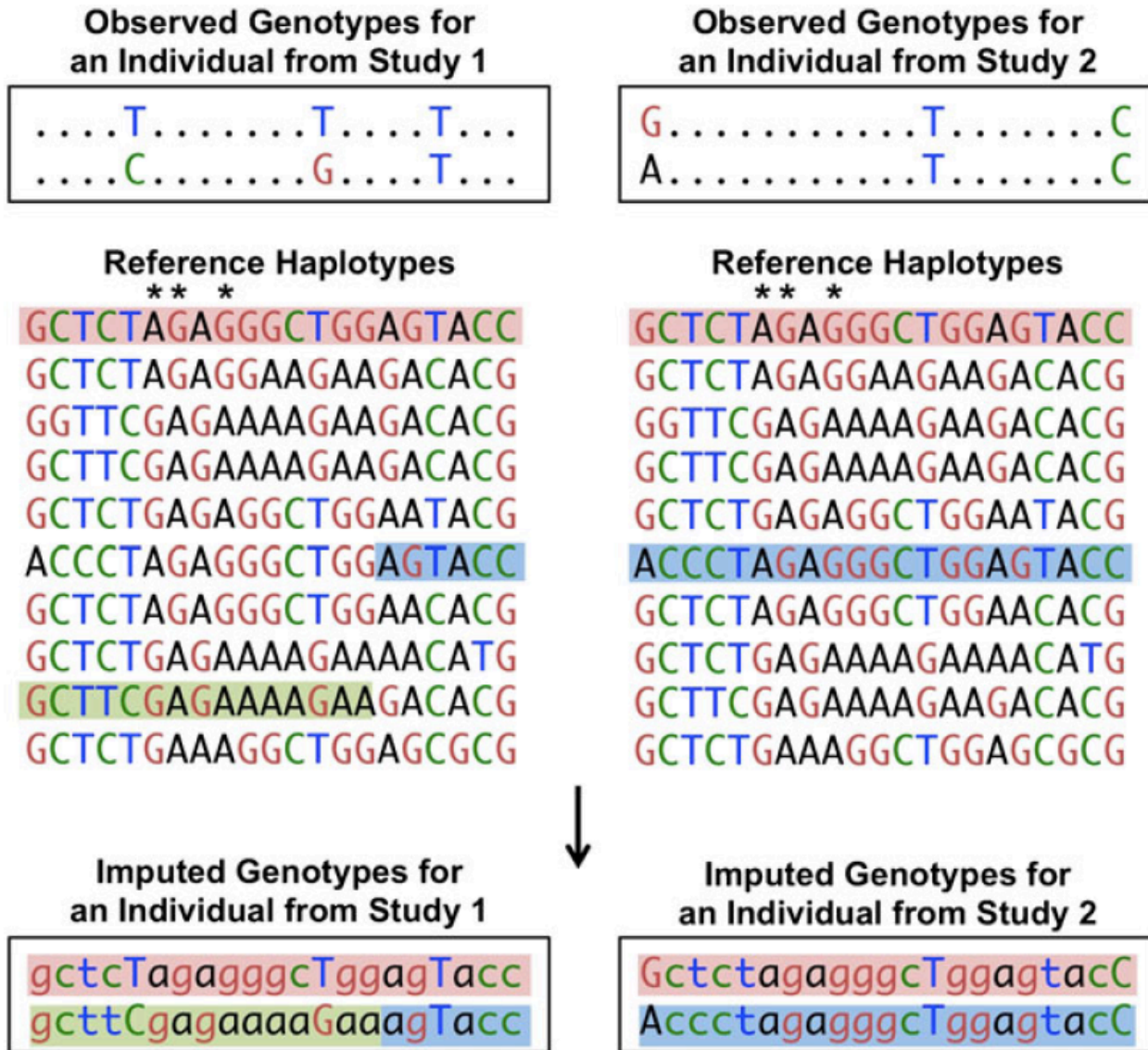Saltiel and Kahn (2001) Nature 414:799

Two applications

Genome-wide association studies

Laboratory studies to identify functional variants at GWAS loci

Example GWAS meta-analysis study design

GWAS Data
57 Cohorts
142,762 Subjects
2,507,022 SNPs

λ_{GC} · · · · · · · Within-study co...

Meta-Analysis

λ_{GC} · · · · · · · Among-studies c...

**Joint GWAS+MC
Meta-Analysis**

210,088 Subjects
2,542,447 SNPs (*union*)
93,057 SNPs (*intersection*)

Platform
Illumina CNV370 and Illumina Hap300
Illumina Human370CNV
Illumina OmniExpress
Illumina 318K, Illumina 370K and Affymetrix 250K
Illumina 550K, Human 610K, Illumina 1M
Illumina Cyto SNP12 v2
Illumina Human660W, Illumina OmniExpress
Affymetrix 500K
Perlegen custom

Shungin, Winkler, Croteau-Chonka, Ferreira, Locke, Magi *et al*, GIANT consortium

Imputation for GWAS meta-analysis

Observed Genotypes for an Individual from Study 1

Observed Genotypes for an Individual from Study 2

Reference Haplotypes

Reference Haplotypes

Imputed Genotypes for an Individual from Study 1

Imputed Genotypes for an Individual from Study 2

Mohlke and Scott (2012) Curr Diabetes Rpts 12:643

Novel loci identified using 1000 Genomes imputation

Title: Genome-wide association study imputed to 1000 Genomes reveals 18 novel associations with type 2 diabetes.(53) (01:30PM-01:45PM on Sun) **(Platform)**
Author(s): R. A. Scott, R. Magi, A. P. Morris, L. Marullo, K. Gaulton, M. Boehnke, J. Dupuis, M. I. McCarthy, L. J. Scott, I. Prokopenko, DIAGRAM+ consortium

Title: Genome-wide association study imputed to 1000 genomes identifies novel loci associated with lung function.(999T) (2:00PM-3:00PM on Tue) **(Poster)**
Author(s): M. Soler Artigas, L. V. Wain, N. Shrine, J. Huffman, I. Sayers, D. Strachan, I. P. Hall, M. D. Tobin, UK BiLEVE consortium, SpiroMeta consortium

Title: Genome-wide association of 44,714 subjects of African ancestry imputed to the 1000 Genomes reference panel identified two novel loci influencing body mass index.(993T) (2:00PM-3:00PM on Tue) **(Poster)**
Author(s): M. C. Y. Ng, M. Graff, A. Justice, Y. Lu, P. Mudgal, K. Rand, Y. Li, B. E. Cade, J. Brody, M. K. Wojczynski, L. R. Yanek, J. Smith, M. A. Nalls, L. A. Lange, S. Vedantam, X. Guo, D. Siscovick, S. R. Patel, B. M. Psaty, I. B. Borecki, D. M. Becker, L. F. Bielak, Y. Liu, J. G. Wilson, J. N. Hirschhorn, J. I. Rotter, C. A. Haiman, R. J. F. Loos, K. E. North, African Ancestry Anthropometry Genetic Consortium

Novel adiponectin signal at ADIPOQ

Novel signal minor allele frequency < .05; variants not present in HapMap

Multiple known signals shown in dark blue

1KG LD

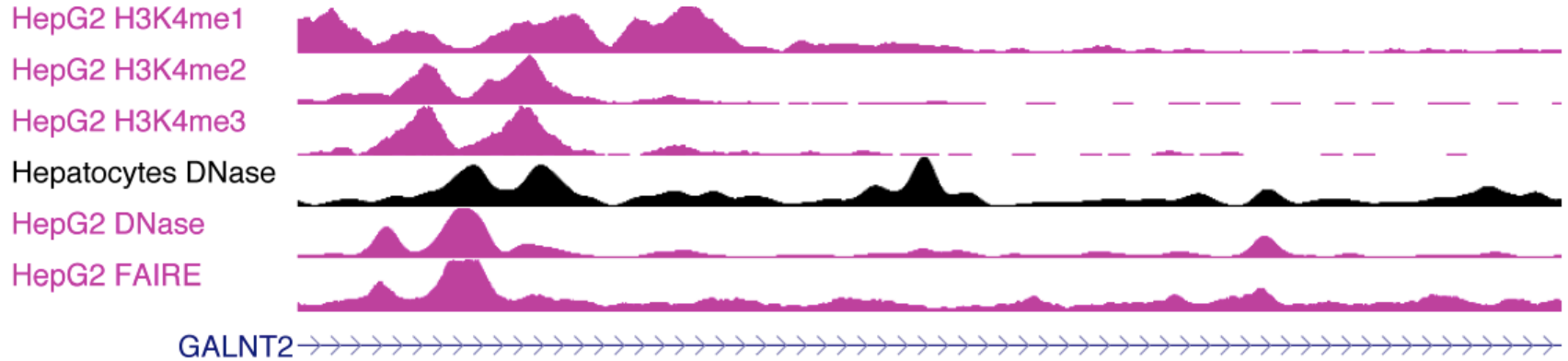METSIM study of ~10,000 Finns; Cassandra Spracklen
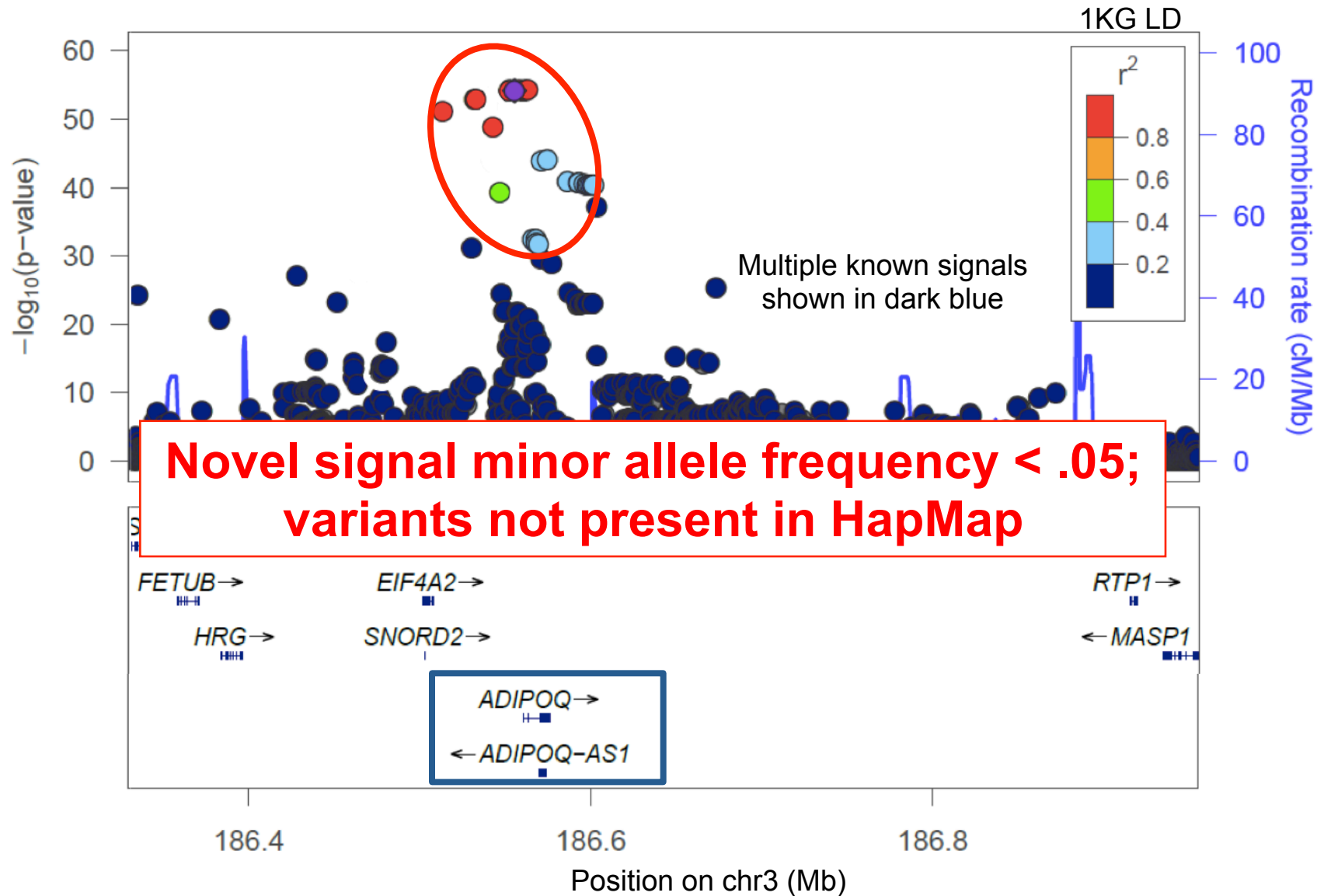
Novel loci identified using 1000 Genomes imputation

- **Variants have a lower allele frequency**

- **Imputation is more accurate across studies in a meta-analysis**

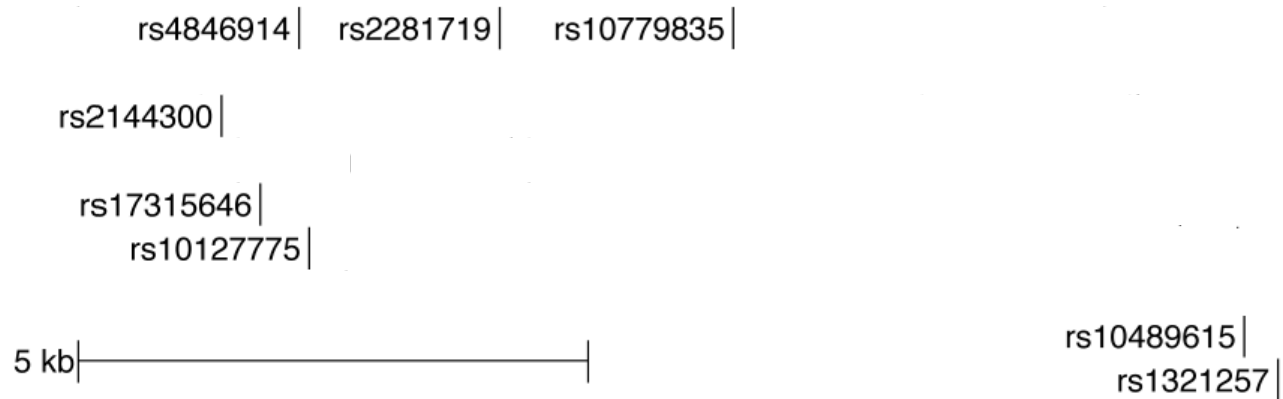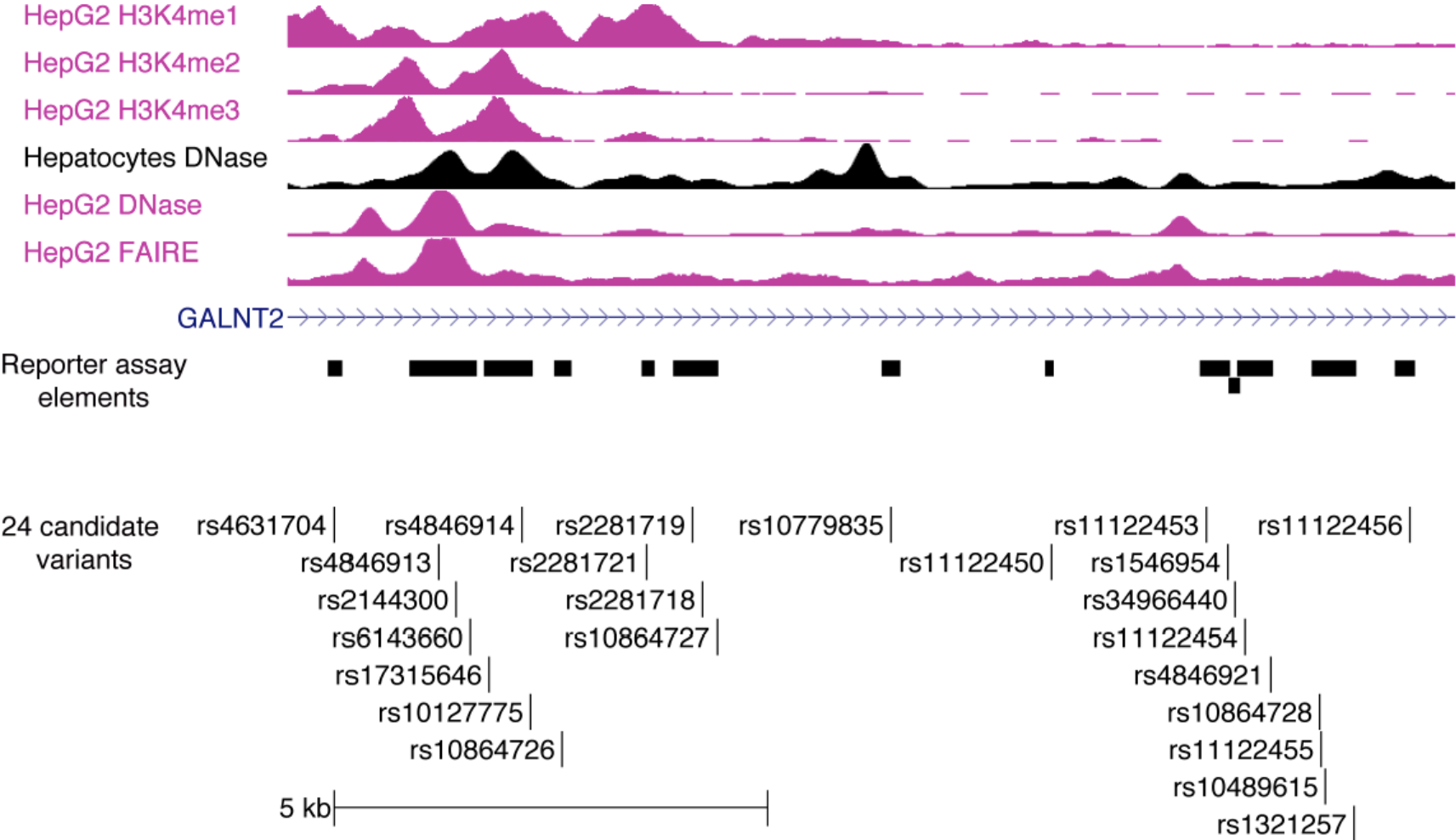- **(Increased sample size because meta-analyses are larger)**

GWAS for HDL and LDL Cholesterol
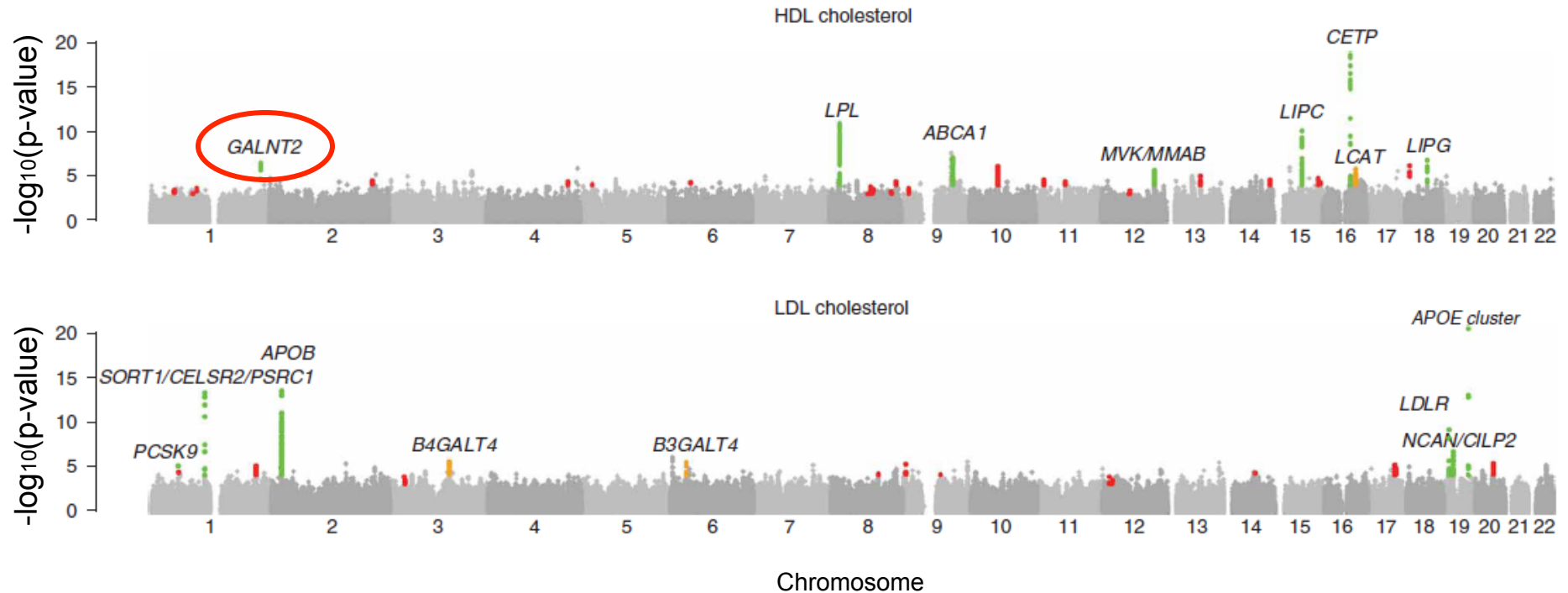
Candidate variants

HapMap: 8 candidate variants based on LD ($r^2 > .8$, CEU)

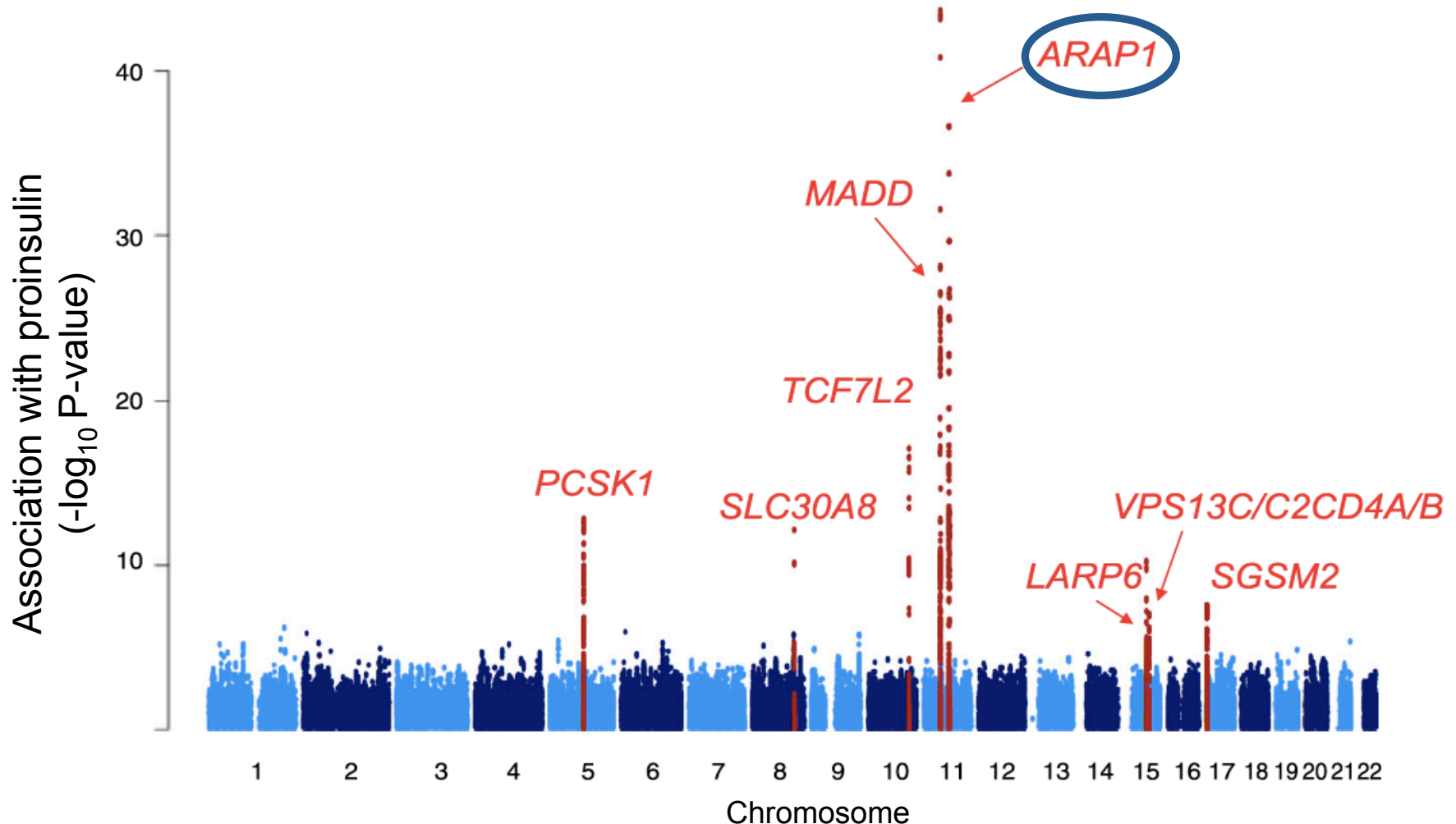# *Identify more candidate variants*

HapMap: 8 candidate variants based on LD ($r^2>.8$, CEU)
1000 G: 24 candidate variants based on LD ($r^2>.8$, EUR)
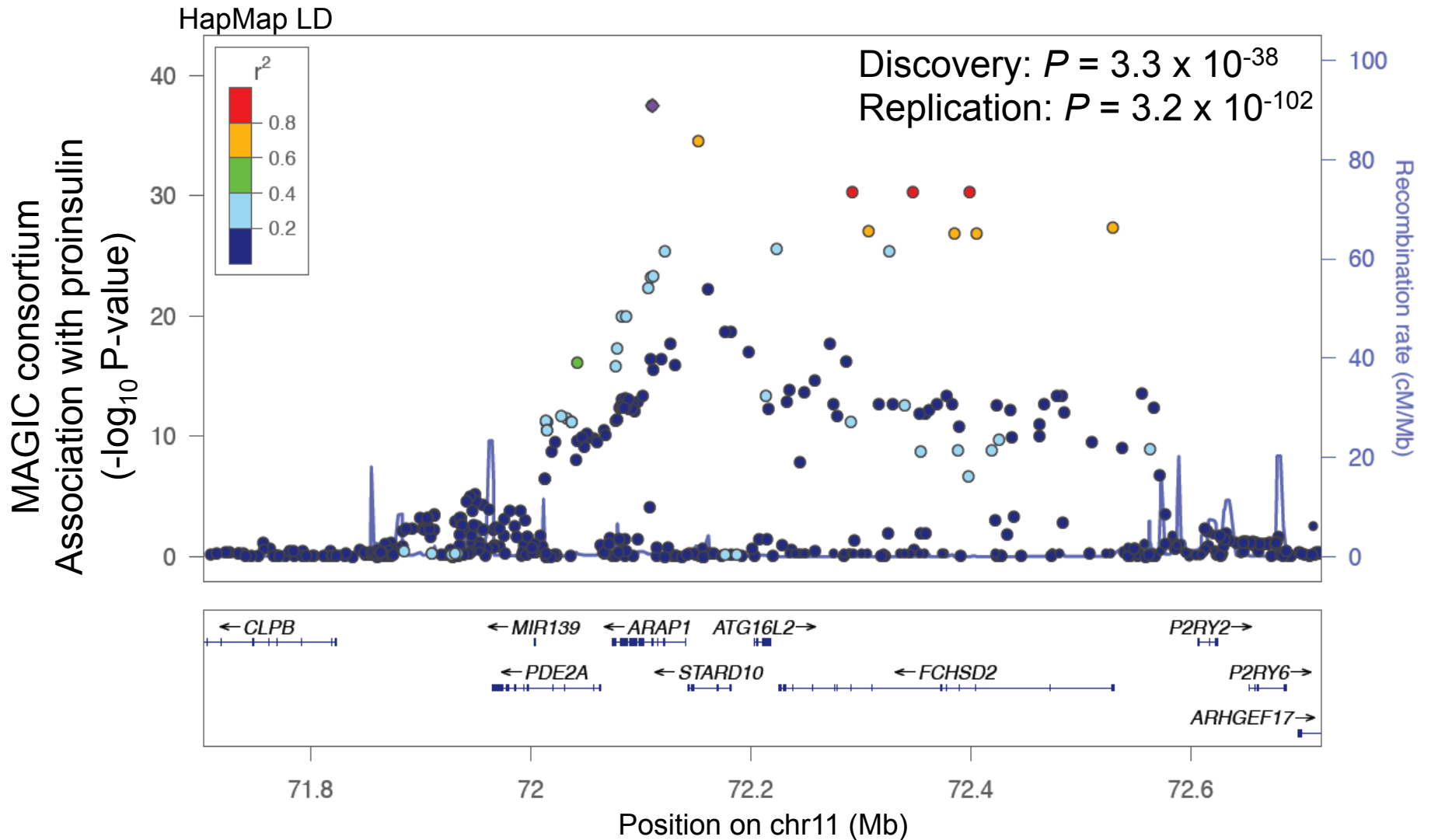
GWAS meta-analysis for proinsulin

~2.3 M markers in HapMap
Log-transformed proinsulin levels adjusted for fasting insulin, sex, and age.
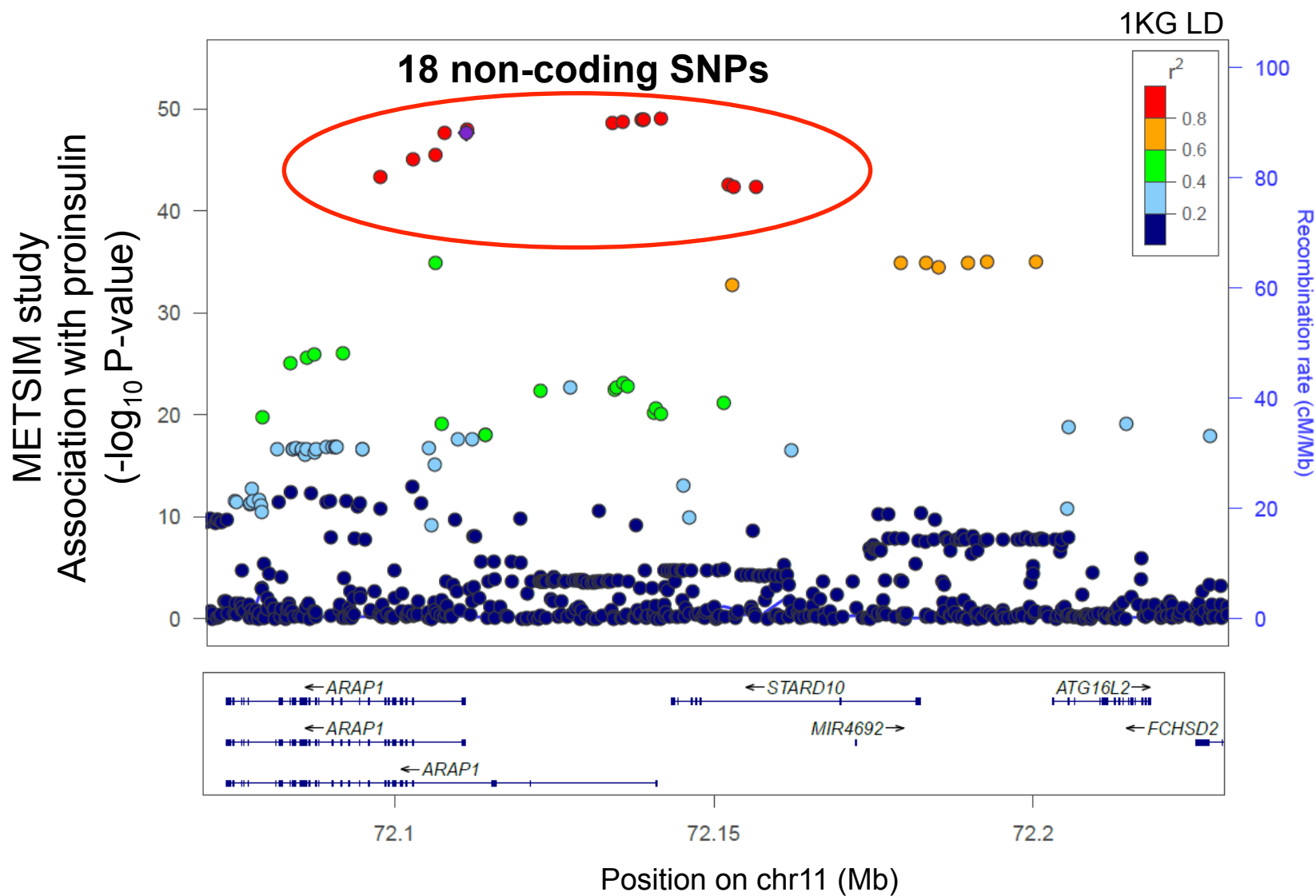Discovery set of 10,701 subjects and replication set of 16,378 subjects.

Strawbridge (2011) Diabetes 60: 2624

ARAP1 locus – HapMap variants

Which genes and variants are responsible?

Strawbridge (2011) Diabetes 60: 2624

ARAP1 locus – 1000 Genomes variants

18 non-coding SNPs

Variants tested for differential transcriptional activity

18 non-coding SNPs r^2 ≥ 0.8 with index SNPs

HapMap & 1000 Genomes; Human islet data from ENCODE & Human Epigenome Atlas; UCSC Genome Browser

Variants tested for differential transcriptional activity

18 non-coding SNPs r^2 ≥ 0.8 with index SNPs

HapMap & 1000 Genomes; Human islet data from ENCODE & Human Epigenome Atlas; UCSC Genome Browser
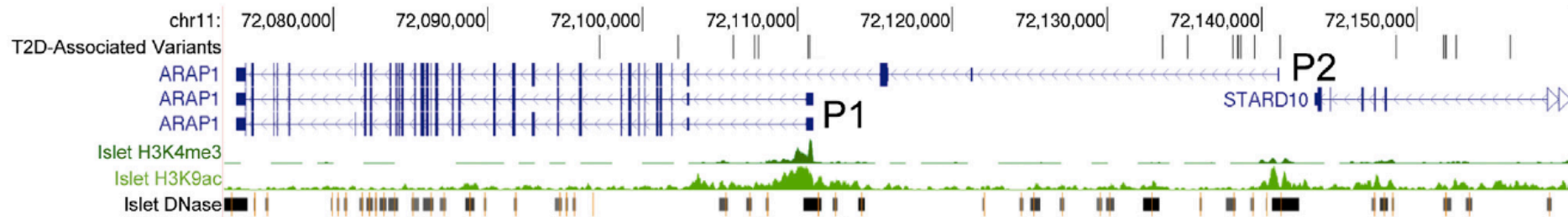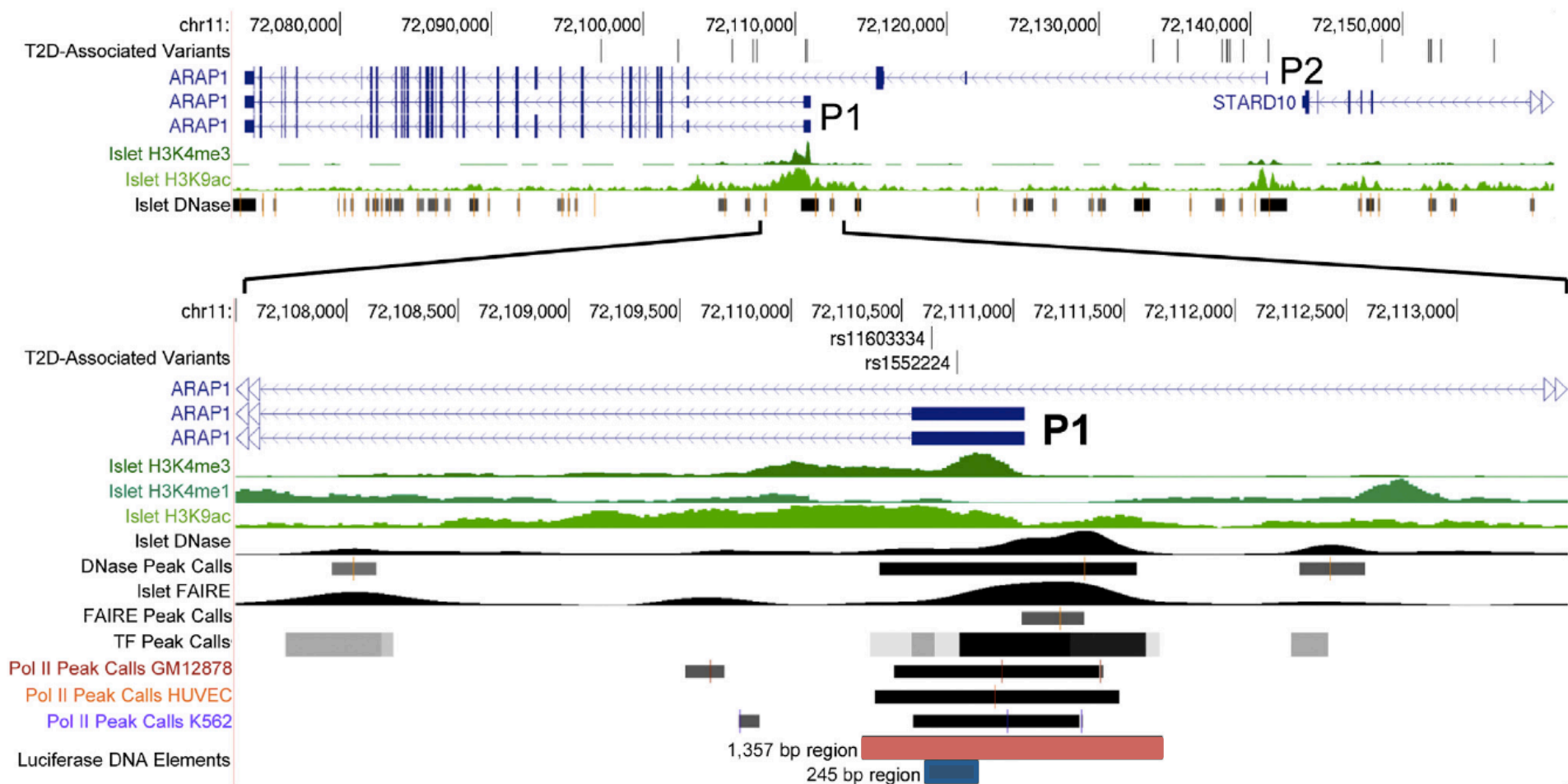
ARAP1 T2D risk alleles affect transcription factor binding and transcriptional activity

Electrophoretic mobility shift assay
rs11603334
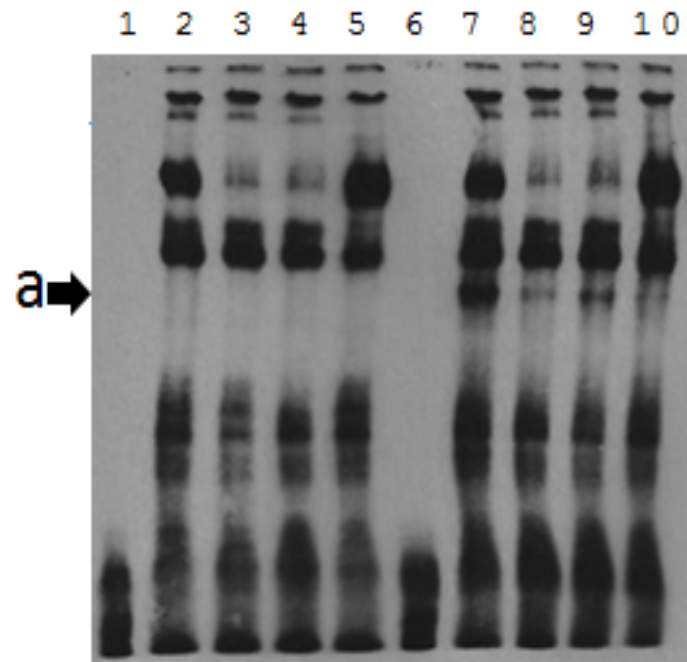
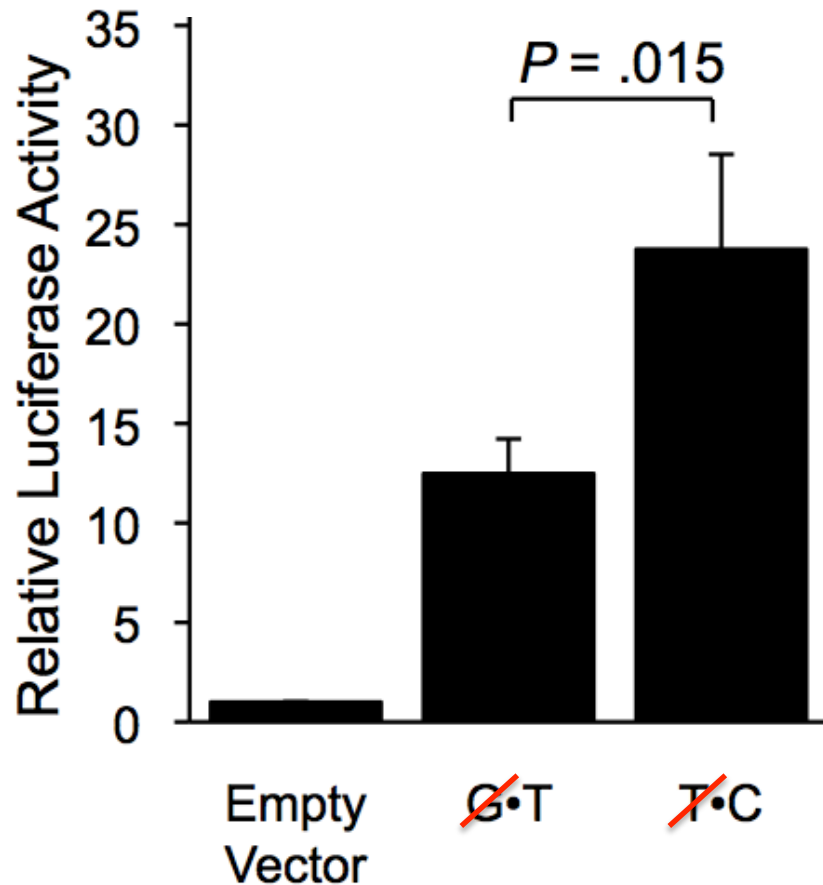| Biotin-DNA	C	C	C	C	C	T	T	T	T	T
MIN6 lysate	–	+	+	+	+	–	+	+	+	+
Unlabeled C-oligo	–	–	+	–	–	–	–	–	+	–
Unlabeled T-oligo	–	–	–	+	–	–	–	+	–	–
Pax6 Ig	–	–	–	–	+	–	–	–	–	+

ARAP1 T2D risk alleles affect transcription factor binding and transcriptional activity


Electrophoretic mobility shift assay rs11603334

Transcriptional reporter assay rs1552224 – rs11603334

1357 nt element; 832/13 rat beta cell line, n ≥ 3 independent clones per haplotype

Kulzer (2014) *AJHG* 94:186

ARAP1 T2D risk alleles affect transcription factor binding and transcriptional activity

Transcriptional reporter assay rs1552224 – rs11603334

1357 nt element; 832/13 rat beta cell line, n ≥ 3 independent clones per haplotype

Kulzer (2014) *AJHG* 94:186
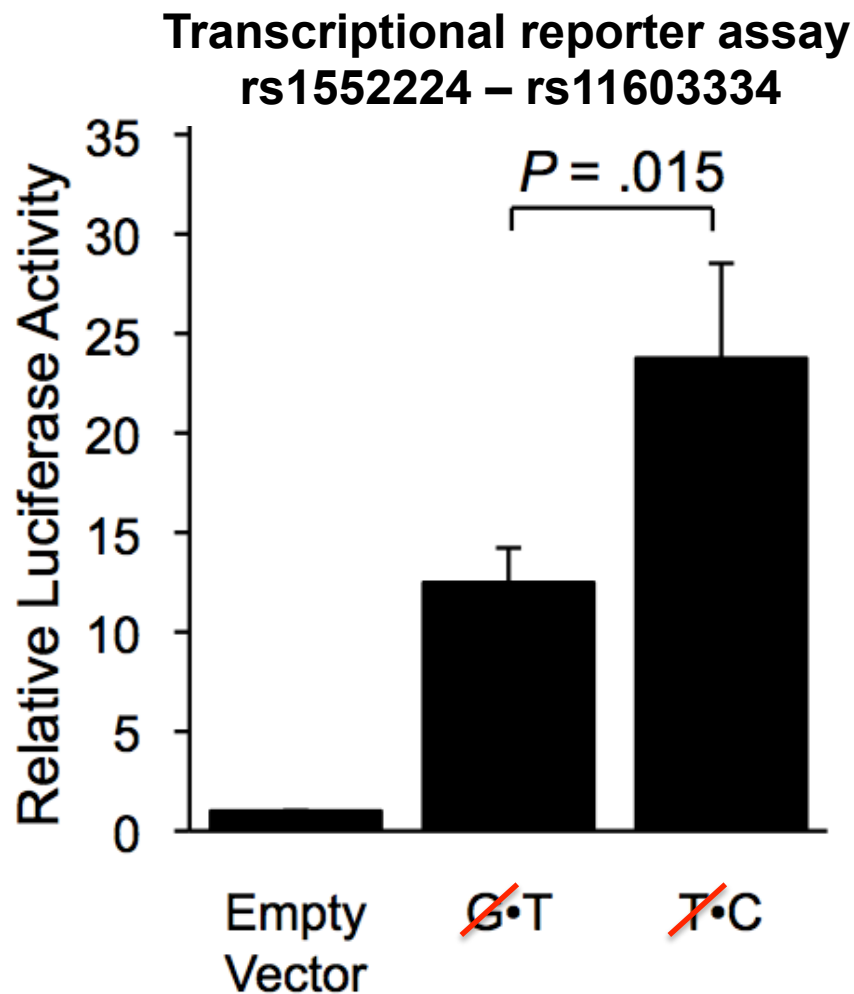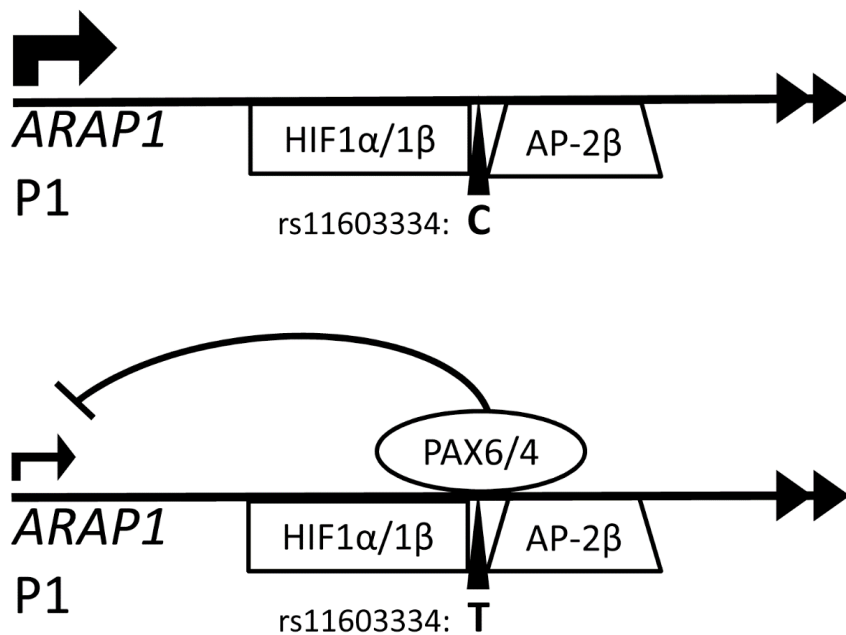
Conclusions

- **In genome-wide association studies**
  - **Identify loci when variants were not genotyped or included in HapMap**
  - **Improved detection of signals**
  - **Future: trans-ancestry meta-analyses**

- **In laboratory studies of GWAS loci**
  - **Improved linkage disequilibrium measures**
  - **Identify the fuller set of variants that may have a functional consequence**

Acknowledgements

UNC Chapel Hill
Martin Buchkovich
Maren Cannon
Jim Davis
Jennifer Kulzer
Chelsea Raulerson
Tamara Roman
Cassanadra Spracklen
Swarooparani Vadlamudi
Ying Wu

GIANT Consortium

MAGIC Consortium

DIAGRAM Consortium

University of Michigan
Michael Boehnke
Christian Fuchsberger
Jeroen Huyghe
Anne Jackson
Laura Scott
Heather Stringham

University of Eastern Finland
Markku Laakso
Johanna Kuusisto

NHGRI/NIH
Francis Collins
Peter Chines
Narisu Narisu
Michael Stitzel