

# Exploring 1000 Genomes with Bioconductor

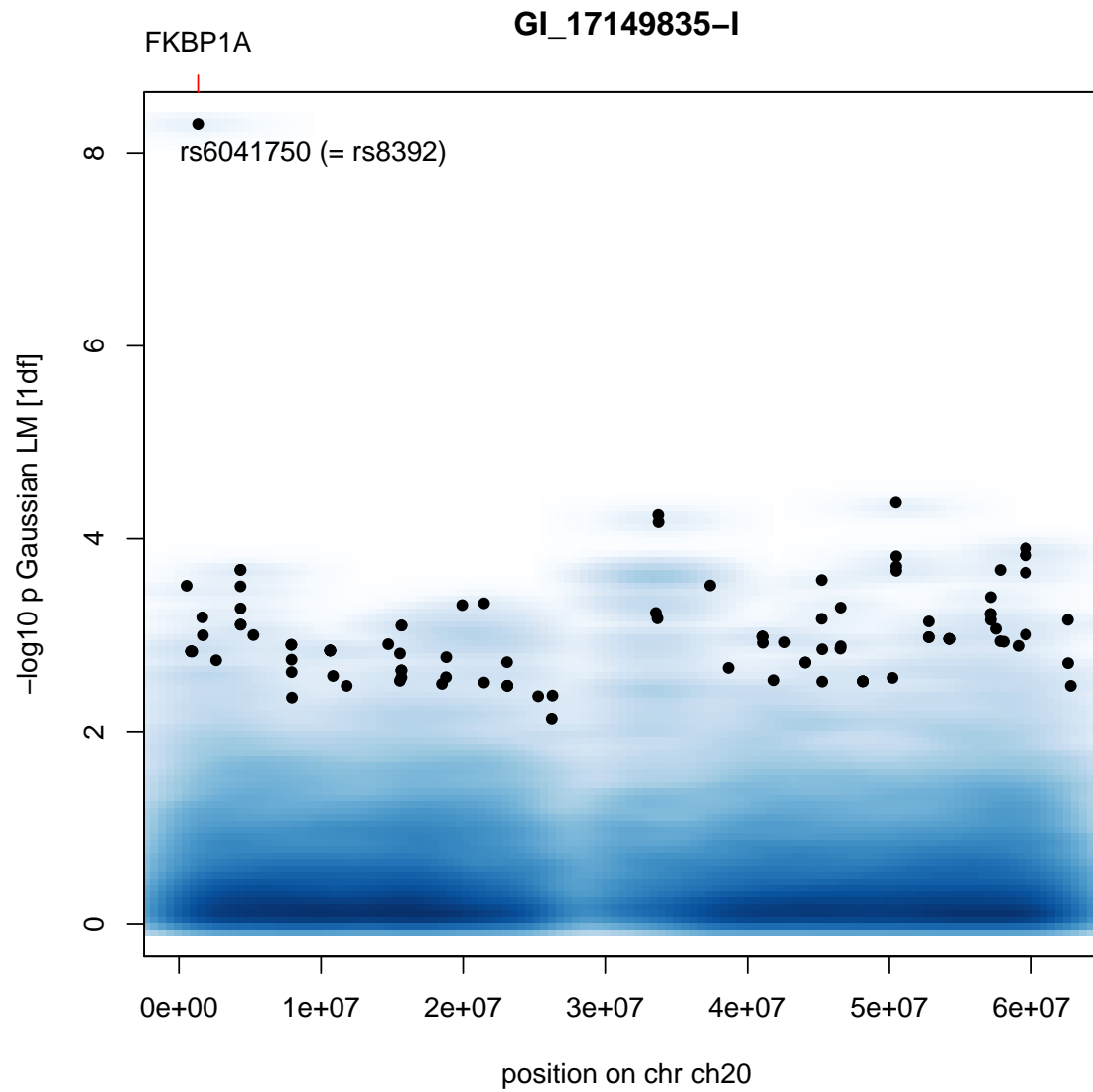
Vince Carey  
Channing Lab  
Harvard Medical School

- Prologue: What is an eQTL?
- Rationales: 1000 genomes; Bioconductor
- Imputation to the 1000 genomes SNP panel
- Expression arrays, RNA-seq, and eQTL identification

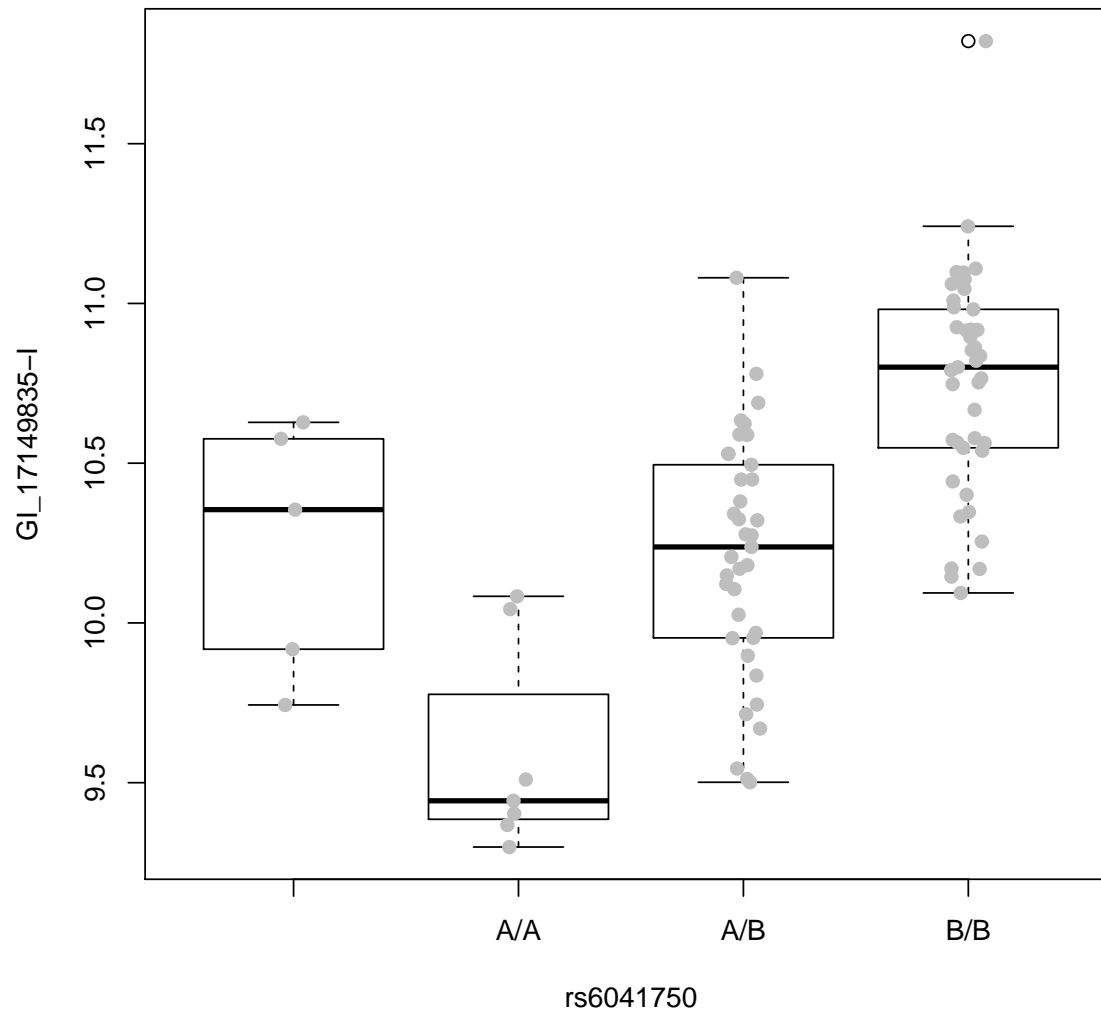
## Prologue: What is an eQTL (expression quantitative trait locus)?

- Arises from a basic form of integrative genome-scale data analysis
- On a cohort of  $N$  individuals
  - SNP-chip yields allele counts for  $S$  SNP,  $S \approx 10^6$
  - Expression array yields mRNA abundance measures for  $G$  genes,  $G \approx 20000$
- perform  $G \times S$  association tests of  $H_{ogs}$  : mean expression of  $g$  is independent of allele count for  $s$
- the best hits are eQTL

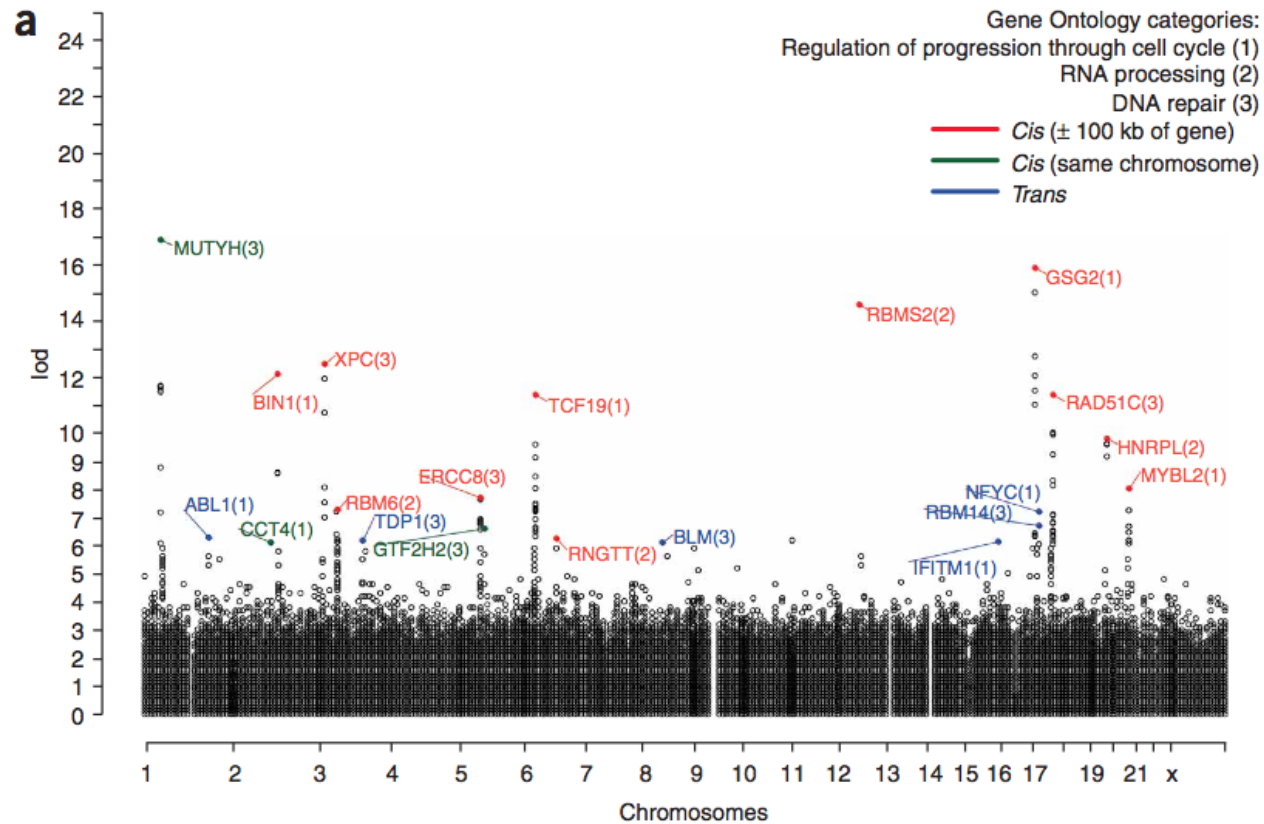
## A chromosome-wide scan for a single gene



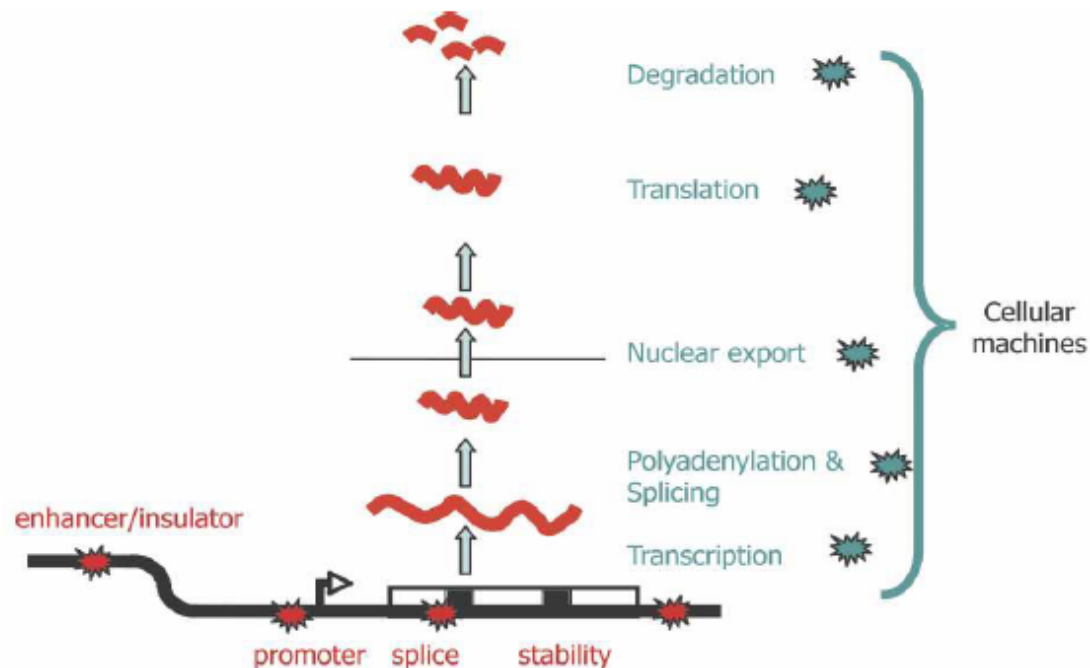
## The 'best SNP' discriminates mean expression



## Dixon 2007 Nat Genet 'global map'



## Why do this? 1: Mechanisms of transcriptional control



**Figure 1.** Plausible sites of action for genetic determinants of mRNA levels. Genetic variations influencing gene expression may reside within the regulatory sequences, promoters, enhancers, splice sites, and secondary structure motifs of the target gene and so be genetically in *cis* (red stars), or there may be variations in the molecular machinery that interact with *cis*-regulatory sequences and so act genetically in *trans* (blue stars).

**Rationale: 1000 genomes xyz**