

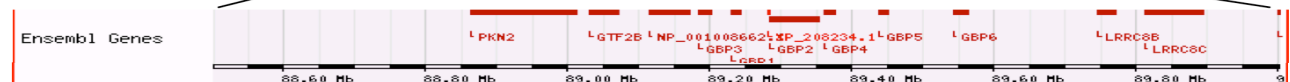
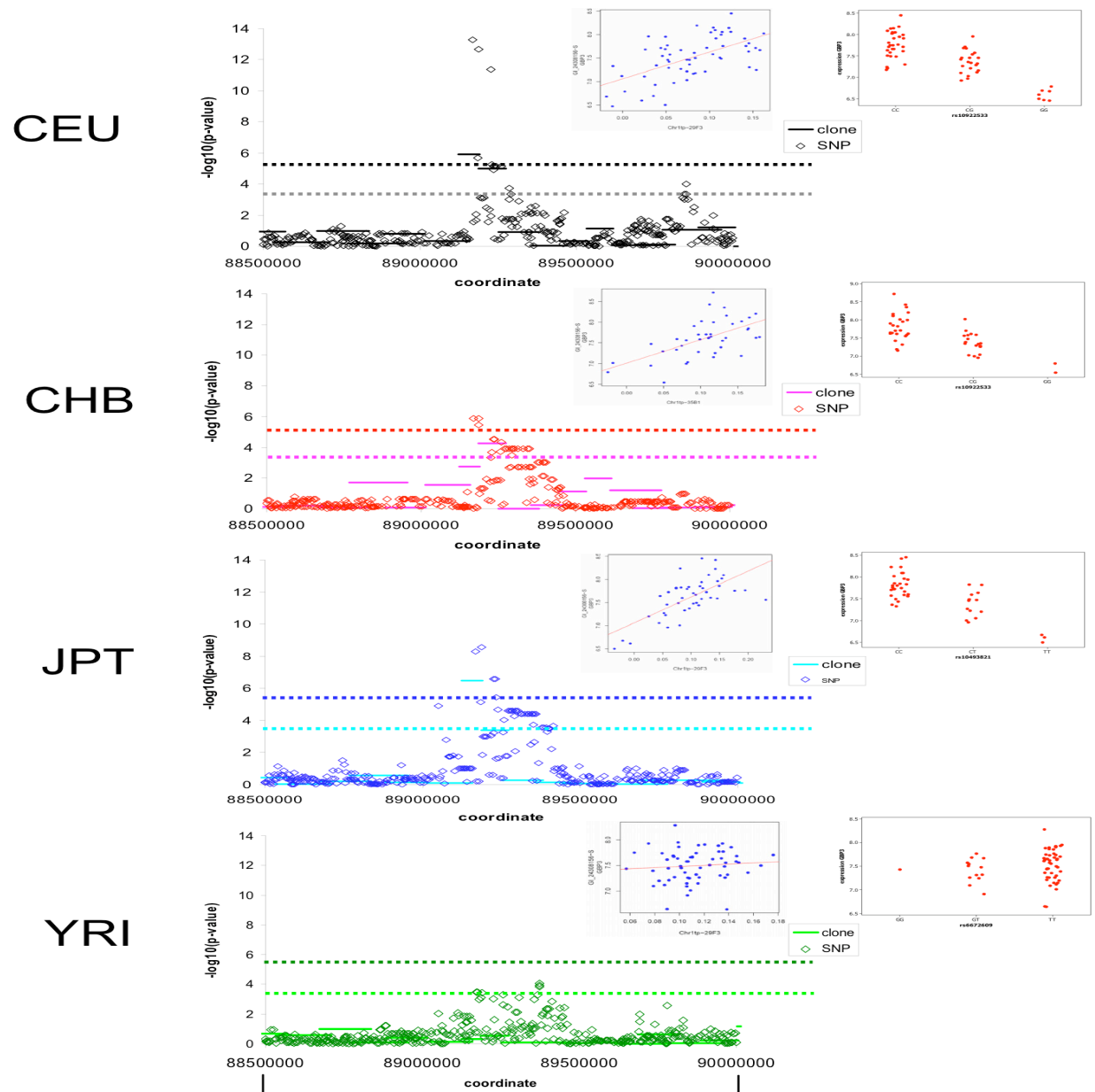
eQTL analysis for Brain exon array data

Crystal March 8th, 2011

Goal

- Find SNP changes that associate with the individual mRNA expression level variation (eSNPs)
- Find the potential causal SNP the cause the gene expression variation across individuals
- Understand the mechanisms of the eSNPs: cis- vs. trans-
- Current status: many eQTL analysis have been done in many tissues, organisms; some potential hotspots have been found; overlap among studies is not high, so the spurious false positive calls prevalent.

Figure S2C



eQTL calculation method review

- SRC
 - LR
 - T-test
 - R^2
 - ANOVA
-
- ...any statistical test can help find the associations...

Dataset

- Exon array expression level for 16 different brain tissues of 57 individuals:
 - 30 Males & 27 Females
 - 32 Europeans; 3 Asians; 2 African/European;
3 Mexicans; 17 Africans;
- 15 Developmental Stage from Embryonic to Late Adulthood

Variables in the dataset

- 1. Individual difference: ✓
- 2. Stage: age difference
- 3. Race: population difference
- 4. Gender
- 5. Tissue specificity: ✓

Ways to remove DEX associated with 2,3,4 factors

- 1. Find DEX across individuals in the same tissue
 - --~50 samples/individuals per tissue; rank gene expression level for all samples, get the highest exp/lowest exp ≥ 2 fold
- 2. Remove gene expression that correlates with age, gender, and race
 - -caveats: cannot remove DEX that doesn't have a general trend in across different stage
- 3. Correlate DEX with SNPs

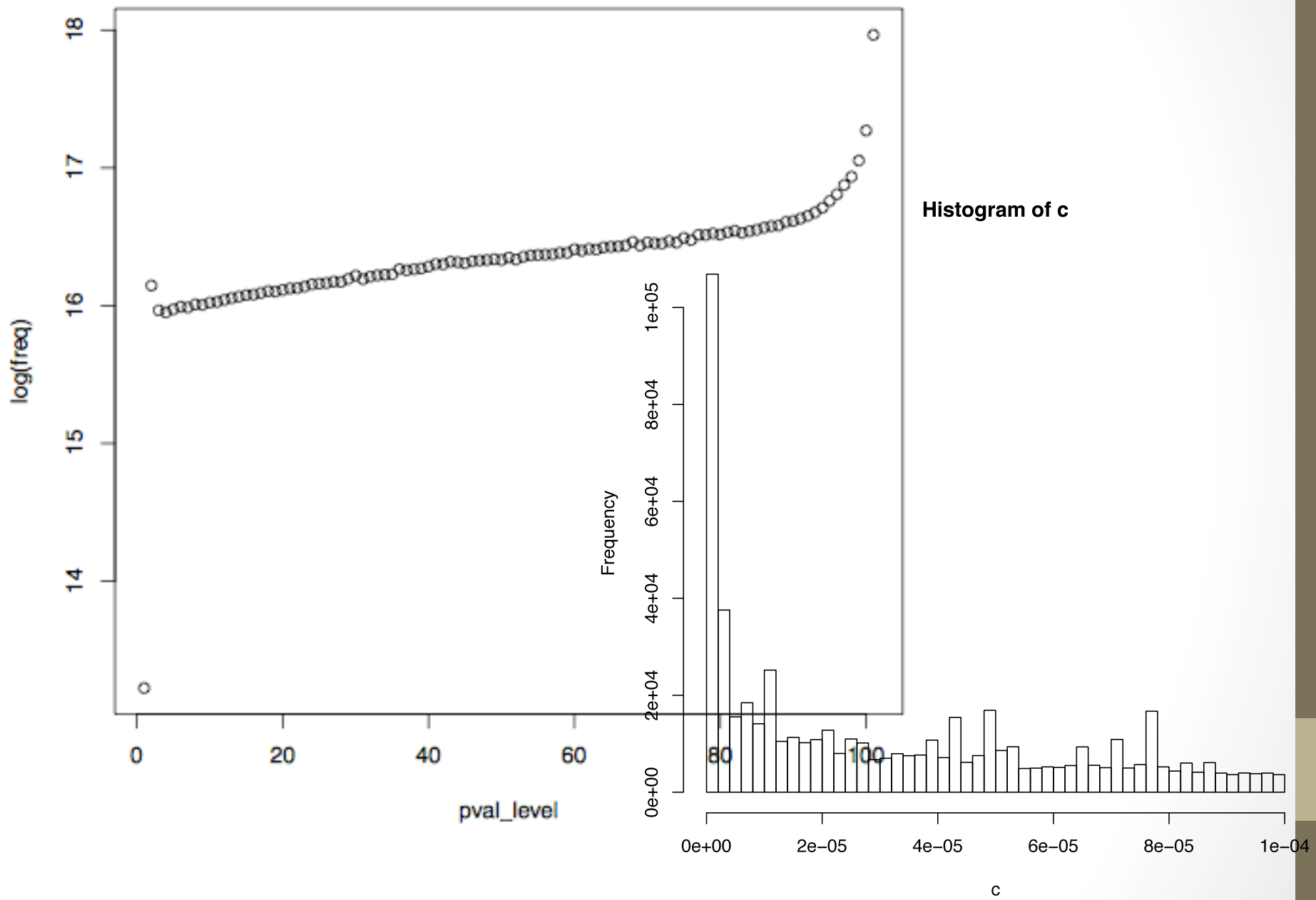
DEX for each sample

Tissue Name	DEX No.	Tissue Name	DEX No.
A1C	765	MFC	1237
AMY	970	OFC	1192
CBL	1539	S1C	838
DFC	1345	STC	983
HIP	1415	STR	997
IPC	849	THM	1660
ITC	881	V1C	1083
M1C	795	VFC	1099

eQTL for each sample/ by eMAP(LR)

Tissue Name	eSNP No. (p-value \leq 1e-4)	eQTL (gene No.)	eSNP No. (p-value \leq 1e-10)	eQTL (gene No.)
A1C	293,966	765	48,819	72
AMY	317,624	749	11,750	86
	107,053	379	104	7

P_value_distribution_A1C



Challenges

- Time consuming: large number of association calculation --- use C or Perl to do calculation
- P-Value cutoff---Permutation; use known dataset to find a better P-Value cutoff

- Future direction:
 - 1. sQTL
 - 2. Tissue-specific eQTL & sQTL; common ones