

IMIDIA

eQTL Analysis

Mickaël Canouil
mickael.canouil@cnrs.fr

November 14, 2016



Expression Data QC

- 1 Expression data were imported from the **GEO** database (private until the 14th of January, 2017).
- 2 Gene annotation from **GEO** (GPL570) for *Affymetrix Human Genome U133 Plus 2.0 Array*.
- 3 Data imported using **affy::ReadAffy** [Gautier et al., 2004].
- 4 Data were normalised using **affy::rma** [Irizarry et al., 2003].
- 5 Batch effect was corrected using **sva::ComBat** [Leek et al., 2012], without the model parameter due to the unbalanced design [Nygaard et al., 2016].

Genotype Data QC

- 1 Genotype data generated using *Illumina 2.5M Omniarray beadchip*.
- 2 Quality control not performed in Lille (MAF>0.05; HWE>0.001, Call Rate>0.95) :
 - 421 samples
 - 1,233,520 SNPs
- 3 Imputation performed with **ShapeIT** (v2.r790)[Delaneau et al., 2012] and **Impute2** (v2.3.2)[Howie et al., 2009] using 1,000 genomes panel (Phase 3)[The 1000 Genomes Project Consortium, 2015] (Imputation Quality >0.9) :
 - 421 samples
 - 7,574,416 SNPs

eQTL Analysis

- 1 eQTL analysis was performed using **FastQTL** [Ongen et al., 2015] on four datasets :

OD (Organ Donors) All samples (n=100)

OD (Organ Donors) Control samples (n=81)

LCM (Laser Capture Micro-dissection) All samples (n=103)

LCM (Laser Capture Micro-dissection) Control samples (n=32)

- 2 Settings :

Cis Window 1 Mb => 500 Kb

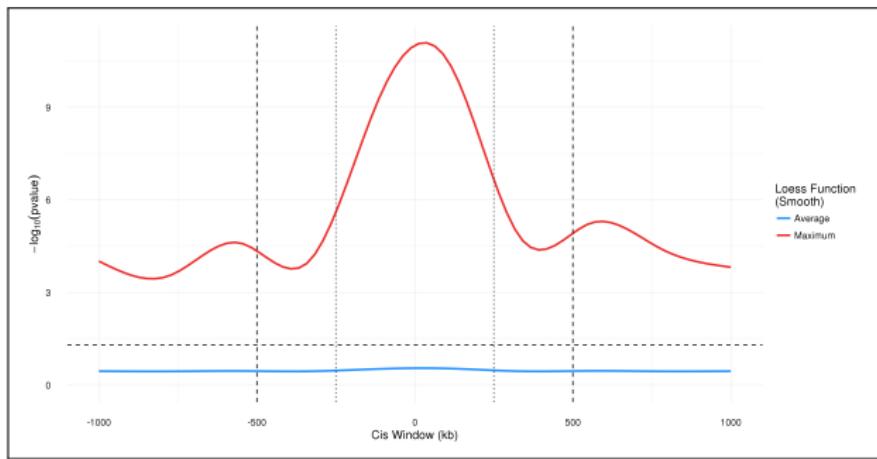
Covariates gender and age

Permutation 1,000 - 10,000 (adaptative procedure)

eQTL Analysis

Cis Window

We set the cis-window of **500 Kb** based on the eQTL results (p-values) for a cis-window of **1 Mb**.



eQTL Analysis

Results

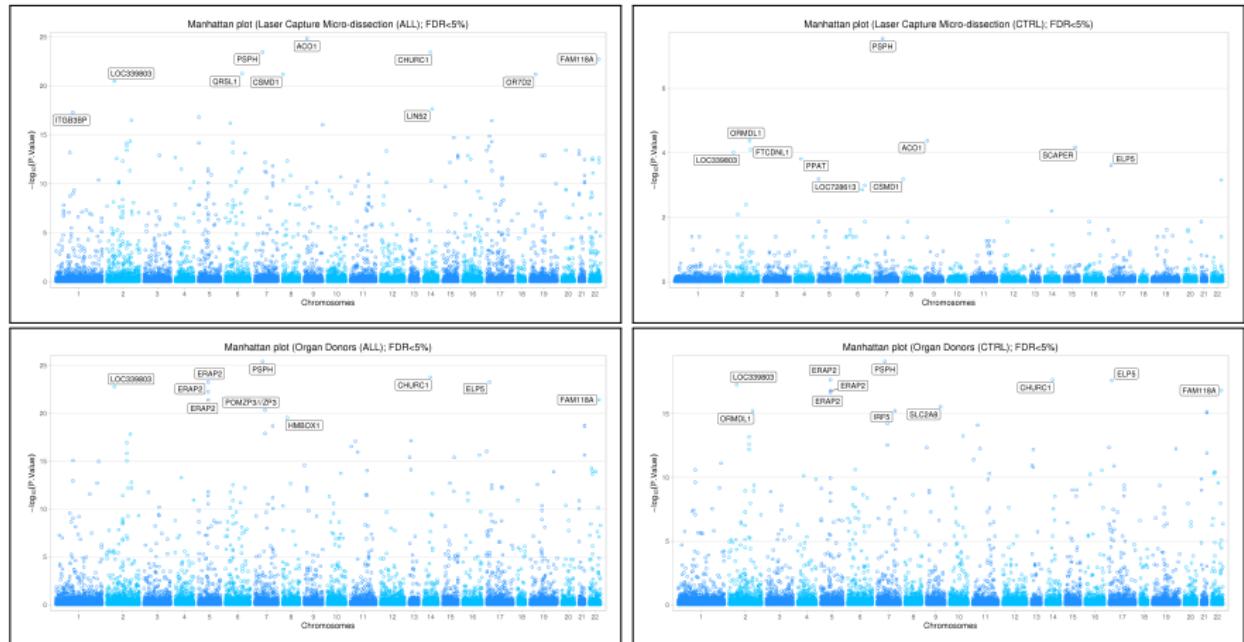
Number of significant eQTL pairs using the nominal or "*permutation pass*" p-value from **FastQTL** with $\alpha = 0.05$.

eQTL	N	nNominal	nFDR
LCM_ALL	42,422	4,381	1,021
OD_ALL	42,422	5,100	1,173
LCM_CTRL	42,422	2,660	48
OD_CTRL	42,422	4,489	842

N is the number of Affymetrix probes.

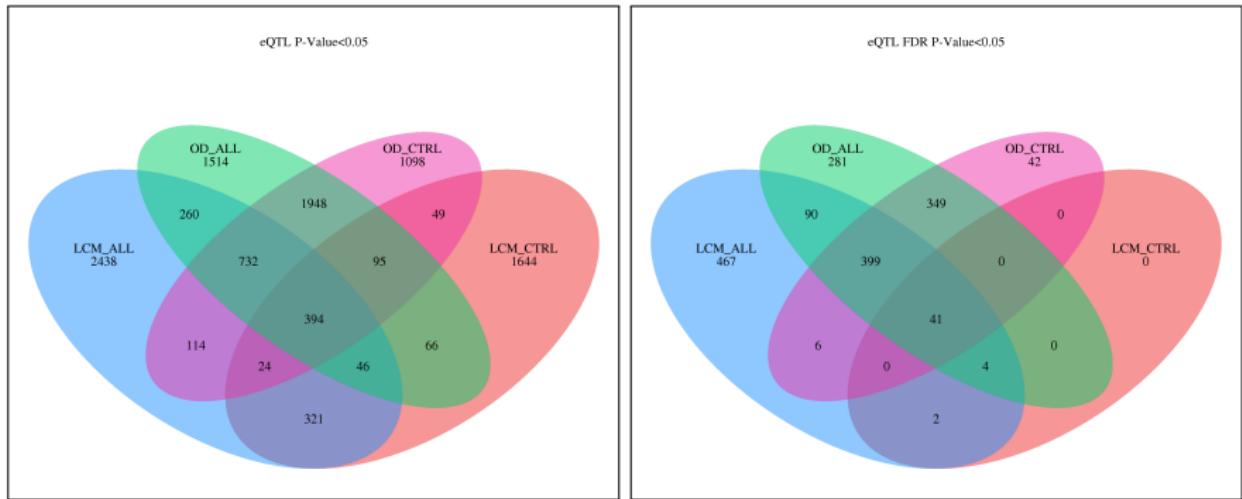
eQTL Analysis

Manhattan Plots



eQTL Analysis

Venn Diagram



Enrichment in GWAS

GWAS data

GWAS results from **MAGIC** and **DIAGRAM** downloaded from their respective website.

Then annotated with **dbSNP** (144) and **refSeqGenes** (hg19) for chromosome, position and gene symbol.

MAGIC "Fasting Glucose": 2,467,443 SNPs and 24,897 genes;

MAGIC "In Fasting Insulin": 2,458,109 SNPs and 24,877 genes;

MAGIC "In HOMA-B": 2,453,970 SNPs and 24,864 genes;

MAGIC "In HOMA-IR": 2,455,091 SNPs and 24,865 genes;

DIAGRAM "Stage 1 GWAS": 2,431,532 SNPs and 24,789 genes.

Enrichment in GWAS

Analysis

Enrichment analysis in GWAS results, was performed using the eQTL dataset from the "*permutation pass*" of [FastQTL](#), and the permutation p-value obtained via beta approximation.

First, data were **aggregated by genes** using the **lowest p-value** in both GWAS results and eQTL results.

Second, enrichment were obtained using **Fisher test** for nominal **p-value<0.05** in both GWAS and eQTL results (no multiple testing correction for now).

Enrichment in GWAS

Analysis: Laser Capture Micro-dissection (LCM)

LCM enrichment for all pancreatic samples:

GWAS	estimate	p.value	conf.low	conf.high	Top eQTL Gene
DIAGRAM	1.1	0.0003788	1.064	1.241	ACO1; CHURC1; QRSL1; CSMD1; FTCDNL1
MAGIC.FG	1.1	0.000414	1.062	1.239	ACO1; FAM118A; QRSL1; CSMD1; OR7D2
MAGIC.FI	1.1	0.05979	0.9967	1.162	ACO1; FAM118A; QRSL1; CSMD1; LIN52
MAGIC.HOMAB	1.1	0.00246	1.042	1.215	ACO1; CSMD1; LIN52; LOC728613; FTCDNL1
MAGIC.HOMAIR	1.1	0.0179	1.016	1.185	ACO1; FAM118A; CSMD1; LIN52; LOC728613

LCM enrichment for control samples:

GWAS	estimate	p.value	conf.low	conf.high	Top eQTL Gene
DIAGRAM	1.2	0.0001593	1.088	1.308	ACO1; SCAPER; FTCDNL1; PPAT; ELP5
MAGIC.FG	1.3	3.948e-10	1.219	1.467	ACO1; FTCDNL1; PPAT; ELP5; LOC728613
MAGIC.FI	1.1	0.04395	1.002	1.204	ACO1; SCAPER; LOC728613; CSMD1; FAM118A
MAGIC.HOMAB	1.1	0.09087	0.9875	1.187	ACO1; SCAPER; FTCDNL1; PPAT; LOC728613
MAGIC.HOMAIR	1.1	0.003268	1.046	1.257	ACO1; SCAPER; LOC728613; CSMD1; FAM118A

Enrichment in GWAS

Analysis: organ Donors (OD)

OD enrichment for all pancreatic samples:

GWAS	estimate	p.value	conf.low	conf.high	Top eQTL Gene
DIAGRAM	1.1	0.003902	1.034	1.198	CHURC1; ELP5; ERAP2; HMBOX1; IRF5
MAGIC.FG	1.1	0.0004836	1.058	1.224	ELP5; ERAP2; FAM118A; POMZP3; B3GLCT
MAGIC.FI	1.1	0.01334	1.019	1.18	ERAP2; FAM118A; B3GLCT; KRTAP5-AS1; C17orf97
MAGIC.HOMAB	1.1	0.001006	1.05	1.215	IRF5; POMZP3; B3GLCT; KRTAP5-AS1; HSD17B12
MAGIC.HOMAIR	1.1	0.01901	1.014	1.174	FAM118A; IRF5; B3GLCT; KRTAP5-AS1; HSD17B12

OD enrichment for control samples:

GWAS	estimate	p.value	conf.low	conf.high	Top eQTL Gene
DIAGRAM	1.1	0.0009281	1.053	1.227	CHURC1; ERAP2; ELP5; IRF5; HMBOX1
MAGIC.FG	1.1	0.01314	1.019	1.187	ERAP2; ELP5; FAM118A; LDHC; POMZP3
MAGIC.FI	1.1	0.01052	1.023	1.191	ERAP2; FAM118A; MRPL43; C17orf97; HSD17B12
MAGIC.HOMAB	1.2	0.0002752	1.067	1.242	IRF5; MRPL43; POMZP3; HSD17B12; B3GLCT
MAGIC.HOMAIR	1.1	0.01309	1.02	1.187	FAM118A; IRF5; MRPL43; HSD17B12; B3GLCT

References I

- Delaneau, O., Marchini, J., and Zagury, J.-F. (2012). A linear complexity phasing method for thousands of genomes. *Nature methods*, 9(2):179–181.
- Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy—analysis of affymetrix genechip data at the probe level. *Bioinformatics*, 20(3):307–315.
- Howie, B. N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*, 5(6):e1000529.
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264.

References II

- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, 28(6):882–883.
- Nygaard, V., Rødland, E. A., and Hovig, E. (2016). Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics*, 17(1):29–39.
- Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T., and Delaneau, O. (2015). Fast and efficient qtl mapper for thousands of molecular phenotypes. *Bioinformatics*, page btv722.
- The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.