Novel and concordant eQTLs from analysis of iPSC-derived megakaryocytes and platelets in the Genetic Studies of Atherosclerosis Risk (GeneSTAR) project

Kai Kammers

The Johns Hopkins University School of Medicine
Baltimore, MD, USA

November 6, 2018

Introduction

- GeneSTAR Research Center: Genetic Studies of Atherosclerosis Risk
  `http://www.genestarstudy.com/`
  PIs: Rasika Mathias, Lewis Becker, Nauder Faraday, ...

- **Goal:** Understanding the biology of platelet aggregation.

Introduction

- GeneSTAR Research Center: Genetic Studies of Atherosclerosis Risk
  `http://www.genestarstudy.com/`
  PIs: Rasika Mathias, Lewis Becker, Nauder Faraday, ...

- **Goal:** Understanding the biology of platelet aggregation.

- Platelets in the circulating blood mediate normal hemostasis and initiate repair at sites of vascular injury, but may also initiate pathological arterial thrombosis that produces heart attacks and strokes.

- GWAS studies have identified common variants associated with platelet aggregation.

- The biological mechanism has remained largely undefined because most signals have occurred in introns or intergenic regions rather than in protein coding regions of known genes.

Introduction

- Platelets are derived from megakaryocytes (large platelet-producing bone marrow cells) in the bone marrow, but themselves are anucleate.

Introduction

- Platelets are derived from megakaryocytes (large platelet-producing bone marrow cells) in the bone marrow, but themselves are anucleate.
- The importance of studying both platelets and their precursor megakaryocytes cannot be overstated given the anucleate state of the platelet but nonetheless its ability to translate and make protein in its adult state.

Introduction

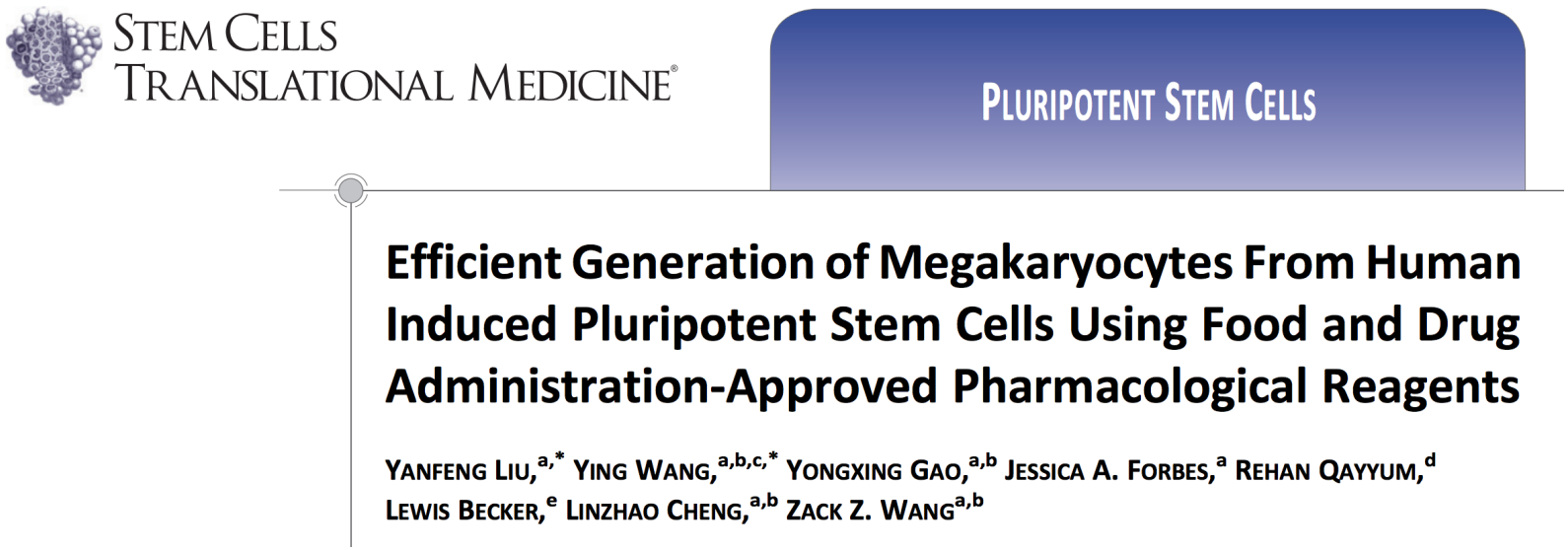
- Platelets are derived from megakaryocytes (large platelet-producing bone marrow cells) in the bone marrow, but themselves are anucleate.
- The importance of studying both platelets and their precursor megakaryocytes cannot be overstated given the anucleate state of the platelet but nonetheless its ability to translate and make protein in its adult state.
- As there is tremendous difficulty in obtaining megakaryocytes in sufficient numbers from large numbers of subjects (they reside in low levels in bone marrow and available only by invasive bone marrow aspirate or biopsy) we rely on validated induced pluripotent stem cell (iPSC)- derived megakaryocytes (MKs) in this study.

Introduction

- Platelets are derived from megakaryocytes (large platelet-producing bone marrow cells) in the bone marrow, but themselves are anucleate.

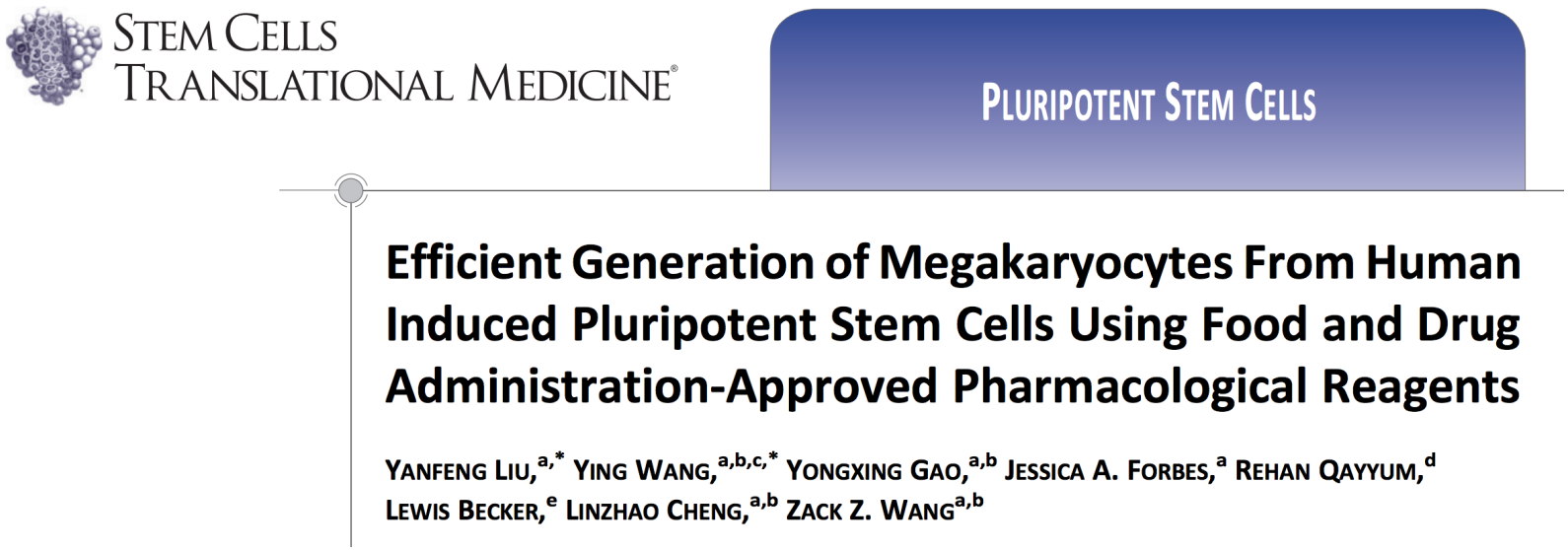
Introduction

- Platelets are derived from megakaryocytes (large platelet-producing bone marrow cells) in the bone marrow, but themselves are anucleate.



Introduction

- Platelets are derived from megakaryocytes (large platelet-producing bone marrow cells) in the bone marrow, but themselves are anucleate.



- Induced Pluripotent Stem Cells $\rightarrow$ Megakaryocytes $\rightarrow$ Platelets

[Liu et al. 2015, *Stem Cells Transl Med*, PMID: 25713465]

Introduction

Platelet mRNA: the meaning behind the message

Rowley, Jesse W.[a,b]; Schwertz, Hansjörg[a,c]; Weyrich, Andrew S.[a,b]

Current Opinion in Hematology: September 2012 - Volume 19 - Issue 5 - p 385–391
doi: 10.1097/MOH.0b013e328357010e
HEMOSTASIS AND THROMBOSIS: Edited by Joseph E. Italiano and Jorge A. DiPaola

[Rowley et al. 2012, *Curr Opin Hematol*, PMID: 22814651]

Introduction

Cancer Cell
Article

RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics

Myron G. Best,[1,2] Nik Sol,[3] Irsan Kooi,[4] Jihane Tannous,[5] Bart A. Westerman,[2] François Rustenburg,[1,2] Pepijn Schellen,[2,6] Heleen Verschueren,[2,6] Edward Post,[2,6] Jan Koster,[7] Bauke Ylstra,[1] Najim Ameziane,[4] Josephine Dorsman,[4] Egbert F. Smit,[8] Henk M. Verheul,[9] David P. Noske,[2] Jaap C. Reijneveld,[3] R. Jonas A. Nilsson,[2,6,10] Bakhos A. Tannous,[5,12] Pieter Wesseling,[1,11,12] and Thomas Wurdinger[2,5,6,12,*]

[Best et al. 2015, *Cancer Cell*, PMID: 26525104]

iPSC - MK: differential expression

RESEARCH ARTICLE

Integrity of Induced Pluripotent Stem Cell (iPSC) Derived Megakaryocytes as Assessed by Genetic and Transcriptomic Analysis

Kai Kammers[1,2], Margaret A. Taub[2], Ingo Ruczinski[2], Joshua Martin[3], Lisa R. Yanek[3], Alyssa Frazee[2], Yongxing Gao[4], Dixie Hoyle[4], Nauder Faraday[3], Diane M. Becker[3], Linzhao Cheng[4], Zack Z. Wang[4], Jeff T. Leek[2], Lewis C. Becker[3]*, Rasika A. Mathias[3]

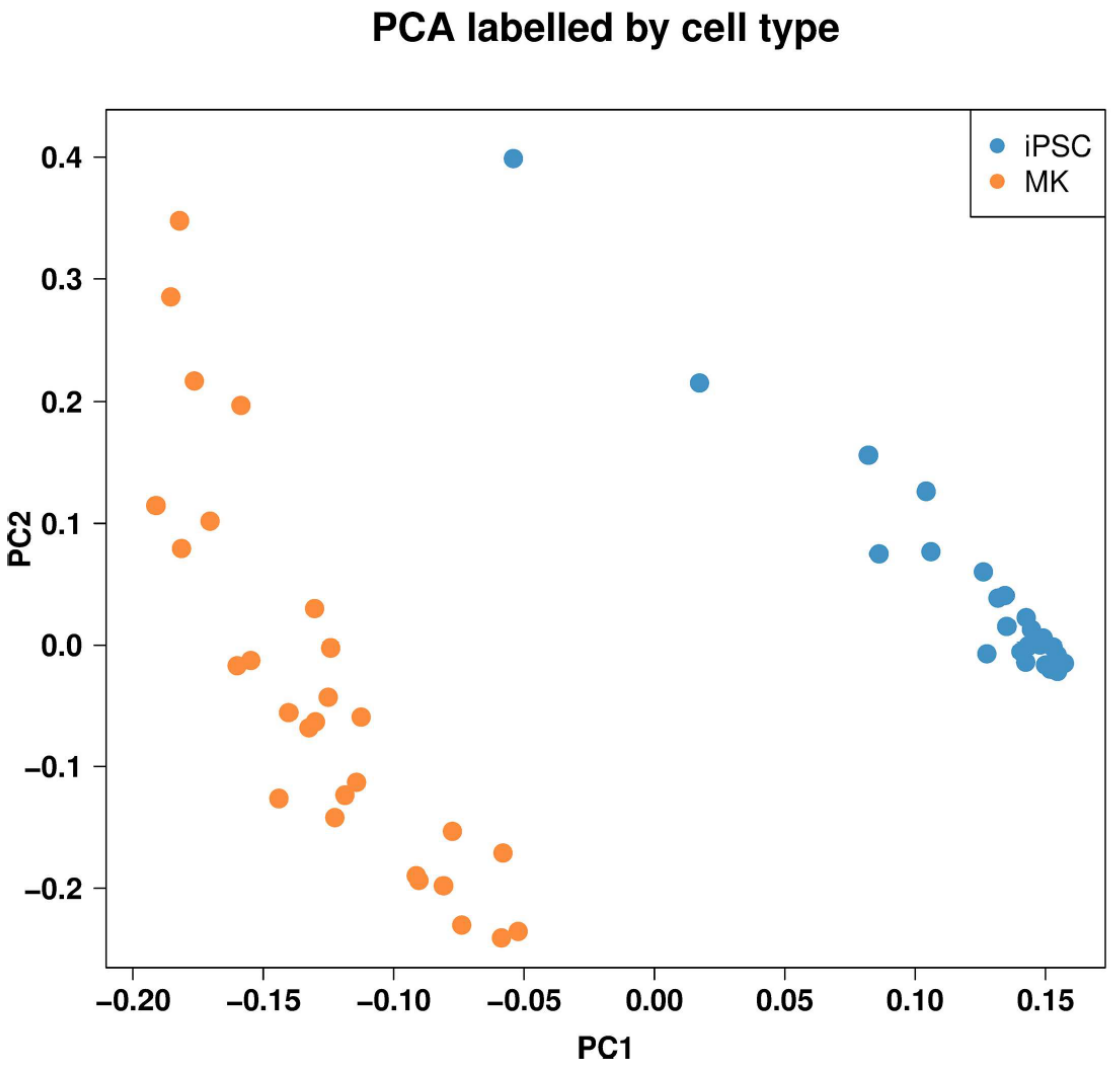
1 Division of Biostatistics and Bioinformatics, Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **2** Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **3** The GeneSTAR Research Program, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America, **4** Division of Hematology and Institute for Cell Engineering, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America

[Kammers et al. 2017, *PloS one*, PMID: 28107356]

iPSC - MK: data and model

- 14 biological replicates.
- Each biological replicate has RNA-sequencing data for two technical replicates of iPSCs and two technical replicates of MKs.
- In total: 56 samples for this study.
- 33,287 transcripts after filtering [robust variability filter].
- **Goal:** Detect differentially expressed transcripts and perform gene set enrichment analysis.
- **Model:** $\Delta^{(iPSC-MK)} = \alpha + \beta \cdot PC1 + \gamma \cdot PC2 + \epsilon.$

iPSC - MK: PCA


PCA labelled by cell type
iPSC
MK
PC2
PC1
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.20
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15

iPSC - MK: gene set enrichment

Genes *turned on* in MKs following differentiation from the iPSCs, we observed the following highly biologically relevant gene sets in the list of statistically significant gene sets ($q < 0.001$):

- Platelet activation [GO:0030168]
- Blood coagulation [GO:0007596]
- Megakaryocyte development [GO:0035855]
- Platelet formation [GO:0030220]
- Platelet degranulation [GO:0002576]
- Platelet aggregation [GO:0070527].

[Kammers et al. 2017, *PloS one*, PMID 28107356]

eQTL

- **eQTL (expression Quantitative Trait Locus):** Genomic locus (genotype, SNP) that contribute to variation in expression levels of mRNAs (gene/transcript expression).

eQTL

- **eQTL (expression Quantitative Trait Locus):** Genomic locus (genotype, SNP) that contribute to variation in expression levels of mRNAs (gene/transcript expression).

- **Approach:** Fit a linear model for each gene-SNP pair to test the association between gene expression E and genotype S (and additional covariates C_k, $k = 1, ..., K$)

$$E = \alpha + \beta S + \sum_k \gamma_k C_k + \epsilon.$$

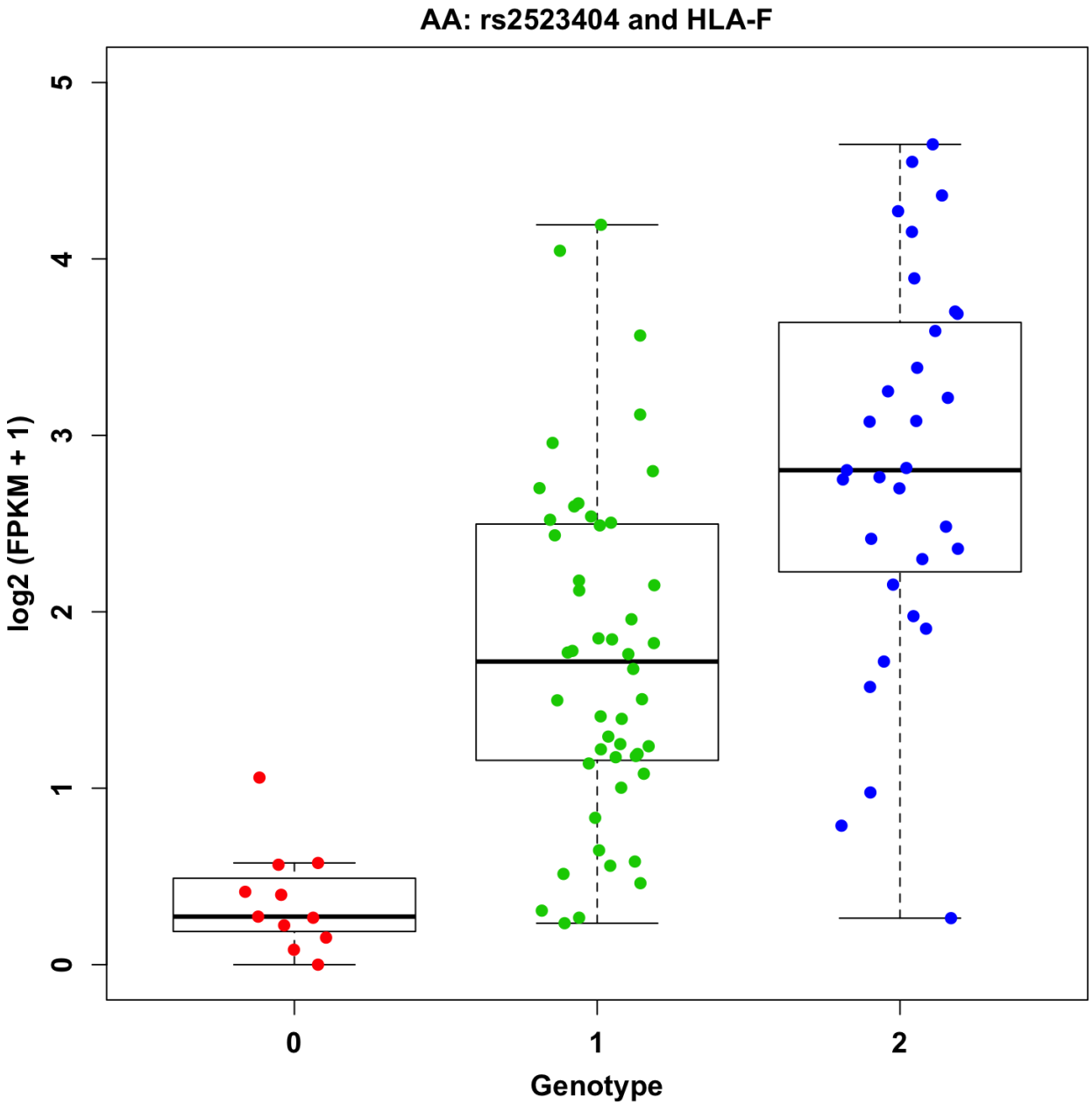
eQTL

- **eQTL (expression Quantitative Trait Locus):** Genomic locus (genotype, SNP) that contribute to variation in expression levels of mRNAs (gene/transcript expression).
- **Approach:** Fit a linear model for each gene-SNP pair to test the association between gene expression E and genotype S (and additional covariates C_k, $k = 1, ..., K$)

$$E = \alpha + \beta S + \sum_k \gamma_k C_k + \epsilon.$$

- MatrixEQTL package in R.

MKs: eQTL example



eQTL

- **local eQTLs [or cis eQTLs]:**

  eQTLs that map to the approximate location of their gene-of-origin.

- **distant eQTLs [or trans eQTLs]:**

  eQTLs that map far from the location of their gene-of-origin, often on different chromosomes.

[details: Albert and Kruglyak, 2015, *Nat Rev Genet*, PMID: 25707927]

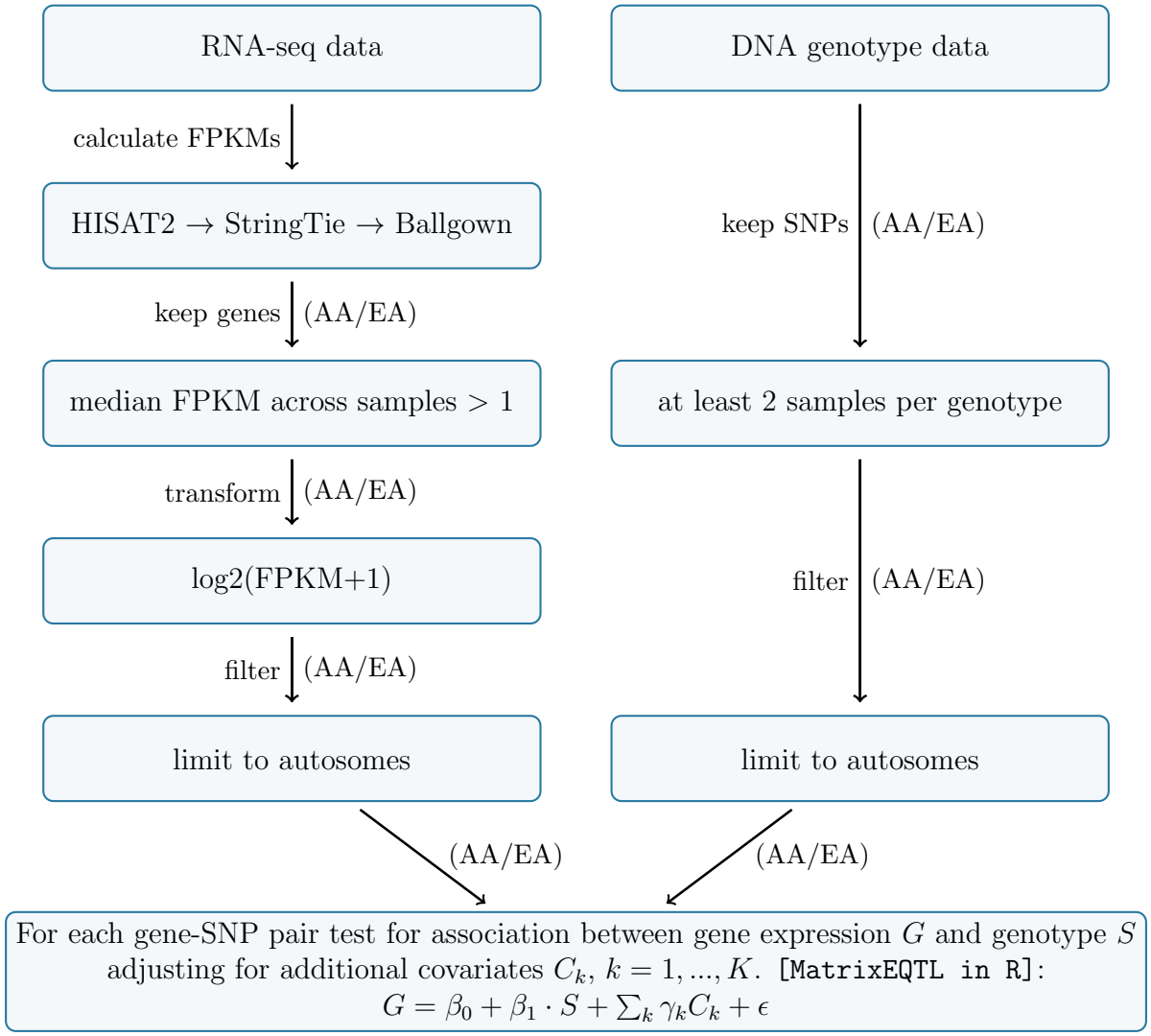
eQTL data sets

- Pheno data set (IDs, covariates).
- Gene expression data set.
- Genotype data set.
- Genomic positions of genes.
- Genomic positions of SNPs.

eQTL data sets

- Pheno data set (IDs, covariates).
- Gene expression data set.
- Genotype data set.
- Genomic positions of genes.
- Genomic positions of SNPs.
- eQTL analyses were carried out stratified by ancestry and cell type:
  - African Americans [AA] and European Americans [EA],
  - MKs and Platelets.

eQTLs


RNA-seq data
DNA genotype data
calculate FPKMs
HISAT2 → StringTie → Ballgown
keep SNPs (AA/EA)
keep genes (AA/EA)
median FPKM across samples > 1
at least 2 samples per genotype
transform (AA/EA)
log2(FPKM+1)
filter (AA/EA)
filter (AA/EA)
limit to autosomes
limit to autosomes
(AA/EA)
(AA/EA)
For each gene-SNP pair test for association between gene expression G and genotype S adjusting for additional covariates C_k, $k = 1, ..., K$. [MatrixEQTL in R]:
$G = \beta_0 + \beta_1 \cdot S + \sum_k \gamma_k C_k + \epsilon$

MKs: eQTL data

- Genotype and expression data of 194 MKs.

MKs: eQTL data

- Genotype and expression data of 194 MKs.
- African Americans (AA):
  - 87 samples
  - 9,509 genes
  - 625,933 SNPs

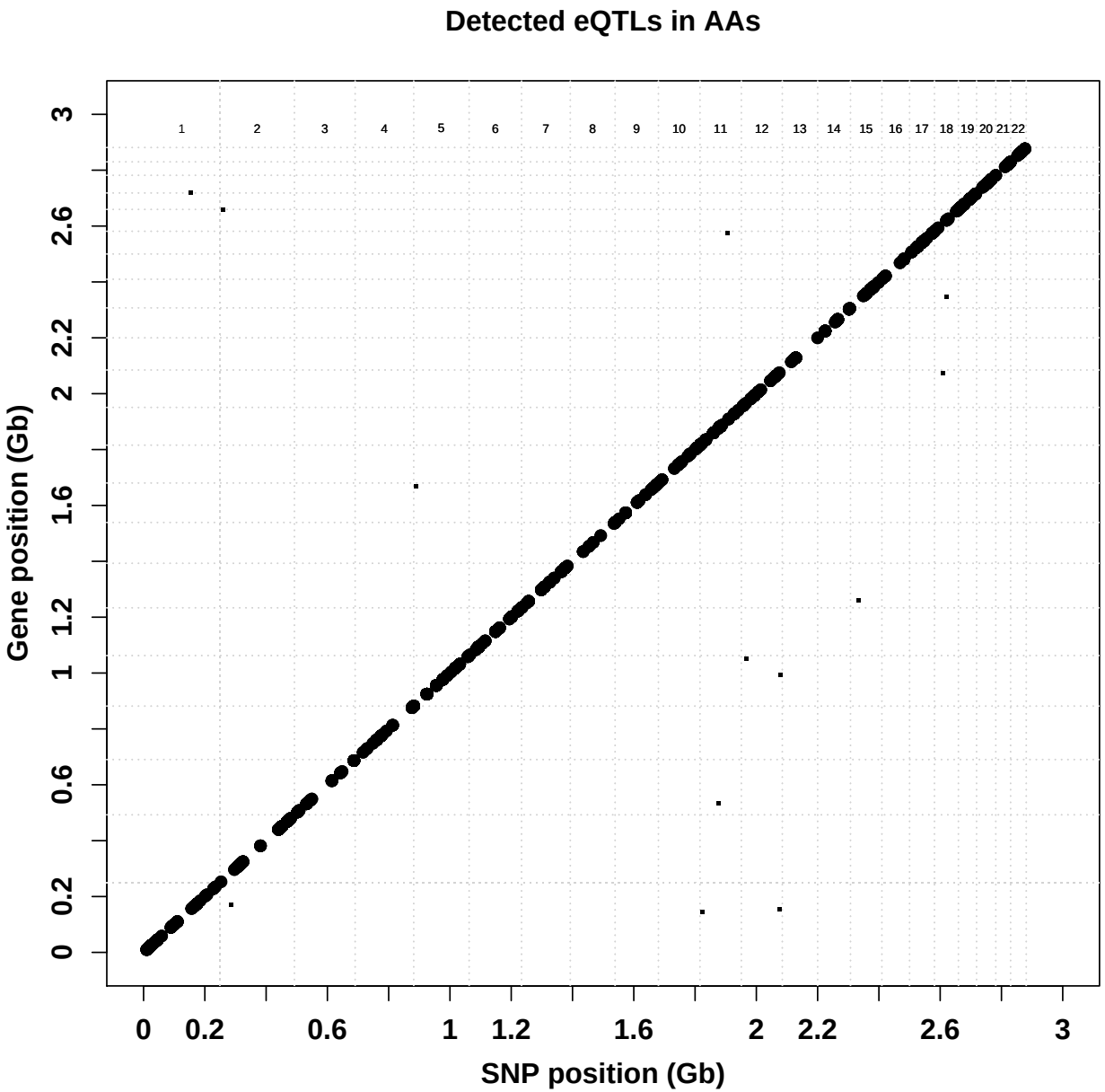
MKs: eQTL data

- Genotype and expression data of 194 MKs.
- African Americans (AA):
  - 87 samples
  - 9,509 genes
  - 625,933 SNPs
- European Americans (EA):
  - 107 samples
  - 9,660 genes
  - 619,318 SNPs

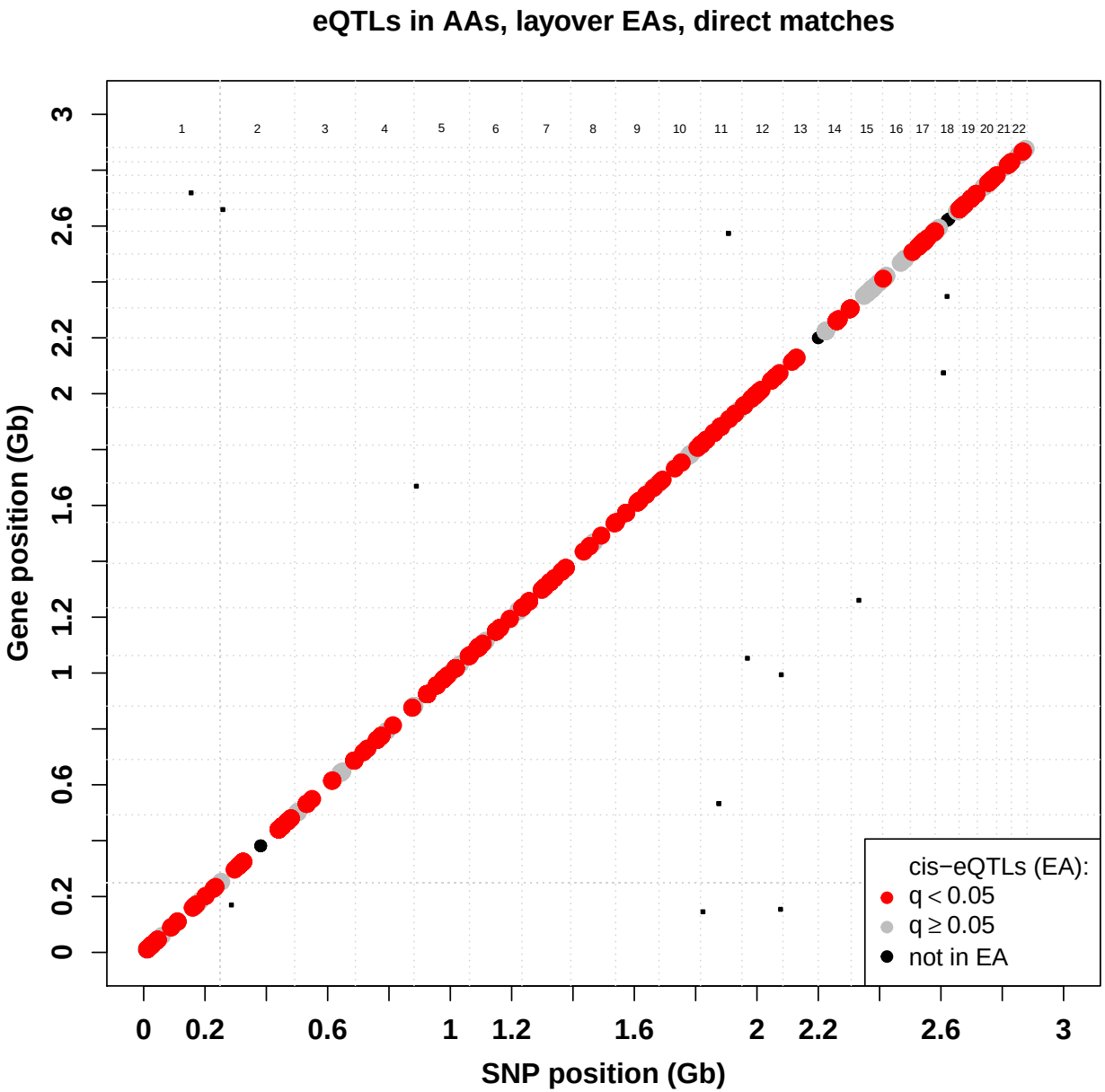
MKs: eQTL data

- Genotype and expression data of 194 MKs.
- African Americans (AA):
  - 87 samples
  - 9,509 genes
  - 625,933 SNPs
- European Americans (EA):
  - 107 samples
  - 9,660 genes
  - 619,318 SNPs
- Approx. 6 billion tests for each subgroup!

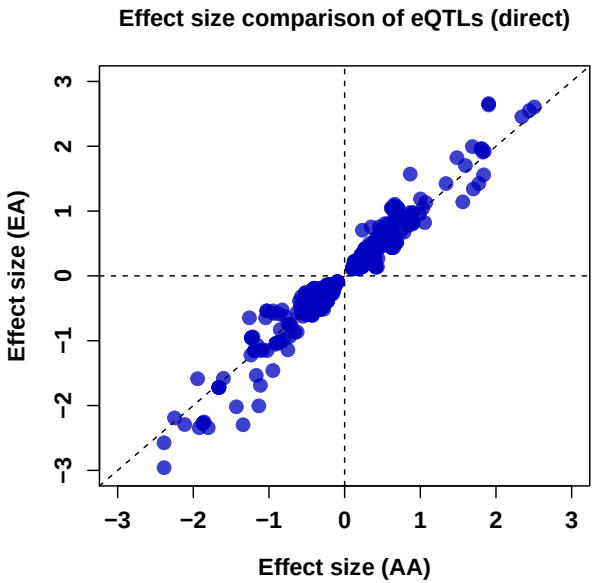
MKs: eQTL analysis

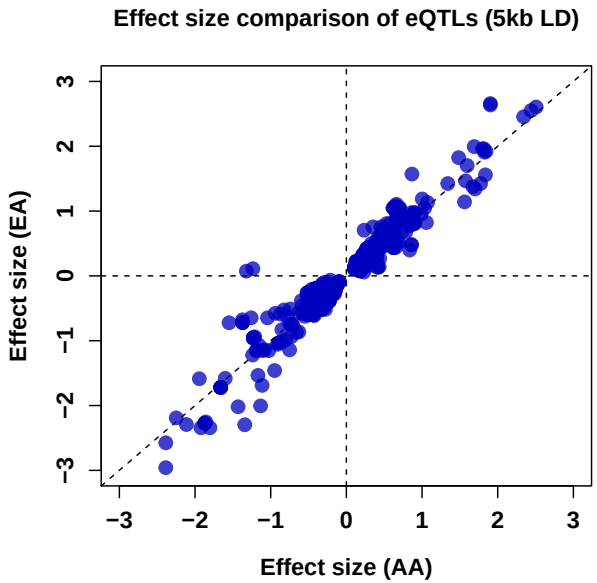


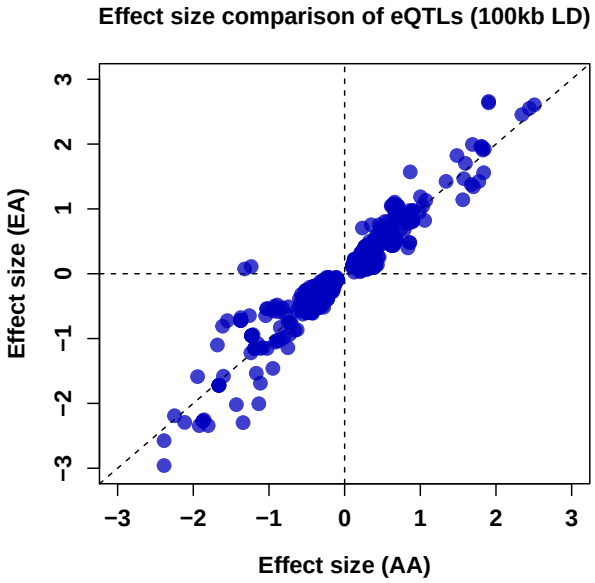
MKs: eQTL analysis

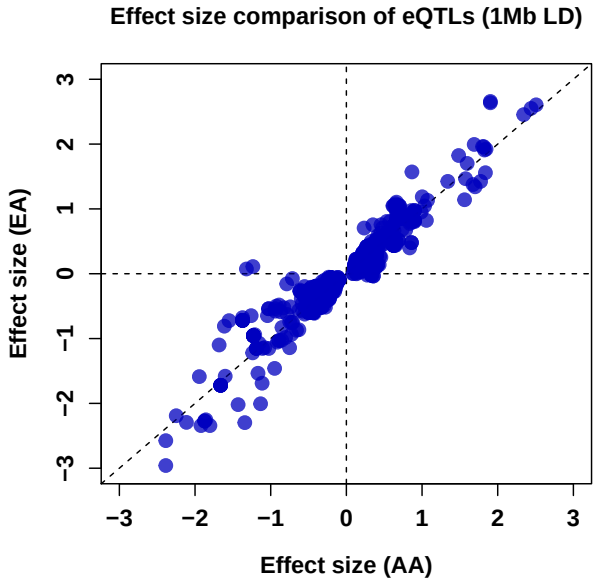


MKs: eQTL analysis

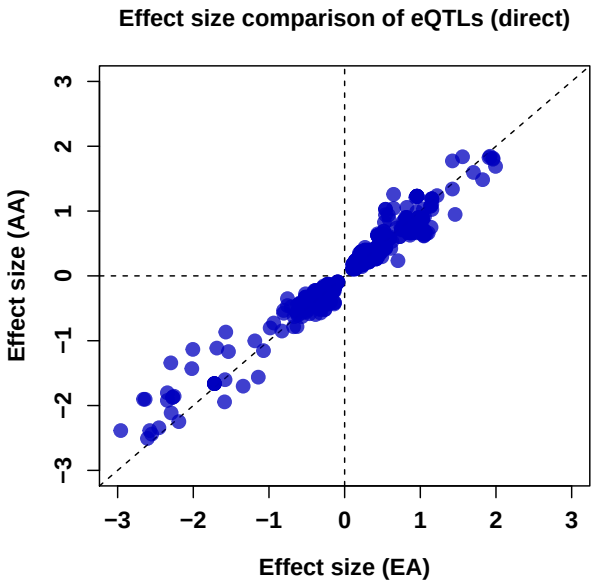


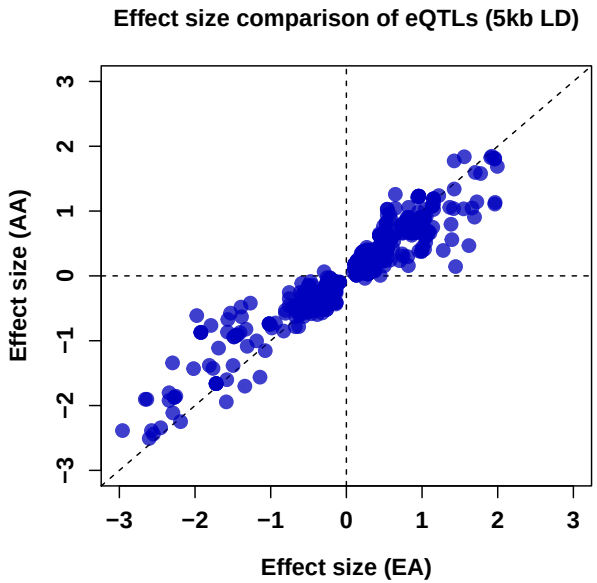


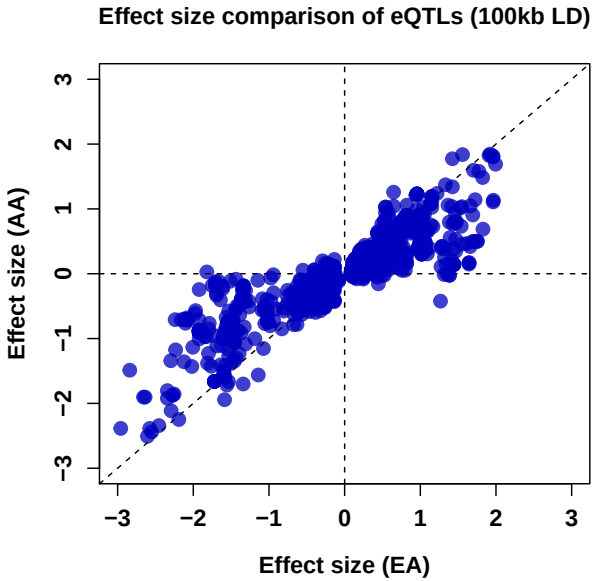


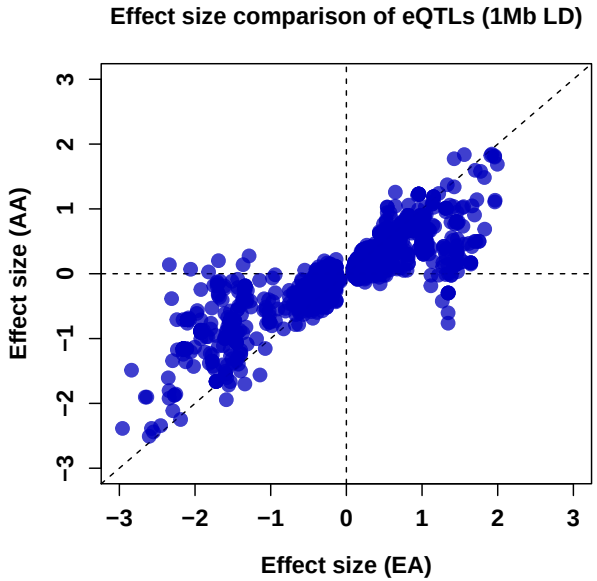


MKs: eQTL analysis









Platelets: eQTL data

- Genotype and expression data of 297 Platelets.

Platelets: eQTL data

- Genotype and expression data of 297 Platelets.
- African Americans (AA):
  - 112 samples
  - 2,836 genes
  - 660,363 SNPs

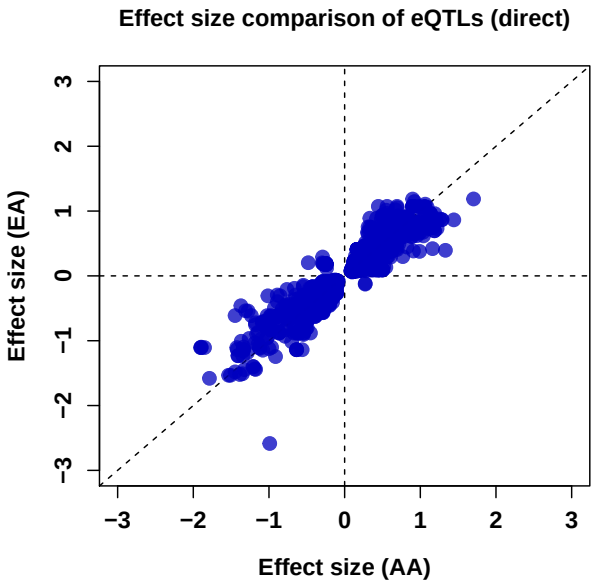
Platelets: eQTL data

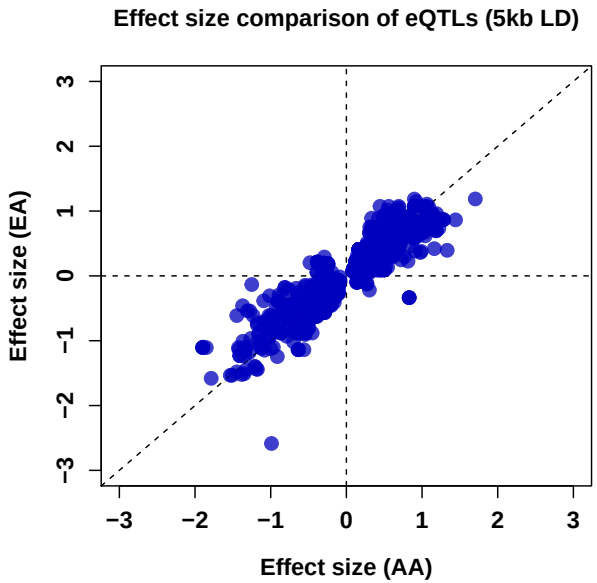
- Genotype and expression data of 297 Platelets.
- African Americans (AA):
  - 112 samples
  - 2,836 genes
  - 660,363 SNPs
- European Americans (EA):
  - 185 samples
  - 2,579 genes
  - 676,242 SNPs

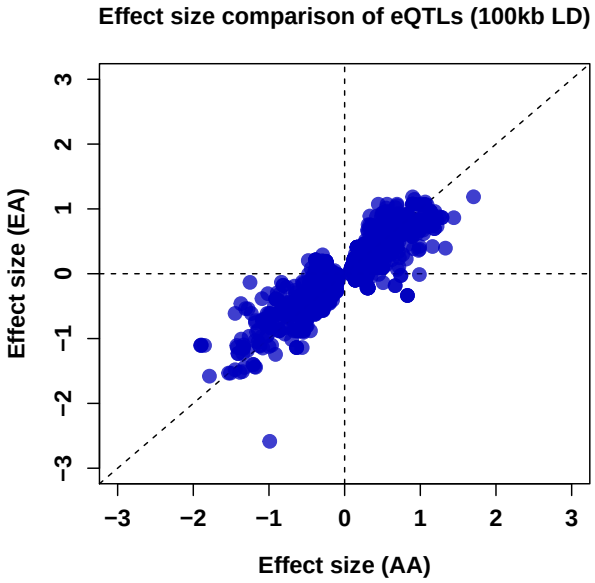
Platelets: eQTL data

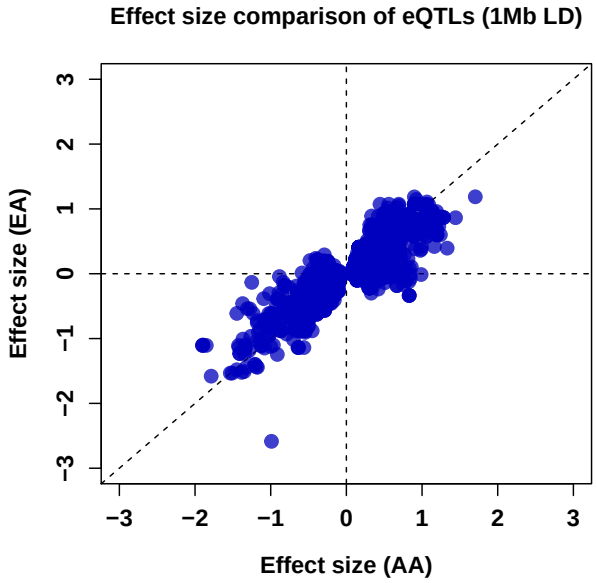
- Genotype and expression data of 297 Platelets.
- African Americans (AA):
  - 112 samples
  - 2,836 genes
  - 660,363 SNPs
- European Americans (EA):
  - 185 samples
  - 2,579 genes
  - 676,242 SNPs
- Approx. 2 billion tests for each subgroup!

Platelets: eQTL analysis

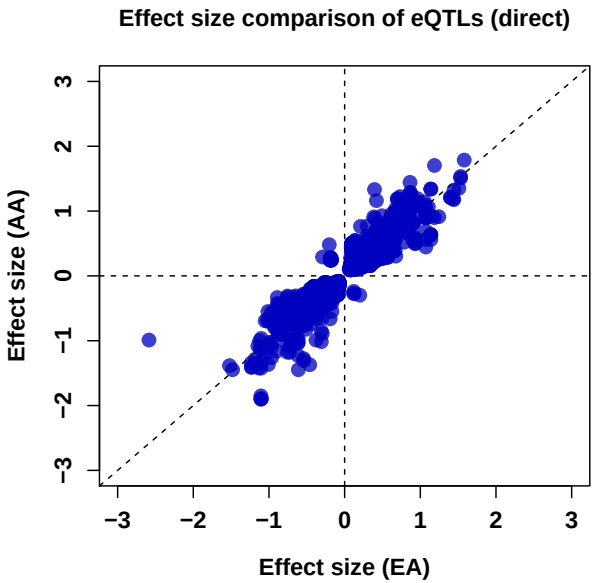


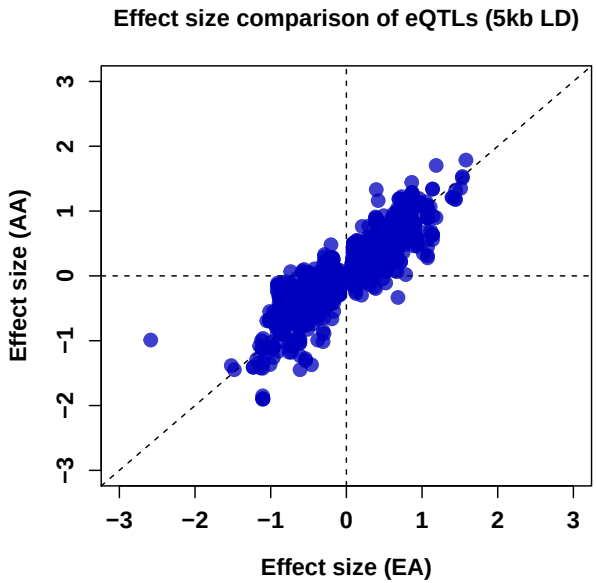


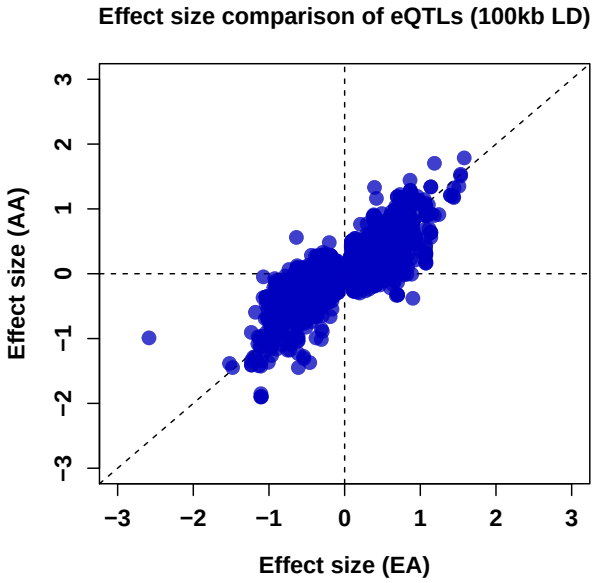


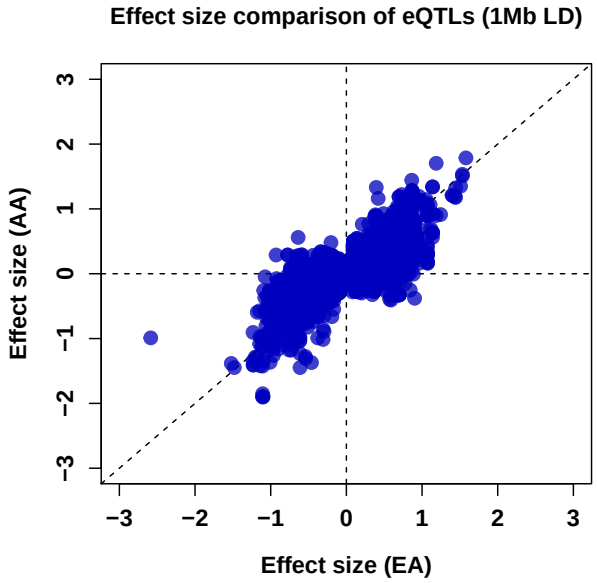


Platelets: eQTL analysis

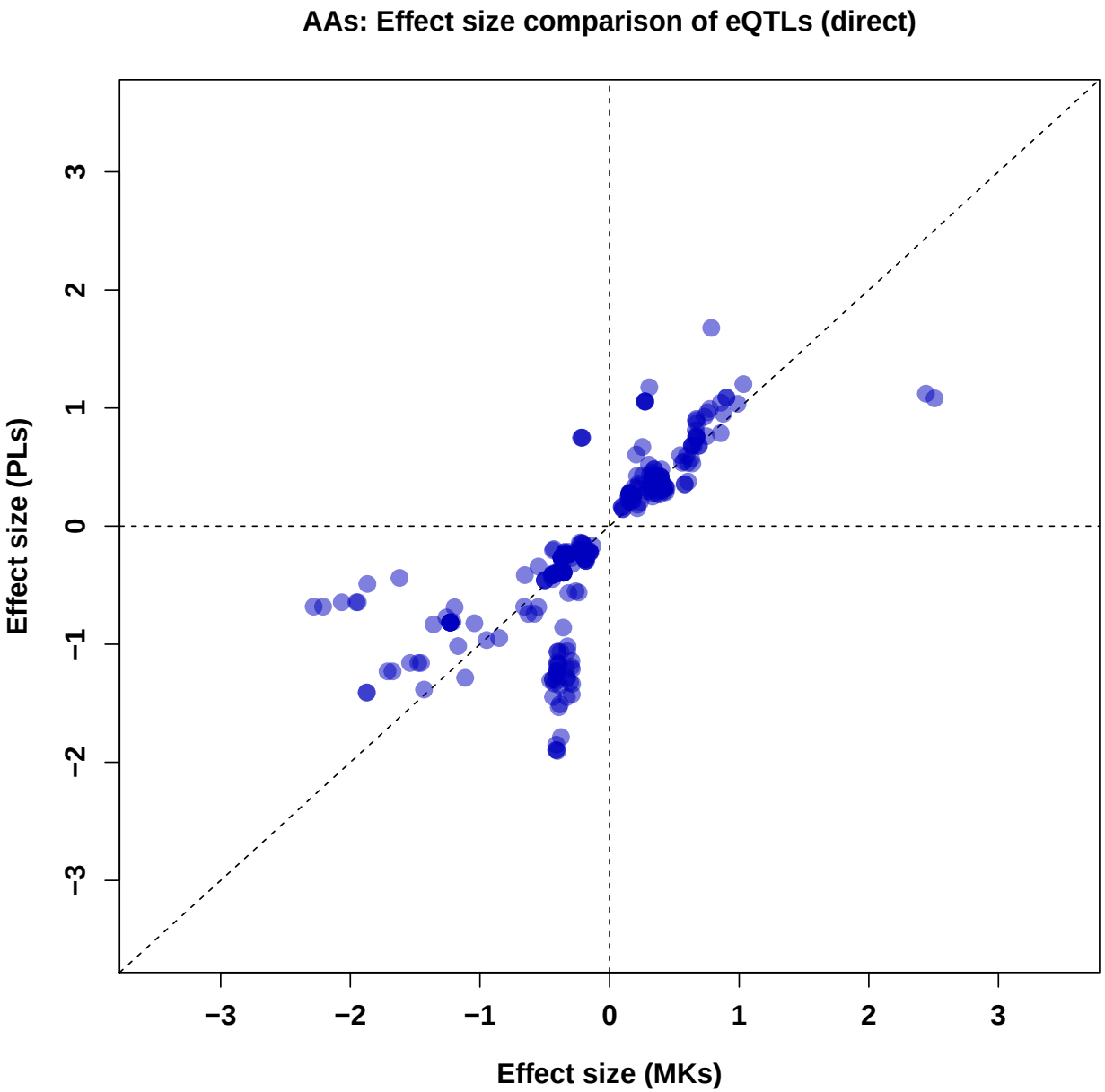




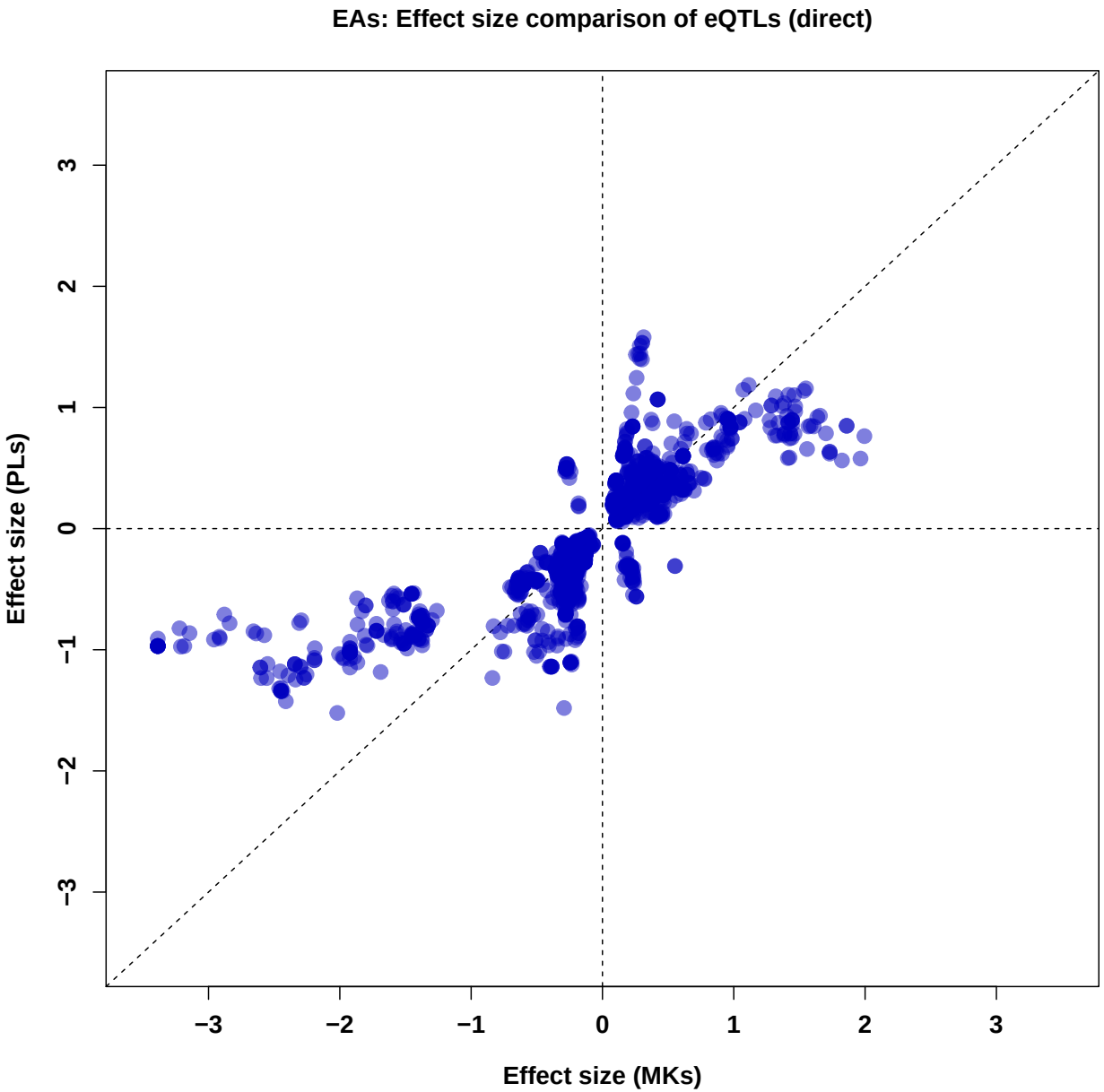




Platelets: eQTL analysis



Platelets: eQTL analysis



Gene Set Enrichment

We found that the eQTL associated genes regulate key pathways in platelet physiology:

- Platelet degranulation.
- Platelet activation and aggregation.
- KLF1 targets down.
- Hemostasis.
- Cell projection.
- Positive regulation of cell communication.

Summary eQTLs

- RNA-sequencing data show strong patterns of differentiation that strongly support the differentiation of the MKs from the iPSCs.

Summary eQTLs

- RNA-sequencing data show strong patterns of differentiation that strongly support the differentiation of the MKs from the iPSCs.
- For MKs [Platelets] we found 8,629 [23,355] local eQTLs in European Americans and 1,280 [5,740] local eQTLs in the African Americans with $p.adj < 0.05$.

Summary eQTLs

- RNA-sequencing data show strong patterns of differentiation that strongly support the differentiation of the MKs from the iPSCs.
- For MKs [Platelets] we found 8,629 [23,355] local eQTLs in European Americans and 1,280 [5,740] local eQTLs in the African Americans with $p.adj < 0.05$.
- The differences between ancestry groups decrease as we transition from a SNP-SNP to a distance-based replication approach indicating the SNP-SNP differences are likely due to LD.

Summary eQTLs

- RNA-sequencing data show strong patterns of differentiation that strongly support the differentiation of the MKs from the iPSCs.
- For MKs [Platelets] we found 8,629 [23,355] local eQTLs in European Americans and 1,280 [5,740] local eQTLs in the African Americans with $p.adj < 0.05$.
- The differences between ancestry groups decrease as we transition from a SNP-SNP to a distance-based replication approach indicating the SNP-SNP differences are likely due to LD.
- We identified 294 [1,423] eQTLs in AAs [EAs] that are present in both tissues MKs and Platelets considering direct matches. Among these shared eQTLs 99.0% [95.8 %] have the same direction of the estimated effect sizes.

Summary eQTLs

- RNA-sequencing data show strong patterns of differentiation that strongly support the differentiation of the MKs from the iPSCs.
- For MKs [Platelets] we found 8,629 [23,355] local eQTLs in European Americans and 1,280 [5,740] local eQTLs in the African Americans with $p.adj < 0.05$.
- The differences between ancestry groups decrease as we transition from a SNP-SNP to a distance-based replication approach indicating the SNP-SNP differences are likely due to LD.
- We identified 294 [1,423] eQTLs in AAs [EAs] that are present in both tissues MKs and Platelets considering direct matches. Among these shared eQTLs 99.0% [95.8 %] have the same direction of the estimated effect sizes.
- Comparing to GTEx Portal (44 different tissue types, version 6), we identified that approx. 36% [55%] of these eQTLs are specific to MKs [Platelets].

Current Work

- Samples now have whole genome sequence data (WGS), and we are transitioning to approx. N=5,000,000 SNPs after filtering instead of relying on a GWAS array, that was not adequate for the AA samples.

Current Work

- Samples now have whole genome sequence data (WGS), and we are transitioning to approx. N=5,000,000 SNPs after filtering instead of relying on a GWAS array, that was not adequate for the AA samples.

- iPSC derived MKs: gene and protein expression landscape.

Current Work

- Samples now have whole genome sequence data (WGS), and we are transitioning to approx. N=5,000,000 SNPs after filtering instead of relying on a GWAS array, that was not adequate for the AA samples.
- iPSC derived MKs: gene and protein expression landscape.
- Identification of molecular mechanisms through which gender and sex hormones impact platelets and MKs.

Team

- Margaret Taub and Ingo Ruczinski
  [JHU Biostatistics]
- Rasika Mathias, Lewis Becker, Nauder Faraday, and Lisa Yanek
  [GeneSTAR, JHU Medicine]
- Benjamin Rodrigez and Andrew Johnson
  [NHLBI]