Integrity of induced pluripotent stem cell derived megakaryocytes as assessed by genetic and transcriptomic analysis

K. Kammers[1], J.T. Leek[1], I. Ruczinski[1], J. Martin[2], M.A. Taub[1], L.R. Yanek[2], A. Frazee[1], D. Hoyle[3], N. Faraday[2], D. Becker[2], L. Cheng[2], Z.Z. Wang[2], L. Becker[2], R.A. Mathias[2]

[1]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health
[2]The GeneSTAR Program, Johns Hopkins School of Medicine
[3]Johns Hopkins School of Medicine

Background

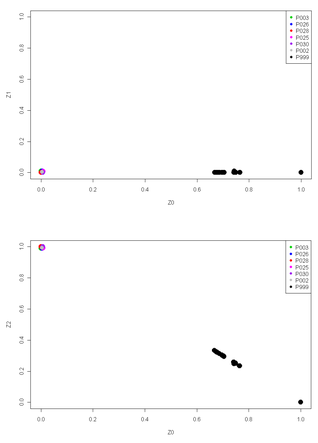
- ✓ The aggregation of activated platelets on ruptured or eroded atherosclerotic plaques is a critical step in the initiation of thromboses of the arterial system, which subsequently results in acute thrombosis-mediated ischemic syndromes such as myocardial infarction, stroke, and peripheral arterial occlusions.
- ✓ The propensity of platelets to aggregate and initiate thromboses is thought to be dependent on local vascular factors, systemic factors which may change over time (such as circadian rhythms, inflammatory processes, smoking, and neuroendocrine stress), and genetic factors which modify platelet aggregability.
- ✓ Platelets are anucleate cells generated from megakaryocytes in the bone marrow, with a life span of 7-10 days in the circulation. Platelets contain mRNA transcripts derived from the parent megakaryocyte that they are able to translate to form new proteins.
- ✓ Platelets on the other hand are readily available in peripheral blood, and as many as 14,000 different transcripts have been identified in platelets by sequencing its mRNA (RNAseq) or by using cDNA microarrays to identify known transcripts of known genes.
- ✓ The importance of studying both platelets and their pre-curser megakaryocytes cannot be overstated given the anucleate state of the platelet but nonetheless its ability to translate and make protein in its adult state.
- ✓ Unfortunately, there is tremendous difficulty in obtaining megakaryocytes in sufficient numbers from large numbers of subjects as they reside in low levels in bone marrow and available only by invasive bone marrow aspirate or biopsy.
- ✓ In this study we aim to study the transciptomic signature of induced pluripotent stem cell (iPSC)-derived megakaryocytes (MKs) with the intention of ultimately integrating these results with our GWAS studies on platelet aggregation to identify functional determinants of platelet aggregation.

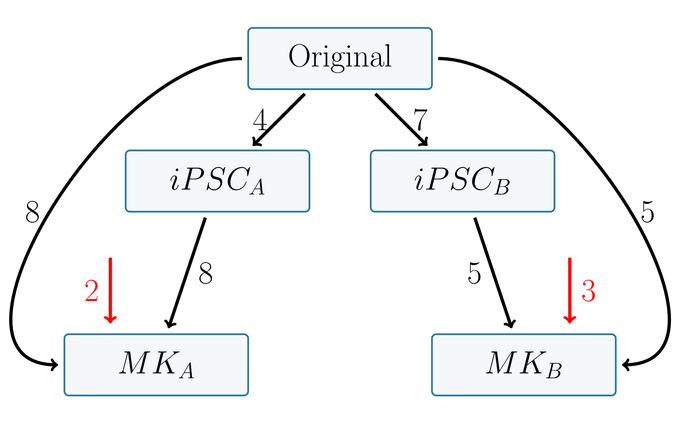
Study Design

To evaluate the integrity of the MKs that we derived from iPSCs that were generated on our GeneSTAR subjects, we performed the following specific experiments:

1. The OMNI Express GWAS array was run on source DNA from PBMCs, up to two cell lines of iPSCs from the PBMCs, and up to two derived MK lines corresponding to the two iPSC lines. We used these data to evaluate the presence of genetic mutations occurring in the iPSCs that were then 'transmitted' onto the derived MKs.
2. We relied on the OMNI GWAS array data to look for copy number variation between the PBMC, iPSC and MK lines to evaluate the presence of larger structural variants occurring in the iPSCs that were then 'transmitted' onto the derived MKs.
3. We performed RNA-seq on RNA derived from the iPSCs and MKs in a paired experiment to look for genes/transcripts that represent the differentiation process.
4. We tested for eQTLs in MKs to identify genetic determinants of transcript level.

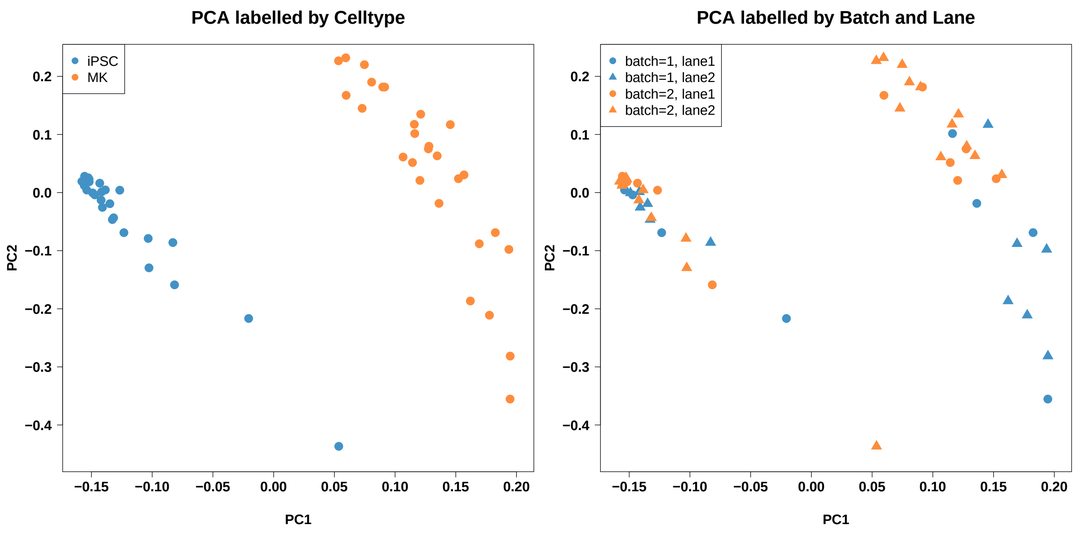
Genomic Integrity





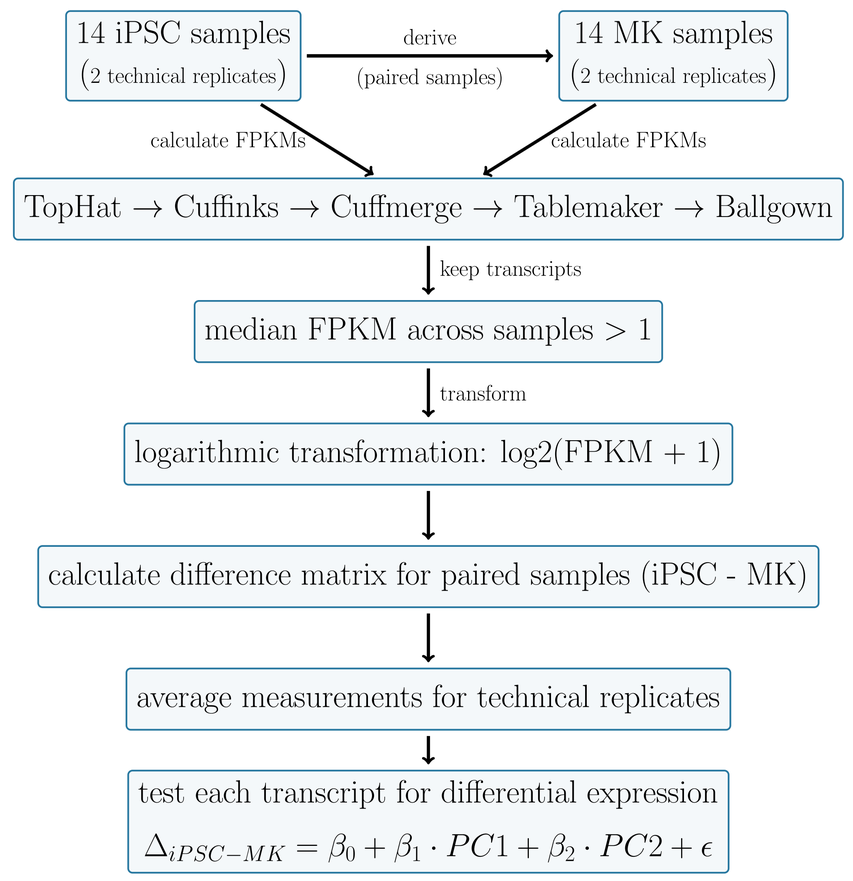
- ➔ **Genotyping IBS summary:** All within sample pairs represent in effect a monozygotic pair (they have $Z0 = 0$, $Z1 = 0$, and $Z2 = 1$). In contrast all between-sample pairs represent in effect unrelateds in a dataset.
- ➔ **Mutation rates** comparing parent cell DNA to iPSC DNA and onward to the differentiated MK DNA (one exemplary sample): Very minimal discordance when looking at all site calls in both members of a pair, well within genotyping error rate. 2-3 mutations appear to pass from iPSC on to MK.
- ➔ **In addition:** We observed no CNVs in the iPSC lines that were not pre-existing in the donor DNA supporting a high degree of structural integrity in our laboratory process of iPSC generation.

Batch / PCA



This work was sponsored by NHLBI projects U01 HL72518-05 and U01 HL107446

RNA-seq Data



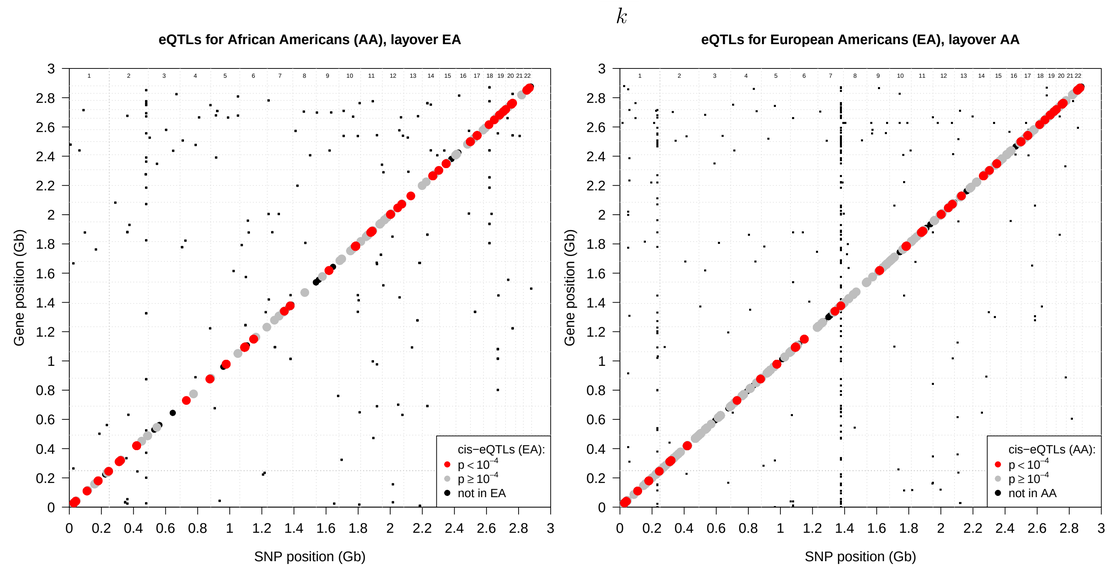
Significant Gene Sets for MKs > iPSCs

- ✓ Platelet activation [GO:0030168]
- ✓ Megakaryocyte development [GO:0035855]
- ✓ Platelet aggregation [GO:0070527]
- ✓ Regulation of cell proliferation [GO:0042127]
- ✓ Platelet formation [GO:0030220]
- ✓ Immune response [GO:0006955]
- ✓ Inflammatory response [GO:0006954]

eQTL Analysis of MKs

For each gene-SNP pair test for association between gene expression G and genotype S (and additional covariates C_k, $k = 1, ..., K$) `[MatrixEQTL in R]`:

$$G = \beta_0 + \beta_1 \cdot S + \sum_k \gamma_k C_k + \epsilon$$



Summary

- ✓ Genotype concordance in the cell lines is excellent and well within genotyping error rates.
- ✓ A few mutations do pass along from one to the next in going to iPSC and then to MK (one sample is a bit higher than the others, but still within error rates).
- ✓ While we observe some evidence for structural variants, this was not a systematic process related to reprogramming of iPSCs or derivation of MKs.
- ✓ RNA-seq data show strong patters of differentiation that strongly support the differentiation of the MKs from the iPSCs.
- ✓ We found 3380 cis-eQTLs in European Americans and 595 cis-eQTLs in the African Americans ($p < 10^{-4}$). There appears to be considerable replication between the two.