Dynamic genetic regulation of gene expression during cellular differentiation

Katherine Rhodes*[1], Reem Elorbany*[2], Benjamin Strober*[3], Nirmal Krishnan[4], Karl Tayeb[4], Alexis Battle[3,4], Yoav Gilad[1,2]

1 Department of Human Genetics, University of Chicago. 2 Department of Medicine, University of Chicago. 3 Department of Biomedical Engineering, Johns Hopkins University. 4 Department of Computer Science, Johns Hopkins University.

Background

- Gene regulation is dynamic, changing through time during differentiation and development.
- Variation in gene regulation during development can contribute to complex traits and disease.

We aim to:

- characterize patterns of temporal variation in gene expression and regulation during differentiation of iPSC-derived cardiomyocytes.
- detect genetic variants associated with the change in gene expression over time, called “dynamic eQTLs”.
- explore the relationship of dynamic eQTLs, including those with transient effects, to human traits and disease.

Time is the primary source of variation in gene expression during cardiomyocyte differentiation

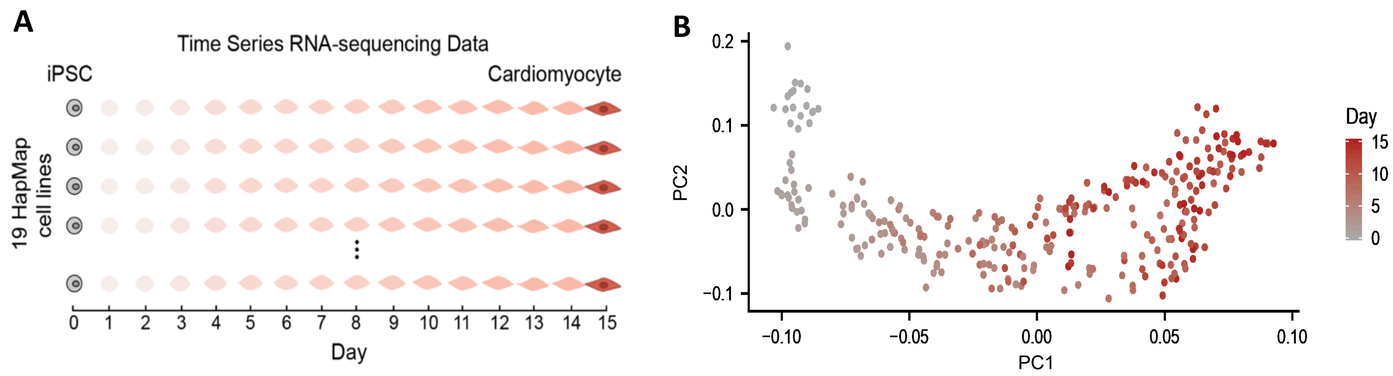

Fig. 1 (A) For 19 individuals, iPSCs were differentiated into cardiomyocytes and bulk RNA-seq data was collected every 24 hours for 16 days of differentiation. **(B)** Principal component analysis shows differentiation day as the primary source of gene expression variation between samples (PC1).

Non-dynamic eQTLs exhibit temporal patterns during differentiation

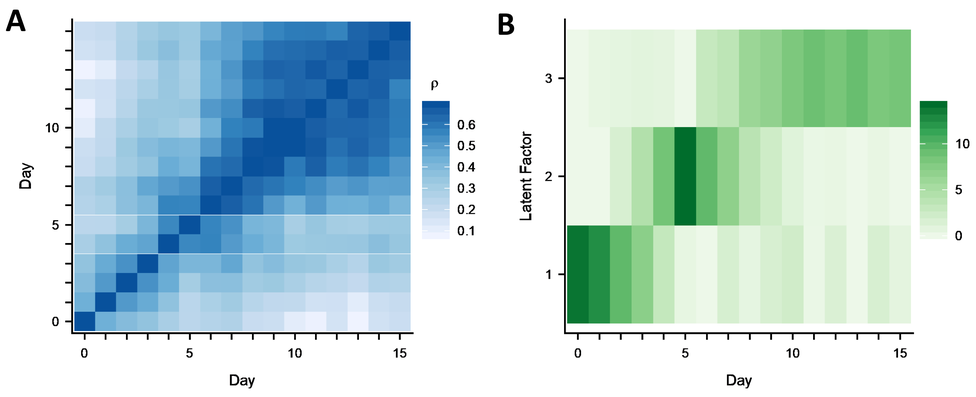

Fig. 2 For eQTLs identified independently at each time point: **(A)** Spearman correlation of eQTL p-values for each pair of days. Correlation between eQTL p-values is higher for closer time points. **(B)** Sparse matrix factorization of eQTL p-values using 3 latent factors. Learned factors capture genetic signal specific to a subset of differentiation time.

Dynamic eQTLs detect genetic regulatory changes caused by cardiomyocyte differentiation

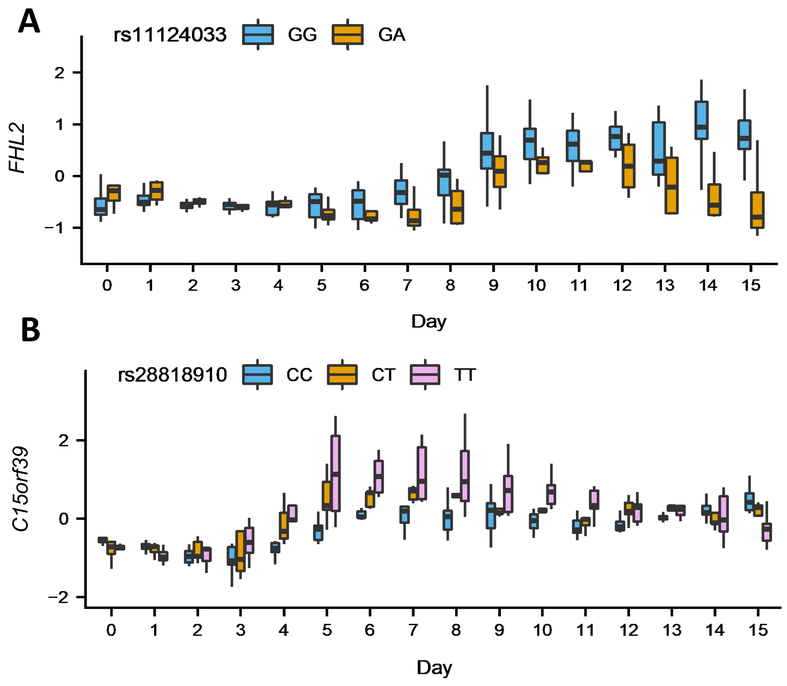

Fig. 3

(A) Example of a linear dynamic eQTL : rs11124033 is associated with expression of *FHL2* gene, which is associated with dilated cardiomyopathy[1]. In total we found 550 linear dynamic eQTLs whose effect varies significantly over time (eFDR <= .05).

(B) Example of a nonlinear dynamic eQTL: rs28818910 is associated with expression of the gene *C15orf39* . We identified 693 genes with a nonlinear dynamic eQTL effect (eFDR <= .05). These nonlinear dynamic eQTLs have transient effects in the middle of the timecourse.

Dynamic eQTLs are relevant to cardiac traits and human disease

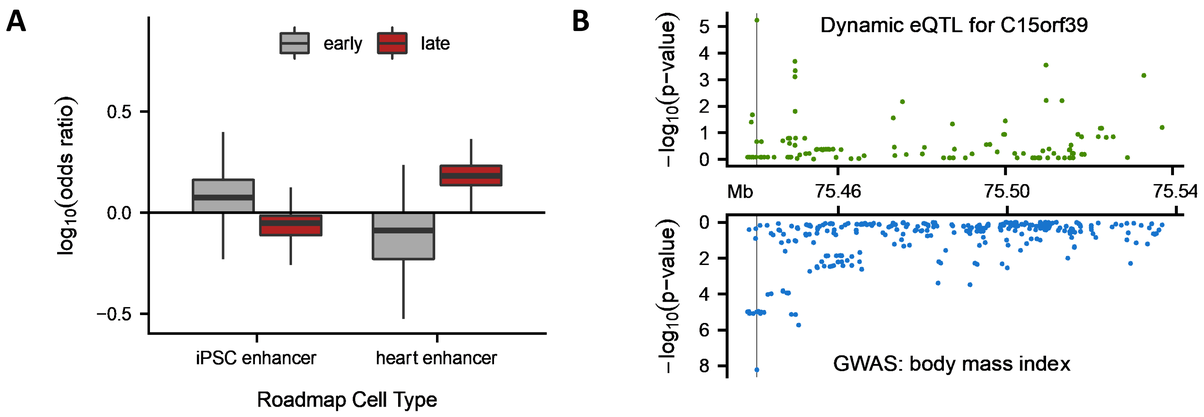

Fig. 4 (A) Among linear dynamic eQTLs, those with greater effect earlier in the timecourse are enriched for chromHMM enhancer elements annotated in iPSCs, while those with an effect later in the timecourse are enriched for enhancer elements in heart-related cell types[2,3]. Enrichment is relative to 1000 sets of randomly selected matched background variants. **(B)** A nonlinear dynamic eQTL affecting *C15orf19* is also associated with body mass index (BMI). In green, we show the association significance of all variants tested within 50 kb of the C15orf39 transcription start site with expression of C15orf39. In blue, we show the GWAS significance for BMI of variants in the same window.[4]. The vertical line marks the most significant nonlinear dynamic eQTL for *C15orf39*.

Conclusions

Our timecourse study design allowed us to identify hundreds of dynamic eQTLs, or eQTLs whose effect changes through time. Dynamic eQTLs may be transient and may not be found in studies using only stem cells or mature tissues. These fleeting genetic associations may represent a new mechanism to explain complex traits and disease, and are candidates to be followed up with further functional validation in relevant intermediate time points

In future studies, we will assess the dynamics of other molecular phenotypes through this developmental timecourse. We will also collect single cell gene expression data during the timecourse, which will allow us to investigate the effects of cell composition changes, in addition to gene regulatory changes during differentiation.

Read the full paper in *Science* →

References:

1. M. A. Burke, S. A. Cook, J. G. Seidman, C. E. Seidman, Clinical and Mechanistic Insights Into the Genetics of Cardiomyopathy. J. Am. Coll. Cardiol. 68, 2871–2886 (2016).
2. J. Ernst, M. Kellis, Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. 12, 2478–2492 (2017).
3. Roadmap Epigenomics Consortium et al., Integrative analysis of 111 reference human epigenomes. Nature. 518, 317–330 (2015).
4. C. Churchhouse, B. Neale, “Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank” (2017); www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank.

Acknowledgements:

KR was funded by the Genetics and Regulation Training Grant (T32 GM09197), the American Heart Association Predoctoral Fellowship (18PRE34030197), and the Ruth L. Kirschstein Predoctoral Individual National Research Service Award (1 F31 HL146171-01). RE was supported by the Medical Scientist Training Program NIH Training Grant (T32 GM007281). YG and AB were supported by NIH/NIGMS R01GM120167. Computational resources were provided by the University of Chicago Research Computing Center.

Contact

Email: klrhodes@uchicago.edu

Twitter: @KatieRho14