Single-cell Analysis of the Genetic Effect on the Cardiac Differentiation of iPSCs

Maika Matsumoto[1], Tianqi Ma[1], Quan Nguyen[1], Nathan Palpant[1], Joseph Powell[2].

1 Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD, Australia
2 Garvan Institute of Medical Research, Garvan-Weizmann Centre for Cellular Genomics, Sydney, NSW, Australia

Introduction

- The genetic loci underlying common diseases are known to act through changing genome regulation activities.
- The crucial missing step is the evidence for the effect of the genetic variation on the dynamic genetic regulation during cell differentiation.
- We performed the single-cell transcriptomic analysis on the >120,000 cells that are derived from 6 iPS cell lines of the distinct genotypes and undergoing cardiac-directed differentiation.
- We have observed the difference in the expression pattern of the genes associated with cardiovascular diseases (CVD) between cell lines.

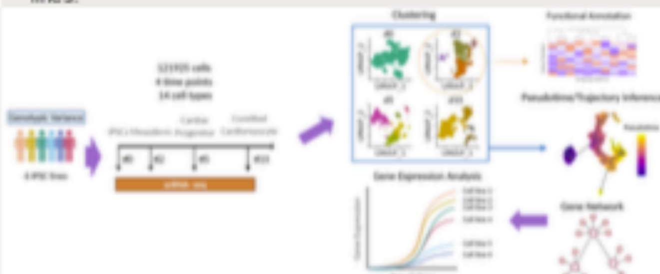

Figure 1: Experimental Design and Analysis Workflow

Method

Data Generation

The cells derived from 6 iPSC lines were captured at four in-vitro differentiation timepoints which correspond to the stage-specific transitions in cell states; pluripotent state (day 0), germ layer speciation (day 2), progression to progenitor(day 5) and committed (day 15) cardiac cell states (Figure 1).

Data Integration and Clustering

The batch effect between different libraries and cell lines were removed by Canonical Correlation Analysis method implemented in Seurat v3(Stuart et al., 2018). The cells were clustered into sub-populations using Louvain algorithm (Blondel et al., 2008).

Functional Annotation of Cell Subpopulations

The identities of cell subpopulations were determined by automatic annotation methods and manual annotation. We used Seurat-label transfer method and SingleR (Aran et al., 2019) for automatic annotation. Manual annotation was based on the 98 known marker genes.

Lineage Tracing of Cardiac lineage

Trajectory inference analysis was performed using Monocle 3 (Cao et al., 2019) and Slingshot (Street et al., 2018) on the cells from all four time-points. Gene regulatory network (GRN) analysis was performed on the lineages using SCENIC (Van de Sande et al., 2020).

Analysis of Genotypic Effect on the Gene Regulation

The gene modules which were positively and negatively correlated with differentiation time was determined by the weighted correlation network analysis (Langfelder and Horvath, 2008). The genes were filtered based on GWAS catalogue (Buniello et al., 2019) to select genes associated with CVD.

Results

Identification of subpopulations at four differentiation time points

- We have identified 14 subpopulations across the four differentiation time points (Fig 2a).
- The expression profiles of subpopulations were similar on day 2 and became more distinct on day 15 (Fig 2b).
- There were small subpopulations on day 2 and day 5, in which all marker genes were lowly expressed.

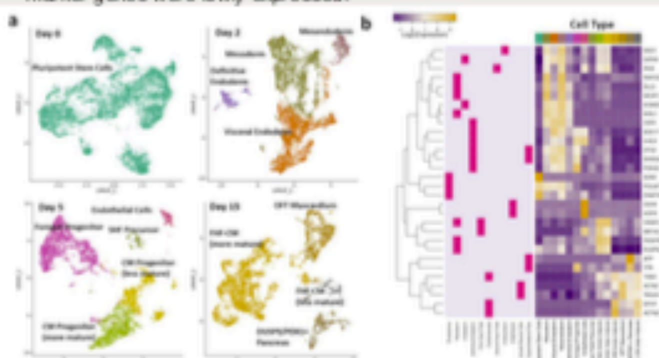

Figure 2: Functional Annotation of Sub-populations. a, UMAP plots for cells from four time points coloured by the cell type annotation. **b,** Heatmap comparing the gene expression of 30 known gene markers of the cell types across 14 subpopulations from the four timepoints.

Cardiac lineage tracing and gene regulation analysis

- We have identified 106 lineages by Monocle 3 (Fig 3b) and 8 lineages by Slingshot (Fig 3d).

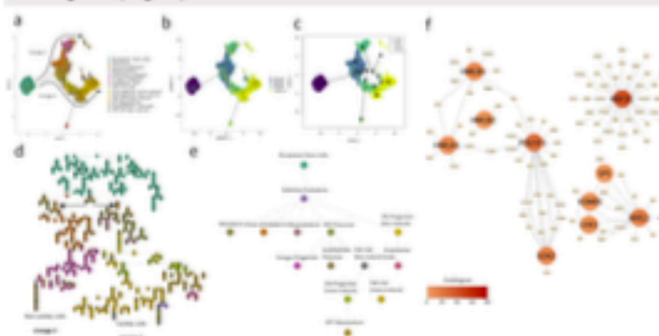

Figure 3: Comparison of Lineages inferred by Monocle3 and Slingshot. a, UMAP plot of cells from all four timepoints coloured by the cell type. The black arrows indicating the developmental direction of lineage 1 and 2. **b and c,** The same UMAP plots as (a) coloured by differentiation time points. The black line indicates the edges of trajectories defined by Monocle3 (b), and the black line and dots indicate the edges and nodes of trajectories defined by Slingshot (c). **d,** 106 lineages defined by Monocle 3. Each pie represents the node with the cell type composition. **e,** 8 lineages defined by Slingshot. The node represents the subpopulation. **f,** GRN of cardiac lineage. The each central node represents the regulating transcription factor. The epi-nodes represent co-expressed genes. The nodes are coloured by outdegree measure.

- The lineages determined by Monocle 3 shows the cells are branched into two major developmental pathways on day 2. Lineage 1 had developed into the non-cardiac cells, foregut progenitor and pancreatic cells (Fig 3a). Lineage 2 had developed into cardiac cells; cardiomyocytes and outflow tract myocardium (Figure 3d).
- GRN on cardiac lineage identified 10 transcription factors that are regulating cardiac development (Fig 3f).

The genetic effect on the gene expression pattern of CVD-genes

- We have identified two gene modules with > 400 genes, that were highly positively and negatively correlated (|r| > 0.8).
- Among these genes, 4 genes that are associated CVDs are matched with the genetic variants in the genotypes of 6 cell lines.
- The gene expression patterns of these CVD associated genes along the pseudotime were different between cell lines (Fig 4).

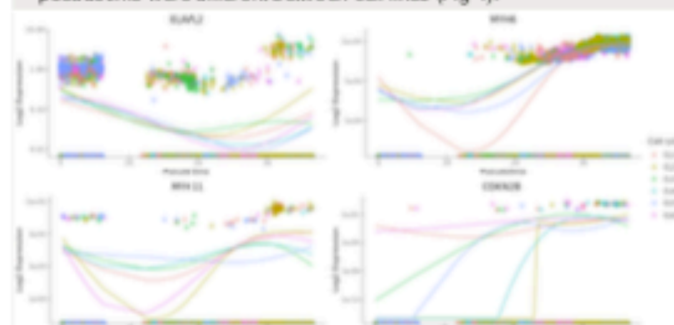

Figure 4. Comparison of gene expression pattern in the cardiac lineage across 6 cell lines. The gene expression of the four CVD-associated along pseudotime. Each dot represents each cell coloured by the cell line. Each line indicate the expression profile of the gene for each cell line.

Conclusions

Our work shows the potential effect of genetic variance on the cardiac differentiation by changing the expression pattern of the regulatory genes. The future direction of this study includes comparing the differentiation efficiencies of iPSCs across different cell lines, and associating the genotypes to the expression levels and patterns of the CVD genes.

Reference

ARAN, D., LOONEY, A. P., LIU, L., WU, E., FONG, V., HSU, A., CHAK, S., NAIKAWADI, R. P., WOLTERS, P. J., ABATE, A. R., BUTTE, A. J. & BHATTACHARYA, M. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature immunology*, 20, 163-172.

BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. & LEFEBVRE, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, P10008.

BUNIELLO, A., MACARTHUR, J. A. L., CEREZO, M., HARRIS, L. W., HAYHURST, J., MALANGONE, C., MCMAHON, A., MORALES, J., MOUNTJOY, E., SOLLIS, E., SUVEGES, D., VROUSGOU, O., WHETZEL, P. L., AMODE, R., GUILLEN, J. A., RIAT, H. S., TREVANION, S. J., HALL, P., JUNKINS, H., FLICEK, P., BURDETT, T., HINDORFF, L. A., CUNNINGHAM, F. & PARKINSON, H. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47, D1005-D1012.

CAO, J., SPIELMANN, M., QIU, X., HUANG, X., IBRAHIM, D. M., HILL, A. J., ZHANG, F., MUNDLOS, S., CHRISTIANSEN, L., STEEMERS, F. J., TRAPNELL, C. & SHENDURE, J. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566, 496-502.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.

STREET, K., RISSO, D., FLETCHER, R. B., DAS, D., NGAI, J., YOSEF, N., PURDOM, E. & DUDOIT, S. 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19, 477.

STUART, T., BUTLER, A., HOFFMAN, P., HAFEMEISTER, C., PAPALEXI, E., MAUCK, W. M., STOECKIUS, M., SMIBERT, P. & SATIJA, R. 2018. Comprehensive integration of single cell data. *bioRxiv*, 460147.

VAN DE SANDE, B., FLERIN, C., DAVIE, K., DE WAEGENEER, M., HULSELMANS, G., AIBAR, S., SEURINCK, R., SAELENS, W., CANNOODT, R., ROUCHON, Q., VERBEIREN, T., DE MAEYER, D., REUMERS, J., SAEYS, Y. & AERTS, S. 2020. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols*, 15, 2247-2276.

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

CREATE CHANGE