High-throughput gene expression analysis of *in vitro* and *in vivo* mammalian cardiogenesis identifies shared developmental gene expression signatures

Andrew Williamson[#], Joseph Dalloul*, Dwayne Tally[#], AJ Farmer*, Laura Cochran*, Hayden Fell*, Garett Oxford*, Rusty Gonser*, Shaad Ahmad*, Jeff Kinne[#], and Kristopher Schwab*

Department of Biology*, Department of Mathematics and Computer Science[#], and The Center for Genomic Advocacy, Indiana State University, Terre Haute, IN 47809

Indiana State University

THE CENTER FOR GENOMIC ADVOCACY AT INDIANA STATE UNIVERSITY

Introduction

Mammalian heart development is regulated by an evolutionarily conserved genetic network that has been defined from numerous studies of model organisms and the genetic investigation of human congenital heart defects. Significant advancements in pluripotent stem cell technology have established new *in vitro* experimental systems allowing for the investigation of complex developmental mechanisms within a single tissue culture dish. An impressive procedure has been developed that utilizes modulation of the Wnt signaling pathways using small molecules to robustly and efficiently differentiate human pluripotent stem cells (hPSCs) into cardiomyocytes in a manner that recapitulates early embryonic heart development (Figure 1). Although this procedure allows for the rapid experimental study of cardiomyocyte differentiation in a tissue culture dish, a gene expression comparison of *in vitro* and *in vivo* cardiomyocyte differentiation has yet to be performed.

To gain a deeper understanding of the differentiating cardiomyocyte gene expression profile and aid our future studies of important cardiac regulatory genes, we have identified shared differentially expressed genes between *in vivo* embryonic mouse heart development and *in vitro* cardiomyocyte differentiation of human pluripotent stem cells from two previously published data sets. Li et al. (2014) completed expression microarray analysis at several important developmental time points of the embryonic mouse heart mouse to the adult heart and included important developmental samples including embryonic stem cells (mESCs) and the embryonic day (E) 7 embryo. Busser et al. (2015) performed RNA seq expression analysis of *in vitro* differentiation of human embryonic stem cell (hESCs) into cardiomyocytes using a similar protocol as described in Figure 1.

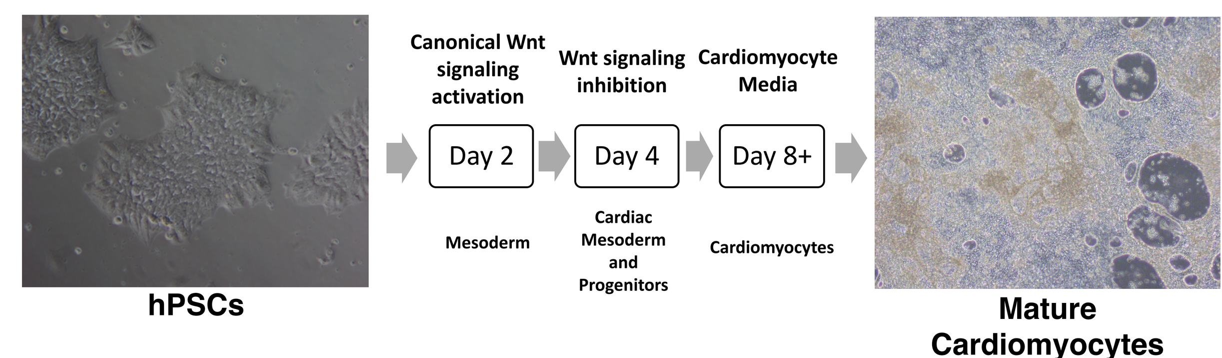

Figure 1. *In vitro* cardiomyocyte differentiation of hPSCs (human pluripotent stem cells) via small molecule modulation of Wnt signaling. Experimental overview of cardiomyocyte differentiation of hPSCs (Lian et al., 2012). Left image: hPSC colonies are cultured under conditions that maintain pluripotency and proliferation. hPSC media is withdrawn and replaced with a media that induces mesodermal differentiation via canonical Wnt signaling activation. Cardiac progenitor specification and proliferation is then induced by inhibiting Wnt signaling molecule secretion at day 4. Proliferating cardiomyocyte cultures are maintained in cardiomyocyte media for the duration of the experiment. Right image: Beating cardiomyocyte syncytia can be observed after eight days of differentiation.

Methods and Discussion

Our data analysis was performed in R using Bioconductor packages. Raw data were downloaded from the Gene Expression Omnibus (GSE51483 [Li et al.] and GSE69618 [Busser et al.]) using the GEOquery package. Each dataset contains over 40,000 rows with each row corresponding to a different gene transcript measured via mRNA sequencing; columns represent different samples taken at different time points. Metadata for the genes were downloaded using the R packages biomaRt, GO.db, and org.Hs.eg.db.

Raw counts for each gene transcript were normalized and converted to log2-counts per million expression values using the edgeR package functions DGEList, calcNormFactors (with method trimmed mean of M-values), and voom.

Differential expression calculations were performed using the limma package functions lmFit, eBayes, and topTable. For analysis, the data from GSE51483 mouse development were grouped based on age - ESC, day 7 whole embryo, heart tissue from days 8 / 9 / 12, adult heart tissue - with each age containing three replicates (columns) in the data. For GSE69618 hiESC induced cardiomyocyte development, only wild type samples were used for analysis, two replicates at day 10 (R1 and R2) were removed due to irregularities with those replicates mentioned by the authors, and analysis was performed by grouping based on age (day 0, 2, 6, 10) with two replicates per age.

The following statistical tests were performed. For GSE51483 data, comparison of day 8/9/12 samples versus ESC and adult heart samples. For GSE69618 data, comparison of day 2 versus day 0, day 6 versus day 2, day 10 versus day 6. For each comparison, gene transcripts were selected as having statistically significant differential expression using the criteria - fold change of at least 2.0 and a pValue of at most 0.01. This resulted in 4304 significant gene transcripts from GSE51483 and 2213 significant gene transcripts from GSE69618.

Gene transcripts that were selected as significant were further grouped by clustering with the kmeans R function and visualized using the heatmap.2 function from the R package gplots. Gene Ontology (GO) and GO Enrichment Analysis was performed on each cluster using the tools available at geneontology.org.

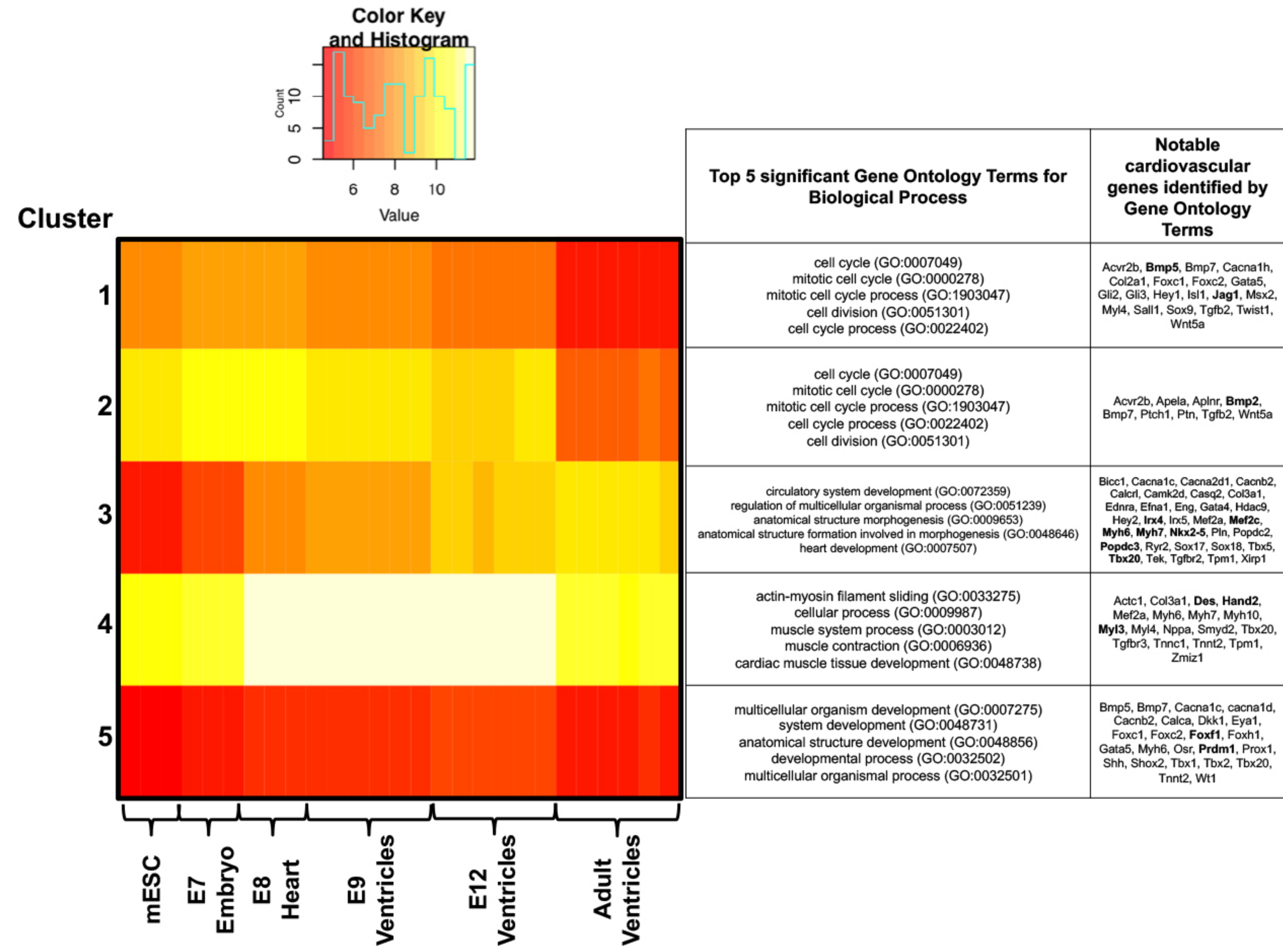

Figure 2. Cardiac gene expression profiling of *in vivo* early embryonic mouse heart development. 4,003 differentially expressed cardiac genes were identified by individually comparing the heart developmental time point (E8-E12) to either the mESC or Adult Ventricle samples using the following criteria: P value > 0.01 and Fold Change > 2.0. The upregulated cardiac genes from all comparisons were then combined and clustered using k-means identifying unique developmental gene expression signatures. Note the following interesting developmental profiles: Interestingly, clusters 1 and 2 identify genes upregulated the early E8 heart which steadily reduce their expression during early heart development. Clusters 3 and 4 identify genes upregulated within the developing heart sample (E8-E12).

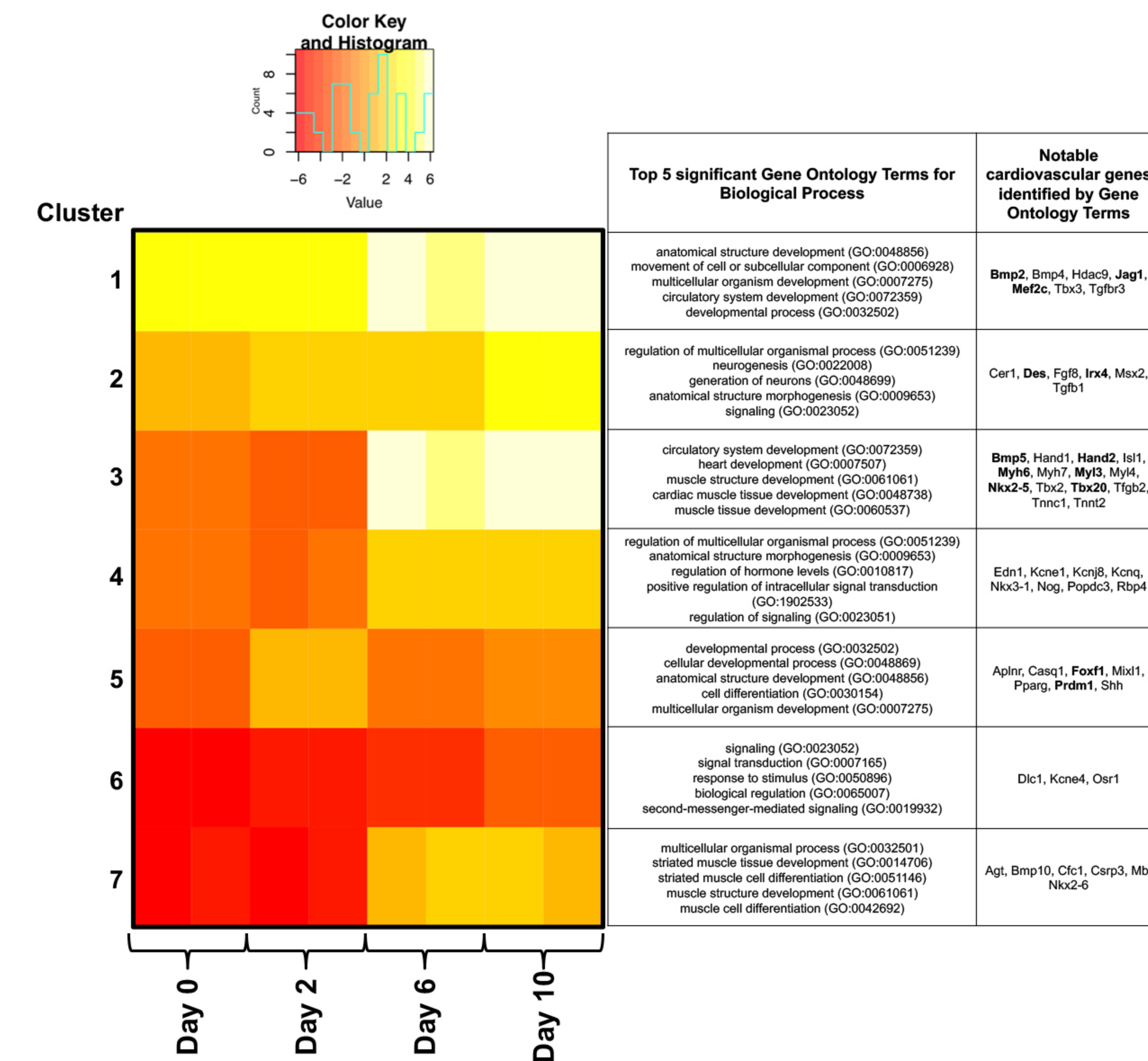

Figure 3. Cardiac gene expression profiling of *in vitro* cardiomyocyte differentiation of hESCs. 2212 differentially expressed cardiomyocyte differentiation genes were identified by performing sequential comparisons of Day 0 (pluripotent stem cells), Day 2 (mesoderm-specified cells), 6 (cardiac progenitor-specified cells), and 10 (functional cardiomyocytes) using the following criteria: P value > 0.01 and Fold Change > 2.0. The upregulated cardiac genes from all comparisons were then combined and clustered using k-means identifying unique cardiomyocyte differentiation gene expression signatures. Note the following interesting cardiac signatures: Clusters 1, 3, 6, and 7 demonstrate significant gene activation at Day 6 of differentiation. Interestingly, cluster 5 identifies genes briefly upregulated at Day 2 suggesting these genes represent early mesodermal and cardiac specification genes. Also, cluster 2 identifies genes that consistently increase during cardiomyocyte differentiation.

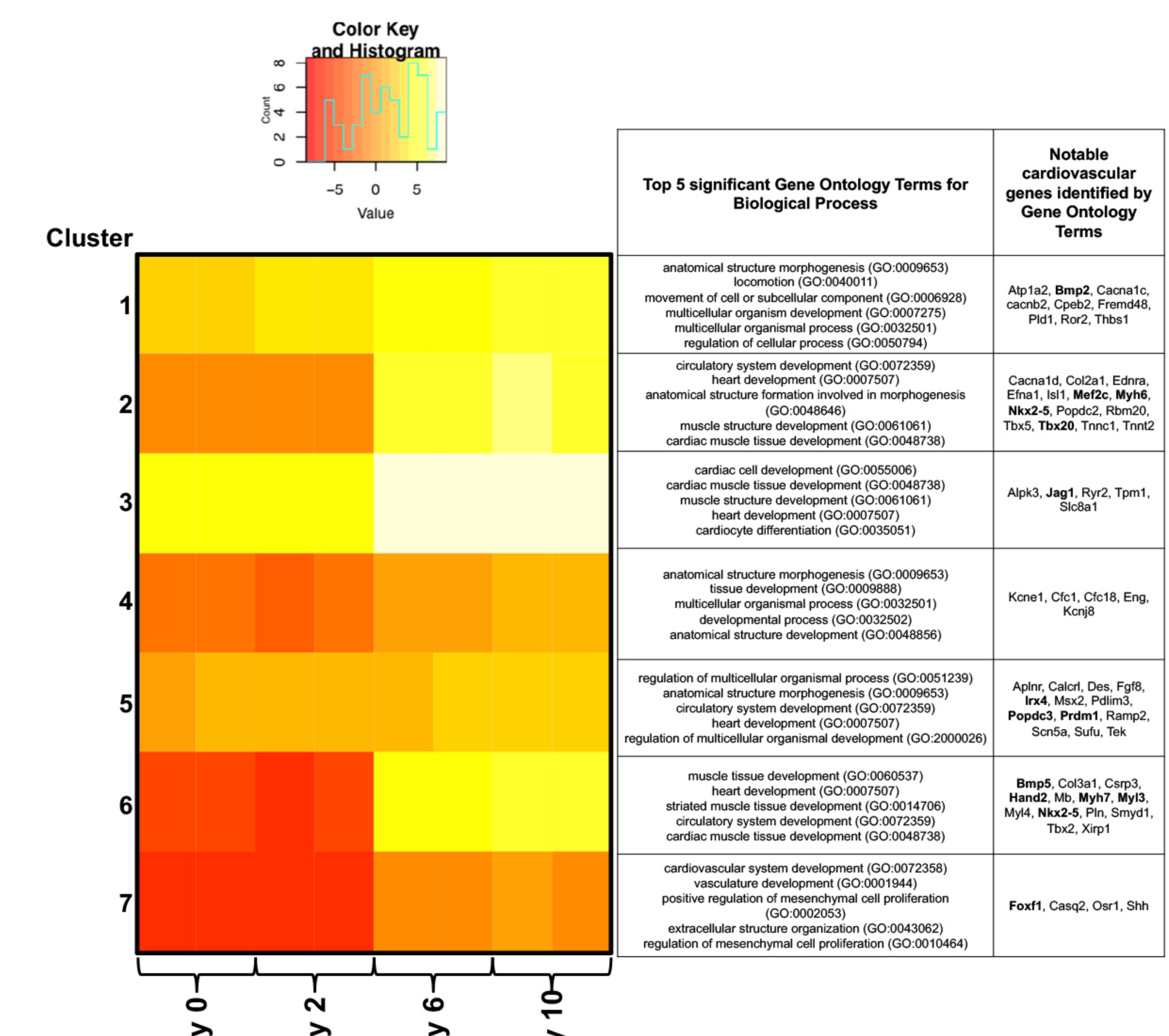

Figure 4. Expression profile comparison of *in vivo* early embryonic mouse heart development and *in vitro* cardiomyocyte differentiation of hESCs identifies common cardiomyocyte differentiation genes. 1152 differentially expressed genes commonly shared between the early embryonic mouse heart (Figure 2) and *in vitro* cardiomyocyte differentiation (Figure 3) were clustered using k-means generating unique expression signatures using the *in vitro* cardiomyocyte differentiation data set. Note the following interesting developmental profiles: The majority of clusters (clusters 2, 3, 6, and 7) identify cardiomyocyte genes strongly activated at Day 6 when cardiac progenitor cells are undergoing specification and proliferation compared to the early timepoints. Interestingly, clusters 1 and 5 identify probable early cardiac genes that are activated on Day 2 which continue to increase in expression as differentiation proceeds.

Results and Discussion

The analysis of the *in vivo* early embryonic heart development data comparing each early developmental heart stage to the mESC or Adult stage identified approximately 4,000 differentially expressed genes that were then clustered based on expression profile using k-means (Figure 2). Each cluster identifies a unique cardiac developmental expression pattern. Gene Ontology enrichment analysis of each cluster identifies several notable cardiovascular genes and reveals significant biological process functions involving cell proliferation, heart development, and developmental processes.

The analysis of the *in vitro* cardiomyocyte differentiation of hESCs data set comparing each time point of differentiation (Day 2, 6, 10) identified over 2000 genes differentially expressed which were similarly clustered accordingly (Figure 3). This analyses reveals that a majority of the clusters were upregulated at Day 6 of differentiation when cardiac and cardiomyocyte progenitors are specified, proliferating, and maturing into cardiomyocyte. Interestingly, cluster 5 identifies several genes that are upregulated specifically at Day 2 representing an early time point of differentiation which involves the specification of the mesoderm and cardiac mesoderm.

Next, genes shared between both the *in vivo* mouse embryonic heart (4003 genes) and *in vitro* cardiomyocyte differentiation (2212 genes) gene lists were selected producing a list of over 1100 genes which were clustered based upon the *in vitro* cardiomyocyte differentiation data set (Figure 4). Again, the majority of the clusters identified expression patterns that are upregulated at Day 6. However, clusters 1 and 5 identify genes that are slightly upregulated at Day 2 during mesoderm and cardiac mesoderm specification which may represent early cardiac regulatory genes. Additionally, this analysis identifies severally known cardiac regulatory genes such as Isl1, Mef2c, Nkx2-5, Tbx2, Tbx20.

This analysis provides a deeper understanding of the gene expression profile of the differentiating cardiomyocyte. This data set has identified both the unique expression profiles of known cardiac regulatory genes as well as genes that remain uncharacterized in cardiac development and function. Furthermore, the shared gene expression profiles identified in this work allow for the creation of a "cardiac reference gene set" for the evaluation of gene function in our future genetic function studies using our *in vitro* cardiomyocyte differentiation system.

References

B. W. Busser et al., An Orthologous Epigenetic Gene Expression Signature Derived from Differentiating Embryonic Stem Cells Identifies Regulators of Cardiogenesis. *PLoS One* 10, e0141066 (2015).

X. Li et al., Transcriptional atlas of cardiogenesis maps congenital heart disease interactome. *Physiol Genomics* 46, 482-495 (2014).

X. Lian et al., Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci U S A* 109, E1848-1857 (2012).

Acknowledgments

This work was supported by: The Center for Genomic Advocacy (Ahmad, Schwab), University Research Council (Schwab), AHA (Ahmad), ISU Summer Undergraduate Research Experience (SURE), and NIH #5R25MD011712 (Gonser, Kinne).