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

Mirian Alvarez-Dubon[1*], Ishmeet Kaur[1*], Dwayne Tally[3], Joseph Dalloul[1], Andrew Williamson[3], Rusty Gonser[1,2], Jeff Kinne[2,3], and Kristopher Schwab[1,2]

[1]Department of Biology, Indiana State University, Terre Haute, IN [2]The Center for Genomic Advocacy, Indiana State University, Terre Haute, IN
[3]Department of Mathematics and Computer Science, Indiana State University, Terre Haute, IN *These authors contributed equally to the work

Background

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) 7.5 embryo (Figure 1) . Liu et al. (2017) performed RNA seq expression analysis of *in vitro* differentiation of human embryonic stem cell (hESCs) into cardiomyocytes with data taken at time points given in Figure 2.

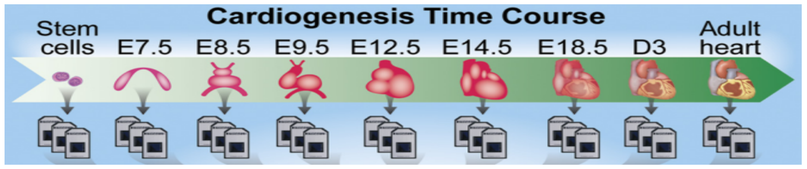


Figure 1. Embryonic Mouse Heart Atlas. The Li et al. gene expression microarray data set investigates several important development stages from the early developing embryo (mESC, E7.5 embryo), early developing embryonic heart, and adult heart. Adapted from X, Li et al. (2014).

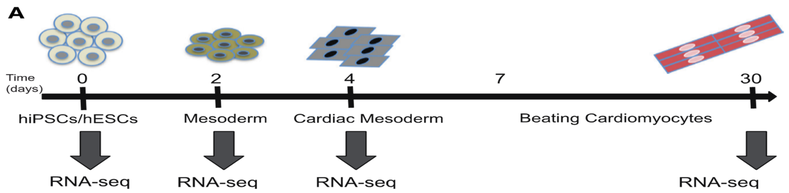


Figure 2. *In vitro* cardiomyocyte differentiation using human pluripotent stem cells (hiPSCs). The Liu et al. RNA seq data set investigates important time points in the cardiomyocyte differentiation of human induced and embryonic stem cells (hiPSCs and ESCs). Samples were taken at days 0, 2, 4, and 30, with expression measured using RNA seq data from each sample.

Methods

Research Aim A survey of the literature (Spater et al. 2014) suggests a list of genes to look for at key time points in embryonic heart development in vivo and during cardiomyocyte differentiation in vitro. Figure 3 lists key stages in heart development and cardiomyocyte differentiation along with genes that are expected to be up-regulated at each stage. One of our main research aims is to determine whether these genes are indeed upregulated in the Li et al. and Liu et al. datasets.

Data Preparation Microarray expression data from Li et al. was downloaded from the Gene Expression Omnibus (GEO id GSE51483) for analysis. For the Liu et al. data, raw Fastq RNA seq files were downloaded from the Sequence Read Archive (SRA id SRP081103). Raw Fastq data files were trimmed for quality and adapter trimming using Cutadapt version 2.3, checked for quality using FastQC version v0.11.8, aligned to human reference genome GRCh38.p12 using HISAT2 version 2.1.0, files converted using SAMtools version 1.9, and counts per gene computed using htseq-count from HTSeq version 0.11.2.

Differential Expression Microarray expression values from the Li et al. dataset and gene counts from the Liu et al. dataset were imported into the CarDGEA (Cardiac Development Gene Expression Analysis) R programming app that has been developed at ISU. CarDGEA was used to convert the counts from the Liu et al. dataset into log2 expression values and to perform statistical differential expression analysis for all genes in both datasets.

Data Set Validation

Our in-house differential gene expression analysis tool, CarDGEA, was utilized to complete the analysis of several time points. Our first task was to investigate the robustness of the gene expression microarray and RNA-seq data sets that describe both in vivo and in vitro differentiation. Several differential gene expression contrasts (P < 0.01, fold-change > 2.0) were completed that investigated unique developmental expression utilizing several "developmental marker" genes that characterize: pluripotency, mesoderm, cardiac progenitors, and differentiated cardiomyocytes (Figure 3). Our analysis of these developmental markers verifies the integrity of the data sets: the majority of developmental gene validators are upregulated and demonstrate their highest ("Peak") expression at the appropriate time points in the embryonic mouse and human differentiation data sets.

Pluripotency and Mesoderm Markers

		Mouse Embryonic Heart Data Set			Human Cardiomyocyte Differentiation Set		
Stages	Gene Symbol	Significantly Upregulated?	First Timepoint Upregulated	Peak Upregulation	Significantly Upregulated?	First Timepoint Upregulated	Peak Expression
Pluripotent	NANOG	Yes	mESC	mESC	Yes	Day 0	Day 0
	POU5F1	Yes	mESC	mESC	Yes	Day 0	Day 0
	SOX4	No	-	-	No	-	-
Mesoderm	MESP1	Yes	E7.5 Embryo	E7.5 embryo	Yes	Day 2	Day 4
	MIXL1	Yes	mESC	mESC	Yes	Day 2	Day 2
	TBXT	Yes	mESC	mESC	Yes	Day 2	Day 2

Cardiac Progenitor and Cardiomyocyte Developmental Markers

		Mouse Embryonic Heart Data Set			Human Cardiomyocyte Differentiation Data Set		
Stages	Gene Symbol	Significantly Upregulated?	First Timepoint Upregulated	Peak Expression	Significantly Upregulated?	First Timepoint Upregulated	Peak Expression
Cardiac Progenitor / Cardiac Mesoderm	HCN4	No	-	-	Yes	Day 30	Day 30
	ISL1	Yes	E7.5 Embryo	E8.5 Heart Tube	Yes	Day 4	Day 4
	KDR	Yes	E7.5 Embryo	Adult Heart	Yes	Day 4	Day 4
	NKX2-5	Yes	E8.5 Heart tube	E8.5 Heart Tube	Yes	Day 4	Day 30
	PDGFRA	Yes	E7.5 Embryo	E12.5 Heart	Yes	Day 2	Day 4
	TBX5	Yes	E8.5 Heart tube	E8.5 Heart Tube	Yes	Day 4	Day 30
	ISL1	Yes	E7.5 Embryo	E8.5 Heart Tube	Yes	Day 4	Day 4
Cardiomyocyte	MYH6	Yes	E8.5 Heart tube	Adult Heart	Yes	Day 4	Day 30
	MYH7	Yes	mESC	E12.5 Heart	Yes	Day 4	Day 30
	NKX2-5	Yes	E8.5 Heart tube	E8.5 Heart Tube	Yes	Day 30	Day 30
	NPPA	Yes	E8.5 Heart tube	E12.5 Heart	No	-	-
	RYR2	Yes	E8.5 Heart tube	Adult Heart	Yes	Day 2	Day 30
	SHOX2	Yes	E8.5 Heart tube	E12.5 Heart	Yes	Day 2	Day 2
	TNNT2	Yes	E7.5 Embryo	Adult Heart	No	-	-

Figure 3. Gene expression analysis of developmental markers. CarDGEA was used to validate the expression of developmental markers at their appropriate stage or time point in both the Li et al. mouse embryonic heart data set and the Liu et al. cardiomyocyte differentiation using hPSC data set. Multiple contrasts were completed between time points identifying significantly upregulated genes that met the following criteria: P value < 0.01 and fold change > 2.0. Genes meeting or exceeding this criteria are denoted "Yes." Next, the developmental stage or differentiation time point that first meets the criteria and exceeded the criteria is identified. Note that many of the developmental markers correctly correspond between the *in vivo* and *in vitro* data (e.g. mesodermal genes are highest in the mESC or E7.5 mouse embryo and are upregulated at Day 2 in differentiating hPSCs).

References

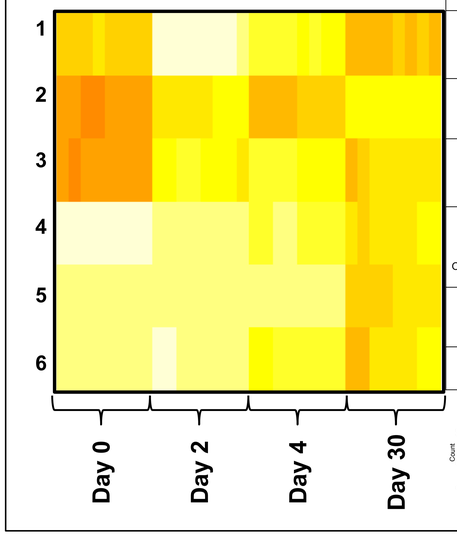
Liu Q, Jiang C, Xu J, Zhao MT et al. 2017. Genome-Wide Temporal Profiling of Transcriptome and Open Chromatin of Early Cardiomyocyte Differentiation Derived From hiPSCs and hESCs. *Circ Res* 121:376-391.

Spater et al. 2014. How to make a cardiomyocyte. *Development* 141: 4418-4431.

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

Clustering Results

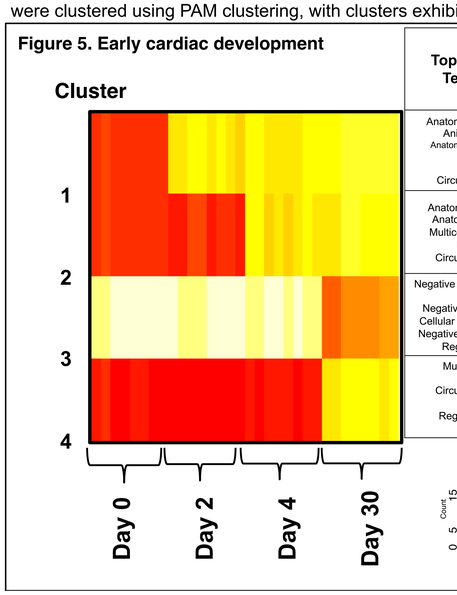
Figure 4. Early embryonic development



Top 5 significant Gene Ontology Terms for Biological Process	Notable cardiovascular genes identified by Gene Ontology Terms
Embryo development GO:0009790 Regionalization GO:0003002 Pattern specification process GO:0007389 Animal organ development GO:0048513 Anterior/posterior pattern specification GO:0009952	Tbx6, WntA
None	None
Epithelium development GO:0060429 Tissue morphogenesis GO:0048729 Tissue development GO:0009888 Morphogenesis of an epithelium GO:0002009 Epithelial tube morphogenesis GO:0060562	CXCR4, TBX3, **MESP1**
Mitotic cell cycle process GO:1903047 Cell cycle G1/S phase transition GO:0044843 Cellular aromatic compound metabolic process GO:0006725 Organic cyclic compound metabolic process GO:1901360 G1/S transition of mitotic cell cycle GO:0000082	CAD, OLFM1, TP53
None	FZD7, FGFR1, CACNG8
Multicellular organism development GO:0007275 Developmental process GO:0032502 Anatomical structure development GO:0048856	STIL, LEFTY1, FGF19

Figure 4. Early embryonic development gene expression signature shared between early stages of early mouse embryo development and *in vitro* cardiomyocyte differentiation. A total of 180 genes were found to be both (a) upregulated in the E7.5 mouse embryo compared to mESCs and heart samples (b) upregulated within cardiomyocyte-differentiation day 2 compared to other time points and (c) not significantly upregulated in later mouse embryo heart development and cardiomyocyte-differentiation (with all comparisons taken with P value < 0.01 and log-fold-change > 2.0). These possible early embryonic development genes were clustered using PAM clustering, with clusters exhibiting interesting patterns worthy of further study.

Figure 5. Early cardiac development



Top 5 significant Gene Ontology Terms for Biological Process	Notable cardiovascular genes identified by Gene Ontology Terms
Anatomical structure morphogenesis GO:0009653 Animal organ Morphogenesis GO:0009887 Anatomical structure formation involved in morphogenesis GO:0048646 System development GO:0048731 Circulatory system development GO:0072359	ADAMTS9, ALDH1A2, ALPK2, ALPK3, BMP2, BMP5, COL11A1, DKK1, FOXC1, FOXF1, GATA4, GATA6, GF1, GREB1L, KCNA5, KCNJ5, KCNJ8, MEIS2, **PDGFRA**, PROX1, SEMA3C, SGCG, SOX17, TENM4, WNT5A
Anatomical structure morphogenesis GO:0009653 Anatomical structure development GO:0048856 Multicellular organism development GO:0007275 Developmental process GO:0032502 Circulatory system development GO:0072359	ADORA1, CAMK2A, CPE, FGFR2, FOXC2, GATA3, GATA5, HAND2, HEY1, IL6ST, **ISL1**, MESI1, MYL4, MYLK3, MYOCD, PDE4B, PDGFRB, RARB, RGS4, SCN4B, SOX5, TBX20, TGFBR3, TWIST1, ZFPM2
Negative regulation of macromolecule metabolic process GO:0010605 Negative regulation of cellular process GO:0048523 Cellular amino acid biosynthetic process GO:0008652 Negative regualtion of metabolic process GO:0009892 Regulation of gene expression GO:0010468	FGF8, FOXH1, MDK, TIAM1, TNNT1
Muscle structure development GO:0061061 Muscle system process GO:0003012 Circulatory system development GO:0072359 Muscle contraction GO:0006936 Regulation of multicellular organismal process GO:0051239	ABCC9, ACTN2, AKAP6, ANK3, ANKRD1, ATP1B1, ATP2A2, BMP7, CACNA1C, CACNA1D, CACNB2, CASQ1, CASQ2, CAV1, CORIN, CSRP3, CTGF, CTNNA3, DHRS3, EDN1, ENG, FGF9, FHL2, FHOD3, HRC, JPH2, KCND3, KCNE1, KCNE1B, KCNH2, KCNK2, KCNQ1, MEF2A, MEF2C, MLIP, MYBPC3, **MYH6**, **MYH7**, MYL1, MYL2, MYL3, MYOM1, **NKX2-5**, NRAP, PDLIM5, PKP2, RNF207, **RYR2**, S1PR1, SCN5A, **SHOX2**, SLC8A1, SMARCD3, SNTA1, SORBS2, TBX5, TCAP, TGFB2, TNNC1, TNNI1, TNNI3, TNNI3, TNNT3, TPM1, TRDN, TRIM63, VCAM1, WNT2, WT1

Figure 5. Genes selected which were upregulated in both (a) early mouse heart, (b) day 4 or day 30 cardiomyocyte-differentiation, resulting in 951 genes. Clustering performed using PAM clustering.

Acknowledgments

This work was supported by: The Center for Genomic Advocacy (Kinne, Schwab), University Research Council (Kinne, Schwab), Indiana Academy of Sciences (Schwab), ISU Summer Undergraduate Research Experience (SURE) and NIH #5R25MDO11712 (Gonser, Kinne).