Differential Gene Expression Analysis of ZNF503 Loss-of-Function in Human Cardiogenesis

Andrew Williamson, Rusty A Gonser, Jeffrey J Kinne, and Kristopher R Schwab

THE CENTER FOR GENOMIC ADVOCACY AT INDIANA STATE UNIVERSITY

Introduction

Heart development is a complex process that requires the step-wise activation of a complex genetic regulatory network. Perturbation of the cardiac regulatory network can result in severe developmental malformations of the heart. Previous published high-throughput gene expression data sets of normal and abnormal mammalian heart development provide the research community with the opportunity to re-investigate important biological processes and develop new hypotheses and experiments.

Znf503, the mammalian ortholog of the *Drosophila noc* gene, encodes a transcription factor that is a putative cardiac regulatory gene and is a gene of interest that is undergoing genetic loss-of-function experiments in both Drosophila and mammal experimental systems (Busser et al. 2015). By using two previously published data sets, we developed a stringent selection criteria to identify downstream cardiac genes that are activated by Znf503 during cardiomyocyte differentiation.

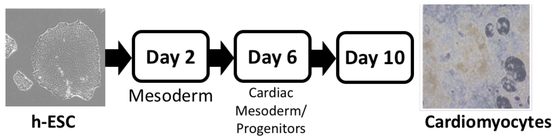

Figure 1: Cardiomyocyte differentiation of hESC using a directed differentiation protocol. RNA-seq expression profiling was performed at Day 0, 2, 6, and 10 corresponding to the mesoderm, cardiac progenitor, and cardiomyocyte stage of differentiation (Busser et al. 2015).

Method and Discussion

Busser and colleagues (2015) utilized a directed differentiation protocol to differentiate hESCs into cardiomyocytes which generates cardiomyocytes at high efficiencies in a manner that replicates in vivo heart development. Wildtype (WT), Nkx2.5-shRNA knockdown, and Znf503-shRNA knockdown hESC cell lines were differentiated into cardiomyocytes. WT hESC differentiated into cardiomyocytes at a high efficiency, but mature cardiomyocytes were not detected in Nkx2-5 and ZNF503 knockdown samples. RNA-seq high-throughput transcript profiling was evaluated for days 0, 2, 6, and 10 of cardiomyocyte differentiation (Figure 1) using the Illumina HiSeq 2500 sequencing system determining expression values for > 41,000 genes.

Raw data was obtained from the Gene Expression Omnibus (GSE69618) repository and was analyzed using GeneSpring software and statistical packages within R and Bioconductor customized for normalization and differential analysis. The normalization process was performed using the edgeR and limma packages in R. A trimmed-mean of M values (TMM) method was used to determine the normalization factors, which were then applied using a voom transformation. The resulting matrix contained normalized log2 counts-per-million expression values.

Differential expression analyses were performed on this data using a pipeline of Bioconductor packages. To identify significantly changed genes between the WT and shRNA-knockdown hESCs, a combination of statistical criteria (Significance of <0.05 with FDR) and fold change criteria of 1.5X was applied to comparisons. Genes significantly down regulated in the Znf503-shRNA knockdown samples were further filtered on two criteria: 1) Genes must be downregulated in the Nkx2-5-shRNA knockdown samples since Nkx2-5 expression is required for cardiomyocyte differentiation and 2) Human gene orthologs must be upregulated in the mouse early embryonic heart samples. This analysis narrowed the list of candidate cardiac genes positively regulated by Znf503 to 488 genes (Figure 2). Normalized data was imported into Genespring for data visualization and cluster analysis (Figure 3). All analysis scripts and a more detailed discussion are available at github.com/awilliamson10.

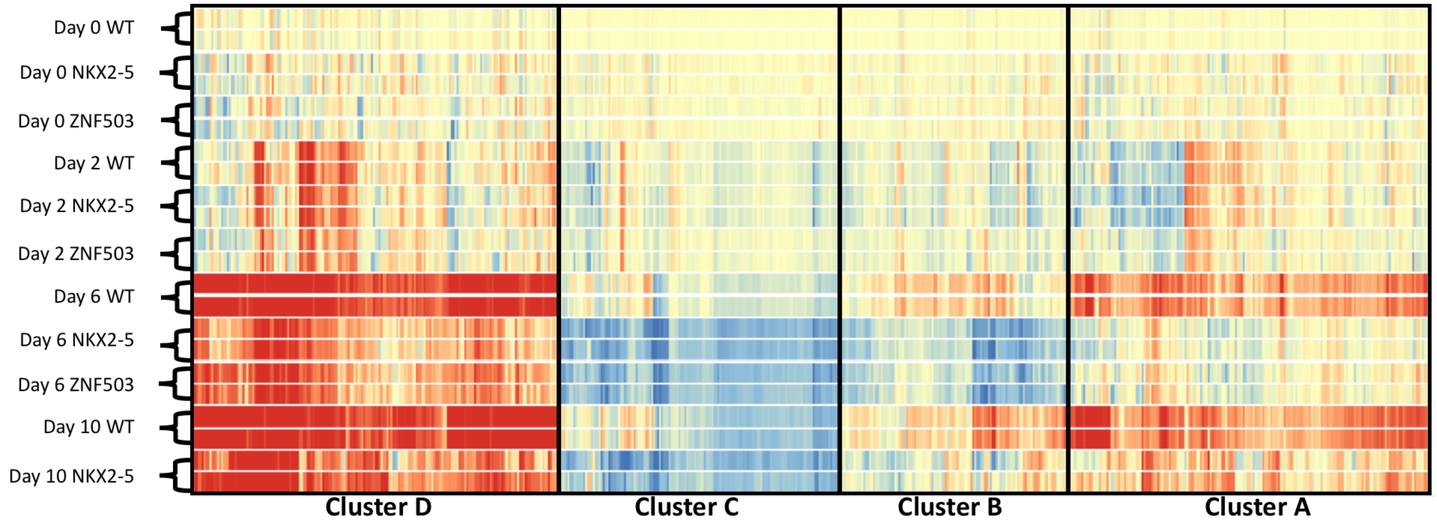

Figure 3: Hierarchical clustering of the 488 genes downregulated in the Day 6 differentiating ZNF503-shRNA knockdown cells compared to wildtype cells. Hierarchical clustering organized these genes into 4 distinct expression signatures based on the expression from Day 0 to Day 10. Cluster A and cluster D are the most notable clusters comprised of a high gene expression signature at Day 6 and Day 10 that is reduced from WT levels in the Znf503 or Nkx2-5 Day 06 samples. A select number of interesting candidate genes are identified in Table 2.

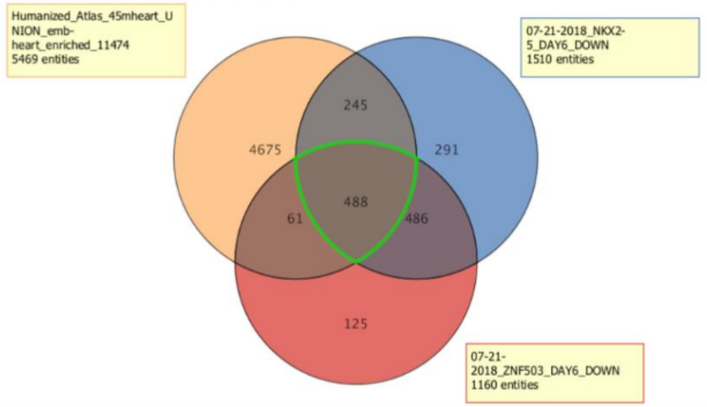

Figure 2: Venn diagram displays total genes present within each subgroup of analysis. Top Left contains the Embryonic Heart Atlas Genes, Top Right contains genes downregulated between NKX2-5 day 6 and WT, Bottom contains genes downregulated between ZNF503 day 6 and wild-type.

		Raw counts		Normalized log2 counts per million		WT d6 vs. ZNF503_ko d6	
GENE	REFSEQ	ZNF503_ko d6	WT d6	ZNF503_ko d6	WT d6	log fold change	P value
RYR2	NM_001035	224	18393	2.26	8.11	5.86	9.08E-08
HMGB2	NM_001130688	2031	10776	5.41	7.30	1.88	2.68E-07
HMGB2	NM_002129	2028	10750	5.41	7.29	1.88	2.70E-07
ACTC1	NM_005159	7073	58717	7.23	9.77	2.53	2.89E-07
HMGB2	NM_001130689	1862	9801	5.29	7.16	1.87	3.17E-07
LMOD1	NM_012134	55	2234	0.21	5.06	4.86	8.19E-07

Table 1: Raw and normalized data and differential expression for select genes in the Busser data (genes in first row of Table 2). Raw counts are counts of RNA-seq fragments counted in the sample. Normalized values are after processing using edgeR and limma packages in R. Fold change and P value are for differential expression of the gene between the two samples types.

References

Busser, Brian W. et al. "An Orthologous Epigenetic Gene Expression Signature Derived from Differentiating Embryonic Stem Cells Identifies Regulators of Cardiogenesis." Ed. Michael Schubert. PLoS ONE 10.10 (2015): e0141066. PMC.
Li, Xing et al. "Transcriptional Atlas of Cardiogenesis Maps Congenital Heart Disease Interactome." Physiological Genomics 46.13 (2014): 482–495. PMC.

Cluster D		Cluster C		Cluster B		Cluster A	
ACTC1	actin alpha cardiac muscle 1	HMGB2	high mobility group box 2	RYR2	ryanodine receptor 2	LMOD1	leiomodin 1
TNNT2	troponin T2, cardiac type	RCOR2	REST corepressor 2	TNNI3	troponin I3, cardiac type	SGCA	sarcoglycan alpha
MYH6	myosin heavy chain	TRIB2	tribbles pseudokinase 2	KCNH2	potassium voltage-gated channel	SNTA1	syntrophin alpha 1
TBX5	T-Box 5	E2F2	E2F transcription factor 2	ESR1	estrogen receptor 1	DES	desmin
WNT2	wnt family member 2	NFIA	nuclear factor I A	LBH	limb bud and heart development	GATA2	GATA binding protein 2
NKX2-5	NK2 homeobox 5	SIX1	SIX homeobox 1	CAND2	cullin associated and neddylation dissociated 2	CREG1	cellular repressor of E1A stimulated genes
ZEB2	zinc finger e-box binding homeobox 2	FGF8	fibroblast growth factor 8	WNT5A	wnt family member 5A	MEF2A	myocyte enhancer factor 2A
ZNF503	zinc finger protein 503	GAP43	growth associated protein 43	GLI3	FLI family member zinc finger 3	COL16A1	collagen type XVI alpha chain 1

Table 2: The table displays genes identified in each cluster of Figure 3.

Conclusions

This preliminary analysis of the Day 6 Znf503-shRNA knockdown using these filtering criteria develops a better picture of the cardiac gene network perturbations and potential downstream candidate genes regulated by Znf503.

Notable observations:

- At differentiation Day 6, both Znf503 and Nkx2-5 are both downregulated in the shRNA knockdown cells demonstrating effective loss of function.
- Several important cardiac genes were downregulated in the Znf503 and Nkx2-5 shRNA knockdowns. These genes include both critical cardiac regulatory genes and genetic markers of functional cardiomyocytes identified in Clusters A and D (see Figure 3 and Table 2).
- The genes identified in these analyses will serve as a starting point as we perform our own Znf503 loss-of-function analysis in our cardiomyocyte differentiation system.

Acknowledgments

SURE Program
The Center for Genomic Advocacy
BD4ISU, NIH 1R25MD011712-01
Department of Mathematics and Computer Science