ETV2 functions as a pioneer factor and regulates the endothelial lineage

Wuming Gong1, Satyabrata Das1, Javier Sierra-Pagan1, Erik Skie1, Mary G. Garry1,2,3, Kenneth S. Zaret4, Daniel J. Garry1,2,3\*

Affiliations:

1Department of Medicine, University of Minnesota, Minneapolis, MN 55455

2Stem Cell Institute, University of Minnesota, Minneapolis, MN 55455

3Paul and Sheila Wellstone Muscular Dystrophy Center, University of Minnesota, Minneapolis, MN 55455

4Institute for Regenerative Medicine, Department of Cell and Developmental Biology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104-5157

\*Corresponding Author

Daniel J. Garry, M.D., Ph.D.

Lillehei Heart Institute

2231 6th St SE

University of Minnesota

Minneapolis, MN 55455

Phone: 612-626-2178

E mail: [garry@umn.edu](mailto:garry@umn.edu)

**Introduction**

During the development, pioneer transcription factors have the unique role of acting on the epigenome to open new regulatory chromatin landscapes. Compared with the non-pioneer transcription factors, the pioneer factors can recognize their target DNA sequences in closed heterochromatin and trigger remodeling of adjacent chromatin landscape to provide accessibility to non-pioneer transcription factors1. By relaxing the chromatin state, these pioneer enable other transcription factors to sequentially access their binding motifs through histone modifications and collectively activating gene expression resulting in reprogramming the cell fate. More than a dozen of pioneer factors have found to play important roles in various biological processes. Examples of pioneer factors include pluripotent factors OSK (Oct4, Sox2, and Klf4) for iPSC reprogramming, FoxA for hepatocyte reprogramming2, GATA3 for mesenchymal to epithelial transition (MET)3, PU.1 and C/EBPa for macrophage differentiation4, Pax7 for pituitary development5, Ebf1 for B cell lineage specification6 and Ascl1 for neurogenesis7.

Studies from our laboratory and others have identified ETV2 as an essential transcription factor for the development of cardiac, endothelial and hematopoietic lineages. Our laboratory has previously shown, using the 3.9kb *Etv2-Cre* and *Rosa-EYFP* reporter alleles, that *Etv2*-EYFP progenitors daughter endocardial/endothelial and hematopoietic lineages in the WT background. Furthermore, when crossed into the *Etv2* mutant background, the *Etv2-*EYFP progenitor cells daughtered cardiomyocytes in the absence of *Etv2*. *Etv2* has been shown to be responsive to BMP, WNT and NOTCH signaling pathways and synergizes with FOXC2 to regulate the endothelial program by directly targeting *Lmo2,*  *Cdh5*, *Tie2* and *Flk1.* Importantly, the ability of ETV2 to reprogram fibroblasts to endothelial cells points to its master regulatory role as a pioneer factor in endothelial development.

In this study, by using two distinct biological systems: ES/EB differentiation and MEF reprogramming, we investigated the role Etv2 as a pioneer factor to drive the endothelial development. Even though these two systems have very different global expression, chromatin accessibility and epigenetic profiles, we found surprisingly similar molecular programs and downstream genes that are regulated by Etv2 induction. We showed that Etv2 cooperates with Brg1, an ATPase of the SWI/SNF chromatin remodeling enzyme has been shown to facilitate other pioneer factors such as Gata3 to relax the closed chromatin, to relax the closed chromatin and convert the nucleosome-enriched to the nucleosome-depleted region during the endothelial lineage development in both MEF and EB. To our knowledge, this is the first study that highlights the pioneer function of Etv2 in two distinct systems.

**Results**

**Etv2 reprogram the fibroblast into endothelial cell population**

Using the *iHA-Etv2* ES cell line, we engineered a mouse that inducibly overexpresses ETV2. We isolated embryonic fibroblasts from this mouse line and demonstrated that these cell populations uniformly expressed fibroblast markers (Thy1.2, CD44 and CD29) and lacked HE expression by FACS (in the absence of Dox) (Supplementary Figure 1a). Using western blot analysis, we further demonstrated that ETV2 (3 hrs post-Dox treatment) was robustly expressed (Supplementary Figure 1b). ETV2 overexpression resulted in more than a 2,000 fold increase in cells expressing FLK1/TIE2 by FACS (Supplementary Figure 1c - 1f). At day 7 post Etv2 induction, we found Ac-LDL uptake, positive sprouting assays and an induction of ETV2 downstream endothelial target expression (Supplementary Figure 1g - 1m).

To further investigate the molecular dynamics of Etv2 inducible reprogramming, we captured and sequenced 3,539, 2,936 and 7,202 high quality single cells from 24 hours, 48 hours and 7 days post induction of Etv2 in mouse embryonic fibroblasts (MEFs). We also performed the scRNA-seq of 948 undifferentiated MEF cells as well as the 827 sorted Flk1+ cells from day 7 reprogrammed cells. The dimension reduction analysis by scVI8, followed by uniform manifold approximation and projection (UMAP) and k-means clustering identified seven distinct cell clusters (Figure 1a, and Figure 1c-1d). We noted that during the first 48 hours of reprogramming, though Etv2 was significantly up-regulated as early as 24 hours, endothelial markers such as Kdr, Lmo2, Emcn, Cdh5 and Sox18 were significantly activated only in subpopulation of cells (cluster 1 for MEF and day 1 cells and cluster 2 for day 2 cells) (Figure 1d, Supplementary Figure 2a and 2b). The pathway analysis suggested that the cell cycle process and chromatin organization related genes were significantly up-regulated in cluster 1 MEF cells, and may be required for the initiation of Etv2 induced reprogramming process (Supplementary Figure 2c). At day 7 of reprogramming, the FACS showed that 17% cells were Flk1+ cells (Supplementary Figure 1c and 1d). The scRNA-seq showed that the Flk1+ cells at day 7 formed a unique cell population (cluster 7). The endothelial marker genes such as Lmo2 and Emcn were robustly expressed in the cluster 7 cells, while the fibroblast markers such as Cd44 and Fosl1 were down-regulated (Figure 1f). The pathway analysis confirmed that genes with functions related to vasculature development and blood vessel development were more abundantly expressed in cluster 7 compared with the remaining clusters (Figure 1g). In summary, the scRNA-seq analysis suggested that inducing Etv2 in MEFs activated the downstream endothelial marker genes and promoted the development of endothelial lineages.

The embryoid body (EB) formation from ES cells is broadly used for studying the role of Etv2 in endothelial development9. To investigate the molecular programs that are commonly or differently responsive to the Etv2 induction in EB and MEF, we induced Etv2 at day 2.5 developing EB and performed bulk RNA-seq of Flk1+ cells at 3 hours post induction (Figure 1b and Supplementary Figure 3). Comparing with the gene expression profiles in D2.5 EB without induction, we identified 2,324 and 2,328 genes that were up- and down-regulated, respectively. Among them, 554 and 1,507 genes were up- and down-regulated in both EB and MEF (Supplementary Figure 4a and 4b). As expected, the commonly up-regulated genes were closed related to the vasculature development, including endothelial markers Myct1 and Sox18, while the commonly down-regulated genes were significantly associated with the translation and RNA processing (Supplementary Figure 4c - 4f).

To examine the chromatin accessibility changes in Etv2 induced cell differentiation, we also performed ATAC-seq (Assays for Transposase-Accessible Chromatin with high-throughput sequencing) at 24 hours, 48 hours and 7 days post induction in MEF and 3 hours post induction in EB (Figure 1a and 1b). The analysis of transcription factors (TF) associated chromatin accessibility analysis suggested that though there existed significant batch effects between MEF and EB (PC1 in Figure 1h), the Flk1+ cells from day 7 MEF and day 2.5 EB post Etv2 induction share the common global chromatin accessibility pattern (PC2 in Figure 1h), including 113 and 246 transcription factors whose chromatin accessibility commonly up-regulated and down-regulated in MEF and EB, respectively (Supplementary Figure 5a - 5d). By integrating the RNA-seq and ATAC-seq data from MEF and EB, we identified 13 TFs whose expression levels and TF associated accessibility consistently increased in Flk1+ cells in both MEF and EB, and 18 TFs decreased in Flk1+ cells (Figure 1i and Supplementary 5e and 5f). It is interesting that mesodermal factors such as Msx2, Eomes and Foxk2 were among the genes that were consistently down-regulated in both MEF and EB, suggesting the critical role of Etv2 on suppressing the non-endothelial lineage development.

Taken together, these results demonstrated that inducing Etv2 was able to promote the cell differentiation toward endothelial lineage and suppressing the non-endothelial lineages, in two very distinct cellular environment: MEF and EB, supporting the notion that Etv2 functioned as a master regulator to drive the endothelial lineage development.

**Etv2 targets nucleosomes during reprogramming**

To investigate how Etv2 binding reshapes the genomic accessibility landscape and drive the endothelial lineage differentiation in different cellular environment, we performed Etv2 ChIP-seq at 24 hours, 48 hours and 7 days post-Etv2 induction in reprogramming MEF, and Etv2 ChIP-seq at 3 hours and 12 hours post-Etv2 induction in developing EB at day 2.5. The initial Etv2 binding events were captured by the Etv2 ChIP-seq at 24 hours in MEF and 3 hours in EB, resulting in 18,024 and 131,001 peaks, respectively (Figure 2a). The 11,751 common Etv2 peaks overlapped between MEF and EB represented the majority (65.2%) of the peaks EB, and located predominately in the promoter region. In contrast, the EB and MEF specific Etv2 peaks were more likely distributed at the distal intergenic regions (Figure 2b). One of the key feature of pioneer factors is their capability of targeting the nucleosomes. To test whether Etv2 is able to target the nucleosome, we first examined the nucleosome profiles of 24 hours Etv2 ChIP-seq peaks in undifferentiated MEF using published histone H3 ChIP-seq and MNase-seq10. We divided the Etv2 ChIP-seq peaks into four quartiles based on the mean MNase-seq signals of the central 200-bp region, and we used the first (lowest mean signal) and the fourth quartile (highest mean signal) to represent the nucleosome free region (NFR) and nucleosome (Figure 2c and Supplementary Figure 6a). Thus, the nucleosome quartile represents the Etv2 peaks which summits located at the nucleosome centers, suggesting that Etv2, like other reported pioneer factors, is also able to target the nucleosome during the MEF reprogramming. Similarly, we divided the Etv2 ChIP-seq peaks at 3h post induction in day 2.5 EB into NFR (5,291 peaks) and nucleosome (8,843 peaks) groups according to the local V-plot and fragment size profiles of ATAC-seq day 2.5 EB without Etv2 induction (Supplementary Figure 6b-6d). We found that similar to the MEF reprogramming, Etv2 also targets the nucleosome centers at day 2.5 EB (Figure 2d). Our results in MEF and EB suggested Etv2 is intrinsically able to target and bind the nucleosomes, and this capability appears to be independent of cellular context. Sequence motif analysis identified a common GGAAAT motif that are significantly more enriched in NFR regions compared with the nucleosomes in both MEF and EB (Fisher's adjusted p-value=6.0E-05 and 6.8E-5) (Figure 2e and Supplementary Figure 7a). This motif has an additional terminal "AT" nucleotides compared with the canonical motifs identified from the canonical Etv2 motifs. These findings suggested that Etv2, similar to other pioneer factors, such as Oct4, Sox2 and Klf4, were able to target nucleosome-enriched sites using partial or degenerate motifs, and targeted their full canonical motif in nucleosome-depleted sites1.

Moreover, we found that Etv2 targeted nucleosome region were characterized by low Brg1 ChIP-seq signals at 24 hours post induction in MEF and 3 hours post induction in EB, suggesting that initial recognition of nucleosome may not require the existence of local Brg1 (Figure 2c and 2d). Interestingly, the Etv2 targeted nucleosomes had distinct H3K27ac surrounding pattern in MEF and EB: Etv2 targeted H3K27ac depleted nucleosomes in MEF and H3K27ac enriched ones in EB, suggesting that Etv2 recognize the nucleosomes regardless of the H3K27ac enrichment.

**Brg1 maintains and stabilizes the binding of Etv2**

To further elaborate how Etv2 induction drive the endothelial lineage development during the MEF reprogramming and ES/EB differentiation, we performed Etv2 ChIP-seq at 48 hours and 7 days post Etv2 induction in reprogramming MEF and Etv2 ChIP-seq at 12 hours post induction in EB. We identified in total 154,468 and 19,651 non-overlapping Etv2 peaks in MEF and EB, respectively. Similar to OSK-binding sites during the reprogramming toward iPS, the majority (more than 80%) of the unique Etv2 peaks were only present at early stages (day 1 for MEF and 3 hour for EB) but absent from late stages (day 7 for MEF and 12 hour for EB), suggesting that most of the initial Etv2 binding events were short-lived and were not sustainable during the differentiation (Supplementary Figure 8a and 8b). We divided the Etv2 peaks into "early", "late" and "persist" groups according to whether Etv2 peaks were present in early stage, late stage or both stages (Figure 3a and 3b). Though we previously found that the majority of initial Etv2 binding events were independent Brg1, the "late" Etv2 peaks were associated with increasing levels of Brg1 in late stage of EB and MEF, while the "persist" Etv2 peaks were associated with the highest levels of Brg1 in the early stage. Note that the establishment of "late" Etv2 peaks were also coupled with the increasing H3K27ac enrichment, in contrast to the "persist" Etv2 peaks that initially located at high H3K27ac environment. These results suggested a significant role of Brg1 on maintaining and stabilizing the binding of Etv2 during the reprogramming and facilitating Etv2 activating the downstream endothelial programs. Indeed, we found there were significant overlap between the "late" Etv2 peaks, as well as the nearby genes (Figure 3c and 3d). Moreover, the "late" Etv2 peaks located near higher proportion of endothelial genes, in contrast to the "early" Etv2 peaks (Figure 3e). Similar to Ascl1 induced neural reprogramming, we also found the significant phasing events around the Etv2 binding motifs in the Flk1+ cells in both MEF and EB (Supplementary Figure 9)11. In summary, these results suggested that Brg1 helps maintain and stabilize the Etv2 binding, promoter the local H3K27ac enrichment, and activated the downstream endothelial genes (Figure 3f).

**Discussion**

**Figure Legends**

**Figure 1. Etv2 promotes the endothelial program in both MEF and EBs. (a)** Schematic of the reprogramming strategy in iHA-Etv2 MEFs by overexpression of Etv2 with doxycycline (Dox). **(b)** Schematic of the differentiation of embryoid body (EB) and induction of Etv2 at day 2.5 of differentiation. **(c-d)** The UMAP plot shows the scRNA-seq of 948 undifferentiated MEFs, 3,539 reprogrammed cells at 24 hours, 2,936 cells at 48 hours and 7,202 cells at 7 days and 827 Flk1+ cells at 7 days post-induction of Etv2 in MEFs. The dimension reduction analysis by scVI, followed by uniform manifold approximation and projection (UMAP) and k-means clustering identified seven distinct cell clusters. The color represents **(c)** the cell sources and **(d)** cell clusters. **(e)** The expression profiles of Etv2 and Kdr (Flk1). **(f)** The volcano plot of genes differentially expressed between cluster 1 and cluster 7. The *p*-values were determined by Wilcoxon ran sum test of the normalized read counts. **(g)** The biological process that are significantly associated with the up-regulated in genes in cluster 7 (Flk1+ cells at day 7 of reprogramming) compared with cluster 1 (undifferentiated MEFs). **(h)** The PCA of the TF deviations of the ATAC-seq of MEF reprogramming (MEF, 24 hours, 48 hours and 7 days post induction) and EB differentiation (2.5 day and 3 hours post induction). The TF deviations were inferred by chromVAR. **(i)** The 31 transcription factors which expression levels and motif associated chromatin accessibility showed consistently directional change in both EB and MEF (13 up-regulated TFs and 18 down-regulated TFs).

**Figure 2. Etv2 targets nucleosomes during reprogramming. (a)** The Venn diagram shows 131,001 and 18,024 Etv2 ChIP-seq peaks at 24 hours post induction in MEF reprogramming and 3 hours post induction in day 2.5 EB, respectively. There are 11,751 common Etv2 peaks overlapped between MEF and EB. **(b)** The genomic distribution of EB specific, MEF specific and common Etv2 peaks. The EB and MEF specific Etv2 peaks were more likely distributed at the distal intergenic regions. **(c)** The heatmap shows the read density of MNase-seq, Brg1 ChIP-seq and H3K27ac ChIP-seq in MEFs, surrounding 131,001 Etv2 ChIP-seq peaks at 24 hours post induction in MEF reprogramming. The Etv2 peaks were divided into four quartiles based on the mean MNase-seq signals of the central 200-bp region. The first (lowest mean signal) and the fourth quartile (highest mean signal) were used to represent the nucleosome free region (NFR) and nucleosome. **(d)** The heatmap shows the ratio of NFR / nucleosome reads density, reads density of Brg1 ChIP-seq and H3K27ac ChIP-seq at EB day 2.5, surrounding 18,024 Etv2 ChIP-seq at 3 hours post induction. The Etv2 peaks were divided into NFR (5,291 peaks) and nucleosome (8,843 peaks) groups according to the local V-plot and fragment size profiles of ATAC-seq day 2.5 EB without Etv2 induction. **(e)** Sequence motif analysis by DREME and CentriMo identified a common GGAAAT motif that are significantly more enriched in NFR regions compared with the nucleosomes in both MEF and EB (Fisher's adjusted p-value=6.0E-05 and 6.8E-5).

**Figure 3.** **Brg1 maintains and stabilize the binding of Etv2. (a)** The heatmap shows the fold enrichment of Etv2 ChIP-seq, Brg1 ChIP-seq and H3K27ac ChIP-seq, at 24 hours, 48 hours and 7 days post-induction of Etv2. We also included the Brg1 and H3K27ac ChIP-seq at undifferentiated MEF. Each site is centered at the 154,468 non-overlapping Etv2 ChIP-seq summits in MEF reprogramming. The Etv2 peaks into "early", "late" and "persist" groups according to whether Etv2 peaks were present in 24 hours post induction (early stage), 48 hours post induction (late stage) or both stages. **(b)** The heatmap shows the fold enrichment of Etv2 ChIP-seq, Brg1 ChIP-seq and H3K27ac ChIP-seq, at 3 hours and 12 hours post-induction of Etv2 in day 2.5 EB. We also included the Brg1 and H3K27ac ChIP-seq at day 2.5 EB without induction. Each site is centered at the 19,651 non-overlapping Etv2 ChIP-seq summits in EB differentiation. The Etv2 peaks into "early", "late" and "persist" groups according to whether Etv2 peaks were present in 3 hours post induction (early stage), 12 hours post induction (late stage) or both stages. **(c-d)** The Venn diagram shows the overlap between late Etv2 peaks in MEF and EB, as well as the overlap between the genes near the late Etv2 peaks. **(e)** The percent of blood vessel development related genes near the early, late or persist Etv2 peaks in EB and MEF. (f) The model of Etv2 induced reprogramming. Etv2 targets the nucleosome during the early stage of reprogramming. Brg1 help to maintain the stabilize the Etv2 binding near the endothelial genes, coupled with the increasing of local H3K27ac levels.

**References**

1. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).   
  
2. Gualdi, R. *et al.* Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Gene Dev* **10**, 1670–1682 (1996).   
  
3. Takaku, M. *et al.* GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol* **17**, 36 (2016).   
  
4. van Oevelen, C. *et al.* C/EBPα Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem Cell Rep* **5**, 232–247 (2015).   
  
5. Mayran, A. *et al.* Pioneer and nonpioneer cooperation drives lineage specific chromatin opening. *bioRxiv* 472647 (2019) doi:10.1101/472647 .   
  
6. Györy, I. *et al.* Transcription factor Ebf1 regulates differentiation stage-specific signaling, proliferation, and survival of B cells. *Gene Dev* **26**, 668–82 (2012).   
  
7. Wapinski, O. L. *et al.* Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621–635 (2013).   
  
8. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat Methods* **15**, 1053–1058 (2018).   
  
9. Rasmussen, T. L. *et al.* ER71 directs mesodermal fate decisions during embryogenesis. *Development (Cambridge, England)* **138**, 4801–4812 (2011).   
  
10. Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* (2017) doi:10.1016/j.cell.2016.12.016 .   
  
11. Wapinski, O. L. *et al.* Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. *Cell reports* **20**, 3236–3247 (2017).