Transcriptional profile of induced and primary neurons reveals new candidate genes for lineage reprogramming

Diego M Coelho, Sandro José de Souza, Marcos R Costa
Bioinformatics Multidisciplinary Environment, IMD, NeuroCell, Brain Institute; Bioinformatics Multidisciplinary Environment, IMD, NeuroCell, Brain Institute

Abstract
Somatic cells can be directly reprogrammed into neurons through the expression of few transcription factors. However, the precise mechanisms involved in the lineage-conversion are poorly understood. Similarly, it remains unclear how similar lineage-reprogrammed induced neurons (iNs) are to bona fide central nervous system neurons. In this work, we used an unsupervised machine-learning approach to compare the transcriptional profiles of lineage-reprogrammed mouse embryonic fibroblasts (MEFs), mouse embryonic telencephalon neural progenitors and neurons, as well as mouse postnatal cerebral cortex neurons. We show that the transcriptional profile of a subpopulation of lineage-reprogrammed MEFs resembles that of primary neural progenitors and neurons, indicating that the intermediate steps enacted by reprogramming factors within MEFs during the transition to iNs are similar to those observed during primary neuron differentiation. Finally, by comparing the transcriptional profiles of MEFs that undertook a neuronal pathway to that of MEFs adopting a myogenic fate or retaining fibroblast features we identified potential candidates to improve the efficiency for lineage conversion of those cells into neurons.

Introduction
Somatic cells can be directly reprogrammed into neurons through the expression of few transcription factors. Astrocytes isolated from the postnatal cerebral cortex of mice were the first cells to be directly reprogrammed into neurons following expression of the transcription factor Neurogenin 2 (Neurog2) or Mammalian achaete-scute homolog 1 (Mash1/Ascl1) \[1\] \[2\]. Subsequently, the list of cell types reprogrammed into induced neurons grew substantially, including non-neural cells, such as mouse fibroblasts and hepatocytes \[3\] \[4\]. Non-neural cells, however, typically require more than one transcription factor to achieve a full neuronal conversion. Recently, it has been reported that expression of Ascl1 alone, but no other proneural genes such as Neurog2, is sufficient to induce conversion of fibroblasts into induced neurons \[5\], albeit at low efficiency (~10%). Studies using different cell types and neurogenic transcription factors describes a significant failure rate in reprogramming \[6\] \[7\] \[8\] \[9\] \[10\] \[11\] \[5\]. Generally, this incompetence of somatic cells to be lineage-reprogrammed is explained by probable differences in the transcriptional machinery activated by neurogenic transcription factors, but the exact mechanisms involved in this phenomenon remains largely unknown. Some initial attempts seeking a better understanding of the molecular mechanisms involved in the reprogramming of somatic cells into neurons were recently made \[11\] \[5\]. Still, these work were mostly based in the comparison of transcriptional profiles of cell populations transduced with neurogenic transcription factors versus control. As discussed above, however, many cells transduced with neurogenic transcription factors fail to reprogram. Thus, the transcriptional profile obtained from total population of cells transduced with neurogenic transcription factors contains both, i) genes regulated in cells that undergo a complete program of neuronal differentiation, and ii) genes regulated in cells that failed to reprogram. As a consequence, genes weakly regulated upon neurogenic transcription factor expression, which may be pivotal for reprogramming, are likely overlooked. More recently, the first comprehensive transcriptional analysis of neuronal induction was published \[12\]. This work showed that mouse embryonic fibroblasts (MEF) expressing only Ascl1 mostly failed to go through a complete differentiation into iN, going instead towards a myogenic phenotype. This is also observed for MEFs expression a combination of three TFs (Brn1, Myt1 and Ascl1-BAM), although
the frequency of cells adopting a neuronal phenotype increases in this last condition. However, it remains unclear whether the transcriptional regulation involved in the lineage reprogramming of MEFs to induced neurons resembles that observed during bona fide neuron differentiation.

In this work, we used an unsupervised machine-learning approach based on principal component analysis (PCA) to select genes that better correspond to different cell states of MEFs, MEFs 5 days after transduction with Ascl1 (ascl1d5), MEFs 22 days after transduction with Ascl1 (ascl1d22) and MEFs 22 days after transduction with Brn2, Ascl1 and Myt1l (bamd22) (Treutlein B et al. 2016), primary neurons from postnatal mice brains (P7) [13] and neural progenitors and primary neurons from embryonic mice brains [14]. Next, we compared the transcriptional profiles of MEFs undergoing neuronal conversion or MEFs that failed to do so and identified some potential candidate genes to enhance lineage reprogramming.

**Objective**

Analyze whether the transcriptional profiles of cells during the transition from fibroblasts to induced neurons upon expression of Ascl1 or the combination Ascl1/Brn2/Myt1l resembles the transcriptional modifications enacted by primary neural progenitors during the process of differentiation into neurons in the developing cerebral cortex. Based on this analysis, we also aim at identifying new candidate genes for direct neuronal reprogramming.
Figure Legend

Figure 1. 
(A) Pseudo-temporal cell decision map. An unsupervised machine-learning approach based on PCA selected 400 genes whose expression signature was used to separate sample cells. A dimension reduction followed by cell ordering was then plotted to give us a pseudo-temporal positioning of cells reaching divergent phenotypes. Cell types are: MEF (mouse embryonic fibroblasts), ascl1d5 (MEF cells reprogrammed with ascl1 after 5 days), ascl1d22 (same as before with 22 days), bamd22 (same with BAM transcriptional factors), neuronP07 (postnatal mouse cerebral cortex neurons), neuronE14 (embryonic neurons, mouse dorsal telencephalon), NDP (neural differentiated progenitors, mouse
Results & Discussion

To compare the transcriptional profiles lineage-reprogrammed and primary cells, an unsupervised machine-learning approach based on principal component analysis (PCA) was used to select genes that better correspond to different cell states [15]. We analyzed three datasets: 1) Single-cell RNAseq of mouse embryonic fibroblast (MEF), MEF 5 days after transduction with Ascl1 (ascl1d5), MEF 22 days after transduction with Ascl1 (ascl1d22) and MEF 22 days after transduction with Brn2, Ascl1 and Myt1l (bamd22) [12] (GEO: GSE67310); 2) Primary neurons from postnatal mice brains (P7) [13] (GEO: GSE52564); and 3) neural progenitors and primary neurons from embryonic mice brains [14] (GEO: GSE65487). As each gene represents a dimension, a strategy based on a dimensional reduction was applied. Thus, samples were ordered to create a pseudo-temporal map, which places cells towards a cell differentiation state [15] (Fig. 1A). Interestingly, we observed that the transcriptional profiles of reprogrammed induced neurons and primary neurons fully overlapped (Fig. 1A). More interestingly, a continuum was observed from MEFs to a subpopulation of Ascl1- or BAM-transduced cells after 5 and 22 days, where part of the cells overlap with primary neural progenitors and immature neurons, whereas other cells follow a different path (Fig. 1A). This first analysis suggests that the transcriptional programs enacted during the conversion from MEFs to induced neurons and neural progenitor cells to neurons are similar. To further confirm that the observed pseudo-temporal map could represent different fates of reprogrammed MEFs, we first analyzed cell-type ontologies on Mouse Gene Atlas using enrichR (Fig. 1B). We could confirm that genes were enriched for MEF, neural cells and muscle cell ontologies (Fig. 1B). Next, pan-neuronal markers (Tubb3, Map2), myocyte markers (Tnnec2, Myo18b) and MEF markers (S100a4, Col1a2) were selected among the 400 genes to analyze the levels of expression in different cell states (Fig. 1C and D). S100a4 and Col1a2 were more expressed in the branch containing MEFs, whereas Tnnec2 and Myo18b were more expressed in the branch comprising most ascl1d5 and a small subset of bamd22 cells. In contrast, Tubb3 and Map2 expression were enriched in the branch comprising few ascl1d5, the majority of bamd22 and primary neuronal progenitors and neurons. These observations suggest that the node (number one) identified in dorsal telencephalon) and NPP (neural proliferating progenitors, mouse dorsal telencephalon).

(B) Mouse Gene Atlas cell ontologies related to genes selected by the machine-learning approach. Genes selected by unsupervised machine-learning were enriched for fibroblasts cell lines (MEF and NIH/3T3), myoblast cell lineage (C2C12) and several neuronal cell lines (Hypothalamus, amygdala, cerebral cortex and others). The combined score is computed by taking the log of the p-value from the Fisher exact test and multiplying that by the z-score of the deviation from the expected rank. (C) Expression of cell-type specific genes along the pseudo-temporal map. Gene expression levels are positively associated to the diameter of the corresponding circles. Note that genes associated with fibroblast phenotype (S100a4 and Col1a2) are enriched in the branch containing MEFs and Ascl1d5 cells, whereas genes associated to muscle phenotypes (Myo18b and Tnnec2) are enriched in a branch containing ascl1d22, few bamd22 and ascl1d5 cells. Genes associated with neuronal fates (Map2 and Tubb3) are enriched in the third branch, containing most bamd22 cells, few ascl1d5 and primary neurons (neuronP07 and neuronE14).

(D) Same as in B, now showing the relative expression of selected genes throughout ordered cells distribution. (E) Heatmap of differentially expressed genes among ascl1d5 subgroups. Some genes were highly related to specific sub-groups like, for example, Acta1 that is mainly expressed in cells from ascl1d5 muscular branch (ascl1d5-mus). Other cases includes Lum gene, mainly expressed in the undifferentiated ascl1d5 branch (ascl1d5-und) and Crmp1, highly expressed in samples located in ascl1d5 neuronal branch (ascl1d5-neu).

(F) Genes enriched in ascl1d5 cells in the neuronal branch along the pseudo-temporal map. Circles with different sizes illustrates how candidate genes are expressed. These pseudo-temporal maps reassure how those genes are related to samples in neuronal branch.
the pseudo-temporal map pinpoints the divergence between cells following a muscular- or neuronal-cell fate.

Next, we set out to identify genes enriched in the ascl1d5 cells classified in the neuronal branch as compared to the MEF and muscle-cell branches. Thus, gene expression patterns of ascl1d5 cell populations classified in each of those branches were compared. We identified 33 genes differentially expressed (q-value < 0.05, likelihood ratio test) in the three different populations of ascl1d5 cells (Fig. 1E). Among these genes, 4 were highly enriched in ascl1d5 cells in the neuronal branch (Fig. 1F). Interestingly, these genes were also enriched in bamd22 cells and primary neural cells (Fig. 1F).

In this work, we show that lineage-reprogrammed MEFs undergo transcriptional changes towards the generation of induced neurons that resembles those observed in the transition from primary cerebral cortex progenitors to early-differentiated neurons. 5 days after expression of Ascl1 in MEFs, a subset of cells show enriched expression of pan-neuronal genes and are transcriptionally similar to neural progenitors. Similarly, the transcriptional profile of bamd22 cells, which mostly adopt a iN phenotype [12], is closely related to P7 cerebral cortex neurons. In contrast, lineage-reprogrammed MEFs that express low levels of pan-neuronal genes showed enriched expression of fibroblast genes or muscle-cell genes, indicating that those two populations represent MEFs that failed to undergo lineage-conversion or followed an alternative fate [12]. We also show that three different populations of ascl1d5 cells can be distinguished based on their transcriptional profiles. Gene expression patterns of these populations are classified in the unsupervised machine-learning approach in branches containing either undifferentiated MEFs, muscle cells or neurons. Using this classification, we identified 33 genes differentially expressed in ascl1d5 cell populations and four genes specifically enriched in ascl1d5 cells in the neuronal branch. These genes may be interesting candidates or contribute to identify new factors to enhance MEF lineage-reprogramming into iNs.

Collapsin Response Mediator Protein 1 (Crmp1) is part of CRMP family of proteins and is typically associated as mediator of sema3A signaling and axon guidance [16] [17]. Interestingly, some CRMP proteins are differentially expressed in axon and dendrites of distinct neuronal types [16]. Embryonic Lethal, Abnormal Vision, Drosophila-Like 4 (Elavl4), also known as Hu-Antigen D (HuD) is a RNA-binding protein involved in neuronal maturation [18], neurite outgrowth and dendritic maintenance [19]. Stathmin 3 (Stmn3) or SCG10-Like Protein (SCLIP) is also related to dendritic formation [20] and neurite outgrowth [21]. Zinc finger, CCHC domain containing 12 (Zcchc12) or Smad-Interacting Zinc Finger Protein 1 (Szn1) is a protein used in BMP [22], AP-1 and CREB signalling [23] as a co-activator. Possibly, Brn2 and Myt1l may sustain gene expression of those candidate genes longer than Ascl1-only reprogramming which allows MEF cells differentiate into neuron-like cells. Thereby, all candidates are related to neuronal phenotype at some level, which indicate possible proteins to help Ascl1 reprogramming MEFs achieve a neuron-like state.

Conclusions

Our results indicate that somatic cells during the process of lineage reprogramming into induced neurons undergo transcriptional changes resembling those enacted in the transition from bona fide neural progenitors to neuronal states. Comparison of transcriptional profiles of intermediate stages during lineage conversion may contribute to identify new candidate genes to improve neuronal reprogramming.

Limitations

Only MEFs with Ascl1 and BAM were analyzed, which illustrates a limited screen of reprogrammed cells possibilities. Thus, analysis of different cell types reprogrammed with Ascl1 and BAM, as well as MEF lineage reprogrammed with other transcription factors are vital to fully understand those pathways taken by reprogrammed cells. Furthermore, a more diverse set of controls are needed since only cerebral cortex cells were used as naive neuronal cell reference.

It would be interesting to evaluate candidate genes with a system biology approach looking at, for example, gene regulatory networks. Furthermore, miRNAs plays an impor-
tant role on regulation of gene networks and had been already used as reprogramming enhancers, so they should not be neglected in the future analysis.

**Additional Information**

**Methods and Supplementary Material**
Please see https://sciencomatters.io/articles/201704000019.

**Funding Statement**
Diego M. Coelho is supported by a Ph.D. fellowship from CAPES.

**Acknowledgements**
For thoughtful discussions, we gladly thank André Fonseca, Vandeclécio da Silva and Prof. Jorge Estefano De Souza.

**Ethics Statement**
Not Applicable.

**Citations**


10. Liu Meng-Lu et al. "Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons". In: *Nature Communications* 4 (July 2013), p. 2183. DOI: 10.1038/ncomms3183. URL: https://doi.org/10.1038/ncomms3183.


