Received: 29 May 2020 | Accepted: 16 August 2020

DOI: 10.1002/sctm.20-0227

ENABLING TECHNOLOGIES FOR CELL-BASED CLINICAL TRANSLATION

STEM CELLS TRANSLATIONAL MEDICINE

TransSynW: A single-cell RNA-sequencing based web application to guide cell conversion experiments

Mariana Messias Ribeiro[1] | Satoshi Okawa[1,2] | Antonio del Sol[1,3,4]

[1]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Belvaux, Luxembourg

[2]Integrated BioBank of Luxembourg, Dudelange, Luxembourg

[3]CIC bioGUNE, Bizkaia Technology Park, Derio, Spain

[4]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Correspondence
Antonio del Sol, PhD, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6, avenue du Swing, L-4367 Belvaux, Luxembourg.
Email: antonio.delsol@uni.lu

Funding information
Fonds National de la Recherche Luxembourg, Grant/Award Number: C17/BM/11662681

Abstract

Generation of desired cell types by cell conversion remains a challenge. In particular, derivation of novel cell subtypes identified by single-cell technologies will open up new strategies for cell therapies. The recent increase in the generation of single-cell RNA-sequencing (scRNA-seq) data and the concomitant increase in the interest expressed by researchers in generating a wide range of functional cells prompted us to develop a computational tool for tackling this challenge. Here we introduce a web application, TransSynW, which uses scRNA-seq data for predicting cell conversion transcription factors (TFs) for user-specified cell populations. TransSynW prioritizes pioneer factors among predicted conversion TFs to facilitate chromatin opening often required for cell conversion. In addition, it predicts marker genes for assessing the performance of cell conversion experiments. Furthermore, TransSynW does not require users' knowledge of computer programming and computational resources. We applied TransSynW to different levels of cell conversion specificity, which recapitulated known conversion TFs at each level. We foresee that TransSynW will be a valuable tool for guiding experimentalists to design novel protocols for cell conversion in stem cell research and regenerative medicine.

KEYWORDS

cellular therapy, clinical translation, differentiation, direct cell conversion, genomics, reprogramming, synergy, transcription factors

1 | INTRODUCTION

Cell conversion is fundamental to many biological processes. Control of cell conversion has significant relevance in stem cell research. For example, generation of functionally specific cells by cell conversion is of clinical interest for cell replacement therapies. However, several roadblocks need to be overcome for achieving optimal cell conversion, such as the accurate characterization of cell populations and the identification of cell conversion factors. Single-cell RNA-sequencing (scRNA-seq) technologies have made it possible to address these challenges. Due to the greater amount of scRNA-seq data generated across the world, experimental researchers are increasingly expressing their interest in deriving novel functional cell types.

Here, we present TransSynW, a scRNA-seq based web application for identifying cell conversion transcription factors (TFs) applicable in stem cell and clinical research (Figure 1A). It prioritizes pioneer factors (PFs) in the prediction of conversion TFs. Evidence suggests that PFs have a key role in chromatin opening, a process often required for cell conversion.[1] Indeed, including PFs on cell conversion protocols has been shown to improve their outcome.[1] Furthermore, it

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2020 The Authors. STEM CELLS TRANSLATIONAL MEDICINE published by Wiley Periodicals LLC on behalf of AlphaMed Press.

predicts marker genes for each target cell type, enabling researchers to assess the fidelity of experimentally converted cells. In addition, it is user-friendly, and it does not require users' computer programming or computational resources. We also created a comprehensive video tutorial for guiding users through the web interface.

The application of TransSynW to various cell systems well-recapitulated known cell conversion TFs and made novel predictions, including the phenotypic conversion between cells in organoids and their in vivo counterparts. Moreover, predicted marker genes were consistent with experimentally known ones. These results highlight the applicability of TransSynW to a wide range of cell conversion experiments.

Significance statement

The study proposes a computational web application, TransSynW. To the best of the author's knowledge, it is the only computational tool that can identify cell conversion transcription factors (TFs) for any cell population in single-cell RNA-sequencing data. TransSynW does not require prior biological information, computer programming, and users computational resources. In addition, TransSynW prioritizes pioneer factors among predicted conversion TFs to facilitate chromatin opening often required for cell conversion. Furthermore, TransSynW predicts marker genes for assessing the performance of cell conversion experiments. Thus, TransSynW will be a staple tool for guiding experimentalists to design novel protocols for cell conversion in stem cell research and regenerative medicine.

2 | RESULTS

2.1 | Method overview

The TransSynW algorithm first identifies specifically and nonspecifically expressed TFs, and selects the combination that exhibits the highest synergistic interactions among them (see Methods) (Figure 1B). Notably, here we considered for the nonspecific part only PFs that have previously been reported to be involved in cell conversion protocols (Table S1). Predicted conversion TFs are then ranked by the expression fold change between the target and starting cell populations and users can prioritize the TFs for experimental follow-ups based on this ranking.

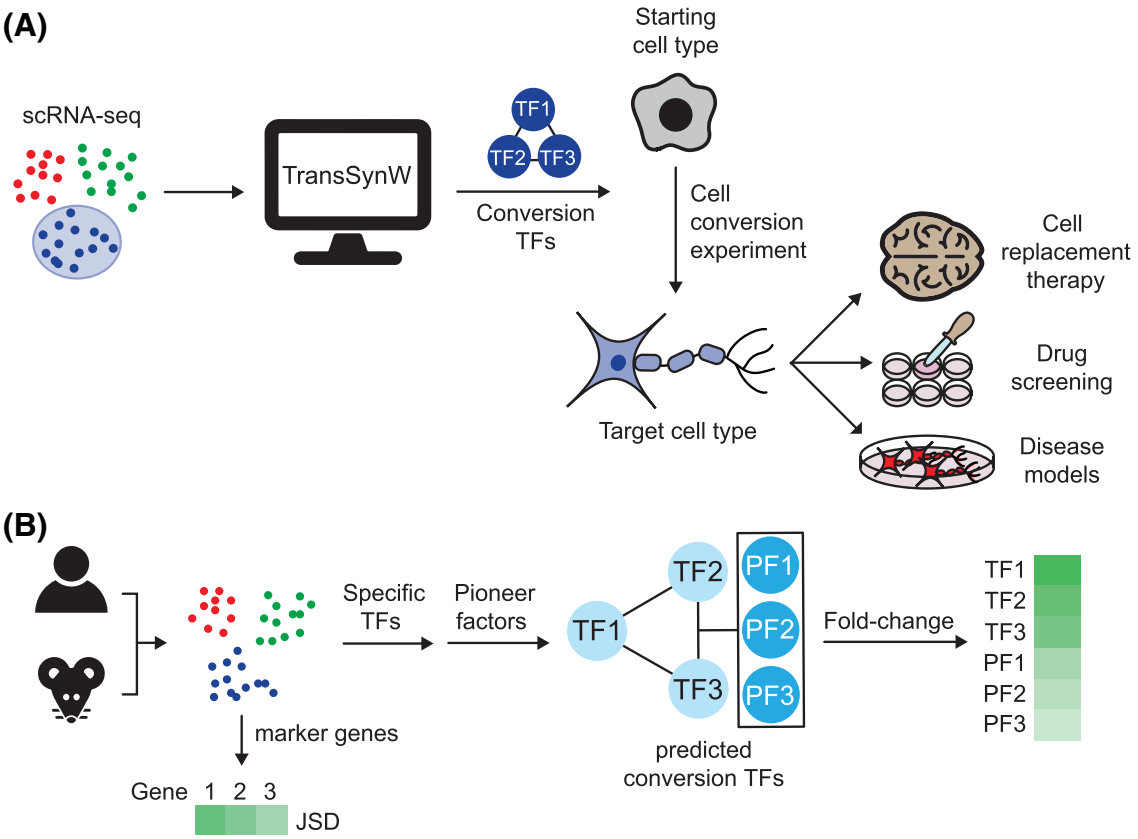

FIGURE 1 A, Application of TransSynW to stem cell research and regenerative medicine. B, Schematic overview of TransSynW algorithm (see also Methods). First, transcription factors (TFs) most specifically expressed in the selected target cell population (specific TFs) and nonspecifically expressed pioneer factors (PFs) are computed. The most synergistic combination of specific TFs and nonspecific PFs is then identified. The predicted set of TFs are ranked by expression fold change between target and starting cell populations. In parallel, top 10 candidate marker genes for target cell population are computed by JSD

We compiled the scRNA-seq data of starting cell types frequently used in cell conversion experiments from various scRNA-seq platforms (Table S2). For optimal results, users are recommended to use starting and target cell type data obtained from the same scRNA-seq platform or, if not available, from the closest sequencing platform. In general, it is recommended to select at least one PF and one specific TF from the predicted conversion TFs. It may be advisable to select more factors if the phenotypic difference between the starting and target cell types is large. Finally, TransSynW also predicts potential marker genes of the target cell populations. This feature enables researchers to select markers for assessing the performance of their cell conversion experiments.

TABLE 1 Predicted specific transcription factors (TFs) and nonspecific PFs

Cell type	Specific TFs	Nonspecific PFs	Annotation in data	Data source (PubMed ID)
(1) Conversion into broad cell type				
Myoblast	**MYF5**, **MYOD1**, **PAX7**, GLIS3, **PAX3**	CEBPB, IRF8, PBX1	1,3,4,5,7	30283141
Keratinocyte	**TRP63**, GATA3, NFIB	**KLF4**, GRHL2, CEBPA	0-16	30283141
Cardiomyocyte	**NKX2-5**, **TBX5**, PROX1, ZFP579, NR0B2	**GATA4**, **MEIS1**, PBX1	9,14	30283141
Hepatocyte	NR1I2, ZFP750, ZFHX4, **HNF1A**, ZBTB48	**HNF4A**, **FOXA3**, **FOXA2**	4,5,10,11,12,15	30283141
HSC	**HLF**, HOXA9, **GATA2**, **TAL1**, **MYCN**	CEBPB, CEBPA, **PBX1**	0,4,8	30283141
Neuron	EOMES, NEUROD6, EGR4, RARB, DLX6	**FOXG1**, **NEUROD1**, PBX1	9,10,12	30283141
Oligodendrocyte/OPC	**NKX6-2**, **OLIG1**, **SOX10**, **OLIG2**, NFE2L3	SOX2	0,6,11	30283141
Macrophage	RUNX3, BATF3, BATF, NFE2, E2F1	**SPI1**, **CEBPA**, ARID3A	Different tissues	30283141
Beta cell	**NKX6-1**, **PDX1**, **MAFA**, OVOL2, MNX1	**NEUROD1**, ISL1, **FOXA2**	0,8,9,11,17	30283141
NSC	ZFP275, **ASCL1**, **TCF3**	**FOXG1**, **SOX2**, PBX1	All young NSCs	30827680
(2) Conversion into subtype				
Dopaminergic neuron	NPAS4, **MYT1L**, EBF3, POU6F1, BNC2	**FOXA2**, **ASCL1**, **GATA3**	hDA	27716510
Medial floorplate progenitor	**LMX1A**, SP2, NR2F6, LMX1B, HMGA2	**FOXA2**, ASCL1, SOX2	hProgFPM	27716510
GABAergic neuroblast	GATA3, SOX14, **MYT1L**, BNC2, ZBTB38	**ASCL1**, **SOX2**, PBX1	hNbGaba	27716510
Oculomotor neuron	**PHOX2B**, **PHOX2A**, **ISL1**, RXRG, NR2F2	FOXA2, ASCL1, "PBX1	hOMTN	27716510
Serotonin neuron	**FEV**, GATA3, SOX1, DPF1, **LMX1B**	**GATA2**, PBX1	hSert	27716510
CD4+ central memory T cell	RBSN, RFX3, NR4A1, KLF9, ID3	GATA3, CEBPB	TCM	29352091
CD8+ memory T cell	**EOMES**, **BACH2**, KLF7, **MYC**, **ID3**	CEBPB, **GATA3**	4,6,11,13	31754020
Memory B cell	KLF13, LMO4, PCBD1, KLF10, ZBTB38	IRF8, SPI1, CEBPB	Memory B cell	31968262
(3) Phenotype conversion				
Primed mESC 1	**LIN28A**, MYC, ID1, FOXP1, ID3	**POU5F1**, **ESRRB**, **KLF4**	FBSLIF	25471879
Naive mESC 1	ZFHX2, MEIS2, ZIC2	**POU5F1**, **ESRRB**, **KLF4**	2iLIF	
Primed mESC 2	**LIN28A**, FOXP1, SOX4	**SOX2**, **POU5F1**, **KLF4**	mES_lif	26431182
Naive mESC 2	SPIC, MITF, MEIS2	**ESRRB**, **KLF4**, **POU5F1**	mES_2i	
Active NSC	CENPS, **EGR1**, INSM1, MXD3, **E2F1**	**ASCL1**, **SOX2**, PBX1	All young aNSCs	30827680
Quiescent NSC	DBP, EPAS1, **ID2**	**FOXG1**, PBX1, **ASCL1**	All young qNSCs	
Fetal hepatocyte	ZGPAT, KLF11, ZBTB20	**GATA4**, **HNF4A**, **CEBPA**	Fetal hepatocyte	30500538
Organoid hepatocyte	HES6, LEF1, THAP8, SOX9, HTT	**FOXA2**, **HNF4A**, MEIS1	Fetal hepatocyte organoid	
Adult hepatocyte 1	KLF9, CEBPD, KLF6	**FOXA2**, **HNF4A**, CEBPB	Hepatocyte	31292543
Adult hepatocyte 2	SCAND1, NR3C1, EDF1	**HNF4A**, **FOXA2**, PBX1	Hepatocyte	30348985
Adult excitatory neuron	MLXIPL, PEG3, HLF, BHLHE40, KLF9	**FOXG1**, CEBPB, PBX1	adult_Ex	31619793
Organoid excitatory neuron	**NEUROG2**, SOX11, SOX4, CSRP2, CARHSP1	**FOXG1**, PBX1	hOrga_EN	
Adult inhibitory neuron	PEG3, MLXIPL, HLF, PPARGC1A, KLF9	**FOXG1**, **SOX2**, PBX1	adult_In	31619793
Organoid inhibitory neuron	SIX3, PAX6, ID4, KLF10, MEIS2	**ASCL1**, **SOX2**, SOX9	hOrga_IN	

Note: Experimentally validated conversion TFs are marked in bold. TFs are ordered from left to right by fold change to MEF/HFF. Cluster IDs annotated to same cell types in PanglaoDB were merged prior to analysis. Macrophage data from different tissues (heart, kidney, lung, muscle, brain, pancreas, skin spleen, trachea) were merged. See Table S3 for literature evidence for predicted conversion TFs.

2.2 | Application to various cell conversions

To demonstrate the applicability of TransSynW, we applied it to different cell systems, which encompassed conversions into broad cell types, subtypes, and phenotypic states (Tables 1 and S3). For example, in the first category, FOXA2, FOXA3, and HNF4A were predicted for the hepatocyte, which, together with HNF1A predicted in the specific part, are known for hepatocyte conversion.[2] The predicted TFs for the beta cells included NKX6-1, MAFA, PDX1, and NEUROD1, which have been shown to induce beta cell conversion.[3-5] Moreover, in both cases the predicted marker genes recapitulated commonly used ones (Tables 2 and S4). Indeed, many predicted conversion TFs are known to regulate each other and the predicted marker genes (Figure 2A,B), supporting the biological relevance of synergistic interactions captured by TransSynW.

Next, we analyzed different subtypes of neurons, as they are one of the most well studied subtypes. Among the predicted TFs for dopaminergic (DA) neurons, MYT1L, ASCL1, FOXA2, and GATA3 have

TABLE 2 Predicted marker genes with documented evidence

Cell type	Predicted marker gene with evidence	Reference (PubMed ID or website)
(1) Conversion into broad cell type		
Myoblast	CALCR, FGFR4, DES, ANKRD1, FITM1	12223412, 26440893, 26492245, 24644428, 8120103
Keratinocyte	KRT5	22028850
Cardiomyocyte	NPPA, MYH6	27123009, https://www.rndsystems.com/cn/research-area/cardiac-stem-cell-markers
Hepatocyte	SRD5A2, FGF21	25974403, 28515909
HSC	ESAM, LHCGR, SLC22A3, TIE1, ANGPT1, RBP1	https://www.rndsystems.com/cn/research-area/hematopoietic-stem-cell-markers 27365425, 27225119
Neuron	HTR2C, NTNG1, HS6ST3	30078709
Oligodendrocyte/OPC	MAG, CLDN11, PLEKHH1, ASPA, TRF	29024657
Macrophage	FOLR2, F13A1, LYZ2, PF4, MGL2, MMP13, CLEC10A	28576768, 29622724, 25477711,
Beta cell	INS1, INS2, G6PC2	22745242, 15133852, 25322827
NSC	NUDC, TUBA1B, TUBA1A	21771589, 29057214, 29281841
(2) Conversion into subtype		
Dopaminergic neuron	ALDH1A1, TH	30096314, http://www.abcam.com/neuroscience/neural-markers-guide
Medial floorplate progenitor	WNT1, MDK	31080111, 24125182, 11750071
GABAergic neuroblast	GAD2	http://www.abcam.com/neuroscience/neural-markers-guide
Oculomotor neuron	PRPH, FGF10, SLIT3, EYA1	24549637, 9221911, 20215354, 31080111
Serotonin neuron	TPH2, SLC6A4	http://www.abcam.com/neuroscience/neural-markers-guide
CD8+ memory T cell	SELL, CXCR5, DRC1	29236683, 18000950, 30243945
Memory B cell	TNFRSF13B, CD27	Company ebioscience, miltenyibiotec
(3) Phenotype conversion		
Primed mESC 1	BMP4	26860365
Active NSC	CENPF	29727663
Quiescent NSC	GJA1	29727663
Fetal hepatocyte	FGB, CYP2E1	28166538, 29622030
Adult hepatocyte 1	CYP3A4	26838674
Adult hepatocyte 2	APOA1	28166538
Adult excitatory neuron	CCK	12815247
Adult inhibitory neuron	CCK, PVALB, CRH	12815247, 2196836, 2843570

Note: See Table S4 for full list of predicted marker genes.

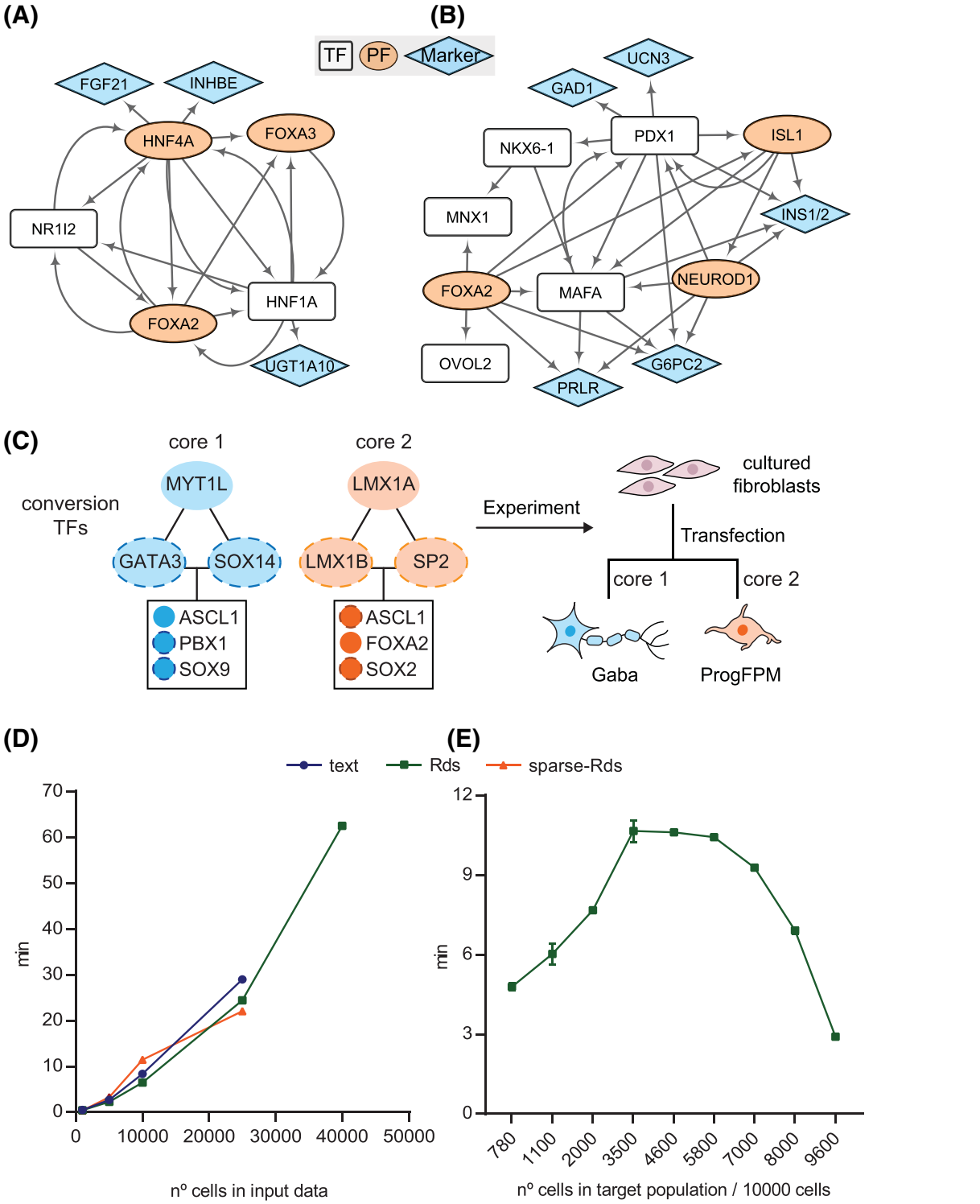

FIGURE 2 Transcriptional regulatory interactions among predicted conversion transcription factors (TFs) and marker genes for, A, hepatocyte and B, beta cell. Interaction data were retrieved from MetaCore from Clarivate Analytic in May/2020. C, Experimental strategy to improve cell conversion protocols for GABAergic neurons (Gaba) and medial floorplate progenitor (ProgFPM) based on TransSynW predicted core TFs. Dashed outlines represent nonvalidated TFs in the literature. D, Processing time vs number of cells in input scRNA-seq file (n = 3). Target population size was fixed to 8% of total size. E, Processing time for Rds files vs number of cells in target population (n = 3). Input population size was fixed to 10 000

been shown to generate DA neurons.[6-8] The predicted TFs for the medial floorplate progenitor, LMX1A and FOXA2, are consistent with the previous attempt to derive this cell subtype.[9] ASCL1 is sufficient to convert fibroblasts into GABAergic neurons.[10] Consistently, the predicted TFs for GABAergic neuroblasts contained ASCL1 and no other TFs known to generate other neuronal subtypes. The predicted

TFs for oculomotor neuron included ISL1, PHOX2A, and PHOX2B which have been reported to generate motor neurons via a synergistic interaction.[11,12] FEV, GATA2, and LMX1B were predicted for serotonergic neurons, which are among the TFs used for deriving this cell subtype.[13] We considered memory T and B cells as subtypes of their naive counterparts. Although a defined set of TFs for generating T cells has not been reported, the nonspecific PFs for both CD4+ and CD8+ T cells contained GATA3 and CEBPB, suggesting that these factors are primary candidates for experimental validation. Indeed, GATA3 is implicated in CD8+ memory T cell conversion.[14] Among the specific TFs, ID3, MYC, BACH2, and EOMES are reported to initiate CD8+ memory T cell conversion.[15-17] The known marker genes, such as SELL and CXCR5, were also identified. Finally, the nonspecific PFs for the memory B cells included IRF8 and SPI1, which together are implicated in the generation of B cell memory.[18]

Another type of cell conversion is phenotypes of a same cell type. The predicted nonspecific PFs for the two mouse embryonic stem cells (mESC) datasets are known to induce pluripotency.[19-21] The specific conversion TFs predicted for both primed mESC populations were LIN28A and FOXP1. LIN28A is known to induce the transition from naive to primed mESCs.[22] FOXP1 is implicated in maintaining pluripotency under non-2i conditions.[23] Whether FOXP1 induces a transition from a naive state to a primed state calls for further investigations. MEIS2 was predicted for both naive mESC populations. Little is known about its role in mESC regulation and hence it constitutes a novel candidate gene. The nonspecific conversion PFs for both active (aNSCs) and quiescent (qNSCs) consisted of known NSC-conversion TFs (eg, ASCL1, SOX2, FOXG1). The specific TFs for aNSCs contained EGR1 known to activate EGFR and accelerate proliferation of NSCs,[24] and E2F1, which is a cell cycle regulator linked to EGFR signaling in NSCs.[25] The conversion TFs for qNSCs included ID2, a BMP effector that has been inferred to regulate qNSCs.[26] Furthermore, CENPF and GJA1 are implicated as markers for late-aNSCs and qNSCs, respectively.[27] Next, the scRNA-seq data of organoid[28] and in vivo hepatocytes[28-30] were analyzed. The nonspecific PFs included general hepatocyte conversion TFs (eg, HNF4A, FOXA2, GATA3). Among the specific TFs for the in vivo hepatocytes were ZBTB20, KLF6, KLF9, CEBPD, and NR3C1. ZBTB20, KLF9 are important for hepatocyte proliferation,[31] whereas KLF6, CEBPD, KLF9, and NR3C1 regulate hepatic glucose and lipid metabolism,[32-34] suggesting that the derivation of in vivo hepatocytes might require sustained cell proliferation and proper metabolization of glucose and lipids. Known hepatocyte marker genes, such as FGB, CYP2E1, CYP3A4, APOA1, were predicted only for the in vivo hepatocytes but none for the in vitro ones. Finally, TransSynW was applied to in vivo and organoid excitatory and inhibitory neurons.[35] TFs predicted only for the in vivo excitatory and inhibitory neurons contained many common TFs (PEG3, KLF9, HLF, and MLXIPL), suggesting a common maturation mechanism. KLF9 is known to be necessary for late-phase maturation of neurons.[36] BHLHE40, which was only predicted for the in vivo excitatory neurons, is implicated in the regulation of neuronal excitability.[37] Moreover, a few known markers (CCK, PVALB, CRH) for excitatory/inhibitory neurons were predicted only for the adult samples. It would be of interest to experimentally test if predicted conversion TFs could indeed convert organoid cells into functional ones.

Taken together, we demonstrated that TransSynW can be effectively applied for identifying conversion TFs for a wide range of cell types. An example experimental strategy for using TransSynW predicted conversion TFs is shown in Figure 2C.

2.3 | Processing speed

The processing speed of TransSynW was assessed using text file, Rds file and a sparse matrix saved as Rds file (sparse-Rds). The time required for the upload of the data was not considered for this analysis. Thus, depending on the users internet connection speed, the overall processing time may vary to a certain degree. Rds files were the most efficient in processing 10 000 cells (6 minutes) (Figure 2D). In addition, up to 40 000 cells were successfully processed with Rds files, whereas only 25 000 cells in the other formats. This is in accordance with the respective file sizes (Table S5). If users wish to use datasets larger than 40 000 cells, we recommend to down-sample them. Next, we benchmarked the execution time against the target cell population size in 10 000 cells. The processing time peaked at 11 minutes for 3500 cells (Figure 2E). Afterwards, it started decreasing due to the reduced size of the background populations. Our general recommendation to users is to use Rds files for datasets with more than 10 000 cells.

3 | DISCUSSION

We have introduced a scRNA-seq based web application, TransSynW, for unbiased identification of cell conversion TFs, following the increasing interest from experimental researchers in generating novel functional cell types identified by scRNA-seq. TransSynW does not require prior biological knowledge, computer programming and computational resources. Moreover, TransSynW identifies potential marker genes for target cell types, which researchers can use for assessing the performance of conversion experiments. Furthermore, prioritization of PFs well recapitulated known conversion TFs in various systems, and predicted novel ones. We foresee that TransSynW will be a valuable tool for the experimental community, particularly for the generation of novel cell populations for stem cell research and regenerative medicine purposes.

4 | MATERIALS AND METHODS

4.1 | Implementation

TransSynW is written in HTML, JavaScript (frontend), PHP and Bash (backend), and runs on a virtual server hosted by Luxembourg Centre of Systems Biomedicine (LCSB, University of Luxembourg).

The frontend allows users to upload all required data, which are then parsed to the backend as different variables. In the backend bash script, the variables are parsed to the TransSynW main R script as different arguments. The output files are compressed into a .zip folder and sent to the user-specified E-mail address.

4.2 | Identification of conversion TFs

The main algorithm is based on the notion that conversion TFs consist of a combination of TFs that are specifically expressed in a target population and TFs that are more broadly expressed in the background population, and that these TFs synergistically interact with each other.[38] The algorithm follows four major steps.

- *Step 1: Identification of candidate TFs*.
  TransSynW first normalizes the data by the total RNA counts. Then TFs whose expression value is 0 across all cells in the target cell population are discarded. Next, it selects top 300 lowest CV (coefficient of variation) TFs as potential candidate TFs, since using more than this number of TFs often resulted in an out-of-memory error during the subsequent computation and conversion TFs usually exhibit low expression variation.
- *Step 2: Identification of most specifically expressed TFs*.
  The set of TFs that are specifically expressed in the target population is determined by Jensen-Shannon Divergence (JSD). JSD is computed for each TF in each cell and the summed JSD value for each TF over all cells is calculated. The top 10 lowest summed-JSD TFs are selected as the most specifically expressed TFs.
- *Step 3: Identification of most synergistic set of specifically expressed TFs*.
  Next, TransSynW identifies the most synergistic subset of TFs among the most specifically expressed TFs by computing MMI.[39]

$$\mathrm{MMI}(S) = -\sum_{T \subseteq S} (-1)^{|T|} H(T),$$

  where $S = \{X_1, X_2, \ldots, X_k\}$, T is a subset of S, $|T|$ denotes the cardinality of T, and H is Shannon's entropies. Negative MMI values imply a synergistic interaction among the TFs.[39] TransSynW first computes MMI of all sets of three TFs among the most specifically expressed TFs. Then a new TF is added to this set and MMI is computed again. If MMI is synergistic, then the next TF is added to the previous set, and so on. This iteration continues until either MMI no longer shows synergy, or when the maximum core size is reached. Here, the maximum core size was set to five.
- *Step 4: Addition of PFs*.
  The specific TF set from step 3 is extended with the nonspecific part, consisting solely of PFs. Every subset of three PFs is added to the specific part. MMI is computed for each set of all TFs and the most synergistic combination is selected as the final conversion TF set.

The final conversion TFs are ranked by the expression fold change calculated between the target cell population and starting cell population.

4.3 | Identification of marker genes

The marker gene set (Table S6) was collected from the following sources; extracellular proteins and membrane receptors,[40] cytoskeletal genes (http://www.informatics.jax.org/), metabolic genes (https://www.vmh.life/#human/all) and CD markers for immune cells (www.abcam.com/CDmarkers). These genes are relatively easily accessible for experimental validation. TransSynW identifies the top 10 candidate marker genes among this compiled set by computing JSD. Literature evidence for predicted markers were collected either manually or from CellMarker (http://biocc.hrbmu.edu.cn/CellMarker/).

4.4 | PF set

Information on PFs that have previously been reported to be involved in cell conversion protocols was manually collected from literature. The list is available in Table S1.

4.5 | scRNA-seq data of starting cell populations

scRNA-seq data of starting cell types were collected from Cell Ranger, GEO and Array Express databases, log 2 transformed and mean gene expression was calculated and compiled in TransSynW (Table S2).

4.6 | scRNA-seq dataset of target cell populations

scRNA-seq data used in this study were obtained from the following sources.[29-31,35,41-48] For References 43, 48, the reprocessed data were retrieved from PangloaDB,[49] as the cell annotation was more accurate than the original one.

ACKNOWLEDGMENTS

We thank Sybille Barvaux for helping gather evidence for pioneer factors. We thank Ernest Arenas, Igor Cervenka, and other anonymous researchers for giving us valuable feedbacks for developing the web application. M.M.R. is supported by Fonds National de la Recherche Luxembourg (C17/BM/11662681).

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

M.M.R., S.O.: collection and/or assembly of data, data analysis and interpretation, manuscript writing; A.d.S.: conception and design, manuscript writing, final approval of manuscript.

DATA AVAILABILITY STATEMENT

TransSynW web application is available at https://transsynw.lcsb.uni.lu/. The code repository is available at https://git-r3lab.uni.lu/mariana.ribeiro/transsynw.

ORCID

Antonio del Sol https://orcid.org/0000-0002-9926-617X

REFERENCES

1. Colasante G, Rubio A, Massimino L, Broccoli V. Direct neuronal reprogramming reveals unknown functions for known transcription factors. *Front Neurosci*. 2019;13:283-290.
2. Huang P, He Z, Ji S, et al. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature*. 2011;475:386-391.
3. Gefen-Halevi S, Rachmut IH, Molakandov K, et al. NKX6.1 promotes PDX-1-induced liver to pancreatic β-cells reprogramming. *Cell Reprogram*. 2010;12:655-664.
4. Guo QS, Zhu MY, Wang L, et al. Combined transfection of the three transcriptional factors, PDX-1, neuroD1, and MafA, causes differentiation of bone marrow mesenchymal stem cells into insulin-producing cells. *Exp Diabetes Res*. 2012;2012:672013-672023.
5. Wang L, Huang Y, Guo Q, et al. Differentiation of iPSCs into insulin-producing cells via adenoviral transfection of PDX-1, NeuroD1 and MafA. *Diabetes Res Clin Pract*. 2014;104:383-392.
6. Hong SJ, Choi HJ, Hong S, Huh Y, Chae H, Kim KS. Transcription factor GATA-3 regulates the transcriptional activity of dopamine β-hydroxylase by interacting with Sp1 and AP4. *Neurochem Res*. 2008;33:1821-1831.
7. Pfisterer U, Kirkeby A, Torper O, et al. Direct conversion of human fibroblasts to dopaminergic neurons. *Proc Natl Acad Sci U S A*. 2011;108:10343-10348.
8. Seok JH, Huh Y, Chae H, Hong S, Lardaro T, Kim KS. GATA-3 regulates the transcriptional activity of tyrosine hydroxylase by interacting with CREB. *J Neurochem*. 2006;98:773-781.
9. Okawa S, Saltó C, Ravichandran S, et al. Transcriptional synergy as an emergent property defining cell subpopulation identity enables population shift. *Nat Commun*. 2018;9:1-10.
10. Chanda S, Ang CE, Davila J, et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. *Stem Cell Rep*. 2014;3:282-296.
11. Mazzoni EO, Mahony S, Closser M, et al. Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat Neurosci*. 2013;16:1219-1227.
12. Mong J, Panman L, Alekseenko Z, et al. Transcription factor-induced lineage programming of noradrenaline and motor neurons from embryonic stem cells. *Stem Cells*. 2014;32:609-622.
13. Vadodaria KC, Mertens J, Paquola A, et al. Generation of functional human serotonergic neurons from fibroblasts. *Mol Psychiatry*. 2016;21:49-61.
14. Wang Y, Misumi I, Di Gu A, et al. GATA-3 controls the maintenance and proliferation of T cells downstream of TCR and cytokine signaling. *Nat Immunol*. 2013;14:714-722.
15. Istaces N, Splittgerber M, Lima Silva V, et al. EOMES interacts with RUNX3 and BRG1 to promote innate memory cell formation through epigenetic reprogramming. *Nat Commun*. 2019;10:3306.
16. Ji Y, Pos Z, Rao M, et al. Repression of the DNA-binding inhibitor Id3 by Blimp-1 limits the formation of memory CD8 + T cells. *Nat Immunol*. 2011;12:1230-1237.
17. Roychoudhuri R, Clever D, Li P, et al. BACH2 regulates CD8 + T cell differentiation by controlling access of AP-1 factors to enhancers. *Nat Immunol*. 2016;17:851-860.
18. Carotta S, Willis SN, Hasbold J, et al. The transcription factors IRF8 and PU.1 negatively regulate plasma cell differentiation. *J Exp Med*. 2014;211:2169-2181.
19. Feng B, Jiang J, Kraus P, et al. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol*. 2009;11:197-203.
20. Hester ME, Song SW, Miranda CJ, Eagle A, Schwartz PH, Kaspar BK. Two factor reprogramming of human neural stem cells into pluripotency. *PLoS One*. 2009;4:e7044.
21. Wernig M, Meissner A, Cassady JP, Jaenisch R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. *Cell Stem Cell*. 2008;2:10-12.
22. Zhang J, Ratanasirintrawoot S, Chandrasekaran S, et al. LIN28 regulates stem cell metabolism and conversion to primed pluripotency. *Cell Stem Cell*. 2016;19:66-80.
23. Gabut M, Samavarchi-Tehrani P, Wang X, et al. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*. 2011;147:132-146.
24. Alagappan D, Balan M, Jiang Y, Cohen RB, Kotenko SV, Levison SW. Egr-1 is a critical regulator of EGF-receptor-mediated expansion of subventricular zone neural stem cells and progenitors during recovery from hypoxia-hypoglycemia. *ASN Neuro*. 2013;5:183-193.
25. Morizur L, Chicheportiche A, Gauthier LR, Daynac M, Boussin FD, Mouthon MA. Distinct molecular signatures of quiescent and activated adult neural stem cells reveal specific interactions with their microenvironment. *Stem Cell Rep*. 2018;11:565-577.
26. Llorens-Bobadilla E, Zhao S, Baser A, Saiz-Castro G, Zwadlo K, Martin-Villalba A. Single-cell transcriptomics reveals a population of dormant neural stem cells that become activated upon brain injury. *Cell Stem Cell*. 2015;17:329-340.
27. Shah PT, Stratton JA, Stykel MG, et al. Single-cell transcriptomics and fate mapping of ependymal cells reveals an absence of neural stem cell function. *Cell*. 2018;173:1045-1057.e9.
28. Hu H, Gehart H, Artegiani B, et al. Long-term expansion of functional mouse and human hepatocytes as 3D organoids. *Cell*. 2018;175:1591-1606.e19.
29. Aizarani N, Saviano A, Sagar LM, et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature*. 2019;572:199-204.
30. MacParland SA, Liu JC, Ma XZ, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun*. 2018;9:1-21.
31. Cvoro A, Devito L, Milton FA, et al. A thyroid hormone receptor/KLF9 axis in human hepatocytes and pluripotent stem cells. *Stem Cells*. 2015;33:416-428.
32. Bechmann LP, Vetter D, Ishida J, et al. Post-transcriptional activation of PPAR alpha by KLF6 in hepatic steatosis. *J Hepatol*. 2013;58:1000-1006.
33. Lai PH, Wang WL, Ko CY, et al. HDAC1/HDAC3 modulates PPARG2 transcription through the sumoylated CEBPD in hepatic lipogenesis. *Biochim Biophys Acta—Mol Cell Res*. 2008;1783:1803-1814.
34. Pei H, Yao Y, Yang Y, Liao K, Wu JR. Krüppel-like factor KLF9 regulates PPARγ transactivation at the middle stage of adipogenesis. *Cell Death Differ*. 2011;18:315-327.
35. Kanton S, Boyle MJ, He Z, et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature*. 2019;574:418-422.
36. Scobie KN, Hall BJ, Wilke SA, et al. Krüppel-like factor 9 is necessary for late-phase neuronal maturation in the developing dentate gyrus and during adult hippocampal neurogenesis. *J Neurosci*. 2009;29:9875-9887.
37. Hamilton KA, Wang Y, Raefsky SM, et al. Mice lacking the transcriptional regulator Bhlhe40 have enhanced neuronal excitability and impaired synaptic plasticity in the hippocampus. *PLoS One*. 2018;13:1-22.
38. Okawa S, Del Sol A. A general computational approach to predicting synergistic transcriptional cores that determine cell subpopulation identities. *Nucleic Acids Res*. 2019;47:3333-3343.
39. Bell AJ. The co-information lattice. In: Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation; 2003. p. 921. http://www.kecl.ntt.co.jp/icl/signal/ica2003/cdrom/data/0187.pdf.

40. Ramilowski JA, Goldberg T, Harshbarger J, et al. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun*. 2015;6:7866.
41. Horns F, Dekker CL, Quake SR. Memory B cell activation, broad anti-influenza antibodies, and bystander activation revealed by single-cell transcriptomics. *Cell Rep*. 2020;30:905-913.e6.
42. Kalamakis G, Brüne D, Ravichandran S, et al. Quiescence modulates stem cell maintenance and regenerative capacity in the aging brain. *Cell*. 2019;176:1407-1419.e14.
43. Kimmel JC, Penland L, Rubinstein ND, Hendrickson DG, Kelley DR, Rosenthal AZ. Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. *Genome Res*. 2019;29:2088-2103.
44. Kolodziejczyk AA, Kim JK, Tsang JCH, et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell*. 2015;17:471-485.
45. Kumar RM, Cahan P, Shalek AK, et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*. 2014;516:56-61.
46. La Manno G, Gyllborg D, Codeluppi S, et al. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell*. 2016;167:566-580.e19.
47. Patil VS, Madrigal A, Schmiedel BJ, et al. Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci Immunol*. 2018;3:1-14.
48. Schaum N, Karkanias J, Neff NF, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. 2018;562:367-372.
49. Franzén O, Gan LM, Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*. 2019;2019:1-9.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ribeiro MM, Okawa S, del Sol A. TransSynW: A single-cell RNA-sequencing based web application to guide cell conversion experiments. *STEM CELLS Transl Med*. 2021;10:230–238. https://doi.org/10.1002/sctm.20-0227