



Integrative and perturbation-based analysis of the transcriptional dynamics of TGF β /BMP system components in transition from embryonic stem cells to neural progenitors

Ruben Dries^{1,2}  | Agata Stryjewska² | Kathleen Coddens² | Satoshi Okawa^{3,4} |
Tineke Notelaers² | Judith Birkhoff¹ | Mike Dekker¹ | Catherine M. Verfaillie² |
Antonio del Sol^{3,5,6} | Eskeatnaf Mulugeta¹ | Andrea Conidi^{1,2} |
Frank G. Grosveld¹ | Danny Huylebroeck^{1,2} 

¹Department of Cell Biology, Erasmus University Medical Center, Rotterdam, The Netherlands

²Department of Development and Regeneration, KU Leuven, Leuven, Belgium

³Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg

⁴Integrated BioBank of Luxembourg, Dudelange, Luxembourg

⁵CIC bioGUNE, Bizkaia Technology Park, Derio, Spain

⁶IKERBASQUE, Basque, Foundation for Science, Bilbao, Spain

Correspondence

Danny Huylebroeck, PhD, Department of Cell Biology, Erasmus MC, Building Ee - room Ee-1040b, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands.

Email: d.huylebroeck@erasmusmc.nl

Ruben Dries, PhD, Department of Pediatric Oncology, Dana-Farber Cancer Institute, Longwood Center, 360 Longwood Avenue, Boston, MA 02115.

Email: rdries@jimmy.harvard.edu

Present addresses

Ruben Dries, Department of Pediatric Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA.
Agata Stryjewska, Neural Development, Plasticity and Repair, Wolfson Institute for Biomedical Research, University College London, London WC1E 6BT, UK.

Funding information

Agentschap voor Innovatie door Wetenschap en Technologie, Grant/Award Number: SB 93097; Belgian Federal Science Policy Office, Grant/Award Number: IAPVII-07; Erasmus Medisch Centrum, Grant/Award Number: start-up fund; Fonds Wetenschappelijk Onderzoek, Grant/Award Numbers: G.0782.14, GA.09411.10; Onderzoekscraad, KU Leuven, Grant/Award

Abstract

Cooperative actions of extrinsic signals and cell-intrinsic transcription factors alter gene regulatory networks enabling cells to respond appropriately to environmental cues. Signaling by transforming growth factor type β (TGF β) family ligands (eg, bone morphogenetic proteins [BMPs] and Activin/Nodal) exerts cell-type specific and context-dependent transcriptional changes, thereby steering cellular transitions throughout embryogenesis. Little is known about coordinated regulation and transcriptional interplay of the TGF β system. To understand intrafamily transcriptional regulation as part of this system's actions during development, we selected 95 of its components and investigated their mRNA-expression dynamics, gene-gene interactions, and single-cell expression heterogeneity in mouse embryonic stem cells transiting to neural progenitors. Interrogation at 24 hour intervals identified four types of temporal gene transcription profiles that capture all stages, that is, pluripotency, epiblast formation, and neural commitment. Then, between each stage we performed esiRNA-based perturbation of each individual component and documented the effect on steady-state mRNA levels of the remaining 94 components. This exposed an intricate system of multilevel regulation whereby the majority of gene-gene interactions display a marked cell-stage specific behavior. Furthermore, single-cell RNA-profiling at individual stages demonstrated the presence of detailed co-expression modules and subpopulations showing stable co-expression modules such as that of the core pluripotency genes at all stages. Our combinatorial experimental approach demonstrates how intrinsically complex transcriptional regulation within a given pathway is during cell fate/state transitions.

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.

©2019 The Authors. STEM CELLS published by Wiley Periodicals, Inc. on behalf of AlphaMed Press 2019

Number: GOA-11/012; ZonMw

KEYWORDScell state transitions, embryonic stem cells, esiRNA, genetic interaction, neural differentiation, signaling pathway, TGF β /BMP**1 | INTRODUCTION**

Embryonic stem (ES) cells provide an ideal *in vitro* system for recapitulating the cell states and subpopulations in early mammalian embryos. Most studies initially focused on how ES cell pluripotency states are maintained and, more recently, how these cells then transit to next cell states or how somatic cells can be reprogrammed to a pluripotent stem cell state. Mouse (m)ES cells can be kept in the pluripotent ground state by inhibition of FGF signaling and GSK3 activity using two small-molecule compounds (named 2i).¹ Their naïve pluripotent condition, which better resembles the inner cell mass of mouse blastocysts, is obtained by adding BMP4 (R&D systems/Bio-Techne, Abingdon, Oxfordshire, United Kingdom, https://www.rndsystems.com/products/recombinant-human-bmp-4-protein_314-bp) and Leukemia Inhibitory Factor (LIF, Sigma-Aldrich, Overijse, Flemish Brabant, Belgium, <https://www.sigmaaldrich.com/catalog/product/mm/esg1107?lang=en&ion=BE>), making them more prone to differentiation cues and hence to exit from pluripotency in cell culture. Removal of all such exogenously added factors results in differentiation to the neural lineage with reasonably high efficiency.² The transitions herein include first the rapid conversion from mES cells to epiblast-like cells (EpiLCs), resembling primed epiblast stem cells (EpiSCs) from early mouse embryos, and can be accomplished by using growth-factor free N2B27 medium.³ When such cells are then cultivated for a prolonged period of time, they will form neural progenitor (NP) cells with increased, transient *Sox1* expression after 96 hours and, later on, the presence of more differentiated (including neuronal) cells.

TGF β family signals, including bone morphogenetic proteins (BMPs) and Activin/Nodal, exert cell-type specific, context-dependent effects, which sometimes are opposite.^{4,5} Signaling occurs through ligand-activated receptors, initiating receptor-activated phospho-Smad (Figure 1A) and non-Smad (kinase-)driven cascades.⁶⁻⁹ It requires strict control and fine-tuning in various ways, including by a multitude of Smad-interacting proteins, for example, transcription factors (TFs).¹⁰⁻¹³ The controls involve transcriptional autoregulation and synexpression, as well as feed-forward or feedback mechanisms, acting on the transcription of the signaling system components.¹⁴⁻¹⁹ Altogether, this mounts the appropriate transcriptional response in target cells,²⁰⁻²² which is crucial for embryogenesis, tissue/organ formation and, in the adult, repair after injury.²³⁻²⁷ Despite significant efforts in identifying and experimentally addressing these regulatory mechanisms, an integrated understanding of the transcriptional dynamics of the TGF β -family system, and their gene-gene interactions, as a whole is not readily available. The often used knockout or knockdown approach of just one system component causes changes in expression of multiple other system components, suggesting compensatory mechanisms that will (co-)determine the

Significance statement

Signaling pathways play pivotal roles during embryogenesis. Extrinsic signaling is transmitted through intracellular proteins that ultimately converge on the transcription regulatory network, which steers cell fate and differentiation. To study the exact role of such top-down pathways during development, researchers have historically used single-component perturbations with readout time points often covering multiple developmental stages. By systematic profiling of multiple components of the transforming growth factor type β /bone morphogenetic protein pathway using time series, perturbations, and single-cell analysis, this study shows that this pathway should be considered an intricate intradependent network of individual components and considerably cell-stage specific. Hence, interpretation of the consequences of single-gene perturbation or knockout in lineage-progressing cells, including for optimizing *ex vivo* stem/progenitor cell differentiation, should occur with caution regarding stage and transition.

phenotype. Appropriate examples are the knockout of *Smad1/5* in mouse embryonic urogenital mesenchyme, resulting in adult gonadal tumors that overproduce Activin/Inhibin and display phospho-Smad2/3 hyperactivation,^{28,29} and ectopic anterior Nodal mRNA expression in *Smad5*-knockout early post-implantation embryos.³⁰ Such genetic perturbations have proven valuable to dissect the developmental role of TGF β -system genes, sometimes including their epistasis, both *in vitro* (in ES cells,^{31,32}) and *in vivo* (mouse embryos,^{30,33}).

Transcriptional regulatory network (TRN) assembly using genome-wide expression data from cells grown and collected in bulk is extensively used to obtain a system's biology type of insight,³⁴ while recent technologies paved the way to perform single-cell analysis combined with CRISPR-based perturbations. However, it remains challenging to determine an unequivocal and direct causal link between the initial perturbation and the observed phenotype(s) over a long period of time. Secondary effects can accumulate rapidly and can lead to inaccurate understanding how each signaling and/or network component is connected. We anticipated to overcome some of these limitations and aimed for a highly focused, but comprehensive and integrative analysis of the TRN underlying the control of mRNA levels of TGF β system components, thereby governing both cell fate(s) and the consecutive transition(s) between cell stages.

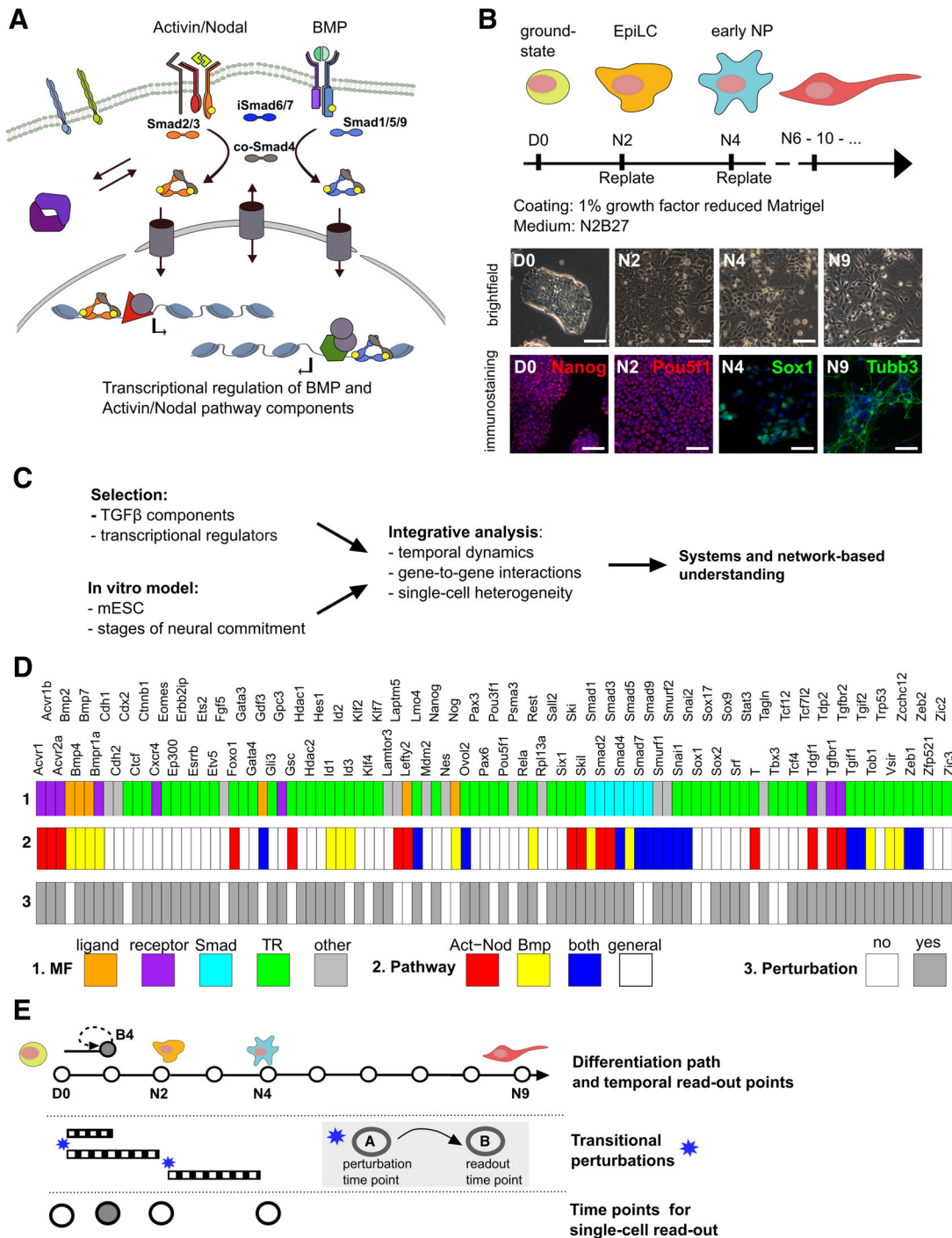


FIGURE 1 A systematic approach to deconstruct dynamic and centered transcriptional networks. A, Simplified overview of Activin/Nodal and bone morphogenetic protein (BMP) ligand receptor binding and Smad-dependent signaling. The latter is further divided in inhibitory Smads and receptor-activated Smads that can form a trimeric complex with the common-mediator Smad Smad4. These activated Smads accumulate in a time and amplitude specific manner in the nucleus. Here, they can team up with cofactors and TFs to stimulate or repress target genes in a chromatin modulated context, and often including genes encoding system components itself. Non-Smad, kinase-driven cascades are not shown. B, Cell states in our modified and optimized embryonic stem (ES) cell system. Brightfield and immunostaining pictures are representative and photographed at same magnification, but not taken from the same cultures. Scale bar: 50 μ m. C, Systems approach consisting of four phases: (a) selection of pathway centric components and critical regulators, (b) establishment of a robust in vitro differentiation system amenable for high-throughput experiments, (c) multilevel and integrative inference, and (d) downstream coupling of inferred network to biological system. D, Overview of selected TGF β -superfamily centric genes and custom annotation for 1. their molecular function (MF), 2. pathway specificity, and 3. perturbation success. E, Diagram visualizing the three levels of transcriptional inference that were assessed in our optimized neural differentiation system. Black and white striped flat rectangles illustrate transition based perturbations, that is, perturbation in one cell state and readout in a subsequent cell state. The gray colored circle is the B4 stage, that is, in BMP4 + LIF

2 | MATERIALS AND METHODS

2.1 | Cell cultures

Sox1-GFP mES cells (A. Smith, Cambridge, UK) were maintained feeder-free on 0.1% gelatin-coated plates in 2i + LIF medium (N2B27 with 1 μ M PD0325901 (Axon Medchem, Groningen, The Netherlands, <https://www.axonmedchem.com/product/1408>), 3 μ M CHIR99021 (Axon Medchem, Groningen, The Netherlands, <https://www.axonmedchem.com/product/1386>), 1000 U LIF/mL (Millipore, ESG1107). mES cells were passaged at high density (10^5 cells/cm²) for at least three passages before the experiments started. Naïve mES cells were obtained by transfer from 2i + LIF medium to N2B27 + LIF (1000 U/mL) + BMP4 (10 ng/mL; R&D Systems).

2.2 | Induction of early NP cells

mES cells were differentiated to NP cells,² but this was optimized for robustness and reproducibility: for neural differentiation the cells were transferred from 2i + LIF to N2B27 on 1% growth-factor reduced (GFR) Matrigel (hESC-qualified, Becton Dickinson, Amsterdam, The Netherlands, <https://ecatalog.coming.com/life-sciences/b2c/NL/en/Surfaces/Extracellular-Matrices-ECMs/Corning%C2%AE-Matrigel%C2%AE-Matrix/p/354230>)-coated 6-well plates (3×10^4 cells/cm²). At day 2, after 48 hours in N2B27, EpiLCs were generated; these were dissociated with Accutase (Thermo Fisher Scientific, Merelbeke, East Flanders, Belgium, <https://www.fishersci.be/be/en/catalog/search/products?keyword=stempro+accutase>) and replated on freshly 1% GFR Matrigel-coated 12-well plates at 4×10^4 cells/cm² in N2B27. After another 48 hours, hence 96 hours in N2B27 in total, these day 4 cells show high Sox1(-driven GFP) levels and are considered early NP cells. To obtain neural cells, these NP cells are replated on freshly 1% GFR Matrigel-coated 12-well plates at 6×10^4 cells/cm² in N2B27 + 10 μ M Rock inhibitor (Y-27632, Stemcell technologies, Cologne, North Rhine-Westphalia, Germany, <https://www.stemcell.com/products/y-27632.html>) for 24 hours. Accutase combined with gentle repetitive pipetting was used to obtain near-single-cell suspensions.

2.3 | Sox1-GFP quantification and indirect immunofluorescence

Sox1-GFP+ cells were measured at day 0, 2, 4, and 10 using the Aria II platform and normalized to day 0 (Figure S1C). For Figure S1B, cells were fixed for 10 minutes with ice-cold paraformaldehyde and blocked for 30 minutes at 24°C with 0.1% Triton X100-1% BSA in PBS. Anti-Oct4 (ab19857, Abcam, Cambridge, Cambridgeshire, United Kingdom, <https://www.abcam.com/oct4-antibody-chip-grade-ab19857.html>), anti-Nanog (ab80892, Abcam, Cambridge, Cambridgeshire, United Kingdom, <https://www.abcam.com/nanog-antibody-ab80892.html>), anti-Tubb3 (ab78078, Abcam, Cambridge, Cambridgeshire, United

Kingdom, <https://www.abcam.com/beta-iii-tubulin-antibody-2g10-neuronal-marker-ab78078.html>), and anti-GFP (ab13970, Abcam, Cambridge, Cambridgeshire, United Kingdom, <https://www.abcam.com/gfp-antibody-ab13970.html>) (all at 1:1000) were used, with DAPI as nuclear counterstain (Life Technologies, D1306).

2.4 | Selection of TGF β -centric network genes

The comprehensive, manually curated gene list ($n = 118$) represented all layers of the signal transduction cascade. To identify and enrich for TFs two available microarray data sets (GSE11523 and GSE29005) were used, for they were generated using similar cells.^{5,35} We performed hierarchical clustering using Pearson correlation scores and average linkage to determine the states and transitions. Differentially expressed genes (DEGs) were called using a SAM-test with fold-change >2 and P -value $<.05$. DEGs were filtered based on GO terms (0003677, 0006355, 0003700, 0006350, 0045449, and 0043565) or presence in AnimalTFDB (<http://www.bioguo.org/AnimalTFDB/>). Remaining genes were prioritized via ToppGene and training parameters *Pathway*, *Pubmed*, and *Interaction*. The curated and filtered gene lists were then combined and a final selection was based on qPCR expression level statistics, considering also detection level, primer efficiency, and variability.

2.5 | esiRNA-mediated perturbation

At day 0: mES cells (10^5 cells/cm²) in 2i + LIF were reverse-transfected with a mix of 200 ng esiRNA (100 ng/ μ L, Sigma) and 2 μ L PowerFect (Tebu-Bio, Boechout, Antwerp, Belgium, https://www.tebu-bio.com/Product/189SL100569-0.1ml/POWERFECT_TM_SIRNA_TRANSFECTION_REAGENT.html) in 110 μ L buffer solution/well of a 0.1% gelatin-coated 12-well plate. After 6 hours at 37°C, the cells were replated in either N2B27 (for differentiation) or N2B27 + LIF + BMP4.

At day 2: EpiLCs (6×10^4 cells/cm²) were reverse-transfected with 100 ng (100 ng/ μ L) esiRNA and 1 μ L PowerFect in 55 μ L buffer per well of a GFR Matrigel coated 24-well plate. After 6 hours at 37°C, the cells were replated in N2B27 medium.

In both transfection conditions, the cells were lysed and RNA was harvested 48 hours post-transfection, that is, as BMP4 + LIF, EpiLC, or NP cell state. In addition, for each sample the endogenous mRNA levels 6 hours post-transfection were analyzed by RT-qPCR. A knockdown is scored sufficient if at least a 50% reduction of the transcript was observed.

2.6 | High-throughput RT-qPCR

RNA was extracted and processed with the RNeasy Micro Kit (Qiagen, Antwerp, Antwerp, Belgium, <https://www.qiagen.com/us/products/diagnostics-and-clinical-research/sample-processing/rneasy-micro-kit/>)

#orderinginformation). For the higher throughput esiRNA-based screening, the Total RNA Purification 96-Well Kit (Norgen, Thorold, ON, Canada, <https://norgenbiotech.com/product/total-rna-purification-96-ell-kit>) was used. RNA was quantified with Nanodrop (Nanodrop Technologies). mRNA was converted into cDNA using SuperScript III First-Strand Synthesis SuperMix (ThermoFisher, 18080400). Primer sequences were obtained from PrimerBank (<http://pga.mgh.harvard.edu/primerbank/>); if not available here, they were designed according to PrimerBank set of rules with addition of an intron-exon boundary. All primers were validated for efficiency, specificity, and dynamic range at the same annealing temperature (60°C) to ensure accurate parallel amplification (Supplementary Table S2, PCR primers). Primers (15 µL/well, at 10 µM) and perturbation samples (12.5 µL/well at 5 ng/µL) were separately stored frozen in 96-well plates before shipping on dry ice to the TATAA Biocenter (Prague) for high-throughput quantitative PCR (HT-qPCR) on the Biomark-HD platform according to Fluidigm's instructions for SYBR-green-based quantification. Each sample plate contained at least three control wells, that is, a no-template control, gDNA control, and total reference cDNA to account for confounding technical properties during downstream processing (see Supplemental Data).

2.7 | Temporal expression analysis and cluster determination

Time series data were clustered based on spearman correlation distance and gene expression dimensions were reduced with principal component analysis (PCA) on scaled and centered data. To determine dynamic gene expression clusters, the data were first converted to z-scores and stable genes were assigned to one cluster based on the lowest expression range quartile. Next, k-means using three centers was used to create additional clusters.

2.8 | Identification of DEGs in perturbation samples

Significant DEGs between target and Renilla-luciferase negative control perturbation samples were determined with the Limma package in R using a false discovery rate of 10% and a minimal absolute log-fold change of 0.5. Probable off-target effects were flagged and removed if a DEG was called in all perturbation samples per 96*96 Biomark chip.

2.9 | Association between gene expression range and perturbation effect

A simple linear model between range of gene expression and perturbation effect (gene is perturbed or gene as perturbator) was fitted for all genes using the `lm` function in base R.

3 | RESULTS

3.1 | Inferring a dynamic, pathway-centric transcriptional network

Gene transcription displays intrinsic stochasticity and susceptibility to small extracellular changes, and both may lead to variation that might hinder correct interpretation of transcriptional readouts of genes when inferring a dynamic TRN during cell differentiation. Here, we used ES cells whose culture medium and extrinsic factors therein can be strictly controlled and when optimized they provide robust cell cultures suitable for gene/protein perturbation studies. To reduce the overall nested coefficient of variation, associated with both cellular differentiation and subsequent sample processing, we introduced several small modifications to a monolayer neural differentiation protocol² that improved the overall efficiency and reproducibility (Figure 1B). This was a prerequisite for our systems-level approach that covers different consecutive cell stages and requires a systematic high-throughput esiRNA-based perturbation approach combined with a qPCR readout (Figure S1A,B; see Materials and Methods section; see Supplemental Information).

To start, we curated a list of TGFβ-system components and cell-stage-related TFs. First, we used an ensemble approach by aggregating the results of manual literature curation with transcriptome analysis of neural differentiating mES cells and context-driven prioritization (Figure 1C and Figure S1C). The latter was based on known genetic/biochemical interactions, involvement in human disease and/or known for causing phenotypes in knockout mice. Next, we ensured that the chosen set of TFs encompassed all stages/transitions in our ES cell system, that is, the ground (2i + LIF, denoted by D0) and naïve state of pluripotency (48 hours, BMP4 + LIF, B4), Epi-LCs (48 hours, N2B27, N2), and early NP cells (96 hours, N2B27, N4), respectively. The final list covers genes representing the entire cascade of Smad-mediated signaling, including TFs, and has near-equal balance between BMP and Activin/Nodal components (Figure 1D). We could successfully perturb the majority of these components (ie, 73 out of 95), as determined by a loss of at least 50% of the initial steady-state transcript levels at 6 hours post-transfection of esiRNA.^{36,37}

To document the transcriptional dynamics of the selected genes, we used high-throughput RT-qPCR (see Materials and Methods section) and profiled the changes in three complementary ways. The three variables that were considered to influence gene expression were single-cell heterogeneity, differentiation time and epistasis or environmental gene interactions (Figure 1E and Figure S1D). First, to assess changes over time, we performed transcriptional profiling at 24 hours intervals (plus B4, ie, Bmp4 + LIF, naïve state of pluripotency) in unperturbed differentiation conditions. Next, to infer gene-gene interactions during differentiation, we performed transitional perturbations such that perturbation and readout occur at consecutive cell stages. For this, we independently perturbed the ES cells at two different time points, that is, in D0 and N2 state (Figure 1E), respectively, and allowed their transit to the subsequent state/s. For the first time point, this is a

transition toward either B4 or N2 cells; for the second it was toward N4. Altogether, for each gene that was responsive to esiRNA-based perturbation, as determined by our previously defined minimal threshold, we collected three transition perturbation samples (ie, D0→B4,

D0→N2, and N2→N4). Finally, we performed single-cell profiling in unperturbed samples at both the start and end points (D0, B4, N2, and N4) of our perturbation experiments to examine the role of transcriptional heterogeneity at any of these given time points.

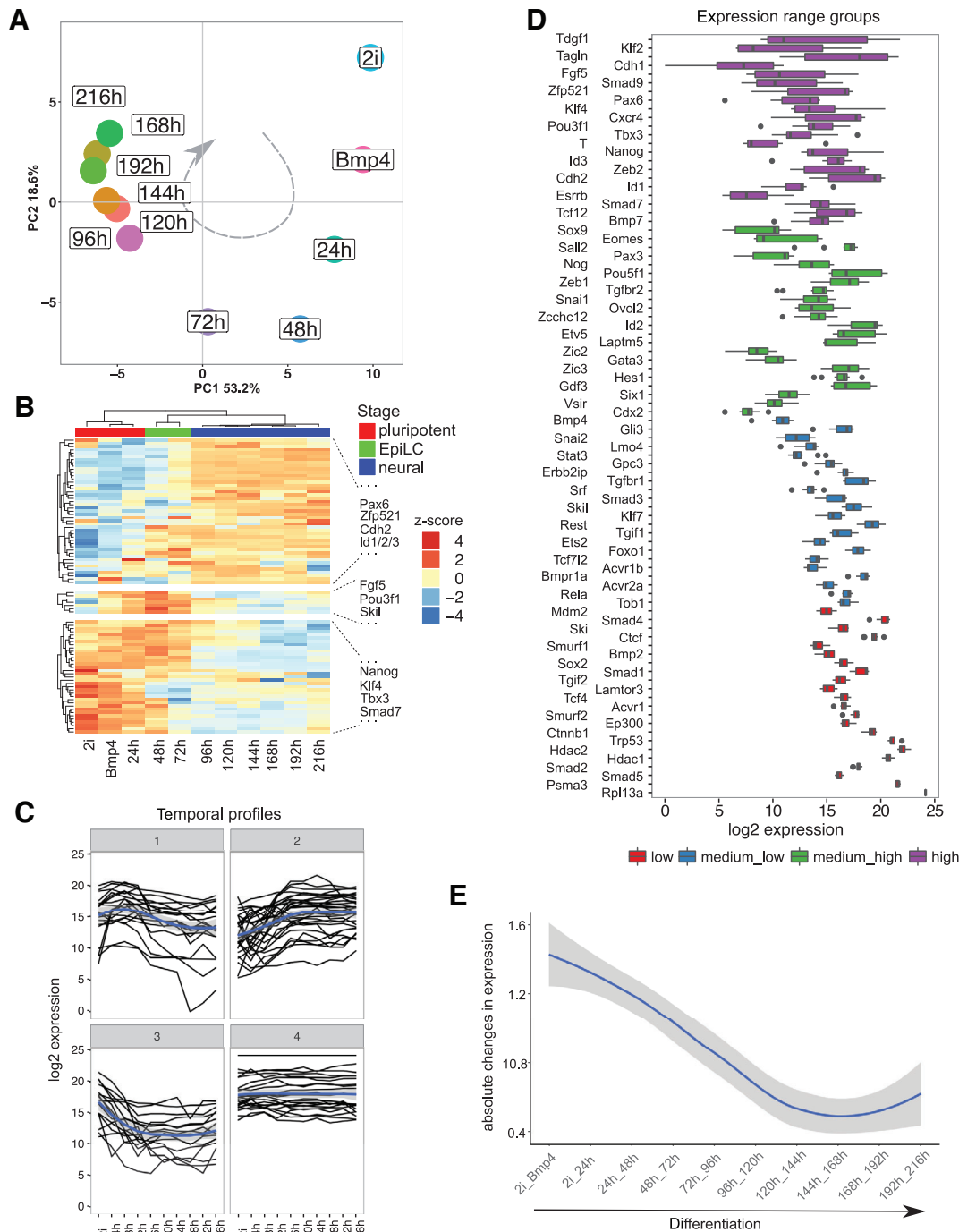


FIGURE 2 Transcriptional dynamics during neural commitment. A, Principal component analysis of 24 hour time interval samples during neural differentiation. Gray arrow indicates direction of differentiation. B, Heatmap showing genes (rows) and samples (columns) using correlation distances. Cell stages are indicated by the colors on top and established marker genes are shown for each cell stage. Heatmap colors represent standardized expression values (z-scores). C, K-means (k = 4) cluster profiles of standardized gene expression values over time. A loess regression curve to show the average profile for each group is shown in blue. D, Boxplots showing the expression range of individual genes during differentiation. Colors represent range-specific subsets based on the quartile values of the overall range distribution. E, Loess regression curve (blue) with confidence band (gray area) to depict the collective transcriptional changes between consecutive time intervals

3.2 | Temporal analyses confirm stage and illustrate subgroup-specific profiles

During differentiation, cells need to rewire their TRN to up or down-regulate the right sets of genes for progressing to the next stage. Despite the continuous nature of differentiation, individual cells can be found enriched for certain states in vivo and/or in cell culture. This can span multiple hours or even days wherein these cells display specific mRNA profiles, morphologic features and fate potential.^{31,38-41} We used a 24-hour time interval to profile mRNA transcript level changes of the genes in our list. PCA showed that time follows the first principal component (PC1), which already explains 53% of the observed variability between genes. Both ground and naïve states of pluripotency were in close proximity, and furthest apart from NP cells (Figure 2A). Hierarchical clustering verified that between the naïve pluripotent (*Nanog*, *Klf4*, *Tbx3* high) and NP (*Pax6*, *Zfp521*, *Cdh2* high) states, a transient population is observed with high *Fgf5* and *Pou3f1* expression resembling EpiLCs (Figure 2B).⁴⁰

To group genes by temporal profile, we performed K-means clustering on standardized expression changes of all samples (except for B4) and used four clusters to identify coarse, robust temporal profile types (Figure 2C and Table S1). The first cluster consisted of genes such as *Pou5f1*, *Fgf5*, and *Pou3f1* and they are moderately or highly induced at the onset of differentiation, usually display high levels in EpiLCs, and are then downregulated when entering neural fate.^{40,42} The second cluster contained almost all neural-related genes (eg, *Pax6*, *Sox9*, *Zeb2*, and *Zfp521*) which gradually increased their mRNA levels.⁴³ To our surprise, this cluster also contained *Id* genes, which are acknowledged downstream targets of *Smad1/5* in BMP-stimulated cells and known to inhibit differentiation, for example, neurogenesis.^{44,45} Genes of the third cluster showed rapid down-regulation and contained most genes related to the naïve core pluripotency network, including not only *Nanog*, *Esrrb*, *Klf4*, and *Tbx3*,⁴⁶⁻⁴⁹ but also *Smad7* and the Nodal receptor complex genes *Acrv1b* and *TdGF1* (*Cripto-1*). The fourth cluster was the most stable cluster and included the two housekeeping genes (*Rpl13a* and *PsmA3*).

We also computed the overall expression range for all individual genes and divided them accordingly to their quartiles (Figure 2D and Figure S2A). While the temporal profiles explain the directionality of the gene expression changes, the expression range might be an estimator for transcriptional responsiveness, that is, how likely does a gene need tight control during differentiation. As expected, genes that had a low range of expression were found exclusively in the aforementioned stable cluster (cluster-4). Interestingly, *Smad1/5*, *Smad2*, and *Smad4* all map to this cluster. *Smad7* and *Smad9* displayed a much larger expression range and fell in the highest-range group. Genes that were documented as up (cluster-2) or downregulated (cluster-3) were mostly genes that also displayed a larger range (Figure S2B). Finally, we also assessed how transcriptional changes—for all genes as a whole—changed during differentiation by assessing how expression changes occurred over consecutive time intervals (Figure 2E and Figure S2C). These changes were the largest at the beginning of differentiation and they gradually decreased until around day 4 (96-120 hours), which was then followed

by a stabilization. And, even though we selected only a limited number of 95 genes, their dynamics showed resemblance with previous transcriptome studies in ES cell states,⁵ indicating that the applied time windows captured all the important cell transitions. Altogether, these transcriptional changes were particularly dynamic in the first 4 days and thus likely subject to more specific regulation in this period.

3.3 | Transient perturbations demonstrate dynamics and specificity of pairwise gene interactions

To understand how the selected components of our TGFβ-system centric network could influence each other's transcript level and dynamic changes that occur during cell differentiation (Figure 2), we performed systematic transitional perturbations of each component that was responsive to esiRNA-mediated degradation (73 out of 95) and documented the effect on the steady-state mRNA level of the remaining 94 genes (Figure 1E). Of note, esiRNA-mediated perturbations were chosen as they result in significantly lower off-target effects while maintaining the same level of on-target effect.³⁷ Altogether a total of 20 805 putative gene-gene interactions originated from 73 knockdowns and 95 readouts at three transitions.

For each perturbation sample, the knockdown efficacy was verified 6 hours post-transfection and for the majority of the components, we still noticed significantly reduced mRNA levels at 48 hours, that is, at the readout time point (Figure S3A). Using an extensive processing framework to reduce possible batch effects, account for off-target effects and impute missing data (see Materials and Methods section), we obtained 488, 457, and 298 significant interactions (adjusted $P < .1$ and absolute log-fold change $>.5$) at time points B4, N2, and N4, respectively (Figure S3B and Table S3). The lower number at N4 may be explained in part by technical factors, such as a more variable start population at N2 and a generally lower knockdown efficacy at the 48 hours readout timepoint (Figure S3A) and thus explaining the observed decreased effect size and significance at N4 (Figure S3C). We retrieved many known interactions at each stage^{32,50-53} and hence consider our method and used parameters sensitive and strict enough to minimize off-target effects and detect true gene-gene interactions, including previously validated interactions (Table S4). Of note, the number of interactions for each perturbed gene resembled a power-law degree distribution, a typical characteristic for many biological networks.^{54,55} In agreement, we observed that *Pou5f1* and *Sox2*, two important TFs in mES cells, occupy the first and third position, respectively (Figure S4A).

3.4 | The majority of gene interactions are cell-stage specific and driven by a limited number of TFs

We observed a roughly equal number of positive and negative interactions. A positive or activating interaction (Figure 3A, green arrow) signifies that a knockdown of a given gene caused a downregulation of one of the genes from our list, while a negative or inhibitory

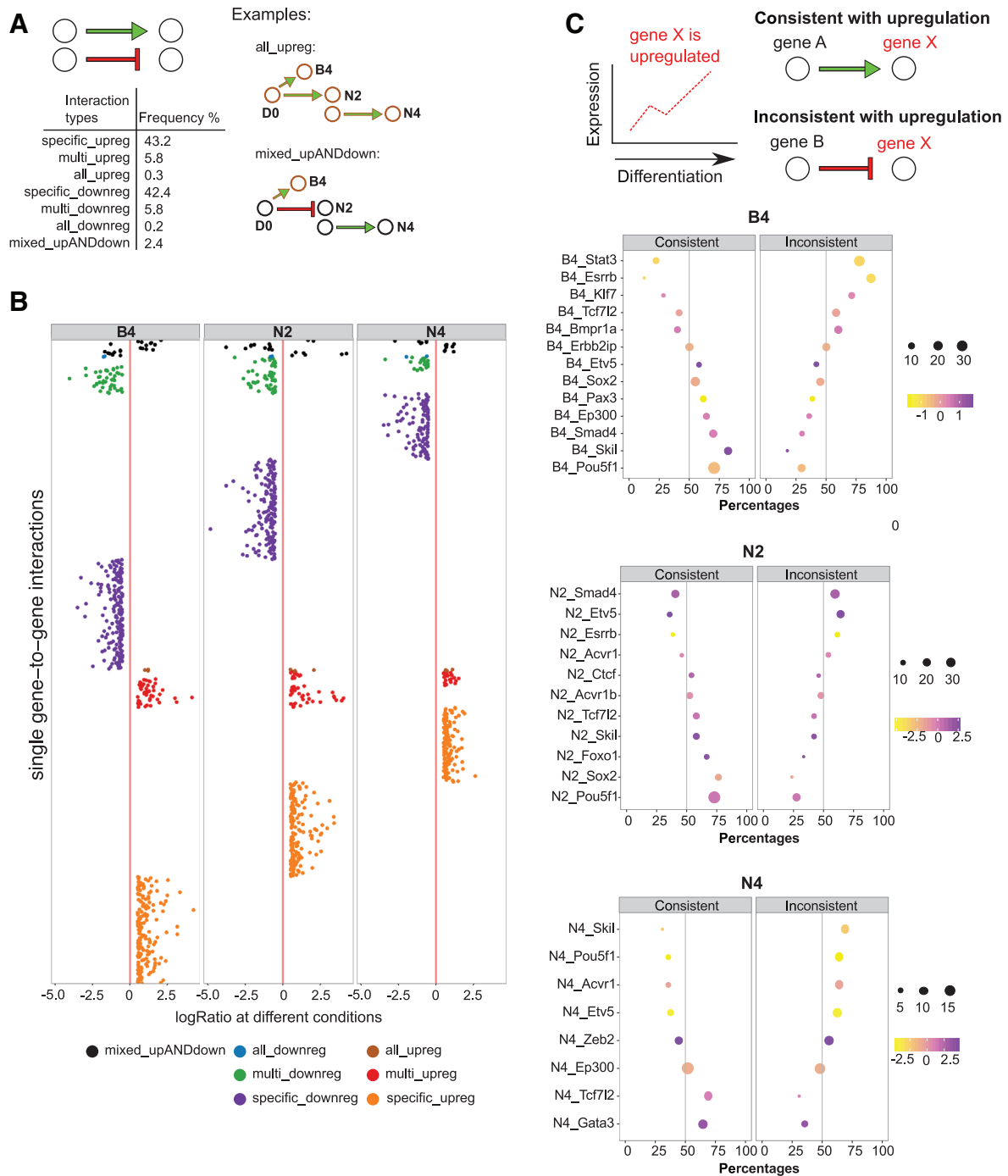


FIGURE 3 Cell-stage specific interactions and regulators. A, Diagram illustrating upregulating/positive (green arrow) and downregulating/negative (red blunt end) interactions. Examples and global frequency of combinations are shown. B, Plot showing the logRatio (log Target/Ctrl) values of all significant interactions (rows) and colors indicate the interaction type class, as shown in (A), they belong to considering all three transition stages. C, Consistent interactions have directions in agreement with the wild-type temporal expression profile for that transition stage and inconsistent do not, as exemplified (top). Dot plot showing the top regulators, that is, most significant interactions upon perturbation, for each transition stage. The size and color of the dot is correlated with respectively the total number of interactions and change of expression levels for the regulator (bottom)

interaction (Figure 3A, red blunt end) means an upregulation of a gene from the list upon a single knockdown. Each gene-gene interaction over all three cell transitions studied here can be classified as one of 27 (=3³) interaction types, that is, at each transition a specific

interaction can be positive, negative or neutral (no significant effect). These types were further aggregated into seven interaction-type clusters based on the stability and direction of an interaction in all three transitions (see Materials and Methods section). Their frequency

percentages are summarized in a table (Figure 3A, insert). Interestingly, the majority (~85%) of all significant interactions were restricted to one specific transition or cell stage (specific_upreg and specific_downreg) (Figure 3B, purple and orange dots). For more than 10% of the interactions, both interaction and direction were maintained in more than one transition, but only a small fraction of interactions was present in all three transitions.

To test whether most interactions were present, but not significant in adjacent transitions, we measured the transcriptional change similarities with pair-wise correlation scores for each gene in all perturbation samples between two transitions. The majority of such correlations were positive, which suggests that most gene interactions are most prominent at only one stage, but that the direction of the interaction, either activating or inhibiting, does not readily change with time (Figure S4B). This was most pronounced when comparing B4 with N2, where the summit of the correlation distribution scores shifted to more positive values ($r \approx 0.5$) (Figure S4B). We also noticed that many genes, including those encoding BMP-system components, were subject to transcriptional regulation in very similar ways in both B4 and N2. Taken together, the perturbations show that the majority of gene-gene interactions are cell-stage specific, driven by a limited number of TFs, and usually the interactions gradually fade out in adjacent cell stages.

3.5 | Epistatic interactions of TFs often display a mixed effect on expression changes during transitions

In order to evaluate whether TFs act as drivers or inhibitors of transitions during neural differentiation, we next assessed whether the direction of interactions of single perturbations was consistent with the direction of expression changes of their target genes in control (ie, unperturbed) differentiation (Figure 3C, top panel, showing as example one interaction for genes A-B). If the interaction direction is consistent with the expression changes during a transition, this regulator is considered to be a driver of that transition. In contrast, when the majority of interactions is inconsistent with how genes behave in unperturbed differentiation conditions, that gene is considered to function as a transition inhibitor.

Remarkably, *Pou5f1* and *Skil* displayed mostly a driver function in the first two transitions (B4 and N2), but this was opposite at N4 (Figure 3C). This agrees with the important role of *Pou5f1* in primed mES cells (B4) and EpiLCs (N2) and the need to downregulate *Pou5f1* to induce neural differentiation (N4). This is in accordance with their declining mRNA levels at N4, which is necessary to commit to differentiation.^{56,57} Other genes like *Zeb2* have roughly an equal number of consistent and inconsistent interactions, suggesting that *Zeb2* protein (and possibly mRNA) levels are important to steer cells into one specific lineage and suppress other lineages, and also promote cell maturation, which is in line with *Zeb2*-knockout studies in nervous systems in vivo^{33,58-61} and ES cells.³² This hypothesis is supported by examination of the target loci/genes to which *Zeb2* binds in neural-differentiating cells, in which *Zeb2* is strongly upregulated³² (Birkhoff, Conidi and Huylebroeck, *data not shown*). The neural-associated genes,

such as *Pax6*, *Cdh2* and *Cxcr4*, had positive interactions with *Zeb2*, whereas known repressors, including for the neural lineage, such as *Id1*, *Id3*, and *Ovol2* had a strong negative interaction with *Zeb2*.

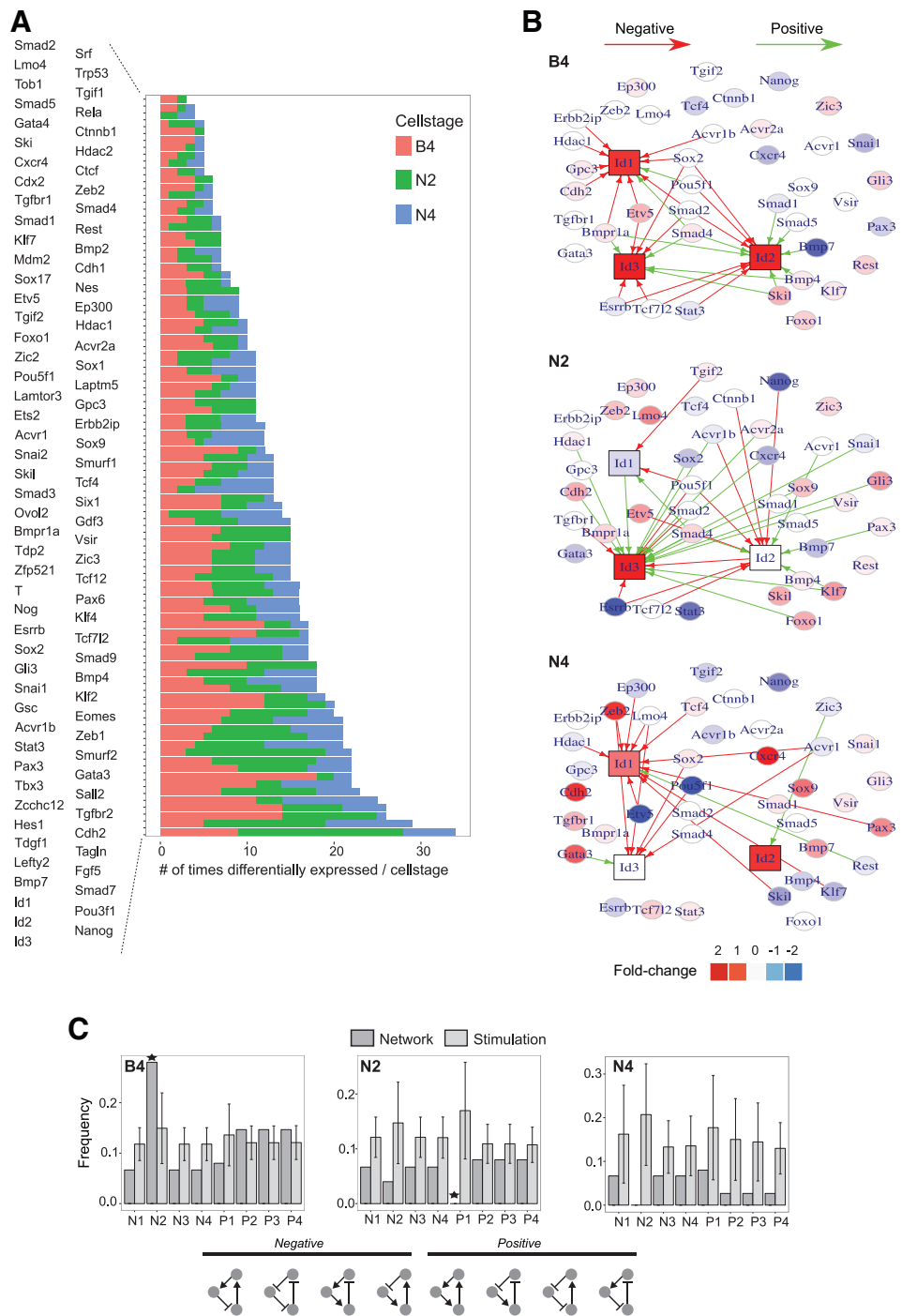
3.6 | Multifaceted transcriptional regulation of *Id* genes

To identify key effector genes in our perturbation setup, we determined the number of times a gene was significantly deregulated upon any of the perturbations. Remarkably, three *Id* family genes were within the top five most deregulated genes following the perturbations (Figure 4A). Given that each of these three *Id* genes displayed medium-high or high range of expression (Figure 2D), we analyzed if in general genes with higher range of expression were more sensitive to deregulation upon perturbations, as we had hypothesized previously (Figure 2D). Doing so, we noted indeed a strong positive association (Pearson's $r = .5$, adjusted $R^2 = 0.25$, and $P = 7.18e-06$) between the number of times a gene was significantly deregulated and its gene expression range during differentiation (Figure S5A). Alternatively, we did not find any strong association between gene expression range and the effect a single perturbed gene has on other genes (Figure S5B).

To further examine the effect on the *Id* genes and assess whether this also implicates that they are co-regulated and expressed in specific and defined modules, we visualized both the interactions and the expression level changes for all three cellular transitions (Figure 4B). All three *Id* genes exhibited increased mRNA levels during the B4 transition. They displayed unique as well as overlapping reactions, while the pluripotency core TF-genes *Sox2* and *Pou5f1* together with *Smad2* and *Smad4* appear to be common regulators at this stage. Unexpectedly, only *Id2* showed to be positively regulated by interaction with (ie, perturbation of) *Bmp4/7* and *Smad1/5*, which encode essential components of the BMP-pathway axis that is known to drive *Id* expression.^{15,62,63} Also the previously Wnt-linked triade *Esrrb*,⁵⁰ *Tcf7l2*, and *Stat3*⁶⁴ showed negative interactions with both *Id2* and *Id3*. During N2 transition, the role of *Id1* was diminished and *Id3* became the dominant target to be regulated. Furthermore, the majority of interactions were positive, which might indicate the necessity of *Id3* upregulation to enter the EpiLC state. Next, during the N4 transition, *Id1* was again elevated and many interactions were negative on both *Id1* and *Id3*. This suggests a model wherein cells in this early progenitor phase start to induce both neural and nonneural genes and that gene cooperation is needed to determine the final fate at later stage.

To interrogate the directionality of inferred networks, we assessed the numbers and types of simple feedback loops (FBLs) consisting of three nodes. We noticed that although most individual comparisons were not statistically significant, both the negative (N1-4) and positive (P1-4) FBLs at N2 and N4 were lower compared with simulated data. This implies that, during differentiation, the TGF β -centric network is rather unidirectional and hence master TFs in our network receive only limited feedback (Figure 4C).

FIGURE 4 Perturbation sensitive genes and network characteristics. A, Barplot demonstrating the perturbation sensitivity of each gene, as measured by how often a gene is deregulated in all perturbation samples over all transition stages. B, Network visualization of all individual *Id* gene (*Id1-3*) interactions in B4 (top), N2 (middle), and N4 (bottom). Node colors represent changes in expression levels during transition, with red and blue depicting, respectively, increase and decrease of expression levels. Red and green edges represent inhibiting and activating interactions, respectively. C, Barplot showing the frequencies in real vs simulated data of all the types of feedback-loops consisting of three nodes (top). Types of negative and positive feedback loops are shown in numerical order (bottom). **P* < .01; Wilcoxon rank-sum test



3.7 | Identification of gene co-expression modules with single-cell RT-qPCR

To further explore the role of transcriptional regulation/variability of our TGFβ-family centric list during ES cell differentiation, we also performed single-cell RT-qPCR (sc-qPCR) at each time point (see also Figure 1E and Supplemental Information). Despite our restricted number of genes, the cells clustered according to their time point as seen with tSNE (Figure S6A) and follow a continuous path from day 0 (D0) to neural differentiation day 4 (N4) using the first two dimensions in PCA space (Figure 5A).

This accurate clustering, despite the aforementioned low number of genes, is expected due to both our gene selection approach (Figure S1C) and the high-quality data generated by our sc-qPCR readout. Indeed, the processed and normalized data matrix (Table S5) for the sc-qPCR data only contains about 8% zeroes, which is at least an order of magnitude lower as seen in most single-cell RNAseq data sets.⁶⁵ As such, calculating gene-gene correlation scores to detect gene co-expression, modules can be performed with high precision (Figure S6B) and resulted in the identification of at least seven notable co-expression modules (Table S6 and Figure S6B). For example, at least two of the

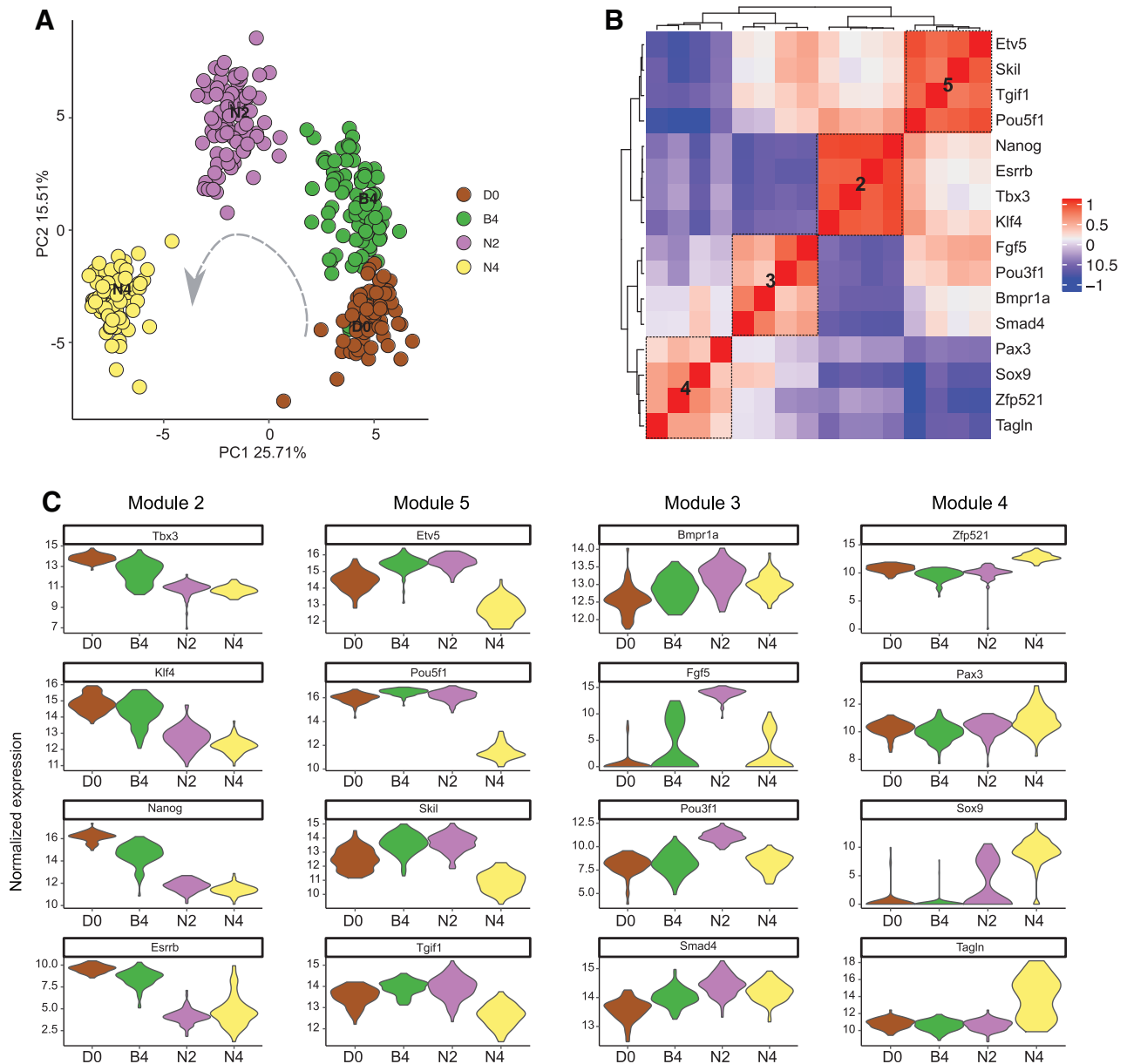


FIGURE 5 Single-cell RT-qPCR and co-expression modules. A, PCA plot showing the first and second principal component to illustrate the distribution of single cells at different time points. Gray arrow indicates direction of differentiation. B, Heatmap depicting the pairwise correlation values between genes (Pearson's r). C, Violin plots showing the expression distribution at different time points for the indicated genes

previously discussed *Id* genes, *Id1* and *Id2*, comprise an interesting co-expression module together with *Zic3*. These genes seem to be both upregulated in BMP4-exposed cells (B4) and more differentiated cells (N4) without addition of any cytokines (Figure S6C, D), which is in line with their proposed roles in both maintenance and exit of pluripotency.^{66,67} To further illustrate the relevance of these co-expression modules, we highlighted a set of four typical genes for four particular modules (Figure 5B) that show distinct expression profiles over all time points (Figure 5C). The first module displays high expression of core pluripotency genes (*Tbx3*, *Klf4*, *Nanog*, and *Esrrb*) that immediately drop in expression between B4 and N2, when cells enter the EpiLC stage. The second module is marked by genes that remain high in both naïve

ES cells and EpiLCs, such as *Pou5f1*, *Tgif1*, *Skil*, and *Etv5*, and which are often associated with early exit of pluripotency.^{57,68,69} Therefore, these genes only decrease in expression between N2 and N4. A third module consists of genes such as *Fgf5*, *Bmpr1a*, *Pou3f1*, and *Smad4*, each showing a spike in expression only in the EpiLC stage and frequently linked to Smad mediated signaling.^{70,71} The last group shows a strong increase at the most differentiated stage, N4, with genes such as *Zfp521*, *Pax3*, *Sox9*, and *Tagln*. The latter genes are known marker genes of different cell progenitors⁷²⁻⁷⁵ and are both co-expressed and upregulated at the N4 stage of differentiation, suggesting that those cells likely still have a multipotent potential to differentiate in different germ layers.

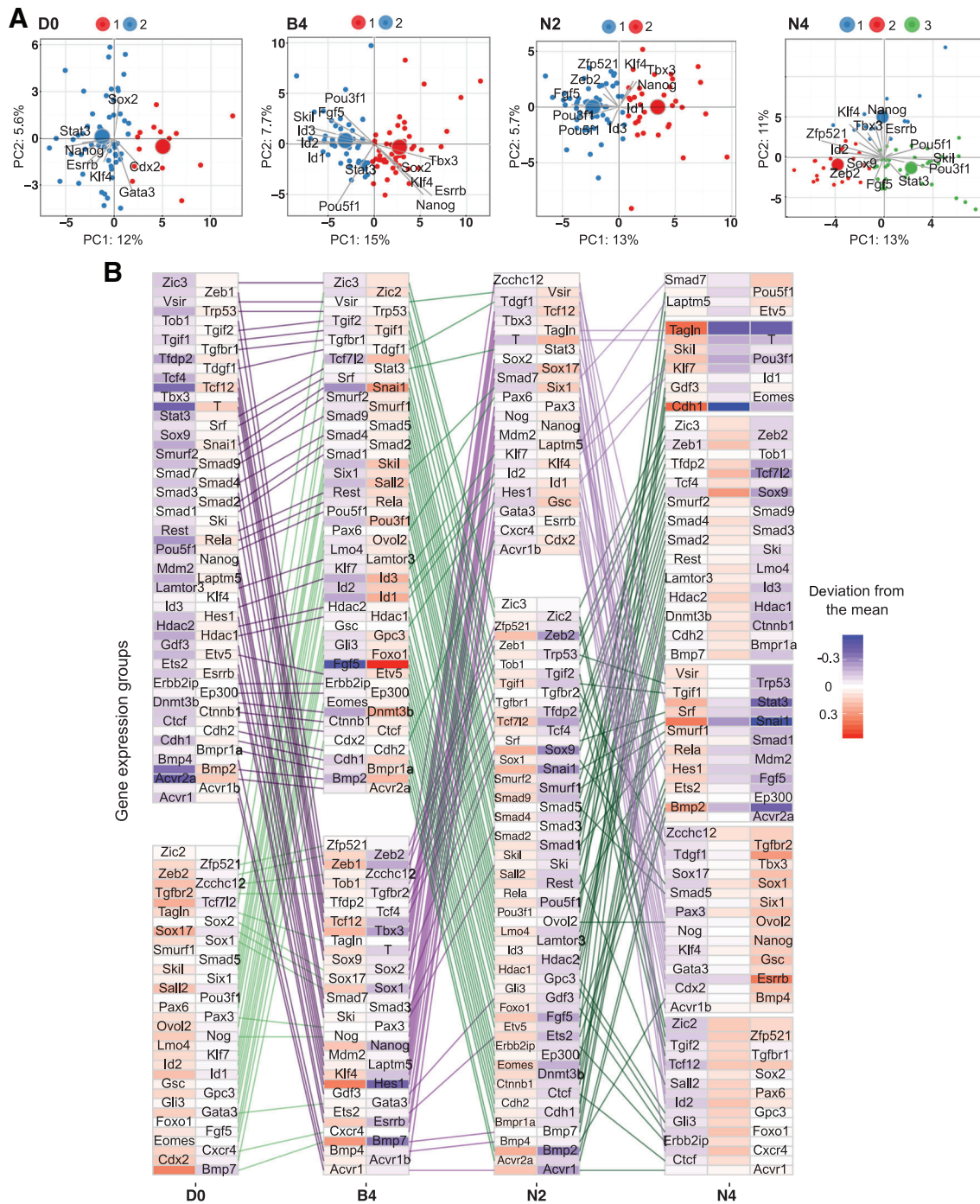


FIGURE 6 Single-cell subpopulations at individual time points. A, 2D tSNE plot showing the distribution of single cells and subclusters in the indicated colors at individual time points. B, Diagram visualizing the number of primed subpopulations in each stage (number of columns) and the gene specific clusters (row groups) together with their average expression deviation from the overall mean. Lines connect the same genes over the different cell stages and line colors correspond with the gene specific clusters from the previous cell stage

3.8 | Transcriptional heterogeneity between single cells denotes subtle subpopulations indicating that differentiation is not synchronous

To see if the observed transcriptional heterogeneity and co-expression modules, as observed in Figure 5 and Figure S6, might also result in more detailed subpopulations within each time point, we further

subclustered the cells. This led to the identification of two subpopulations in each of the first three time points (D0, B4, and N2) and three subpopulations at N4 (Figure 6A).

Within these subpopulations, we calculated the deviation from the population mean for each gene separately (Figure S7A-D). This is a simple and forward approach to assess which sets of genes play a role and what their relationships are. mES cells that are kept in a ground state

exhibit overall lower variability for pluripotency and developmental related genes. However, using our approach we noticed that a small group of cells showed slightly elevated mRNA levels of *Cdx2* (essential for trophectoderm in pre-implantation mouse embryos⁷⁶) and *Sox17* (a marker for extraembryonic visceral endoderm at E6.0 and endoderm of mid- to late-gastrula stage embryos⁷⁷), while mES cells kept in a naïve primed state exhibited reduced levels for the naïve core pluripotency markers *Tbx3*, *Nanog*, *Klf4*, and *Esrrb* (abbreviated here as quartet TNKE). This suggests that some subtle priming occurs and that this stochasticity might allow cell fate switches when the right cues are presented. Interestingly, it appears that the quartet TNKE functions in a co-expression module, which was also observed in Figure 5B, and that this is true in all identified stages. Indeed, they were always enriched together in one of the subpopulations in Figure 6A. This does not indicate that at each of these time points, there were cells that remain in their original pluripotent state; rather it suggests that some cells had irregular kinetics for this set of genes and hence were lagging slightly behind on their differentiation path. The net consequence would then be that single-cell expression analysis shows that in all stages a subset of cells is more resilient to differentiation cues. Similar observations of associations between transcriptional heterogeneity and differences in differentiation or reprogramming have previously been documented for several single factors, including *Esrrb* and *Nanog*, with particular focus on the pluripotent state.⁷⁸⁻⁸¹

A more complex narrative applies to N4 with its three subpopulations (Figure 6A, right, and Figure S7D). Besides the aforementioned delayed subpopulation (ie, 14 cells here in blue), there were two more subpopulations (which can be described by six groups of gene expression dynamics): (a) a more neural-orientated group of 22 cells marked by higher average levels of NP-cell associated genes such as *Zfp521*, *Pax6*, *Cxcr4*, *Zeb2*, *Zic3*, and *Sox9*, and (b) the largest group of 39 cells that showed slightly lower levels of neural mRNAs and higher levels of specific combinations of genes, such as *Stat3*, *Srf*, *Snai1*, *Tagln*, *Skil*, *Pou3f1*, and *Cdh1*. Some are known smooth-muscle related genes, suggesting again that these progenitor cells are not fully restricted to the neural lineage. In any case, groups of genes that define the different subpopulations (the columns at each time point, Figure 6B) remained relatively stable at different time points or cell states, suggesting that the regulatory network that governs transcription at transitions is robust, including to changes, although many transition-specific interactions are formed and removed.

4 | DISCUSSION

We report an experimental strategy and data set allowing the deconstruction of a dynamic TRN that operates in conjunction with components of a selected, prioritized, and well-studied growth factor signaling system in embryogenesis. We further enriched the selection with ES cell state relevant TFs and also included dynamic transitions of cultured ES cells. We further integrate three transcriptional dimensions, that is, temporal dynamics, single-cell expression heterogeneity, and gene-gene interactions as deduced from esiRNA-based perturbations. These

perturbations reduce the target protein amount rather than totally removing the protein (eg, like in most CRISPR/Cas-based experiments). RNAi-based perturbations are fast, simple, and often easier to interpret compared to genetic knockouts.⁸²⁻⁸⁵ It also raises the interesting question as to whether both types of perturbation, when applied to a signaling system, will eventually result in similar conclusions or not in terms of gene-gene interactions. With regard to cell lineage expansion or progression in vitro, where CRISPR-Cas based mutagenesis is being considered to obtain desired cell types with higher efficiency, and the developmental and/or disease relevance of TGF β -system genes, such difference between the perturbation approaches could have important conceptual, mechanistic, and application consequences.

Our temporal profiling confirms the presence and gradual transition of at least three major cell states, that is, pluripotent mES cell, primed EpiLC, and early NP cell. EpiLCs are not to be confused with EpiSCs, which can be maintained for several passages using high doses of Activin and FGF, but resemble primarily the anterior primitive streak.⁴³ EpiLCs are on the contrary rapidly and directly induced from mES cells and perhaps reflect a more transient population, similar to cells allocated to the pre-gastrulating mouse epiblast.³ Since N2 or EpiLCs are in our experiments both a readout and perturbation time point, our data provide unique insight in this transient population. We mapped four robust, temporal expression profile types, and found that *Smad1/5* and *Smad2*, and also *Smad4* transcripts are enriched in the stable gene cluster. This suggests that modifications of the already present protein, rather than transcriptional modulation of the gene encoding such immediate-effector protein, play a more important role in the processing of TGF β -family ligand action and in the fine-tuning of Smad-mounted target gene responses.⁸⁶⁻⁸⁸

Inhibitory-Smad mRNA was strongly downregulated upon differentiation, suggesting that mES cells require this negative feedback to dampen Smad-signaling, which is in line with previous observations. Surprisingly, three of the *Id* family genes end up in the neural cluster, which contains genes that are upregulated during neural differentiation. This is in contrast to studies showing that *Id* genes have a direct negative effect on neural fate acquisition.^{44,45} Hence, this may be a prime example of multipotent cells and how this is assisted by the complex nature of the TRN that controls *Id* family mRNA levels. This is further illustrated by the highly selective induction of *Id3* and many of its positive interactions in transition from mES cells to EpiLCs, suggesting it is essential here (and without BMP added to the cells).

The majority of significant gene-gene interactions are cell-stage specific and combinations of TFs that appear important for different cell states and transitions. *Pou5f1* is the prime regulator of the TGF β -system during the first two stages. Here, *Pou5f1* teams up with *Sox2*, *Skil*, and *Smad4* to drive the establishment of the naïve state, whereas *Stat3* and *Esrrb* contribute to maintaining the ground state. However, in the next stage Smads act mostly as transition inhibitors. Despite our data obtained from perturbation-based interaction calling, we cannot infer whether Smad-dependent activity plays an exclusive role in these processes, and it is likely that there are also Smad-independent interactions. In an even broader context, it would be of great interest to simultaneously document the role and interplay with other

signaling pathways than those of the TGF β (-Smad) family. For example, complementary approaches are being developed that assess the activity of multiple signaling pathways and use a multiplex reporter gene assay combined with next-generation sequencing.⁸⁹ This principle can also be taken to CRISPR-based (knockout) screens, down to the single-cell level, in experiments that involve (co-)barcoding allowing simultaneous identification of the knockout mutation and the single-cell identity in RNA sequencing. However, irrespective to whether a knockdown or knockout strategy (for many individual targets) is used to construct a gene-gene interaction network, it remains challenging to probe the effective loss of activity of each of the targeted proteins. Here, we created a simple Boolean network based on a 50% reduction cutoff; however, one potential way forward would be to discretize the different levels of the knockdown efficiency of each component and see how that further correlates with different downstream effects.

Perturbation remains instrumental for studying the individual gene or protein function, but inferring direct gene functions from long-time experiments such as permanent knockouts should be done with caution, for secondary effects are hard to account for and accumulate fast. Here, we show that cellular responses to perturbations during development are primarily driven by the cell stage. This is even true for seemingly functionally similar cell types, such as the naïve (B4) and primed (N2) cell states, which are about 48 hours apart, believed to be both pluripotent and capable of forming an embryo proper. Second, even within the same cell state, our results indicate that single perturbations should be interpreted with care. For example, the *Id* genes, which are implicated in many cellular differentiation processes either alone or together, respond in both diverse and unique ways to our different perturbations at each stage. Finally, predicting a system or network gene expression, including as response to perturbation, remains a challenging task and will also require concurrent upscaling of multi-dimensional data sets and development of novel multi-omics readout methods at the single-cell level.

ACKNOWLEDGMENTS

This work was supported by KU Leuven (GOA-11/012, to D.H.), Fund for Scientific Research-Flanders FWO-V (GA.09411.10, G.0782.14, to D.H.), Belgian Science Policy IAPVII-07 (to C.V., F.G., and D.H.), Netherlands Organisation for Health Research and Development ZonMW (Off Road, to E.M.), and Erasmus MC start-up funds (D.H.). R.D. was PhD fellow with Flanders Innovation & Entrepreneurship (IWT, SB93097). We thank Frank Buchholz (Dresden), Rudi Balling and Enrico Glaab (Belvaux), Stein Aerts and An Zwijsen (Leuven), Niels Galjart and Harmen van de Werken (Rotterdam), and Austin Smith (Cambridge) for valuable advice, support and materials during the course of this work.

CONFLICT OF INTEREST

The authors indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

R.D.: conception and design, financial support, collection and/or assembly of data, data analysis and interpretation, manuscript writing, final

approval of manuscript; A.S.: financial support, collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript; K.C., M.D.: collection and/or assembly of data, final approval of manuscript; S.O.: data analysis and interpretation, manuscript writing, final approval of manuscript; T.N.: conception and design, collection and/or assembly of data, data analysis and interpretation, final approval of manuscript; J.B., A.C.: collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of manuscript; C.M.V., F.G.: financial support, manuscript writing, final approval of manuscript; E.M.: financial support, data analysis and interpretation, manuscript writing, final approval of manuscript; D.H.: conception and design, financial support, data analysis and interpretation, manuscript writing, final approval of manuscript.

DATA AVAILABILITY STATEMENT

The data sets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

ORCID

Ruben Dries  <https://orcid.org/0000-0001-7650-7754>

Danny Huylebroeck  <https://orcid.org/0000-0003-4862-1079>

REFERENCES

1. Ying QL, Wray J, Nichols J, et al. The ground state of embryonic stem cell self-renewal. *Nature*. 2008;453:519-523.
2. Ying QL, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol*. 2003;21:183-186.
3. Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*. 2011;146:519-532.
4. Mullen AC, Orlando DA, Newman JJ, et al. Master transcription factors determine cell-type-specific responses to TGF- β signaling. *Cell*. 2011;147:565-576.
5. Thomson M, Liu SJ, Zou LN, Smith Z, Meissner A, Ramanathan S. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell*. 2011;145:875-889.
6. Mu Y, Gudey SK, Landström M. Non-Smad signaling pathways. *Cell Tissue Res*. 2012;347:11-20.
7. Matsuzaki K. Smad phospho-isoforms direct context-dependent TGF- β signaling. *Cytokine Growth Factor Rev*. 2013;24:385-399.
8. Macias MJ, Martin-Malpartida P, Massagué J. Structural determinants of Smad function in TGF- β signaling. *Trends Biochem Sci*. 2015;40:296-308.
9. Zhang YE. Non-Smad signaling pathways of the TGF- β family. *Cold Spring Harb Perspect Biol*. 2017;9(2):a022129.
10. Weng Q, Chen Y, Wang H, et al. Dual-mode modulation of Smad signaling by Smad-interacting protein Sip1 is required for myelination in the central nervous system. *Neuron*. 2012;73:713-728.
11. Constam DB. Regulation of TGF β and related signals by precursor processing. *Semin Cell Dev Biol*. 2014;32:85-97.
12. Bier E, De Robertis EM. Embryo development. BMP gradients: a paradigm for morphogen-mediated developmental patterning. *Science*. 2015;348:aaa5838.
13. Xu P, Lin X, Feng XH. Posttranslational regulation of Smads. *Cold Spring Harb Perspect Biol*. 2016;8(12):a022087.
14. Karaulanov E, Knöchel W, Niehrs C. Transcriptional regulation of BMP4 synexpression in transgenic *Xenopus*. *EMBO J*. 2004;23:844-856.

15. Paulsen M, Legewie S, Eils R, Karaulanov E, Niehrs C. Negative feedback in the bone morphogenetic protein 4 (BMP4) synexpression group governs its dynamic signaling range and canalizes development. *Proc Natl Acad Sci USA*. 2011;108:10202-10207.
16. Weiss A, Attisano L. The TGFbeta superfamily signaling pathway. *Wiley Interdiscip Rev Dev Biol*. 2013;2:47-63.
17. Montagner M, Martello G, Piccolo S. Monitoring Smad activity in vivo using the Xenopus model system. *Methods Mol Biol*. 2016;1344:245-259.
18. Budi EH, Duan D, Derynck R. Transforming growth factor- β receptors and Smads: regulatory complexity and functional versatility. *Trends Cell Biol*. 2017;27:658-672.
19. Zinski J, Tajer B, Mullins MC. TGF- β family signaling in early vertebrate development. *Cold Spring Harb Perspect Biol*. 2017;10(6):a033274.
20. Massagué J. TGF β signalling in context. *Nat Rev Mol Cell Biol*. 2012;13:616-630.
21. Gaarenstroom T, Hill CS. TGF- β signaling to chromatin: how Smads regulate transcription during self-renewal and differentiation. *Semin Cell Dev Biol*. 2014;32:107-118.
22. Hill CS. Transcriptional control by the SMADs. *Cold Spring Harb Perspect Biol*. 2016;8(10):a022079.
23. Seuntjens E, Umans L, Zwijsen A, Sampaolesi M, Verfaillie CM, Huylebroeck D. Transforming growth factor type beta and Smad family signaling in stem cell function. *Cytokine Growth Factor Rev*. 2009;20:449-458.
24. Massagué J, Xi Q. TGF- β control of stem cell differentiation genes. *FEBS Lett*. 2012;586:1953-1958.
25. Oshimori N, Fuchs E. The harmonies played by TGF- β in stem cell biology. *Cell Stem Cell*. 2012;11:751-764.
26. Sakaki-Yumoto M, Katsuno Y, Derynck R. TGF- β family signaling in stem cells. *Biochim Biophys Acta*. 2013;1830:2280-2296.
27. Akhurst RJ, Padgett RW. Matters of context guide future research in TGF β superfamily signaling. *Sci Signal*. 2015;8(399):re10.
28. Pangas SA, Li X, Umans L, et al. Conditional deletion of Smad1 and Smad5 in somatic cells of male and female gonads leads to metastatic tumor development in mice. *Mol Cell Biol*. 2008;28:248-257.
29. Middlebrook BS, Eldin K, Li X, Shivasankaran S, Pangas SA. Smad1-Smad5 ovarian conditional knockout mice develop a disease profile similar to the juvenile form of human granulosa cell tumors. *Endocrinology*. 2009;150:5208-5217.
30. Pereira PN, Dobрева MP, Maas E, et al. Antagonism of nodal signaling by BMP/Smad5 prevents ectopic primitive streak formation in the mouse amnion. *Development*. 2012;139:3343-3354.
31. Gomes Fernandes M, Dries R, Roost MS, et al. BMP-SMAD Signaling regulates lineage priming, but is dispensable for self-renewal in mouse embryonic stem cells. *Stem Cell Reports*. 2016;6:85-94.
32. Stryjewska A, Dries R, Pieters T, et al. Zeb2 regulates cell fate at the exit from epiblast state in mouse embryonic stem cells. *STEM CELLS*. 2017;35:611-625.
33. van den Berghe V, Stappers E, Vandesande B, et al. Directed migration of cortical interneurons depends on the cell-autonomous action of Sip1. *Neuron*. 2013;77:70-82.
34. Lu R, Markowitz F, Unwin RD, et al. Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature*. 2009;462:358-362.
35. Aiba K, Nedorezov T, Piao Y, et al. Defining developmental potency and cell lineage trajectories by expression profiling of differentiating mouse embryonic stem cells. *DNA Res*. 2009;16:73-80.
36. Henschel A, Buchholz F, Habermann B. DEQOR: a web-based tool for the design and quality control of siRNAs. *Nucleic Acids Res*. 2004;32:W113-W120.
37. Kittler R, Putz G, Pelletier L, et al. An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature*. 2004;432:1036-1040.
38. Kunath T, Saba-El-Leil MK, Almousailleakh M, et al. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. *Development*. 2007;134:2895-2902.
39. Stavridis MP, Lunn JS, Collins BJ, Storey KG. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development*. 2007;134:2889-2894.
40. Hayashi K, de Sousa Lopes SMC, Tang F, Lao K, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell*. 2008;3:391-401.
41. Turner DA, Trott J, Hayward P, Rue P, Martinez Arias A. An interplay between extracellular signalling and the dynamics of the exit from pluripotency drives cell fate decisions in mouse ES cells. *Biol Open*. 2014;3:614-626.
42. Iwafuchi-Doi M, Matsuda K, Murakami K, et al. Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. *Development*. 2012;139:3926-3937. Erratum in: *Development* 2012;139:4675.
43. Kojima Y, Kaufman-Francis K, Studdert JB, et al. The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell*. 2014;14:107-120.
44. Di-Gregorio A, Sancho M, Stuckey DW, et al. BMP signalling inhibits premature neural differentiation in the mouse embryo. *Development*. 2007;134:3359-3369.
45. Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol*. 2009;27:275-280.
46. Chambers I, Silva J, Colby D, et al. Nanog safeguards pluripotency and mediates germline development. *Nature*. 2007;450:1230-1234.
47. Festuccia N, Osorno R, Halbritter F, et al. Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell*. 2012;11:477-490.
48. Katano M, Ema M, Nakachi Y, et al. Forced expression of Nanog or Esrrb preserves the ESC status in the absence of nucleostemin expression. *STEM CELLS*. 2015;33:1089-1101.
49. Russell R, Ilg M, Lin Q, et al. A dynamic role of TBX3 in the pluripotency circuitry. *Stem Cell Reports*. 2015;5:1155-1170.
50. Martello G, Sugimoto T, Diamanti E, et al. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell*. 2012;11:491-504. Erratum in: *Cell Stem Cell* 2013;12:630.
51. Nishiyama A, Sharov AA, Piao Y, et al. Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells. *Sci Rep*. 2013;3:1390.
52. Xu H, Ang YS, Sevilla A, et al. Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLoS Comput Biol*. 2014;10:e1003777.
53. Xu H, Barouk C, Dannenfelser R, et al. ESCAPE: database for integrating high-content published data collected from human and mouse embryonic stem cells. *Database*. 2013;2013:bat045.
54. Alon U. Network motifs: theory and experimental approaches. *Nat Rev Genet*. 2007;8:450-461.
55. Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. *Genes Dev*. 2007;21:1010-1024.
56. Niwa H, Miyazaki J, Smith AG. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet*. 2000;24:372-376.
57. Tsuneyoshi N, Tan EK, Sadasivam A, et al. The SMAD2/3 corepressor SNON maintains pluripotency through selective repression of mesodermal genes in human ES cells. *Genes Dev*. 2012;26:2471-2476.
58. Quintes S, Brinkmann BG, Ebert M, et al. Zeb2 is essential for Schwann cell differentiation, myelination and nerve repair. *Nat Neurosci*. 2016;19:1050-1059.

59. Wu LM, Wang J, Conidi A, et al. Zeb2 recruits HDAC-NuRD to inhibit notch and controls Schwann cell differentiation and remyelination. *Nat Neurosci.* 2016;19:1060-1072.
60. Li J, Riedt T, Goossens S, et al. The EMT transcription factor Zeb2 controls adult murine hematopoietic differentiation by regulating cytokine signaling. *Blood.* 2017;129:460-472.
61. Watanabe Y, Stanchina L, Lecerf L, et al. Differentiation of mouse enteric nervous system progenitor cells is controlled by endothelin 3 and requires regulation of Ednrb by SOX10 and ZEB2. *Gastroenterology.* 2017;152:1139-1150.e4.
62. Hollnagel A, Oehlmann V, Heymer J, Rütter U, Nordheim A. Id genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *J Biol Chem.* 1999;274(28):19838-19845.
63. Nakahiro T, Kurooka H, Mori K, Sano K, Yokota Y. Identification of BMP-responsive elements in the mouse Id2 gene. *Biochem Biophys Res Commun.* 2010;399(3):416-421.
64. Hao J, Li TG, Qi X, Zhao DF, Zhao GQ. WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Dev Biol.* 2006;290:81-91.
65. Grün D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. *Nat Methods.* 2014;11:637-640.
66. Lim LS, Loh YH, Zhang W, et al. Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Mol Biol Cell.* 2007;18:1348-1358.
67. Yang S-H, Andrabi M, Biss R, Murtuza Baker S, Iqbal M, Sharrocks AD. ZIC3 controls the transition from naive to primed pluripotency. *Cell Rep.* 2019;27:3215-3227.e6.
68. Lee B-K, Shen W, Lee J, et al. Tgif1 counterbalances the activity of core pluripotency factors in mouse embryonic stem cells. *Cell Rep.* 2015;13:52-60.
69. Kalkan T, Bornelöv S, Mulas C, et al. Complementary activity of ETV5, RBPJ, and TCF3 drives formative transition from naive pluripotency. *Cell Stem Cell.* 2019;24:785-801.e7.
70. Zhu Q, Song L, Peng G, et al. The transcription factor Pou3f1 promotes neural fate commitment via activation of neural lineage genes and inhibition of external signaling pathways. *Elife.* 2014;3:e02224.
71. Morikawa M, Koinuma D, Mizutani A, et al. BMP sustains embryonic stem cell self-renewal through distinct functions of different Krüppel-like factors. *Stem Cell Reports.* 2016;6:64-73.
72. Cheung M, Briscoe J. Neural crest development is regulated by the transcription factor Sox9. *Development.* 2003;130:5681-5693.
73. Blake JA, Ziman MR. Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development.* 2014;141:737-751.
74. Tang K, Peng G, Qiao Y, Song L, Jing N. Intrinsic regulations in neural fate commitment. *Dev Growth Differ.* 2015;57:109-120.
75. Elsafadi M, Manikandan M, Dawud RA, et al. Transgelin is a TGF β -inducible gene that regulates osteoblastic and adipogenic differentiation of human skeletal stem cells through actin cytoskeleton organization. *Cell Death Dis.* 2016;7:e2321.
76. Strumpf D, Mao CA, Yamanaka Y, et al. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development.* 2005;132:2093-2102.
77. Choi E, Kraus MR, Lemaire LA, et al. Dual lineage-specific expression of Sox17 during mouse embryogenesis. *STEM CELLS.* 2012;30:2297-2308.
78. Kobayashi T, Kageyama R. Hes1 oscillations contribute to heterogeneous differentiation responses in embryonic stem cells. *Genes.* 2011;2:219-228.
79. Marucci L. Nanog dynamics in mouse embryonic stem cells: results from systems biology approaches. *Stem Cells Int.* 2017;2017:1-14.
80. Yu P, Nie Q, Tang C, Zhang L. Nanog induced intermediate state in regulating stem cell differentiation and reprogramming. *BMC Syst Biol.* 2018;12(22):22.
81. Festuccia N, Halbritter F, Corsinotti A, et al. Esrrb extinction triggers dismantling of naïve pluripotency and marks commitment to differentiation. *EMBO J.* 2018;37:e95476.
82. Stainier DY, Kontarakis Z, Rossi A. Making sense of anti-sense data. *Dev Cell.* 2015;32:7-8.
83. El-Brolosy MA, Kontarakis Z, Rossi A, et al. Genetic compensation triggered by mutant mRNA degradation. *Nature.* 2019;568:193-197.
84. Ma Z, Zhu P, Shi H, et al. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature.* 2019;568:259-263.
85. Wilkinson MF. Genetic paradox explained by nonsense. *Nature.* 2019;568:179-180.
86. Schmierer B, Hill CS. TGF β -SMAD signal transduction: molecular specificity and functional flexibility. *Nat Rev Mol Cell Biol.* 2007;8:970-982.
87. Conidi A, Cazzola S, Beets K, et al. Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGF β /BMP signaling in vivo. *Cytokine Growth Factor Rev.* 2011;22:287-300.
88. Lee KL, Lim SK, Orlov YL, et al. Graded nodal/activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. *PLoS Genet.* 2011;7:e1002130.
89. O'Connell DJ, Kolde R, Sooknah M, et al. Simultaneous pathway activity inference and gene expression analysis using RNA sequencing. *Cell Syst.* 2016;2:323-334.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Dries R, Stryjewska A, Coddens K, et al. Integrative and perturbation-based analysis of the transcriptional dynamics of TGF β /BMP system components in transition from embryonic stem cells to neural progenitors. *Stem Cells.* 2020;38:202–217. <https://doi.org/10.1002/stem.3111>