Title: Intracellular and intercellular gene regulatory networks inference from time-course individual RNA-Seq.

Authors: Makoto Kashima[1*], Yuki Shida[2], Takashi Yamashiro[2], Hiromi Hirata[1], and Hiroshi Kurosaka[2]

[1]College of Science and Engineering, Aoyama Gakuin University, 5-10-1, Fuchinobe, Sagamihara Chuo Ku, Kanagawa 252-5258, Japan;

[2]Department of Orthodontics and Dentofacial Orthopedics, Osaka University, Yamadaoka, Suita, Osaka, 565-0871, Japan.

* Corresponding author

Structured Abstract

Motivation: Recent technical advances in bulk RNA-Seq have enabled time-course RNA-Seq of whole individual embryos to understand the underlying molecular mechanisms. Thus, we hypothesized that gene regulatory networks (GRNs) inferred from time-course individual RNA-Seq during embryonic development reveal intercellular regulatory relationships involved in signaling pathways.

Results: Time-course bulk RNA-Seq of individual mouse embryos in early development, followed by pseudo-time analysis and GRN inference, demonstrated that GRN inference from RNA-Seq with pseudo-time can be applied for individual bulk RNA-Seq similar to scRNA-Seq. Validation using an experimental-source-based database showed that our approach could significantly infer GRN for all transcription factors in the database. Furthermore, the inferred ligand-related and receptor-related downstream genes were significantly overlapped. Overall, the inferred GRN include intracellular as well as intercellular regulatory relationships, which cannot be inferred from scRNA-Seq. Thus, inferring GRN from time-course bulk RNA-Seq is a novel approach for understanding

the regulatory relationships underlying cellular events.

Availability and Implementation: The R source code used in the present study is available at https://github.com/Makoto-Kashima/TIODE

Contact: kashima@chem.aoyama.ac.jp

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Regulation of gene expression is a fundamental factor that controls cellular events such as proliferation and differentiation. Understanding gene regulatory interconnections is an important step to understand the molecular mechanisms underlying cellular events. Recently, gene regulatory network (GRN) inference based on time-course data has attracted much attention in single-cell RNA-Seq (scRNA-Seq). State-of-the-art scRNA-Seq analysis techniques can generate transcriptome information from thousands of cells. (Sasagawa *et al.*, 2018, 2013; Gao *et al.*, 2020; Hayashi *et al.*, 2018; Klein *et al.*, 2015). Dimensional reduction of scRNA-Seq data provides a trajectory of cellular events and enables assignment of pseudo-time to each cell during differentiation and proliferation (Treutlein *et al.*, 2014; Haghverdi *et al.*, 2016). Because scRNA-Seq with pseudo-time is a dense time-course observation of cellular events, gene regulatory networks can be inferred by comparing the timing of gene upregulation and downregulation along pseudo-time (Matsumoto *et al.*, 2017; Aalto *et al.*, 2020).

Originally, GRN inference was applied for gene expression data from tissue and pooled cells (bulk samples), generated using DNA microarray and RNA-Seq [reviewed in (Ko and Brandizzi, 2020; Fernandez-Valverde *et al.*, 2018)]. Compared to steady-state data, time-course data allows GRN inference by comparing the timing

of gene upregulation and downregulation (Iglesias-Martinez *et al.*, 2016; Zhang *et al.*, 2019; Krouk *et al.*, 2010; Ogami *et al.*, 2012). However, GRN inference from time-course data of bulk samples has not been popular due to the following drawbacks: 1) Transcriptome analysis of a large number of bulk samples is expensive (Spies and Ciaudo, 2015), and 2) biological variance results in inconsistency between real sampling time and transcriptome status (Kilfoil *et al.*, 2009). Fortunately, recent technical advances related to bulk RNA-Seq have overcome these difficulties. Advances in sequencing platforms and bulk 3′ RNA-Seq library preparation methods have enabled cost-effective time-course bulk RNA-Seq (Muir *et al.*, 2016; Alpern *et al.*, 2019; Kamitani *et al.*, 2019; Li *et al.*, 2020). Because these methods were developed based on the methods for scRNA-Seq, they can also be applied for individual RNA-Seq (Kashima *et al.*, 2020) (for instance, RNA-Seq of a single embryo or tissue of individuals with several ng of total RNA). Pseudo-time analysis for individual RNA-Seq might capture individual differences of gene expression in the time-course of biological events such as development. After assignment of pseudo-time to each individual RNA-Seq data, GRN can be inferred similar to an inference based on scRNA-Seq.

Theoretically, GRN inferred from whole-body and tissue RNA-Seq are different from those inferred from scRNA-Seq. scRNA-Seq provides transcriptomic information at the cellular level that enables inference of intracellular GRN involved in the progression of proliferation and differentiation. In contrast, bulk RNA-Seq of whole body and tissues could contain transcriptomic information at the level of cell populations. Time course individual RNA-Seq during development would enable inference of both intracellular and intercellular GRN (cell-cell communications)

involved in the developmental process. For instance, during embryonic development, upregulation of ligand-related genes can be followed by upregulation of downstream genes in cells expressing receptor genes.

We thus hypothesized that GRNs inferred from time-course individual RNA-Seq during embryonic development would include intercellular regulatory relationships between ligand genes and downstream genes of related signaling pathways. To test this hypothesis in the present study, we conducted time-course bulk RNA-Seq of individual mouse embryos in early development, followed by pseudo-time analysis and GRN inference.

2 Materials and methods

2.1 Maintenance of mice

Embryos were collected from pregnant female Institute of Cancer Research (ICR) mice (CLEA, Tokyo, Japan) at each stage. All sacrificed female mice were housed under a 12 h dark-light cycle in which the light phase started from 8 am. All animal experiments were performed in accordance with the guidelines of the Animal Care and Use Committee of Osaka University Graduate School of Dentistry, Osaka, Japan.

2.2 RNA extraction

Total RNA extraction was performed using the RNeasy® kit (QIAGEN, Hilden, Germany) according to manufacturer's protocol. The total RNA concentration was measured using Qubit™ RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) and was adjusted to 5 ng/ul, and stored at -80 ℃ until subsequent analysis.

2.3 RNA-Seq library preparation and sequencing

Non targeted RNA-Seq were conducted according to the Lasy-Seq ver. 1.1

97 protocol (https://sites.google.com/view/lasy-seq/) (Kamitani *et al.*, 2019;
98 Kashima *et al.*, 2020). Briefly, 50 ng of total RNA were reverse-transcribed using
99 an RT primer with index and SuperScript IV reverse transcriptase (Thermo
100 Fisher Scientific, Waltham, MA, USA). Then, all RT mixtures of the samples
101 were pooled and purified using an equal volume of AMpure XP beads (Beckman
102 Coulter, Brea, CA, USA) according to the manufacturer's instructions. Second
103 strand synthesis was conducted on the pooled samples using RNaseH (5 U/μL,
104 Enzymatics, Beverly, MA, USA), and DNA polymerase I (10 U/μL, Enzymatics,
105 Beverly, MA, USA). To avoid the carryover of large amounts of rRNAs, the
106 mixture was subjected to RNase treatment using RNase T1 (Thermo Fisher
107 Scientific, Waltham, MA, USA). Then, purification was conducted with a 0.8×
108 volume of AMpure XP beads. Fragmentation, end-repair, and A-tailing were
109 conducted using 5× WGS Fragmentation Mix (Enzymatics, Beverly, MA, USA).
110 The Adapter for Lasy-Seq was ligated using 5× Ligation Mix (Enzymatics,
111 Beverly, MA, USA), and the adapter-ligated DNA was purified twice with a 0.8×
112 volume of AMpure XP beads. After optimisation of PCR cycles for library
113 amplification by qPCR using Evagreen, 20× in water (Biotium, Fremont, CA,
114 USA) and the QuantStudio5 Real-Time PCR System (Applied Biosystems,
115 Waltham, MA, USA), the library was amplified using KAPA HiFi HotStart
116 ReadyMix (KAPA BIOSYSTEMS, Wilmington, MA, USA) on the ProFlex
117 PCR System (Applied Biosystems, Waltham, MA, USA). The amplified library
118 was purified with an equal volume of AMpure XP beads. One microliter of the
119 library was then used for electrophoresis using a Bioanalyzer 2100 with the
120 Agilent High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA)

121 to check for quality. Then, sequencing of 150-bp paired-end reads was performed
122 using HiSeq X Ten (Illumina, San Diego, CA, USA).

123 **2.4 Mapping and gene quantification**

124 Read 1 reads were processed with fastp (version 0.21.0) (Chen *et al.*, 2018)
125 using the following parameters: --trim_poly_x -w 20 --
126 adapter_sequence=AGATCGGAAGAGCACACGTCTGAACTCCAGTCA --
127 adapter_sequence_r2=AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTG
128 T -l 31. The trimmed reads were then mapped to the mouse reference sequences
129 of Mus_musculus.GRCm38.cdna.all.fa, using BWA mem (version 0.7.17-
130 r1188) (Li and Durbin, 2009) with the default parameters. The read count for
131 each gene was calculated with salmon using -l IU, which specifies the library
132 type (version v0.12.0) (Patro *et al.*, 2017). Then, using R (version 4.0.1) (R
133 Core Team, 2017), the sums of read counts per gene were calculated. Genes
134 with read counts greater than zero were used in the subsequent analysis.

135 **2.5 Pseudo-time analysis**

136 Read counts were normalized using the “NormalizeData” function with the
137 default parameters in Seurat (version 4.0.0) (Hao *et al.*, 2020). For principal
138 component analysis (PCA), the normalized read counts were scaled using the
139 “ScaleData” function with the default parameters except for do.scale = F. Then,
140 PCA was conducted using the “RunPCA” function for genes with high
141 dispersion, which were selected using the “FindVariableFeatures” function
142 with default parameters except for selection.method = "mvp". Finally,
143 SingleCellExperiment (version 1.10.1) (Amezquita *et al.*, 2020) and slingshot
144 (version 1.6.1) (Street *et al.*, 2018) were used to calculate the pseudo-time for

each sample.

2.6 Gene regulatory network inference

To infer GRN, A was optimized 20 times using the SCODE algorithm (Matsumoto *et al.*, 2017) with one hundred iterations and $D = 4$. Pearson's correlation coefficients between values of each A from the 20 optimizations and the mean$A$ (the average of each value of A) were calculated. In the following analysis, we used the average values of the top 10 A showing higher correlations with the mean$A$ from 20 optimizations. To define the thresholds for downstream gene selection for each gene, the linear function was regressed using the "nls" function in R for the scatter plot of absolute values of A for downstream genes in the decreasing order, and the X and Y axes showed the integers from one to 28117, and the absolute values of A, respectively. Genes with larger absolute values of A compared to the Y values of the regressed line were defined as the inferred gene downstream of each gene.

2.7 Evaluation of inferred GRN by comparing with a TF-downstream gene database

For validating the inferred GRN, the information regarding the binding motifs and targets for 438 mouse TFs in the TF2DNA database was used (Pujato *et al.*, 2014; Badis *et al.*, 2009; Berger *et al.*, 2006; Mathelier *et al.*, 2014; Matys *et al.*, 2006; Weirauch *et al.*, 2014; Berger *et al.*, 2008; Chen *et al.*, 2011; Jolma *et al.*, 2013; Sebé-Pedrós *et al.*, 2013; Wei *et al.*, 2010; Weirauch *et al.*, 2013). Area under the curve (AUC) values for downstream prediction based on the absolute values of A were calculated using the "performance" function in ROCR (version 1.0-11) (Sing *et al.*, 2005). Statistical analysis for the

enrichment of validated target genes among the inferred genes was conducted using the "fisher.test" function in R. Statistical analysis for the overlap of inferred downstream genes for ligand-related and receptor-related genes was conducted using the "enrichment_test" function in Rvenn (version 1.1.0) (Akyol, 2019). For all statistical tests, BH correction was performed using the "p.adjust" function. The upset plots were drawn using the "upset" function in UpSetR (version 1.4.0) (Conway *et al.*, 2017).

2.8 Data availability

All RNA-Seq sequence data is deposited in PRJNA725414 of SRA.

3 Results

3.1 Pseudo-time analysis of the time-course bulk RNA-Seq for mouse embryos

To examine whether gene regulatory network (GRN) inference based on time-course bulk RNA-Seq works well, we conducted time-course bulk RNA-Seq for individual mouse embryos. RNA was extracted from each individual whole embryo (n = 10 or 11) at seven time points: E7.5, E8.5, E9.5, E10.5, E11.5, E12.5, and E.13.5, followed by 3′ RNA-Seq using the 'Lasy-Seq' method (Kamitani *et al.*, 2019). As a result, we obtained 76 RNA-Seq datasets with an average of 8.5 million reads per sample. The reads were mapped onto the mouse reference sequence, followed by calculation of the read counts of each gene in each sample. We then used 'Seurat' (Stuart *et al.*, 2019), an R package used for single cell omics analysis, normalization of read counts, detection of highly variable genes, dimension reduction of omics date, and so on. On the PC1 and PC2 plane obtained with Seurat, each sample at the same stage was close to each other (Fig. 1A). As expected, the clusters of each stage were ordered according to the developmental process (Fig. 1A). Using an R package 'slingshot'

193 (Street *et al.*, 2018), we inferred the developmental trajectory of mouse embryos and
194 calculated the pseudo-time for each sample (Fig. 1 A and B). Pseudo-time analysis
195 revealed individual differences among embryos in the speed of their developmental
196 processes. For example, the pseudo-time of a sample at E10.5 and of E11.5 were
197 almost the same (Fig. 1 A and B). Using the pseudo-time analysis, the time-course
198 RNA-Seq data of seven time points could be converted into those of 76 time points.

199 **3.2 Gene regulatory network inference based on individual RNA-Seq of whole**
200 **mouse embryos**

201 Next, we inferred a GRN from the dataset of time-course individual RNA-Seq of
202 whole mouse embryos. We used SOCDE, which solves linear ordinary differential
203 equations to infer GRN (Matsumoto *et al.*, 2017). To avoid incorrect emphasis on the
204 technical noise of RNA-Seq, we used a non-scaled normalized gene expression matrix
205 as the input for SCODE. Selection of the size of D, a parameter of SCODE, is
206 important for robust GRN inference, because a needlessly large D results in an
207 unstable inferred GRN (Matsumoto *et al.*, 2017). In this study, we used $D = 4$ similar
208 to the previous study (Matsumoto *et al.*, 2017), with which the residual sum of squares
209 (RSS) was relatively small (Fig. 2A). For 28117 genes whose read counts were greater
210 than zero, SCODE produced A (28117 × 28117 matrix) corresponding to the inferred
211 gene regulatory network, in which the value of $A_{i,j}$ indicates regulatory effects on the
212 downstream gene i from the regulator j. $A_{i,j} > 0$ indicates that the regulator j positively
213 regulates gene i, whereas $A_{i,j} < 0$ indicates the opposite. Because SCODE optimizes
214 A by random sampling, we optimized A 20 times to check for reproducibility. Then,
215 Pearson's correlation coefficients between the values of each A from the 20
216 optimizations and the mean$A$ were calculated (Fig. 2B). Almost all optimizations

217 produced similar *A* with high correlation coefficients (Fig. 2B). In the subsequent
218 analysis, we used the average values of the top 10 *A* showing higher correlations with
219 the mean*A* from 20 optimizations. We then tried to define the thresholds for significant
220 regulatory relationships between regulators and downstream genes. Because the
221 average expression level of downstream genes showed higher correlation with the
222 absolute values of *A* compared to the regulators (Fig. 2 C and D), we independently
223 defined a threshold for each regulator as follows. For example, absolute values of *A*
224 indicating regulatory relationships between *Sox8* and its regulators showed a positive
225 correlation (Pearson's correlation coefficient = 0.76) (Fig. 2E), whereas absolute
226 values of *A* showing regulatory relationships between *Sox8* and its downstream genes
227 showed smaller correlations (Pearson's correlation coefficient = 0.54) (Fig. 2F).
228 Although most of the inferred values of *A* for regulatory relationships between *Sox8*
229 and its downstream genes were around zero, some values were outliers (Fig. 2G). To
230 define the threshold for the downstream genes of *Sox8*, we regressed the linear
231 function for the scatter plot of absolute *A* values for downstream genes in the
232 decreasing order, and defined the threshold (See Material and methods) (Fig. 2G).
233 Finally, in the subsequent analysis, genes with larger absolute values of *A* compared
234 to the threshold were defined as the inferred genes downstream of *Sox8* (Fig. 2G).

235 **3.3 Validation of the inferred network**

236 Next, we evaluated the inferred GRN by comparing the inferred genes downstream
237 of the TFs with the information in the TF2DNA database, which is an experimental-
238 source-based database of the binding motifs and downstream genes for 438 mouse
239 TFs (Pujato *et al.*, 2014; Badis *et al.*, 2009; Berger *et al.*, 2006; Mathelier *et al.*, 2014;
240 Matys *et al.*, 2006; Weirauch *et al.*, 2014; Berger *et al.*, 2008; Chen *et al.*, 2011; Jolma

241 *et al.*, 2013; Sebé-Pedrós *et al.*, 2013; Wei *et al.*, 2010; Weirauch *et al.*, 2013). First,
242 we evaluated the effectiveness of target prediction based on the absolute values of A
243 by calculating the area under the curve (AUC) (Fig. 3A). The average AUC was 0.704,
244 indicating that the inferred regulatory relationships could reflect the true regulatory
245 network. Second, we examined the validity of the defined thresholds by checking the
246 differences between the validated target gene rate (number of validated target genes /
247 number of all inferred downstream genes) above the defined thresholds and the best
248 validated rate (Fig. 3B). Below the thresholds, validated target gene rates for the 438
249 TFs were 0.69% smaller than the best validated target gene rates on average. The
250 validated target genes were 67.5% of the inferred downstream genes on average
251 (Table S1, Fig. 3C and D). Compared with the background rate (number of known
252 target genes / number of all genes), the validated target gene rates of the inferred
253 downstream genes of all TFs were statistically higher (adjusted p-value < 0.01) (Fig.
254 3D). These results suggest that our approach could infer the GRN underlying early
255 development in mice.

256 **3.4 The inferred network contained regulatory relationships involved in cell-**
257 **cell interaction**

258 Our GRN inference was based on bulk RNA-Seq containing the information of all
259 cells in the body. We thus hypothesized that the inferred GRN also included the
260 intercellular regulatory network. To examine this possibility, we checked the overlaps
261 of inferred downstream genes for genes related to the ligands and receptors of nine
262 major signaling pathways (Fig. 4A): Wnt/βcatenin, TNF, TGF-β, Hedgehog, FGF,
263 EGF, Delta/Notch, BMP, and Retinoic acid (RA) signaling pathways (Table S2). As
264 expected, the inferred downstream genes of all pairs of ligand and receptor genes were

265 significantly overlapped (adjusted p-value < 0.01) (Fig. 4A). On average, 94.2% of
266 the inferred downstream genes for the ligand-related and receptor-related genes were
267 overlapped (Fig. 4A). For example, 4510 genes were inferred as the downstream of
268 *Wnt* genes, and 4545 genes were downstream of *Fzd* genes. However, 97.5% of the
269 inferred downstream of *Wnt* genes were also inferred as the downstream of *Fzd* genes
270 (Fig. 4A). Because SCODE can infer whether each downstream gene is positively or
271 negatively regulated (Matsumoto *et al.*, 2017), we checked the overlaps of positively
272 and negatively regulated gene downstream of the representative ligand-related and
273 receptor-related genes of each signaling pathway, i.e., with the largest number of
274 inferred downstream genes among each gene family (Fig. 4 B-D and Supplementary
275 Figure 1,2). For six signaling pathways (TNFβ, TGF, FGF, EGF, Delta/Notch, and
276 BMP signaling pathways), which activate the downstream genes only when ligands
277 bind to the receptors (Gilbert and Barresi, 2017), the regulatory directions of ligand-
278 related and receptor-related genes for most inferred downstream genes were the same
279 (Fig. 4B and Supplementary Figure 1). In contrast, in case of the RA signaling
280 pathway, wherein RA receptors function as transcriptional repressors without RA
281 binding (Supplementary Figure 2A) (Glass and Rosenfeld, 2000), the regulatory
282 directions of a ligand-related gene, *Aldh1a3,* which encodes a protein involved in RA
283 synthesis, and a receptor-related gene, *Rara,* for the downstream genes were opposite
284 (Fig. 4C). In case of the Hedgehog signaling pathway, the regulatory directions of
285 *Ssh2* and *Ptchd4* tended to be opposite (Fig. 4D). The regulatory directions of *Ssh2*
286 and *Gli3* for the downstream genes were also opposite (Fig. 4D). In the absence of
287 Hedgehog ligands, the full-length Gli family proteins are ubiquitinated and function
288 as repressors (Supplementary Figure 2B) (Skoda *et al.*, 2018). With the binding of

289 Hedgehog ligands, PATCHED inhibits the degradation of Gli family proteins and
290 allows them to function as transcriptional activators (Supplementary Figure 2A)
291 (Skoda *et al.*, 2018). Thus, the overlap of downstream genes that are positively and
292 negatively regulated by *Ssh2, Ptchd4,* and *Gli3* is reasonable. In case of the
293 Wnt/βcatenin signaling pathway, the regulatory directions of *Wnt7a* and *Fzd5* for the
294 downstream were opposite (Supplementary Figure 3A). Our time-course RNA-Seq
295 revealed that among the *Fzd* and *Wnt* genes, *Fzd5* and *Wnt8a,* but not *Wnt7a,* were
296 the only genes expressed dominantly in the early developmental stage among the
297 protein families. (Supplementary Figures 4 and 5). Similar regulatory directions were
298 found for most of the inferred downstream genes of *Wnt8a-Fzd5*, which could be a
299 functional pair in the early developmental stages. Furthermore, the regulatory
300 directions for most of the inferred downstream genes of *Wnt7a-Fzd6* (*Fzd* genes with
301 the second most inferred downstream genes) were also the same (Supplementary
302 Figure 3). In conclusion, our approach allowed successful inference the intercellular
303 regulatory relationships related to the major signaling pathways as well as the
304 intracellular pathways related to TFs.

305 **4 Discussion**

306 Recently, GRN inference based on the combination of scRNA-Seq and pseudo-time
307 analysis has attracted much attention (Dai *et al.*, 2020). However, to our knowledge,
308 this study is the first to report GRN inference based on the combination of individual
309 bulk RNA-Seq and pseudo-time analysis. Unlike scRNA-Seq, which can provide
310 transcriptomic dynamics in a certain cellular event such as proliferation and
311 differentiation, bulk RNA-Seq could provide a mixture of various transcriptomic
312 dynamics regarding the cellular events occurring in an embryo (Chasman and Roy,

313 2017). Theoretically, GRN inference based on bulk RNA-Seq cannot provide
314 multiple lineage-specific regulatory relationships. This might be the reason why
315 GRN inference based on bulk RNA-Seq has not been attempted as with scRNA-Seq.
316 In this study, we succeeded in inferring the known regulatory relationships of the TFs
317 in the TF2DNA database with high AUC compared to the GRN inferred from
318 scRNA-Seq (Chen and Mar, 2018) (Fig. 3). This suggests that GRN inference from
319 bulk RNA-Seq can provide insights into the regulatory interconnections of genes,
320 even though GRN inferred from bulk RNA-Seq would have limitations in the
321 comprehensiveness of inferred regulatory relationships compared to those from
322 scRNA-Seq. Furthermore, drastic changes in the cell population occur during early
323 developments. It is possible that our GRN inferred from bulk RNA-Seq merely
324 reflect the changes in cell populations but not the interconnections of genes.
325 Considering the results of the inference for *Wnt* and *Fzd* genes (Supplementary
326 Figure 2), these points should be considered when interpreting the inferred GRN
327 based on the time-course bulk RNA-Seq.

328 Assignment of pseudo-time is an important step in GRN inference from time-
329 course data. In case of scRNA-Seq, the accuracy of pseudo-time assignment is
330 controversial (Tritschler *et al.*, 2019). In contrast, in the case of time-course
331 individual bulk RNA-Seq, the accuracy of pseudo-time assignment could be assured
332 by the actual sampling time, which can be an advantage compared to GRN inference
333 from scRNA-Seq.

334 GRN indicate the intracellular interconnections of genes in a narrow sense;
335 intercellular regulation of genes via cell-cell communication is also a key factor to
336 understand the regulatory mechanisms underlying multicellular organisms. Several

studies have tried to systematically identify cell-cell communications based on single cell gene expression profiles and information regarding ligand-receptor pairs (Kumar *et al.*, 2018; Jin *et al.*, 2021; Cabello-Aguilar *et al.*, 2021; Wang *et al.*, 2019). Because these approaches required prior knowledge, they could only be applied for the major model organisms and could not reveal novel signaling pathways. We demonstrated that GRN inference based on time-course individual RNA-Seq could infer intercellular regulatory relationships related to cell-cell communication via cell signaling pathways. This approach only requires time-course RNA-Seq results and is applicable for non-model organisms without ligand-receptor information. Theoretically, the GRN inferred based on scRNA-Seq cannot include intercellular interconnections of genes. Comparison of GRN from scRNA-Seq and bulk RNA-Seq would be useful for systematic identification of genes involved in cell-cell communications during cellular events of interest. Taken together, our approach is a powerful tool for understanding intracellular and intercellular regulatory relationships of genes, which cannot be achieved using existing GRN inferences based on scRNA-Seq alone. As discussed above, bulk RNA-Seq has limitations in the comprehensiveness of inferred regulatory relationship because multiple lineage-specific regulatory relationships cannot be deduced from GRN inference based on bulk RNA-Seq. A future novel bioinformatic approach that can deconvolute gene expression in each tissue and cell lineage from bulk RNA-Seq will help overcoming this limitation.

Finding: This work was supported by Japan Society for the Promotion of Science [grant number JP19H03858 to H.K.].

Disclosures: Conflicts of interest: No conflicts of interest declared.

Acknowledgments. We would like to thank Editage (www.editage.com) for English language editing.

References

Aalto,A. *et al.* (2020) Gene regulatory network inference from sparsely sampled noisy data. *Nat. Commun.*, **11**, 3493.

Akyol,T.Y. (2019) RVenn: An R package for set operations on multiple sets.

Alpern,D. *et al.* (2019) BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. *Genome Biol.*, **20**, 71.

Amezquita,R.A. *et al.* (2020) Orchestrating single-cell analysis with Bioconductor. *Nat. Methods*, **17**, 137–145.

Badis,G. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science (80-.).*, **324**, 1720–1723.

Berger,M.F. *et al.* (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.*, **24**, 1429–1435.

Berger,M.F. *et al.* (2008) Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell*, **133**, 1266–1276.

Cabello-Aguilar,S. *et al.* (2021) SingleCellSignalR: Inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.*, **48**, 55.

Chasman,D. and Roy,S. (2017) Inference of cell type specific regulatory networks on mammalian lineages. *Curr. Opin. Syst. Biol.*, **2**, 130–139.

Chen,L. *et al.* (2011) hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data. *Bioinformatics*, **27**,

385 1447–1448.
386 Chen,S. *et al.* (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor.
387 *Bioinformatics*, **34**, i884–i890.
388 Chen,S. and Mar,J.C. (2018) Evaluating methods of inferring gene regulatory networks
389 highlights their lack of performance for single cell gene expression data. *BMC*
390 *Bioinformatics*, **19**, 232.
391 Conway,J.R. *et al.* (2017) UpSetR: an R package for the visualization of intersecting
392 sets and their properties. *Bioinformatics*, **33**, 2938–2940.
393 Dai,H. *et al.* (2020) Reconstructing gene regulatory networks in single-cell
394 transcriptomic data analysis. *Zool. Res.*, **41**, 599–604.
395 Fernandez-Valverde,S.L. *et al.* (2018) Inference of developmental gene regulatory
396 networks beyond classical model systems: New approaches in the post-genomic
397 era. *Integr. Comp. Biol.*, **58**, 640–653.
398 Gao,C. *et al.* (2020) The Comparison of Two Single-cell Sequencing Platforms: BD
399 Rhapsody and 10x Genomics Chromium. *Curr. Genomics*, **21**, 602–609.
400 Gilbert,S.F. and Barresi,M.J.F. (2017) DEVELOPMENTAL BIOLOGY, 11TH
401 EDITION 2016. *Am. J. Med. Genet. Part A*, **173**, 1430–1430.
402 Glass,C.K. and Rosenfeld,M.G. (2000) The coregulator exchange in transcriptional
403 functions of nuclear receptors. *Genes Dev.*, **14**, 121–141.
404 Haghverdi,L. *et al.* (2016) Diffusion pseudotime robustly reconstructs lineage
405 branching. *Nat. Methods*, **13**, 845–848.
406 Hao,Y. *et al.* (2020) Integrated analysis of multimodal single-cell data. *bioRxiv*,
407 2020.10.12.335331.
408 Hayashi,T. *et al.* (2018) Single-cell full-length total RNA sequencing uncovers

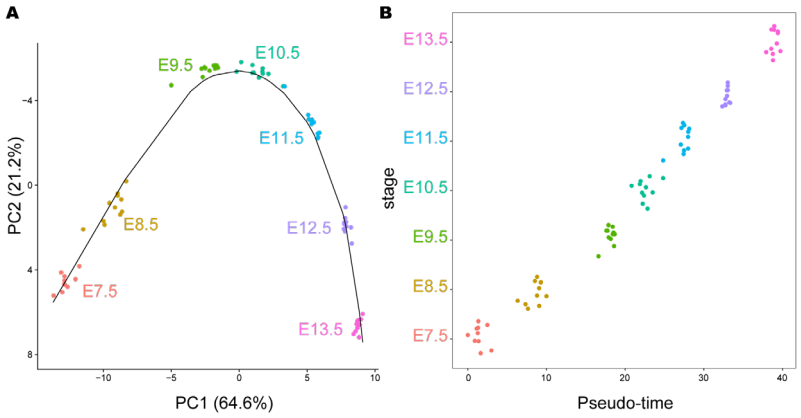
409 dynamics of recursive splicing and enhancer RNAs. *Nat. Commun.*, **9**, 1–16.
410 Iglesias-Martinez,L.F. *et al.* (2016) BGRMI: A method for inferring gene regulatory
411 networks from time-course gene expression data and its application in breast
412 cancer research. *Sci. Rep.*, **6**, 1–12.
413 Jin,S. *et al.* (2021) Inference and analysis of cell-cell communication using CellChat.
414 *Nat. Commun.*, **12**, 1–20.
415 Jolma,A. *et al.* (2013) DNA-binding specificities of human transcription factors. *Cell*,
416 **152**, 327–339.
417 Kamitani,M. *et al.* (2019) Lasy-Seq: a high-throughput library preparation method for
418 RNA-Seq and its application in the analysis of plant responses to fluctuating
419 temperatures. *Sci. Rep.*, **9**, 7091.
420 Kashima,M. *et al.* (2020) DeLTa-Seq: direct-lysate targeted RNA-Seq from crude tissue
421 lysate. (8/15 words) Running Title: Development of direct-lysate targeted RNA-
422 Seq method Corresponding Author. *bioRxiv*, 2020.09.15.299180.
423 Kilfoil,M.L. *et al.* (2009) Stochastic variation: From single cells to superorganisms.
424 *HFSP J.*, **3**, 379–385.
425 Klein,A.M. *et al.* (2015) Droplet barcoding for single-cell transcriptomics applied to
426 embryonic stem cells. *Cell*, **161**, 1187–1201.
427 Ko,D.K. and Brandizzi,F. (2020) Network-based approaches for understanding gene
428 regulation and function in plants. *Plant J.*, **104**, 302–317.
429 Krouk,G. *et al.* (2010) Predictive network modeling of the high-resolution dynamic
430 plant transcriptome in response to nitrate. *Genome Biol.*, **11**, R123.
431 Kumar,M.P. *et al.* (2018) Analysis of Single-Cell RNA-Seq Identifies Cell-Cell
432 Communication Associated with Tumor Characteristics. *Cell Rep.*, **25**, 1458-

433 1468.e4.
434 Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-
435 Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
436 Li,Y. *et al.* (2020) Decode-seq: A practical approach to improve differential gene
437 expression analysis. *Genome Biol.*, **21**, 66.
438 Mathelier,A. *et al.* (2014) JASPAR 2014: An extensively expanded and updated open-
439 access database of transcription factor binding profiles. *Nucleic Acids Res.*, **42**,
440 D142–D147.
441 Matsumoto,H. *et al.* (2017) SCODE: an efficient regulatory network inference
442 algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics*, **33**,
443 2314–2321.
444 Matys,V. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional
445 gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108.
446 Muir,P. *et al.* (2016) The real cost of sequencing: scaling computation to keep pace with
447 data generation. *Genome Biol.*, **17**, 53.
448 Ogami,K. *et al.* (2012) Computational gene network analysis reveals TNF-induced
449 angiogenesis. *BMC Syst. Biol.*, **6**, 3–8.
450 Patro,R. *et al.* (2017) Salmon provides fast and bias-aware quantification of transcript
451 expression. *Nat. Methods*, **14**, 417–419.
452 Pujato,M. *et al.* (2014) Prediction of DNA binding motifs from 3D models of
453 transcription factors; identifying TLX3 regulated genes. *Nucleic Acids Res.*, **42**,
454 13500–13512.
455 R Core Team (2017) R: A language and environment for statistical computing. R
456 Foundation for Statis-tical Computing, Vienna, Austria.

457 Sasagawa,Y. *et al.* (2013) Quartz-Seq: a highly reproducible and sensitive single-cell
458 RNA-Seq reveals non-genetic gene expression heterogeneity. *Genome Biol.*, **14**,
459 R31.
460 Sasagawa,Y. *et al.* (2018) Quartz-Seq2: A high-throughput single-cell RNA-sequencing
461 method that effectively uses limited sequence reads. *Genome Biol.*, **19**, 1–24.
462 Sebé-Pedrós,A. *et al.* (2013) Early eèolution of the T-box transcription factor family.
463 *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 16050–16055.
464 Sing,T. *et al.* (2005) ROCR: visualizing classifier performance in R. *Bioinformatics*, **21**,
465 3940–3941.
466 Skoda,A.M. *et al.* (2018) The role of the hedgehog signaling pathway in cancer: A
467 comprehensive review. *Bosn. J. Basic Med. Sci.*, **18**, 8–20.
468 Spies,D. and Ciaudo,C. (2015) Dynamics in Transcriptomics: Advancements in RNA-
469 seq Time Course and Downstream Analysis. *Comput. Struct. Biotechnol. J.*, **13**,
470 469–477.
471 Street,K. *et al.* (2018) Slingshot: Cell lineage and pseudotime inference for single-cell
472 transcriptomics. *BMC Genomics*, **19**, 477.
473 Stuart,T. *et al.* (2019) Comprehensive Integration of Single-Cell Data. *Cell*, **177**, 1888-
474 1902.e21.
475 Treutlein,B. *et al.* (2014) Reconstructing lineage hierarchies of the distal lung
476 epithelium using single-cell RNA-seq. *Nature*, **509**, 371–375.
477 Tritschler,S. *et al.* (2019) Concepts and limitations for learning developmental
478 trajectories from single cell genomics. *Dev.*, **146**.
479 Wang,S. *et al.* (2019) Cell lineage and communication network inference via
480 optimization for single-cell transcriptomics. *Nucleic Acids Res.*, **47**, 66.

481 Wei,G.H. *et al.* (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and
482 in vivo. *EMBO J.*, **29**, 2147–2160.
483 Weirauch,M.T. *et al.* (2014) Determination and inference of eukaryotic transcription
484 factor sequence specificity. *Cell*, **158**, 1431–1443.
485 Weirauch,M.T. *et al.* (2013) Evaluation of methods for modeling transcription factor
486 sequence specificity. *Nat. Biotechnol.*, **31**, 126–134.
487 Zhang,J. *et al.* (2019) Differential regulatory network-based quantification and
488 prioritization of key genes underlying cancer drug resistance based on time-course
489 RNA-seq data. *PLOS Comput. Biol.*, **15**, e1007435.
490
491

492 Figures



493

494 **Figure 1 Pseudo-time analysis for time-course individual bulk RNA-Seq of mouse**

495 **embryos in early development**

496 (A) Transcriptomic trajectory of mouse embryos from E7.5 to E13.5. Each point on the

497 PC1-PC2 plane indicates each individual RNA-Seq result. The solid line indicates an

498 inferred trajectory by slingshot. (B) A scatterplot of pseudo-time for each sample. Each

499 point indicates each individual RNA-Seq result and is jittered along the Y-axis.

500

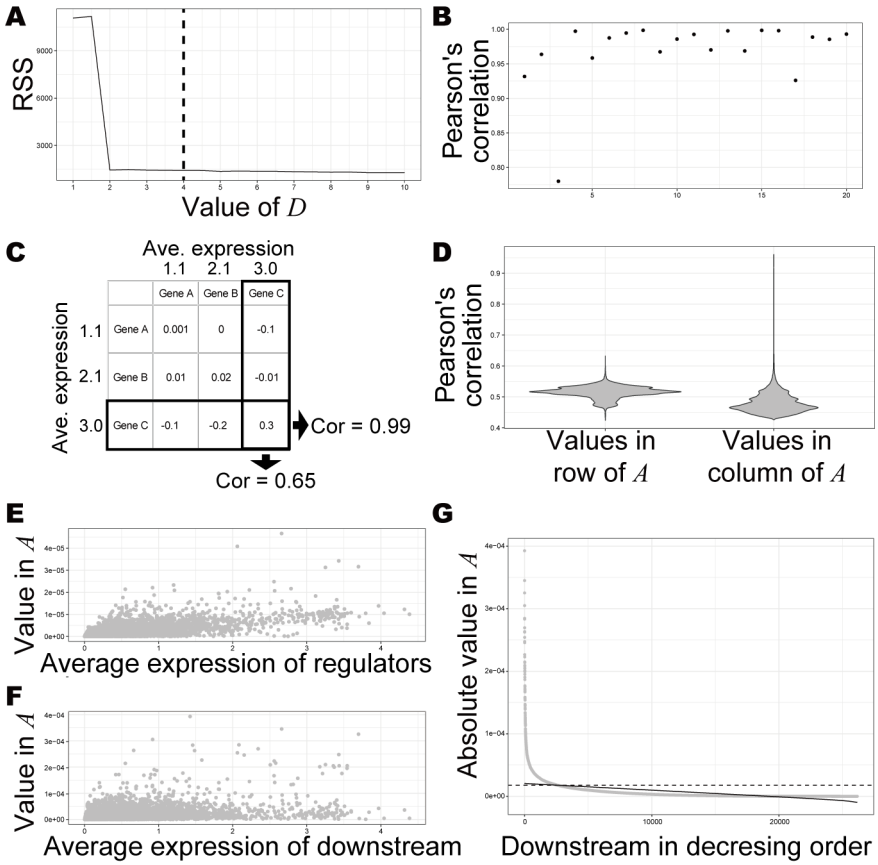


Figure 2. Inference of downstream genes based on the values in *A*

(A) Selection of a parameter *D* for SCODE. The residual sum of squares (RSS) for each *D* (from 1 to 10 by 0.5) was calculated. In this study, we used *D* = 4. (B) Pearson's correlation coefficients between the values of each *A* from 20 optimizations and the mean*A,* which is the average of each value from all optimizations. (C) Calculation of Pearson's correlations between the average expression level of each gene and the absolute values in each row and column of *A*. (D) The violin plots of Pearson's correlations (E, F) Scatter plot of the average expression level of each gene and the absolute values in the column (E) and row (F) of *A* for *Sox8*. (G) An example of the definition of thresholds for significant regulatory relationship between the regulator and downstream genes. A scatter plot of the values of *A* for a regulator, *Sox8,* in the decreasing order. The solid line is a

regression line for the scatter plot. The dashed lines indicate the defined thresholds.

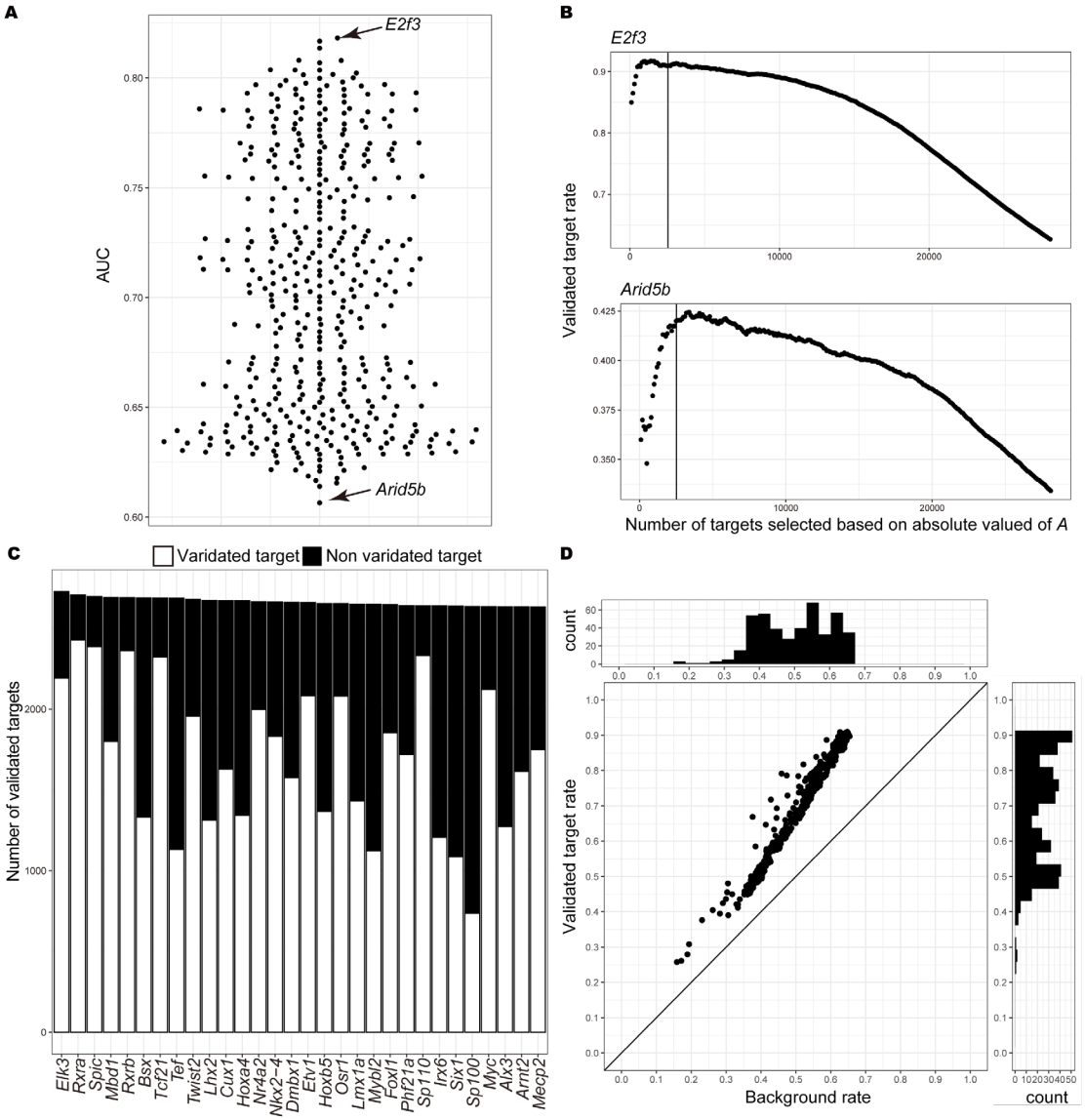


Figure 3. Validation of the inferred regulatory network of transcription factors with the TF2DNA database

(A) Scatter plot of area under the curve (AUC) target selection for each transcription factor (TF) based on the absolute values of *A*. (B) Scatter plots of the validated inferred target gene rate of *E2f3* and *Arid5b* in the TF2DNA database. Downstream genes of TFs were selected based on the absolute values of *A* in decreasing order. Solid lines indicate genes at the thresholds. (C) Bar graph of validated and non-validated downstream genes

523 of each TF in the TF2DNA database, in decreasing order of the total number of inferred
524 downstream genes. Only the top 30 TFs are shown. (D) Scatter plot of the validated target
525 gene rate of the inferred downstream genes of TF*j* and the background rate of target genes
526 in the TF2DNA database. Histograms show the distribution of validated target gene rates
527 of inferred downstream genes and the background rates of downstream genes in the
528 TF2DNA database.
529

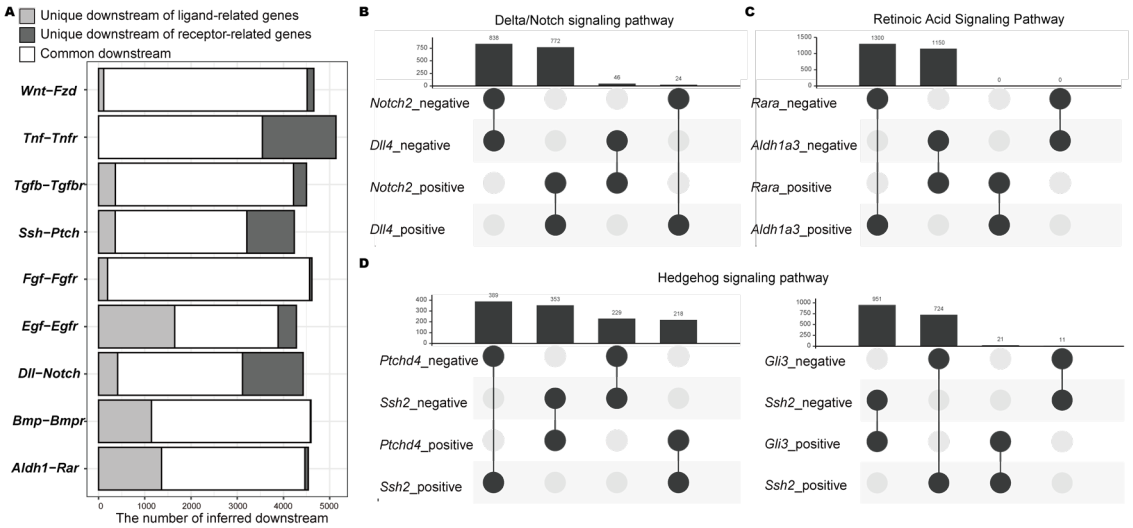
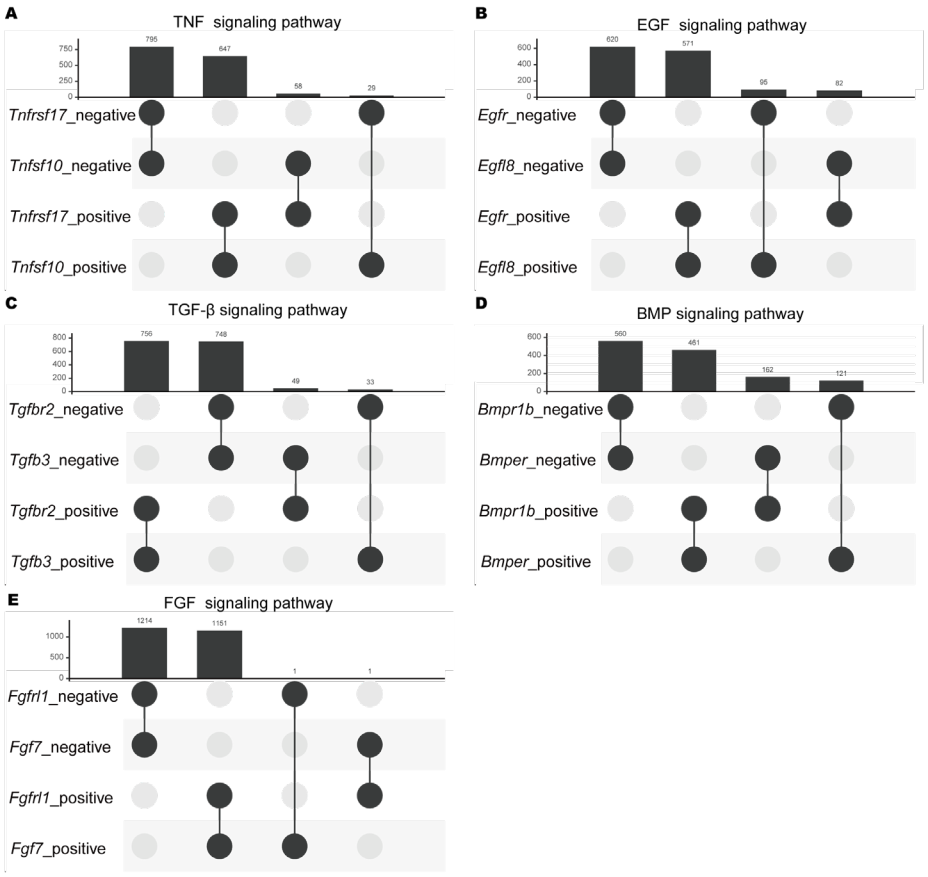


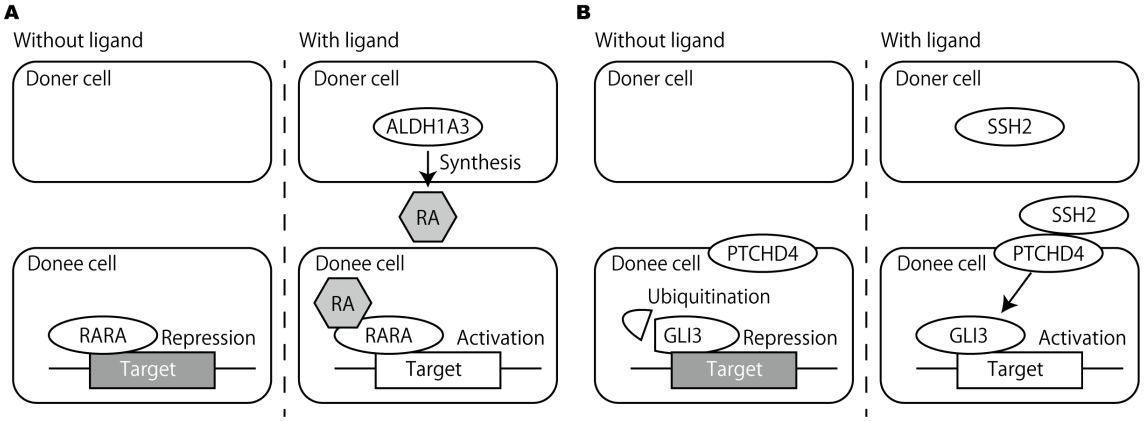
Figure 4. Overlap of inferred genes downstream of ligand-related and receptor-related genes

(A) Bar plot of the number of inferred downstream genes that are common and unique for each ligand-receptor pair. (B-E) Upset plots of inferred downstream genes that are positively and negatively regulated by the representative ligand-related and receptor-related genes. (B) Retinoic acid signaling pathway. (C) Delta/Notch signaling pathway. (D) Hedgehog signaling pathway.



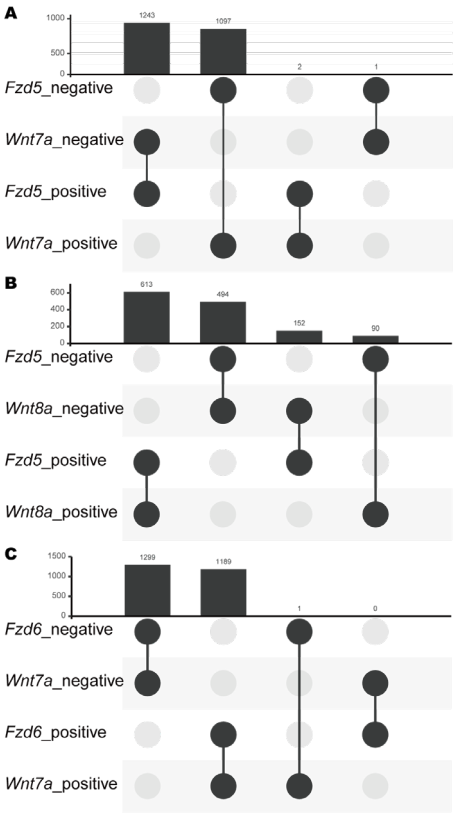
Supplementary Figure 1. Overlaps of inferred downstream genes for ligand-related and receptor-related genes

Upset plots of inferred downstream genes that were positively and negatively regulated by the representative ligand-related and receptor-related genes. (A) TNF signaling pathway. (B) EGF signaling pathway. (C) TGFβ signaling pathway. (D) BMP signaling pathway. (E) FGF signaling pathway.



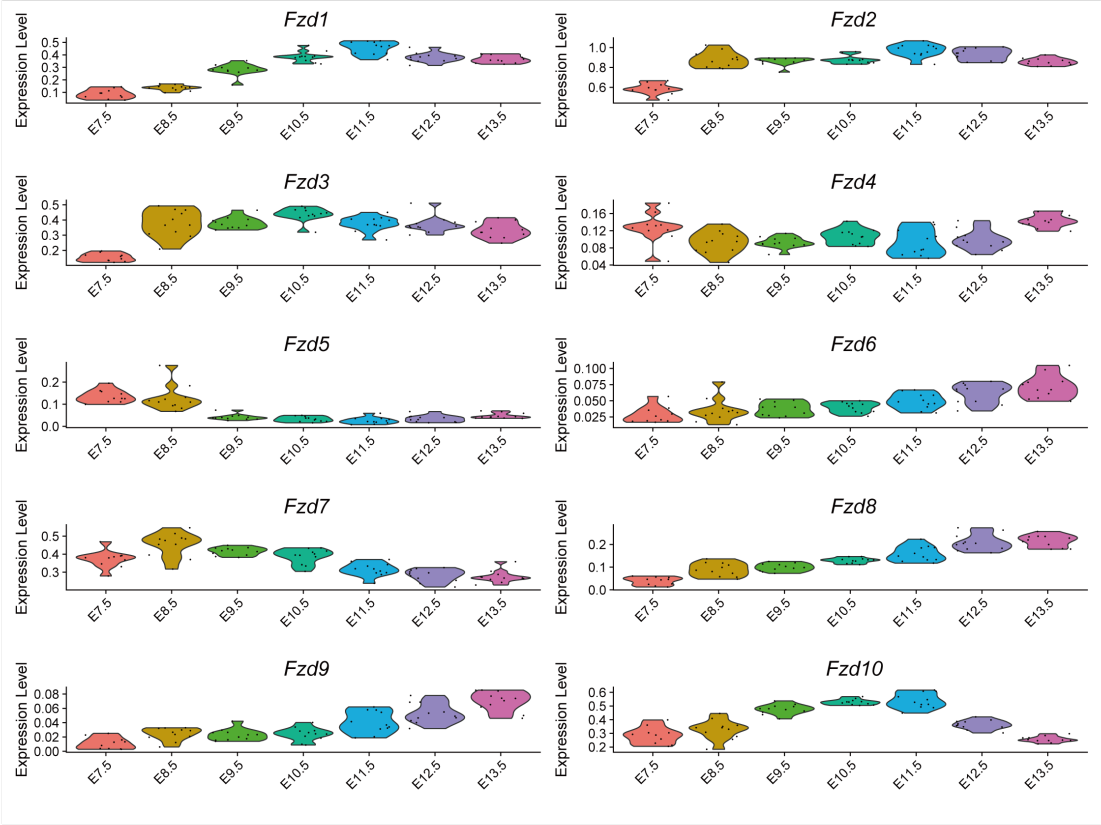
Supplementary Figure 2. Schematic illustration of the retinoic acid and Hedgehog signaling pathway

(A) In the absence of retinoic acid (RA), RA receptor (RARA) represses the target genes. ALDH1A (ALDH1A3) synthesizes RA. In the presence of RA, RA receptor activates the target genes. (B) In the absence of Hedgehog ligand (SSH2), GLI protein (GLI3) is ubiquitinated to function as a repressor of the target genes. In the presence of Hedgehog ligand, PATCHED (PTCHD4) inhibits GLI ubiquitination to activate the target genes.



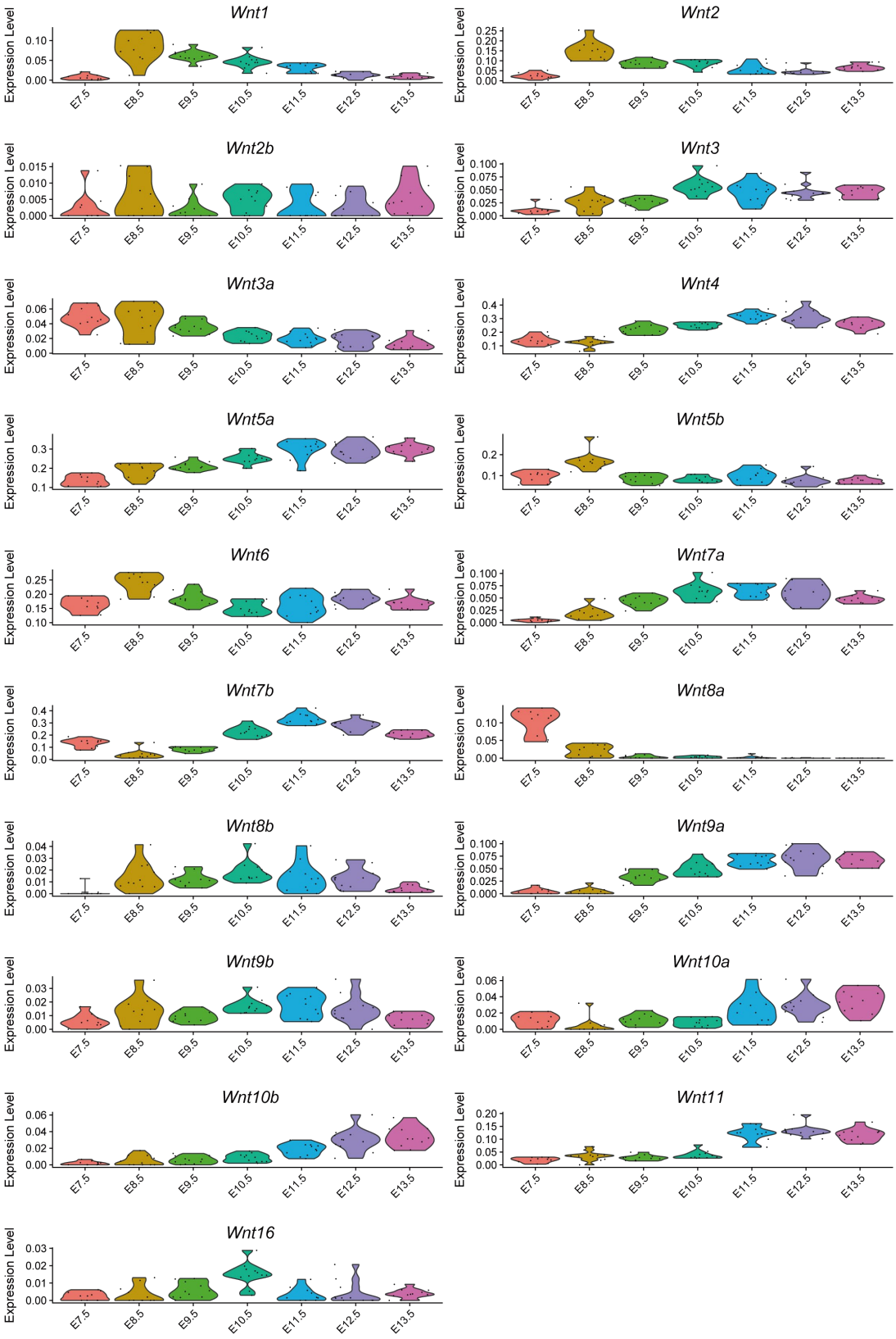
Supplementary Figure 3. Overlaps of inferred downstream genes of *Wnt7a, Wnt8a, Fzd5,* and *Fzd6*

Upset plots of inferred downstream genes that were positively and negatively regulated by the representative *Wnt7a* and *Fzd5* (A), *Wnt8a* and *Fzd5* (B), and *Wnt7a* and *Fzd6* (C).



Supplementary Figure 4. Gene expression dynamics of *Fzd* genes

Violin plots for the relative gene expression of *Fzd* genes from E7.5 to E13.5.


Wnt1
Wnt2
Wnt2b
Wnt3
Wnt3a
Wnt4
Wnt5a
Wnt5b
Wnt6
Wnt7a
Wnt7b
Wnt8a
Wnt8b
Wnt9a
Wnt9b
Wnt10a
Wnt10b
Wnt11
Wnt16
Expression Level
E7.5
E8.5
E9.5
E10.5
E11.5
E12.5
E13.5

564

565 **Supplementary Figure 5. Gene expression dynamics of *Wnt* genes**

566 Violin plots for the relative expression of *Wnt* genes from E7.5 to E13.5.

567