1Predicting placenta transcriptional regulatory interactions based on immunohistochemistry

2images data and convolutional neural network

3Yajun Liu ^{1,2*}, Yilin Guo ^{1,2}, Ya Gao ¹, Guiming Hu ³, Jun Ma ⁴, Jinquan Cui ^{1,2*}

4

51. Department of Obstetrics and Gynecology, The Second Affiliated Hospital of Zhengzhou

6University, Zhengzhou, China; Zhengzhou, China;

72. Henan Clinical Research Center for Gynecology Oncology, the Second Affiliated Hospital of

8Zhengzhou University, Zhengzhou, China;

93. Department of Pathology, the Second Affiliated Hospital of Zhengzhou University, Zhengzhou,

10China;

114. Department of Clinical Laboratory, the Second Affiliated Hospital of Zhengzhou University,

12Zhengzhou, China;

13

14

15*Corresponding authors

16Yajun Liu

17Contact information: Second Affiliated Hospital of Zhengzhou University, Henan No. 2, Jingba

18Road, Zhengzhou, China. e-mail : liuyajun_biology@126.com

19Jinquan Cui

20Contact information: Second Affiliated Hospital of Zhengzhou University, Henan No. 2, Jingba

21Road, Zhengzhou, China. e-mail : jinquan_cui@163.com

22

23ORCID for corresponding author:

24Yajun Liu 0000-0002-8203-5762

25Jinquan Cui 0000-0003-3708-1839

26

27

28Abstract

29Aims: The dysfunction of placenta development is correlated to the defects of pregnancy and
30fetal growth. The detailed molecular mechanism of placenta development is not identified in
31human due to the lack of material in vivo. Image-based reconstructions of GRN are still very
32underdeveloped.

33Methods and Results: In this study, immunohistochemistry images of different transcription
34factor (TF)s in chorionic villus were obtained by a high-resolution scanner. Next, we use
35convolutional neural network and machine learning methods, based on transfer learning
36technology, to extract the corresponding features from these images, obtain the pattern
37discrimination based on the linear regression model, and calculate the corresponding confusion
38matrix. The experimental results show that deep learning models reveals regulatory roles that
39have not yet been fully recognized. The spatial expression data reveal new regulatory
40relationships that traditional experiments have failed to recognize, and has allowed the
41development of gene regulation networks based on the spatial distribution of gene expression.

42Conclusions: We demonstrate the effectiveness of this approach in building networks using
43high-resolution images of human placenta. Our analysis is of certain significance for further
44exploration of the development of placenta and the occurrence of pregnancy-related diseases in
45the future. The datasets and analysis provide a useful source for the researchers in the field of
46the maternal-fetal interface and the establishment of pregnancy.

47

48Keywords : placenta; image processing; convolutional neural network; transcription factor

49

50Introduction

51

52 Implantation failure and insufficient placental development are pregnancy problems. The
53spatial and temporal features of gene expression are crucial for understanding transcriptional
54regulation [1], and the early stages of pregnancy in particular have a significant impact on
55pregnancy outcomes [2].

56 Gene regulatory networks (GRN) are usually represented by graph data structures in which
57nodes and edges denote genes and their interactions. It is notoriously difficult to conclude a GRN,
58and in recent decades a variety of models have been developed to reconstruct GRN and
59gene-gene interactions [3]. The basic knowledge of GRN was obtained from a large number of
60known and unknown genes, such as the genes discovered in the RNA-Seq experiment, motif
61prediction and gene regulation, as well as from other experimental data [4].

62 Advances in high-throughput technologies have led to the development of new methods for
63analyzing GRNs, and have enabled the creation of a wide range of applications in bioinformatics
64and biomedical research [5].The ability to sequence DNase hypersensitive sites and use
65transposase accessible chromatin sequencing has been developed to investigate the regulatory
66relationship between transcription factors. Previously study by us identified 154 TFs for DNase I
67hypersensitive sites (DHSs) based on publicly available chromatin accessibility data of
68BMP4-induced trophoblast (TB) cell lines [6].

69 Numerous statistical and algorithmic approaches have also been developed to draw
70conclusions about the molecular basis and regulatory circuits of gene interaction networks on the
71basis of data from the genome wide association study and other methods [7]. However,
72traditional unsupervised methods, which rely mainly on image similarity, have failed to capture
73the local hidden patterns associated with regulatory interactions, particularly for functional gene
74networks. A major paradigm shift has taken place in the field of biological imaging, with virtually
75all scientific fields, including biological image analysis, being replaced by the use of deep learning

76as the primary tool of choice for quantitative imaging analysis [8]. The scientific literature on
77deep learning has grown explosively in recent years[9] [10].

78 In this study, human first-trimester chorionic villus tissue was collected. Next, we analysis
79the expression of important trophoblastic cell-specific factors (*JUN*, *FOS*, *TFAP2A*, *TFAP2C*, *TEAD1*,
80*TEAD3*, *TEAD4*, *GATA2*, *GATA3*) by study immunofluorescence images of tissue sections obtained
81by a high-resolution scanner to capture the abundance of mRNA. We use image feature
82extraction methods to fulfill image-based GRN reconstruction based on state-of-the-art deep
83learning models. To overcome the problems of imbalanced training data set, transfer learning
84strategies were used to tackle the imbalanced training set problem. The experimental result
85shows the ability to identify the regulatory role of a known GRN during the development of
86human placenta. We characterize the molecular basis and regulatory circuits of gene interaction
87networks in human placenta. The results of our study provide insights into the biological
88mechanisms behind the development of pregnancy and pregnancy outcomes, as well as the
89molecular mechanism of placental development in human pregnancy.

90

91

92

93Results :

94

95The expression level of trophoblast specific transcription factors in Chorionic villus

96 Chorionic villus exhibits typical chorionic villus tissue morphology according to hematoxylin
97and eosin stain (H&E tissue sections) (Figure S1). Immunofluorescence biopsies of antibodies to
98different transcription factors based on the same patient source were obtained by
99high-resolution scanning. The results showed that *JUN*, *FOS*, *TFAP2A*, *TFAP2C*, *TEAD4*, *GATA2*,
100*GATA3* were most strongly expressed in chorionic villus (Figure 2B). *TEAD1*, *TEAD3* can be
101detected, though at lower levels than other genes. As a control, the well-studied chorionic villus
102surface marker *HLA-G* were also expressed in the chorionic villus (Figure 1A).

103 Next, we analyzed the expression of trophoblastic cell-specific transcription factors in

transcriptome data of chorionic villus and decidua from six first-trimester donors were obtained by high-throughput sequencing. *JUN*, *FOS*, *TFAP2A*, *TFAP2C*, *TEAD4*, *GATA2*, *GATA3* and Kisspeptin and its receptors *KISS*, which play an essential role in the establishment of the maternal-fetal dialogue, are highly expressed in chorionic villus. Progesterone-associated endometrial protein (PAEP) is highly expressed in decidua as control (Figure 1B).

109

110

111 Deep learning Experiment Setup

112 The input image is a tissue section taken by a high-resolution scanner. The first step is to
113 split the image into small blocks and enter the patches to calculate the classification vector.
114 Finally, logistic regression is used to classify cell images based on feature vectors. A schematic
115 diagram of the system pipeline is shown in Figure 2B.

116 To ensure other researcher with no coding can implement our neural network, it was based
117 on the open-source Deep Learning library implementation in The Konstanz Information Miner
118 (KNIME) Analytics Platform [11]. In order to train the deep and revolutionary network, we
119 introduced a transfer learning strategy and all experiments were carried out with Keras for the
120 prediction system and the logistic regression for the prediction of gene regulation interactions.
121 (Fig. 2B). For the training process of each data fold, we ran 40 epochs. Of the training sample data,
122 60% and 40% were used for training and validation, respectively.

123 Take VGG16 as an example, the architecture is composed of 5 blocks of 13 convolutional
124 layers (Fig. 2A). For each convolutional layer, convolve with the layer input with a convolution
125 kernel of size 3 by 3, and get an output tensor. The output tensor of the convolutional layer is
126 converted to a finite value through the activation function of the rectifying linear unit (ReLU), that
127 is, $F(x) = x$ when $x > 0$ and 0 otherwise. At the end of each block in the convolutional layer, a
128 max-pooling layer is used for down-sampling, and the output feature map of each block is
129 divided into 2×2 pooling regions, and the maximum value of each region is calculated. The
130 down-sampled feature map of each maximum pooling layer then inputs to the next convolutional
131 layer. In the fine-tuning stage, to improve the generalization ability, the Binary Cross Entropy (BCE)

optimizer driving quantity is selected. A small learning rate of 0.0001 is used to fine-tune the weight without changing the overall weight structure.

134

Classification and evaluation

The last step of our scheme applies a machine learning model to classify the image by generating a likelihood score based on features extracted by convolutional neural network (Figure 3B). In order to make reasoning and internal relationship (or linkage) of the classifier easy explainable or understandable, we trained and implemented a logistic regression classifier into the scheme. Logistic regression is a relatively simple “linear” classifier because its output is only determined by the linear combination of the selected features. The working mechanism of the logistic regression classifier is transparent to the users as the optimized weights of the linear combination can reflect how the features contribute to the classification process.

The classification scores range from 0 to 1, indicating the probability of positive for selective transcriptional factor. This procedure was iteratively performed 10 times using different training and testing datasets. As a result, each of immunohistochemistry images of different TFs in chorionic villus was used as a testing case once and obtained a classification score. Subsequently, the classification scores of all cases were processed by a receiver operating characteristic (ROC) curve fitting program based on the use of a maximum likelihood data analysis method (Fig.3A).

Fig.3B includes confusion matrices of the classification results of logistic regression-based classifiers that were trained/tested using immunohistochemistry images when applying a threshold of 0.5 (at the middle point in the range of the classification scores) to divide between positive and negative groups. The results show that applying the CNN, 79% of immunohistochemistry images of different TFs in chorionic villus were correctly classified in average.

To obtain the intuitive visualization of feature vectors based on features extracted by the convolutional neural network, we used t-distributed stochastic neighbor embedding (t-SNE) for visualization. The image features extracted based on the convolutional neural network can divide all images into ten categories.

160

161 Discussion

162

163 In this study, expression patterns of different transcription factors based on CNN was
164 explored from immunofluorescence images of tissue sections obtained by a high-resolution
165 scanner. Since imbalanced data sets are small and the differences between categories are small,
166 we adopt strategies such as transfer learning, data expansion, and logistic regression to solve
167 these problems. These predictive results reveal new biologically important regulatory
168 interactions in human placenta. The model could also be applied to other functional GRN gene
169 expression images, which we conclude have the potential to be well studied in terms of their
170 potential role in understanding gene regulation.

171 Although the study of spatial expression patterns has led to many new insights, this study
172 has many limitations. A major problem of this study is that many existing studies of images based
173 on GRN conclusions rely exclusively on measuring the similarity of images (provided that similar
174 images indicate a relationship between co-expression or expression) [12]. Due to the complexity
175 of biological systems, gene interaction patterns are generally much more complex than the
176 spatial expression patterns of individual genes, and methods based on image similarity may not
177 fit this task [13]. The way it works would lead to unreliable results, and the results of these
178 studies suggest the need for a more robust model of regulatory interactions. Although CNNs has
179 successfully identified most of the relationships in the independent test sets, it is questionable
180 whether predicting highly likely regulatory interactions is really biologically meaningful.

181 Another limitation of this study is that certain tissues are relatively difficult to obtain and it
182 is difficult to achieve consistency in developmental stages during comparison. Recently,
183 generative adversary networks (GANs) have been successfully used to generate realistic 2D cell
184 images and predict intermediate stages of cells [14]. GANs are used here as a tool to create
185 synthetic images with a variety of biological features such as the shape of the cell, the size of its
186 cells, its shape and molecular structure, etc [15]. In the future, fluorescence histochemical data of
187 villi at different developmental stages can be generated systematically based on this technology,

188thus providing great support and convenience for training corresponding convolutional neural
189networks and constructing transcriptional regulatory networks.

190 This study is a preliminary work for the automatic identification of gene regulation networks
191based on the spatial distribution of gene expression. 3D printing technology provides strong
192support for the ultimate application of stem cell tissue engineering in clinical practice. Hence, it is
193necessary to build a 3D model of the tissue before 3D printing in the future. Currently, deep
194learning tools related to 3d stereoscopic tissue imaging based on light-sheet microscopy systems
195have been developed [16] , and the combination of these technologies will provide powerful
196tools and support for our in-depth understanding of placenta development in the future and is
197expected to provide an important tool for the prevention and treatment of pregnancy-related
198diseases, toxicology and pharmacology screening [17] and birth defects in the future.

199

200

201Materials and Method :

202Collection protocol

203 The first-trimester placenta (six to nine-week of gestation) was collected in the Second
204Affiliated Hospital of Zhengzhou University with the approval of Life Science Ethics Review
205Committee of Zhengzhou University. After being separated ex-vivo, the placenta was immediately
206washed in the ice saline and divided into several parts by a scalpel blade on ice, and then were
207transformed into cryogenic vials which were filled with 1 ml RNA store beforehand. When
208disposed of properly, the samples were put into 4°C refrigerator for 24h to let the RNA store
209immerse them. After 24h the RNA store was abandoned and the samples were sopped up using
210sterile absorbing paper, then the samples were separately collected into new cryogenic vials and
211were store into -80°C refrigerator for further investigations.

212

213HE (Hematoxylin & Eosin) staining

214 Samples collected before were fixed in formalin, embedded in paraffin and sliced up to 4 μm

215sections, and then were deparaffinized and rehydrated. The deparaffin and rehydration protocols
216are xylene I for 5 min, xylene II for 5 min, 100% ethanol for 2min, 95% ethanol for 1min, 80%
217ethanol for 1min, 75% ethanol for 1min, and finally distilled water for 2min. After the process
218above, the sections were stained in hematoxylin for 5 min and rinsed with tap water, then
219differentiated in hydrochloric acid and ethanol for the 30s respectively. At last, the sections were
220soaked into tap water for 5min and sealed in neutral resins with cover glass, then were observed
221under an ordinary optical microscope, 10 pictures were randomized obtained per section.

222

223Immunohistochemistry

224 Incubate sections in 2 changes of xylene, 15 min each. Dehydrate in 2 changes of pure
225ethanol for 5 min, followed by dehydrate in gradient ethanol of 85% and 75% ethanol,
226respectively, 5 min each. Wash in distilled water. Immerse the slides in EDTA antigen retrieval
227buffer (pH 8.0) and maintain at a sub-boiling temperature for 8 min, standing for 8 min and then
228followed by another sub-boiling temperature for 7 min. Be sure to prevent buffer solution
229evaporate. Let air cooling. Wash three times with PBS (pH 7.4) in a Rocker device, 5 min each. Use
230the right antigen retrieval buffer and heat extent according to tissue characteristics. Eliminate
231obvious liquid, mark the objective tissue with liquid blocker pen. Add 3% BSA to cover the
232marked tissue to block non-specific binding for 30 min. Cover objective area with 10% donkey
233serum (for the case of primary antibody originated from goat) or 3% BSA (for the case of primary
234antibody originated from others). Throw away the blocking solution slightly. Incubate slides with
235primary antibody (diluted with PBS appropriately) overnight at 4 °C, placed in a wet box
236containing a little water. Wash slides three times with PBS (pH 7.4) in a Rocker device, 5 min each.
237Then throw away liquid slightly. Cover objective tissue with secondary antibody (appropriately
238respond to primary antibody in species), incubate at room temperature for 50 min in dark
239condition. DAPI counterstain in nucleus: wash three times with PBS (pH 7.4) in a Rocker device, 5
240min each. Then incubate with DAPI solution at room temperature for 10 min, kept in dark place.
241Wash three times with PBS (pH 7.4) in a Rocker device, 5 min each. Add spontaneous
242fluorescence quenching reagent to incubate for 5 min. Wash in running tap water for 10 min.

243 Throw away liquid slightly, then cover slip with anti-fade mounting medium.

244

245 RNA-seq data analysis

246 Transcript abundance was quantified using STAR (Bray et al., 2016) and gene fold changes
247 were generated by comparing gene expression levels between two groups using the limma R
248 package [18].

249

250 Cell Imaging and Image Processing

251 Images were acquired with a Panoramic MIDI Slide scanner (3D HISTECH, Budapest,
252 Hungary). DAPI glows blue by UV excitation wavelength 330-380 nm and emission wavelength
253 420 nm; FITC glows green by excitation wavelength 465-495 nm and emission wavelength
254 515-555 nm; CY3 glows red by excitation wavelength 510-560 nm and emission wavelength 590
255 nm. Each picture was then sliced to get images of 224x224 pixels. Each TF contains about 4000
256 pictures.

257

258 Convolutional neural network

259 To ensure other researchers with no coding experiences can implement our neural network,
260 the analysis was based on the open-source Deep Learning library implementation in The
261 Konstanz Information Miner (KNIME) Analytics Platform [11]. Preparing images is done by
262 resizing each one to be compatible with each pre-trained CNNs image input layer. Each of the
263 previous pre-trained networks has its input layer size: As mentioned earlier, a patch size of 30 was
264 selected, and the images were resized to 224x 224 for VGG16. After the preprocessing step, each
265 cropped patch was fed into a convolutional neural network to extract the image features for
266 classification. The network was constructed using the Keras platform on a server with an Intel
267 CPU i7 7800X machine with 3.6 GHz of the clock and 64 GB RAM. A GeForce RTX 2080 AERO GPU
268 with 11 GB of memory under CUDA version 8.0 was also used in this study.

269

270 Hierarchical clustering and 2D mapping

271 Orange [19] was used for dimensionality reduction, clustering, and visualize cell clusters

272 using, t-SNE.

273

274 Evaluation metrics

275 Overall accuracy was used to evaluate the quantitative performance of different
276 classifiers. Overall accuracy represents the correctly classified areas for the whole image and is
277 calculated by dividing the number of correctly classified pixels by the total number of pixels in the
278 confusion matrix. Precision, also known as positive predictive values, illustrates how many
279 detected pixels for each category are true. Recall, also known as sensitivity, indicates how many
280 actual pixels in each category are detected.

281

$$282 \quad \text{Precision} = \frac{\text{True positives}}{\text{True positive} + \text{False positive}}$$

$$283 \quad \text{Recall} = \frac{\text{True positives}}{\text{True positive} + \text{False negative}}$$

284

285

286 Acknowledgments:

287 All authors contributed to the study conception and design. LYJ conceived the project and
288 completed the core program. LYJ performed the computational analysis. YLG and YG performed
289 the wet experiment. MGH, JM and CJQ provides the necessary software and hardware
290 foundation for this research. LYJ wrote the manuscript. We give thanks to Wuhan ServiceBio
291 technology co.LTD for the assistance on the immunohistochemical experiment. RNA-seq data
292 analyzed in this study could be found in European Nucleotide Archive with accession
293 ERX3472341.

294* Conflict of Interest. We declare that we have no financial and personal relationships with other
295 people or organizations that can inappropriately influence our work, there is no professional or
296 other personal interest of any nature or kind in any product, service and/or company that could
297 be construed as influencing the position presented in, or the review of, the manuscript entitled.

298* Funding: Yajun Liu was supported by “Young scientists startup grand of The Second Affiliated

299Hospital of Zhengzhou University" and "Foundation of Henan Educational Committee (CN)
300(19A320044).

301* Ethical approval: This article is approved by the life science ethics review committee of
302Zhengzhou University and does not contain any studies with human participants or animals
303performed by any of the authors.

304

305

306Data availability

307 Any relevant data are available from the authors upon reasonable request. The data
308produced by the analysis including supplementary fil in this manuscript is freely available in
309zenodo data repository (<https://doi.org/10.5281/zenodo.4285690>). RNA-seq could be found in
310European Nucleotide Archive with accession ERX3472341 for chorionic villus and ERX3472299 for
311decidua.

312

313

314

315Reference

- 316[1] S.G. Rodriques, R.R. Stickels, A. Goeva, C.A. Martin, E. Murray, C.R. Vanderburg, J. Welch,
317 L.M. Chen, F. Chen, E.Z. Macosko, Slide-seq: A scalable technology for measuring
318 genome-wide expression at high spatial resolution, *Science*. 363 (2019) 1463–1467.
319 <https://doi.org/10.1126/science.aaw1219>.
- 320[2] R. Soldatov, M. Kaucka, M.E. Kastriti, J. Petersen, T. Chontorotzea, L. Englmaier, N.
321 Akkuratova, Y. Yang, M. Häring, V. Dyachuk, C. Bock, M. Farlik, M.L. Piacentino, F.
322 Boismoreau, M.M. Hilscher, C. Yokota, X. Qian, M. Nilsson, M.E. Bronner, L. Croci, W.-Y.
323 Hsiao, D.A. Guertin, J.-F. Brunet, G.G. Consalez, P. Ernfors, K. Fried, P.V. Kharchenko, I.
324 Adameyko, Spatiotemporal structure of cell fate decisions in murine neural crest, *Science*.
325 364 (2019). <https://doi.org/10.1126/science.aas9536>.
- 326[3] C.M. Dooley, N. Wali, I.M. Sealy, R.J. White, D.L. Stemple, J.E. Collins, E.M. Busch-Nentwich,
327 The gene regulatory basis of genetic compensation during neural crest induction, *PLOS*
328 *Genetics*. 15 (2019) e1008213. <https://doi.org/10.1371/journal.pgen.1008213>.
- 329[4] K. Davie, J. Jacobs, M. Atkins, D. Potier, V. Christiaens, G. Halder, S. Aerts, Discovery of
330 Transcription Factors and Regulatory Regions Driving In Vivo Tumor Development by

331 ATAC-seq and FAIRE-seq Open Chromatin Profiling, *PLOS Genetics*. 11 (2015) e1004994.
332 <https://doi.org/10.1371/journal.pgen.1004994>.
333[5] B. Verd, N.A. Monk, J. Jaeger, Modularity, criticality, and evolvability of a developmental
334 gene regulatory network, *ELife*. 8 (2019) e42832. <https://doi.org/10.7554/eLife.42832>.
335[6] Y. Liu, D. Ding, H. Liu, X. Sun, The accessible chromatin landscape during conversion of
336 human embryonic stem cells to trophoblast by bone morphogenetic protein 4, *Biol Reprod*.
337 96 (2017) 1267–1278. <https://doi.org/10.1093/biolre/iox028>.
338[7] L. Van den Broeck, M. Gordon, D. Inzé, C. Williams, R. Sozzani, Gene Regulatory Network
339 Inference: Connecting Plant Biology and Mathematical Modeling, *Front. Genet*. 11 (2020).
340 <https://doi.org/10.3389/fgene.2020.00457>.
341[8] M.Y. Lu, M. Zhao, M. Shady, J. Lipkova, T.Y. Chen, D.F.K. Williamson, F. Mahmood, Deep
342 Learning-based Computational Pathology Predicts Origins for Cancers of Unknown Primary,
343 ArXiv:2006.13932 [Cs, q-Bio]. (2020). <http://arxiv.org/abs/2006.13932> (accessed November
344 14, 2020).
345[9] M. Raghu, E. Schmidt, A Survey of Deep Learning for Scientific Discovery, ArXiv:2003.11755
346 [Cs, Stat]. (2020). <http://arxiv.org/abs/2003.11755> (accessed November 14, 2020).
347[10] E. Meijering, A bird's-eye view of deep learning in bioimage analysis, *Comput Struct*
348 *Biotechnol J*. 18 (2020) 2312–2325. <https://doi.org/10.1016/j.csbj.2020.08.003>.
349[11] M.R. Berthold, N. Cebon, F. Dill, T.R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B.
350 Wiswedel, KNIME: The Konstanz Information Miner, in: C. Preisach, H. Burkhardt, L.
351 Schmidt-Thieme, R. Decker (Eds.), *Data Analysis, Machine Learning and Applications*,
352 Springer, Berlin, Heidelberg, 2008: pp. 319–326.
353 https://doi.org/10.1007/978-3-540-78246-9_38.
354[12] J.-B. Lugagne, H. Lin, M.J. Dunlop, DeLTA: Automated cell segmentation, tracking, and
355 lineage reconstruction using deep learning, *PLOS Computational Biology*. 16 (2020)
356 e1007673. <https://doi.org/10.1371/journal.pcbi.1007673>.
357[13] M. Nitzan, N. Karaikos, N. Friedman, N. Rajewsky, Gene expression cartography, *Nature*.
358 576 (2019) 132–137. <https://doi.org/10.1038/s41586-019-1773-3>.
359[14] P. Baniukiewicz, E.J. Lutton, S. Collier, T. Bretschneider, Generative Adversarial Networks for
360 Augmenting Training Data of Microscopic Cell Images, *Front. Comput. Sci*. 1 (2019).
361 <https://doi.org/10.3389/fcomp.2019.00010>.
362[15] A. Osokin, A. Chessel, R.E.C. Salas, F. Vaggi, GANs for Biological Image Synthesis,
363 ArXiv:1708.04692 [Cs, Stat]. (2017). <http://arxiv.org/abs/1708.04692> (accessed November
364 20, 2020).
365[16] J.M. Girkin, M.T. Carvalho, The light-sheet microscopy revolution, *J. Opt*. 20 (2018) 053002.
366 <https://doi.org/10.1088/2040-8986/aab58a>.
367[17] S.L. Beedie, A.J. Diamond, L.R. Fraga, W.D. Figg, N. Vargesson, Vertebrate embryos as tools
368 for anti-angiogenic drug screening and function, *Reproductive Toxicology*. 70 (2017) 49–59.
369 <https://doi.org/10.1016/j.reprotox.2016.11.013>.
370[18] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, limma powers
371 differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids*
372 *Res*. 43 (2015) e47–e47. <https://doi.org/10.1093/nar/gkv007>.

373[19] DemšarJanez, CurkTomaž, ErjavecAleš, GorupČrt, Hočevartomaž, MilutinovičMitar,
374 MožinaMartin, PolajnarMatija, ToplakMarko, StaričAnže, ŠtajdoharMiha, UmekLan,
375 ŽagarLan, ŽbontarJure, ŽitnikMarinka, ZupanBlaž, Orange, The Journal of Machine Learning
376 Research. (2013). <https://dl.acm.org/doi/abs/10.5555/2567709.2567736> (accessed
377 November 6, 2020).

378

379

380Figure legend:

381Figure 1 (A) Immunofluorescence staining of the trophoblastic specific transcription factors (TFs)
382in chorionic villus; (C) Track of the gene expression signal (intensity on Y-axis) of *JUN*, *FOS*,
383*TFAP2A*, *TFAP2C*, *KISS1* and *PAEP* in chorionic villus tissues in the human genome;

384Figure 2 (A) The architecture of the convolutional feature extraction network, where 'Conv N'
385stands for a 3x3 convolutional layers with N channels, 'Max Pool' stands for a 2x2 max-pooling
386layer; (B) Analysis process for machine learning image recognition of for classifying
387immunohistochemistry images of different TFs in in chorionic villus; (B) Figure 1 The experiment
388and analysis workflow of this study;

389Figure 3 (A) ROC curves of classifiers trained and tested using immunohistochemistry images of
390different TF in in chorionic villus. The AUC values are **0.786±0.026**.

391(B) Confusion matrices of prediction results for classifying immunohistochemistry images of
392different TFs in in chorionic villus;

393Figure 4 (A) Hierarchical clustering and heatmap displaying differentially expressed genes
394different TFs in in chorionic villus; (B)The visualization results of image recognition by VGG16
395based t-sne.

396

397

398Figure S1 (A) Tissue section, hematoxylin staining in chorionic villus; (B) Track of the gene
399expression signal (intensity on Y-axis) of *TEAD1*, *TEAD3*, *GATA2*, *GATA3* in chorionic villus tissues
400in the human genome;

401

Figure 1

A

B

Figure 2

Figure 3

B

		Predicted										
		DAPI	FOS	GATA2	GATA3	JUN	TEAD1	TEAD3	TEAD4	TFAP2A	TFAP2C	Σ
Actual	DAPI	203	0	1	0	1	1	1	1	45	10	263
	FOS	0	912	148	36	8	16	22	94	6	2	1244
	GATA2	5	65	1017	204	80	34	98	183	63	28	1777
	GATA3	0	6	123	1463	10	14	129	2	14	7	1768
	JUN	8	5	82	24	931	14	30	23	74	74	1265
	TEAD1	0	6	5	4	4	1519	11	3	13	25	1590
	TEAD3	12	5	58	36	39	7	1278	13	191	16	1655
	TEAD4	1	29	219	16	51	10	40	807	24	19	1216
	TFAP2A	6	4	17	2	18	2	95	16	1501	27	1688
	TFAP2C	10	0	6	3	28	3	11	7	71	1581	1720
Σ		245	1032	1676	1788	1170	1620	1715	1149	2002	1789	14186

Figure 4

A

B

Figure S1

A

B