1

2 Intuitive Graphical Visualization of Transcriptomes by Nonlinear 3 Dimensionality Reduction Exposes Relatedness between Human 4 Placenta Tissues

5

6 Yajun Liu ^{1,5}, Yi Zhang ^{1,5}, Shiwen Li ¹, Jinqun Cui^{1,2}

7

8 **Affiliations**

9

10 1. Department of Obstetrics and Gynecology, the Second Affiliated Hospital of Zhengzhou
11 University, Zhengzhou, China;

12 2. Henan Clinical Research Center for Gynecology Oncology, the Second Affiliated Hospital of
13 Zhengzhou University, Zhengzhou, China;

14 3. Department of Clinical Laboratory, the Second Affiliated Hospital of Zhengzhou University,
15 Zhengzhou, China;

16 4. Department of Pathology, the Second Affiliated Hospital of Zhengzhou University, Zhengzhou,
17 China;

18 5. Academy of Medical Sciences of Zhengzhou University Translational Medicine platform,
19 Zhengzhou University, No.100 Science Avenue, Zhengzhou, China;

20

21 **Keywords:** t-SNE, RNA-seq, placenta

22 **Running title:** RNA-seq of human placenta

23 **Corresponding author(s)**

24

25 Yajun Liu

26 e-mail : liuyajun_biology@126.com

27 Jinqun Cui

1 e-mail : jinquan_cui@163.com

2

3 ORCID for corresponding author:

4 Yajun Liu 0000-0002-8203-5762

5 Jinquan Cui 0000-0003-3708-1839

6

7

8

1 **Abstract**

2 The establishment of a complex multi-scale model of biological tissue is of great significance for
3 the study of related diseases, and the integration of relevant quantitative data is the premise to
4 achieve this goal. Whereas, the systematic collation of data sets related to placental tissue is
5 relatively lacking. In this study, 18 published transcriptomes (a total of 425 samples) datasets of
6 human pregnancy-related tissues (including chorionic villus and decidua, term placenta,
7 endometrium, in vitro cell lines, etc.) from public databases were collected and analyzed. We
8 compared the most widely used dimensionality reduction (DR) methods to generate a 2D-map
9 for visualization of these data. We also compared the effects of different parameter settings and
10 commonly used manifold learning methods on the results. The result indicates that the nonlinear
11 method can better preserve the small differences between different subtypes of placental tissue
12 than linear method. It led the foundation for the study on accurate computational modeling of
13 placental tissue development in the future. The datasets and analysis provide a useful source for
14 the researchers in the field of the maternal-fetal interface and the establishment of pregnancy.

15

16

1 Introduction

2

3 Placental dysplasia is associated with diseases such as gestational eclampsia and
4 hypertension. Placental tissue is mainly composed of fetal-derived chorionic villus tissue and
5 maternal-derived decidua. Chorionic villus tissue is developed from the trophoblast of the
6 blastocyst, and decidua is transformed by the mother's endometrium after implantation (Roberts
7 et al. 2016). Implantation failure and insufficient placental development are important causes of
8 female infertility, recurrent miscarriage, and other pregnancy-related problems.

9 The understanding of morphogenesis is of great significance for the study of related diseases.
10 The preliminary work of establishing a multi-scale and morphological dynamics model of
11 placental tissues is to obtain a multi-dimensional and high-quality data set. Several research
12 groups including us have conducted detailed studies on the molecular profile of trophoblast cells
13 derived from both in vitro and placenta tissue in vivo (Telugu et al. 2013) (Jain et al. 2017) (Liu et
14 al. 2017) based on high-throughput sequencing technology. Whereas, most of the current
15 research is to study a specific tissue or cell line separately. At present, there is a lack of integration
16 and systematic analysis of data sets related to placental tissue subtypes.

17 In this study, we compared 18 published datasets of transcriptome data of
18 pregnancy-related tissues (including fetal-derived placenta, maternal-derived decidua and
19 endometrium, and amniotic fluid), and systematically analyzing the gene expression profiles of
20 these cell lines and tissues.

21 We employed dimensionality reduction algorithm (mainly including Principal Component
22 Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE)) to produce
23 visualizations that reveal both local and long-range relationships within a dataset in a single
24 mapping on a variety of transcriptome datasets. This study provides a comprehensive data set
25 and the basis for further biological modeling of placental tissue in the future.

26

27

28

1

2 **Method**

3 RNA-seq data analysis

4 Transcript abundance was quantified using kallisto(Bray et al. 2016) and gene fold changes
5 were generated by comparing gene expression levels between two groups using the limma R
6 package(Ritchie et al. 2015).

7

8 2D mapping

9 Orange (Demšar et al. 2013), which provided a wrapper for scikit-learn algorithms
10 (Pedregosa et al. 2011), was used for batch effect remove, filtering (by cells and genes), scaling,
11 normalization, clustering, dimensionality reduction, clustering and visualize cell clusters using,
12 t-SNE, and PCA.

13 In detail, the original data was compared to the genome and by gene annotation. We
14 obtained a gene expression matrix (a total of 425 samples of 35238 human gene expression cases)
15 after the alignment of raw data to the human genome and combined the data from each article
16 listed in Table 1.

17 We started by eliminating the batch effect by the Batch Effect Removal widget that uses a
18 linear regression model to decorrelate batch variables from gene expressions. After that, Genes
19 widget was used to match gene names in the data to their IDs. Continuing with filter (with 26678
20 genes and 308 cells) and preprocess widget (the choice and the order of the preprocessing steps
21 follows that from Seurat), we selected 1000 most variable genes, normalized the samples on
22 counts per million and put them on a logarithmic scale. Once the data has been normalized, the
23 Louvain Clustering widget was used to cluster the cells. We adjusted the resolution so that the
24 clustering produces 6 clusters. Dimensionality techniques (including t-sne, PCA, and other
25 manifold learning with different parameter settings) were used to generate intuitive
26 visualization of the gene expression spectrum clustering map. Figure 1A provided detailed data
27 analysis process.

28

29 Computational resources and repeatability of experiments

30 Data analysis were performed on a server with an Intel CPU i7 7800X machine with 3.6 GHz

1 of the clock and 64 GB RAM. A GeForce RTX 2080 AERO GPU with 11 GB of memory under CUDA
2 version 8.0 was also used in this study. We provided parameter setting related program (Figure
3 S2-5) and analysis workflow and datasets (Supplementary file 2) based on Orange software in
4 this study for the reader to do repeatedly analysis.

5

1 Results

2 To compare and distinguish the transcriptome of components associated with
3 pregnancy-related tissues, transcriptome data from the published report including placenta,
4 decidua, intrauterine, amniotic fluid, and in vitro established cell lines based on Illumina's
5 high-throughput sequencing technology platform were collected. The detailed information on the
6 public data obtained in this study could be found in Table 1. Figure 1A provided detailed data
7 analysis process. Results of analysis of public data could be found in supplementary file 1.

8 We first compare the visual results of the 2d map of linear and nonlinear dimensional
9 reduction methods on this data set. The results indicated that for a large number of samples, PCA
10 could not correctly distinguish different tissues and cell lines. PCA can make a general
11 classification of the sample, whereas, various cells and tissues were almost confounded in one
12 region, and different subtypes of placental tissue cannot be accurately distinguished by PCA
13 (Figure 1B). In general, t-SNE can distinguish the sample as a fetal derived (cluster 1, 2, 3, 4, 7)
14 and mother derived 5, 6) (perplexity= 30) (Figure 1C). Cluster 1 includes JEG3 (GSE79779), Bewo
15 (GSE66962), HTR8-SVneo (GSE85995), and mid-gestation chorionic villus (GSE104348),
16 chorionic villus (this study). Cluster 2 includes syncytiotrophoblast from fresh-frozen sectioned
17 tissue (GSE87726), chorionic villus (GSE109082). Cluster 3 includes hiPSC line cells after BMP4
18 treatment (GSE119265), Cluster 4 includes human endometrium (GSE57182), Cluster 5, 6
19 includes decidual and peripheral blood CD8 T cells (GSE105064), endometrium (GSE97395,
20 GSE86491, GSE98386). Table S1 provided more detailed information on the Louvain Cluster
21 result.

22 In order to compare the effects of different parameters on the nonlinear dimension
23 reduction results. First, we compare the different effects of perplexity on the results. The result
24 indicates perplexity value has a complex effect on the generated images. In general, when
25 perplexity value is small (n=10), the "distances" between each sample are larger. The sample
26 points are tightly clustered closer when the setting of is large (n=100), the "distances" between
27 each sample is smaller (Figure 2A). We next compared the effects of different metrics including
28 Euclidean, Manhattan, Chebyshev, Jaccard on the results. Our results show that Manhattan metric
29 has similar effects to that Euclidean metric. And the distance between the samples based on

1 Chebyshev and Jaccard metric is increased, in particular, Jaccard metric could not reflect the real
2 distance between the samples (Figure 2B).

3 We also compare the results of visualization of dimensionality reduction result by different
4 manifold learning methods including MDS, Isomap, Locally Linear Embedding and Spectral
5 Embedding. Then we use a scatter plot to draw the embedded graph. Although manifold learning
6 algorithms are based on different assumptions of the data sample space, compared with linear
7 dimension reduction, these nonlinear dimension reduction methods can distinguish the data set
8 correctly (Figure S1).

9 At last, we carried out cluster analysis on these sample data. After examining the distances
10 between data instances, the distance matrix is passed to hierarchical clustering (Figure 2C). The
11 results show that different subtypes of cells in placental tissue are basically distinguishable.
12 Whereas, factors such as different laboratory sources and patients' genetic background can also
13 have a significant impact on the results. The result indicates that trophoblastic cells from
14 different sources including can cluster closely including immortalized cells in vitro (JEG3, Bewo,
15 HTR8-SVneo), hESC line cells after BMP4 treatment.

16
17
18

Discussion

In summary, we used nonlinear dimensionality reduction to construct a 2D visualization of based on transcriptome data of chorionic villus, decidua, term placenta and endometrial cells at different developmental stages and different cultural environments.

Both nonlinear and linear dimensionality reduction techniques can distinguish villi tissue from decidual tissue. Dimension reduction of transcriptome data based on a large number of samples to achieve data visualization is a critical means of critical informatic discovery. PCA has been widely used in the field of reproduction biology to identify and classify patterns of gene expression behavior in transcriptome data of tissues derived from different developmental stage (Yabe et al., 2016a), different technology (Blakeley et al., 2015) or tissue from both healthy and disease origin (Altmäe et al., 2017) (Sigurgeirsson et al., 2017). Whereas, gene expression data derived from cellular state space, behavior, and gene regulation are inherently non-linear (Moon et al., 2018), PCA often fails to capture enough information in two dimensions to be useful for visualization. Because of these limitations, nonlinear dimensionality reduction (DR) methods have been developed to preserve local structure in the data. The nonlinear dimension reduction technique was superior to the linear dimension reduction technique in distinguishing different subtypes of cells within these two tissues. After the successful implantation of the embryo, the trophoblast cells from the fetus and the uterus established the mother-fetus dialogue, thereby activating the related signaling pathways, and then the related cytokines were established and dispersed in the trophoblast cells and uterine cells respectively, thus realizing the development of the villi and decidua, the two main components of the placental tissue. Recent studies have shown that this signal diffusion tends to be nonlinear in biological tissue (Landge et al., 2020), so this could explain why nonlinear dimensionality reduction techniques are more suitable for distinguishing different subtypes of cells within tissues.

The datasets we analyzed here provides useful resources of maternal-fetal interface study models for obstetricians and gynecologists (Biancotti Juan-Carlos et al. 2010). Recently developed generative adversarial networks have made great breakthroughs in generating biomedical images (Tschuchnig et al. 2020) and single cell sequencing data (Li et al. 2020). As the acquisition of placental tissue, especially the chorionic villus in early pregnancy, is restricted by

medical ethics, the relevant data sets from different laboratories comprehensively analyzed here provide a preliminary basis for the generation of relevant data sets with different genetic backgrounds and developmental stage (Yang et al. 2020) based on generative adversarial networks in the future, which can effectively alleviate the difficulties of funding and ethical limitations in biological research.

Authors contributions

All authors contributed to the study conception and design. LYJ conceived the project and completed the core program. LYJ performed the computational analysis. ZY and LSW performed the wet experiment. HGM, RJL and MJ provide the necessary software and hardware foundation for this research. LYJ wrote the manuscript. All authors analyzed and discussed the results.

Acknowledgments

We give thanks to Qunying Wei of Department of Obstetrics and Gynecology, the Second Affiliated Hospital of Zhengzhou University for technical support.

Data availability

Any relevant data are available from the authors upon reasonable request. The data produced by the analysis including supplementary file in this manuscript is freely available in zenodo data repository (<https://doi.org/10.5281/zenodo.4283076>).

Disclosure Statement

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

* Funding: Yajun Liu was also supported by "Young scientists startup grand of The Second Affiliated Hospital of Zhengzhou University" and "'Foundation of Henan Educational Committee

- 1 (CN) (19A320044).
- 2 * Ethical approval: This article is approved by life science ethics review committee of Zhengzhou
- 3 University and does not contain any studies with human participants or animals performed by
- 4 any of the authors.
- 5
- 6
- 7

1

2 Reference

3

4 Biancotti Juan-Carlos, Narwani Kavita, Buehler Nicole, Mandefro Berhan, Golan-Lev Tamar, Yanuka

5 Ofra, Clark Amander, Hill David, Benvenisty Nissim and Lavon Neta (2010) Human Embryonic Stem

6 Cells as Models for Aneuploid Chromosomal Syndromes. STEM CELLS 28:1530–1540. doi:

7 10.1002/stem.483

8 Blakeley P, Fogarty NME, del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P and

9 Niakan KK (2015) Defining the three cell lineages of the human blastocyst by single-cell RNA-seq.

10 Development 142:3151–3165. doi: 10.1242/dev.123547

11 Bray NL, Pimentel H, Melsted P and Pachter L (2016) Near-optimal probabilistic RNA-seq quantification.

12 Nature Biotechnology 34:525–527. doi: 10.1038/nbt.3519

13 Cliff TS, Wu T, Boward BR, Yin A, Yin H, Glushka JN, Prestegard JH and Dalton S (2017) MYC Controls

14 Human Pluripotent Stem Cell Fate Decisions through Regulation of Metabolic Flux. Cell Stem Cell

15 21:502-516.e9. doi: 10.1016/j.stem.2017.08.018

16 Demšar J, Curk T, Erjavec A, Gorup Č, Hočevar T, Milutinovič M, Možina M, Polajnar M, Toplak M, Starič

17 A et al. (2013) Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research

18 14:2349–2353.

19 Ferreira LMR, Meissner TB, Mikkelsen TS, Mallard W, O'Donnell CW, Tilburgs T, Gomes HAB, Camahort

20 R, Sherwood RI, Gifford DK et al. (2016) A distant trophoblast-specific enhancer controls HLA-G

21 expression at the maternal–fetal interface. PNAS 113:5364–5369. doi: 10.1073/pnas.1602886113

22 Gonzalez TL, Sun T, Koepfel AF, Lee B, Wang ET, Farber CR, Rich SS, Sundheimer LW, Buttle RA, Chen

23 Y-DI et al. (2018) Sex differences in the late first trimester human placenta transcriptome. Biology of

24 Sex Differences 9:4. doi: 10.1186/s13293-018-0165-y

25 Jain A, Ezashi T, Roberts RM and Tuteja G (2017) Deciphering transcriptional regulation in human

26 embryonic stem cells specified towards a trophoblast fate. Scientific Reports 7:17257. doi:

27 10.1038/s41598-017-17614-5

28 Juiz NA, Torrejón I, Burgos M, Torres AMF, Duffy T, Cayo NM, Tabasco A, Salvo M, Longhi SA and

29 Schijman AG (2018) Alterations in Placental Gene Expression of Pregnant Women with Chronic Chagas

30 Disease. The American Journal of Pathology 188:1345–1353. doi: 10.1016/j.ajpath.2018.02.011

31 Kamath-Rayne BD, Du Y, Hughes M, Wagner EA, Muglia LJ, DeFranco EA, Whitsett JA, Salomonis N and

32 Xu Y (2015) Systems biology evaluation of cell-free amniotic fluid transcriptome of term and preterm

33 infants to detect fetal maturity. BMC Medical Genomics 8:67. doi: 10.1186/s12920-015-0138-5

34 Krendl C, Shaposhnikov D, Rishko V, Ori C, Ziegenhain C, Sass S, Simon L, Müller NS, Straub T, Brooks KE

- 1 et al. (2017) GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell
2 differentiation to trophoctoderm with repression of pluripotency. PNAS 114:E9579–E9588. doi:
3 10.1073/pnas.1708341114
- 4 Landge AN, Jordan BM, Diego X and Müller P (2020) Pattern formation mechanisms of self-organizing
5 reaction-diffusion systems. Developmental Biology 460:2–11. doi: 10.1016/j.ydbio.2019.10.031
- 6 Lee B, Kroener LL, Xu N, Wang ET, Banks A, Williams J, Goodarzi MO, Chen Yi, Tang J, Wang Y et al.
7 (2016) Function and Hormonal Regulation of GATA3 in Human First Trimester Placentation. Biol
8 Reprod. doi: 10.1095/biolreprod.116.141861
- 9 Li H, Sharma A, Luo K, Qin ZS, Sun X and Liu H (2020) DeconPeaker, a Deconvolution Model to Identify
10 Cell Types Based on Chromatin Accessibility in ATAC-Seq Data of Mixture Samples. Front Genet. doi:
11 10.3389/fgene.2020.00392
- 12 Liu Y, Ding D, Liu H and Sun X (2017) The accessible chromatin landscape during conversion of human
13 embryonic stem cells to trophoblast by bone morphogenetic protein 4. Biol Reprod 96:1267–1278. doi:
14 10.1093/biolre/iox028
- 15 Lucas ES, Dyer NP, Murakami K, Lee YH, Chan Y-W, Grimaldi G, Muter J, Brighton PJ, Moore JD, Patel G
16 et al. (2016) Loss of Endometrial Plasticity in Recurrent Pregnancy Loss. STEM CELLS 34:346–356. doi:
17 10.1002/stem.2222
- 18 Nguyen LH and Holmes S (2019) Ten quick tips for effective dimensionality reduction. PLOS
19 Computational Biology 15:e1006907. doi: 10.1371/journal.pcbi.1006907
- 20 Pavličev M, Wagner GP, Chavan AR, Owens K, Maziarz J, Dunn-Fletcher C, Kallapur SG, Muglia L and
21 Jones H (2017a) Single-cell transcriptomics of the human placenta: inferring the cell communication
22 network of the maternal-fetal interface. Genome Res 27:349–361. doi: 10.1101/gr.207597.116
- 23 Pavličev M, Wagner GP, Chavan AR, Owens K, Maziarz J, Dunn-Fletcher C, Kallapur SG, Muglia L and
24 Jones H (2017b) Single-cell transcriptomics of the human placenta: inferring the cell communication
25 network of the maternal-fetal interface. Genome Res 27:349–361. doi: 10.1101/gr.207597.116
- 26 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss
27 R, Dubourg V et al. (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning
28 Research 12:2825–2830.
- 29 Rekker K, Altmäe S, Suhorutshenko M, Peters M, Martinez-Blanch JF, Codoñer FM, Vilella F, Simón C,
30 Salumets A and Velthut-Meikas A (2018) A Two-Cohort RNA-seq Study Reveals Changes in Endometrial
31 and Blood miRNome in Fertile and Infertile Women. Genes 9:574. doi: 10.3390/genes9120574
- 32 Renaud SJ, Chakraborty D, Mason CW, Rumi MAK, Vivian JL and Soares MJ (2015) OVO-like 1 regulates
33 progenitor cell fate in human trophoblast development. PNAS 112:E6175–E6184. doi:
34 10.1073/pnas.1507397112
- 35 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK (2015) limma powers differential

- 1 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47–e47. doi:
2 10.1093/nar/gkv007
- 3 Roberts RM, Green JA and Schulz LC (2016) The evolution of the placenta. *Reproduction* 152:R179-189.
4 doi: 10.1530/REP-16-0325
- 5 Sharpe J (2017) Computer modeling in developmental biology: growing today, essential tomorrow.
6 *Development* 144:4214–4225. doi: 10.1242/dev.151274
- 7 Sheridan MA, Yang Y, Jain A, Lyons AS, Yang P, Brahmasani SR, Dai A, Tian Y, Ellersieck MR, Tuteja G et
8 al. (2019) Early onset preeclampsia in a model for human placental trophoblast. *PNAS* 116:4336–4345.
9 doi: 10.1073/pnas.1816150116
- 10 Sigurgeirsson B, Åmark H, Jemt A, Ujvari D, Westgren M, Lundeberg J and Gidlöf S (2017)
11 Comprehensive RNA sequencing of healthy human endometrium at two time points of the menstrual
12 cycle. *Biol Reprod* 96:24–33. doi: 10.1095/biolreprod.116.142547
- 13 Telugu BP, Adachi K, Schlitt JM, Ezashi T, Schust DJ, Roberts RM and Schulz LC (2013) Comparison of
14 extravillous trophoblast cells derived from human embryonic stem cells and from first trimester
15 human placentas. *Placenta* 34:536–543. doi: 10.1016/j.placenta.2013.03.016
- 16 Tschuchnig ME, Oostingh GJ and Gadermayr M (2020) Generative Adversarial Networks in Digital
17 Pathology: A Survey on Trends and Future Potential. *ArXiv abs/2004.14936*:
- 18 Yabe S, Alexenko AP, Amita M, Yang Y, Schust DJ, Sadovsky Y, Ezashi T and Roberts RM (2016)
19 Comparison of syncytiotrophoblast generated from human embryonic stem cells and from term
20 placentas. *PNAS* 113:E2598–E2607. doi: 10.1073/pnas.1601630113
- 21 Yang KD, Damodaran K, Venkatachalapathy S, Soylemezoglu AC, Shivashankar GV and Uhler C (2020)
22 Predicting cell lineages using autoencoders and optimal transport. *PLOS Computational Biology*
23 16:e1007828. doi: 10.1371/journal.pcbi.1007828
- 24 Yockey LJ, Jurado KA, Arora N, Millet A, Rakib T, Milano KM, Hastings AK, Fikrig E, Kong Y, Horvath TL et
25 al. (2018) Type I interferons instigate fetal demise after Zika virus infection. *Science Immunology*
26 3:eaao1680. doi: 10.1126/sciimmunol.aao1680
- 27 Zeng W, Liu X, Liu Z, Zheng Y, Yu T, Fu S, Li X, Zhang J, Zhang S, Ma X et al. (2018) Deep Surveying of the
28 Transcriptional and Alternative Splicing Signatures for Decidual CD8+ T Cells at the First Trimester of
29 Human Healthy Pregnancy. *Front Immunol*. doi: 10.3389/fimmu.2018.00937
- 30 Zeng W, Liu Z, Liu X, Zhang S, Khanniche A, Zheng Y, Ma X, Yu T, Tian F, Liu X-R et al. (2017) Distinct
31 Transcriptional and Alternative Splicing Signatures of Decidual CD4+ T Cells in Early Human Pregnancy.
32 *Front Immunol*. doi: 10.3389/fimmu.2017.00682

33 **Figure legend**

34

Figure 1 . Visual results of the 2d map of linear and nonlinear dimensional reduction

methods on this data set. (A) This flow chart of our analysis is based on Orange. The combined data undergoes batch effect removal, filtering and normalization. Cluster analysis reveals cell types and gene markers. (B) Projection of 425 samples (including chorionic villus, decidua, endometrium and amniotic fluid) in a 2D-map using PCA: Each point represents a sample which is colored according to the cluster; (C) Projection of 425 samples in a 2D-map using t-SNE (Perplexity=100; metric was set as Euclidean).

Figure 2 . The effects of different parameters on the nonlinear dimension reduction

results. (A) Projection of 425 samples in a 2D-map using t-SNE based on different parameter settings (Perplexity=10, 50, 100); (B) Projection of 425 samples in a 2D-map using t-SNE based on different metrics; (C) Clustering results of transcriptome data from 425 samples.

Supplementary figure

Figure S1 Dimension reduction result based on different manifold learning techniques;

Figure S2 Parameter setting of Figure 1 related program based on Orange software;

Figure S3 Parameter setting of Figure 2B related program based on Orange software;

Figure S4 Parameter setting of Figure 2C related program based on Orange software.

Figure S5 Parameter setting of Figure S1 related program based on Orange software;

Supplementary file 1: Results of analysis of public data;

Supplementary file 2: The analysis workflow and datasets in this study which are based on

Orange software (Demšar et al. 2013). Interested readers could repeat our data analysis based on these files.

A

B

C

A

Perplexity=10

Perplexity=50

Perplexity=100

B

Manhattan

Chebyshev

Jaccard

C