Development of quantitative direct prediction algorithm for the human target organ similarity of human pluripotent stem cell-derived organoids and cells

Mi-Ok Lee
Korea Research Institute of Bioscience and Biotechnology
Sugi Lee
Korea Research Institute of Bioscience & Biotechnology
Cho-Rok Jung
Korea Research Institute of Bioscience and Biotechnology
Jea-Woon Ryu
Korea Research Institute of Bioscience and Biotechnology
Ye Seul Son
Korea Research Institute of Bioscience and Biotechnology
Kwang Bo Jung
Korea Research Institute of Bioscience and Biotechnology
Jun-Ho Ahn
Korea Institute of Toxicology
Jung-Hwa Oh
Korea Institute of Toxicology https://orcid.org/0000-0003-4729-9153
Janghwan Kim
Korea Research Institute of Bioscience and Biotechnology https://orcid.org/0000-0003-3227-958X
Insu Jang
Korea Research Institute of Bioscience and Biotechnology
Jinhyuk Choi
Korea Research Institute of Bioscience and Biotechnology
Jaeeun Jung
Korea Research Institute of Bioscience and Biotechnology
Kunhyang Park
Korea Research Institute of Bioscience and Biotechnology
Byungwook Lee
KRIBB
Dae-Soo Kim (kds2465@kribb.re.kr)
Kribb https://orcid.org/0000-0002-4999-1613

Mi-Young Son (myson@kribb.re.kr)
Korea Research Institute of Bioscience and Biotechnology https://orcid.org/0000-0001-7590-8812
Hyun-Soo Cho (chohs@kribb.re.kr)
Korea Research Institute of Bioscience and Biotechnology

Article

Keywords: Gene panel, calculation algorithm, hPSC-derived organoids

Posted Date: August 25th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-60397/v1

Version of Record: A version of this preprint was published at Nature Communications on July 23rd, 2021. See the published version at https://doi.org/10.1038/s41467-021-24746-w.

Development of quantitative direct prediction algorithm for the human target organ similarity of human pluripotent stem cell-derived organoids and cells

Mi-Ok Lee[1, 2†], Su-gi Lee[1, 2†], Cho-Rok Jung[1, 2†], Jae-Woon Ryu[1], Ye Seul Son[1, 2], Kwang Bo Jung[1, 2], Jun-Ho Ahn[3], Jung-Hwa Oh[3], Janghwan Kim[1,2], Insu Jang[1], Jinhyuk Choi[1], Jaeeun Jung[1], Kunhyang Park[1], Byungwook Lee[1], Dae-Soo Kim [1, 2*], Mi-Young Son[1,2*], Hyun-Soo Cho[1, 2*]

[1]Korea Research Institute of Bioscience and Biotechnology, Daejeon, 34141, Republic of Korea;[2]Korea University of Science and Technology, Daejeon, 34113, Republic of Korea; [3]Korea Institute of Toxicology(KIT), Daejeon, 34114, Republic of Korea

[†]***Authors share co-first authorship***

Running title: Organ similarity calculation system for hPSC-derived organoids and cells

Keywords: Gene panel, calculation algorithm, hPSC-derived organoids

Address correspondence to: Hyun-Soo Cho (chohs@kribb.re.kr), Mi-Young Son (myson@kribb.re.kr), Dae-Soo Kim (kds2465@kribb.re.kr); *Korea Research Institute of Bioscience and Biotechnology, Daejeon, 305-333, Republic of Korea*

Abstract

Human pluripotent stem cell (hPSC)-derived organoids and differentiated cells have similar characteristics, such as cell types, structure, and functions, to human organs and tissues. Thus, *in vitro* human organoids and tissue-specific cells serve as a superior alternative to conventional cell lines and animal models in drug development and regenerative medicine. However, since hPSC-derived organoids and differentiated cells show fetal-like features, further differentiation and maturation methods have been developed for the generation of high-quality *in vitro* models of the corresponding human organs and tissues. Therefore, for a simple and reproducible analysis of the quality of organoids and cells to compensate for the shortcomings of existing experimental validation studies, a quantitative evaluation method should be developed. In this study, using the GTEx database (a total of 8,555 samples in 53 tissues), we constructed a quantitative calculation system (organ-specific panels and calculation algorithm) to assess the similarity to the human lung, stomach, and heart and confirmed the algorithm using in-house RNA-seq data (total RNA from 20 tissues). To evaluate our system, we generated hPSC-derived lung organoids, gastric organoids, and cardiomyocytes and detected 33.4%, 51.7%, and 83.4% similarity, respectively, to the corresponding human target organs. To facilitate access and use of our system for researchers, we developed the web-based user interface (Web-based Similarity Analysis System, W-SAS; for liver, lung, stomach, and heart) presenting similarity to the appropriate organs as percentages and specific gene expression patterns. Thus, the W-SAS system could provide valuable information for the generation of high-quality and readily available organoids/cells differentiated from hPSCs and a strategy to guide proper lineage-oriented differentiation.

Introduction

Animal models have served as essential tools for elucidating the pathogenesis of human disease and investigating potential therapeutic targets before entering clinical trials. However, animal models cannot completely reproduce human pathophysiology and have frequently failed to predict human responses because of several human-specific characteristics not present in animal species, such as immune-related responses and pharmacokinetics [1, 2]. Human primary cells are considered the gold standard *in vitro* models for studying human physiology, disease, and drug response [3, 4]. However, isolation or *in vitro* expansion of primary cells is difficult, limiting their use in research for human disease and drug development. Therefore, robust and representative *in vitro* human organ/tissue models are urgently needed to overcome these limitations.

Recent stem cell biology studies have focused on the *in vitro* generation of tissue-specific functional cells or tissue analogs by inducing a transition of cellular states [5, 6]. The most common method is to differentiate human pluripotent stem cells (hPSCs) into tissue-specific cells by regulating developmental signaling [5]. The advancement of stem cell differentiation technology has led to the efficient differentiation of hPSCs into various cell types, such as neurons [7], cardiomyocytes (CMs) [8], hepatocytes [9], and lung airway epithelial cells [10]. Alternatively, differentiated somatic cells such as fibroblasts and peripheral blood cells can be directly converted into target cells by combinational expression of tissue-specific transcription factors [11]. Furthermore, advances in three-dimensional (3D) culture systems have enabled the development of complex organotypic models using 3D organoids that differentiate from stem cells into tissue-like miniature analogs recapitulating complex tissue-specific cell compositions, architectures, and functions [12, 13, 14]. These 3D organoid technologies provide an

opportunity to study human development and disease in depth by assessing cellular interactions, location, and structural changes [15, 16]. Overall, technologies to manipulate cell fate provide various options that can produce *in vitro* models with similar properties to the corresponding tissues, and these strategies can be broadly applicable for *in vitro* disease modeling, especially human infectious disease modeling, an important topic recently, drug development, and cell-based therapies.

However, multiple studies of *in vitro* generation of tissue-specific cells or organoids from hPSCs have demonstrated critical limitations of current technology, including immature characteristics [17, 18] and variation in quality [19, 20]. Transcriptome analysis of diverse tissue organoids demonstrated that organoids from hPSCs have fetus-like characteristics even in long-term culture [20, 21, 22]. In the case of lung-bud organoids, long-term culture up to 7 months results in lung-like organoids composed of tissue-specific cells, but they persist in a fetus-like state [23]. Thus, the development of additional maturation methods is required for advanced *in vitro* human models that more closely mimic human tissue. Moreover, heterogeneous production critically limits their utility in various biomedical applications [19, 20].

Currently, the development of tissue-specific differentiation methods and quality control of differentiated cells/organoids rely on expression analysis of tissue-specific markers by histology and gene expression analysis. Evaluation of differentiation using key tissue-specific markers can be an efficient strategy for the design and optimization of differentiation methods, but evaluating the similarity between human tissue and differentiated cells/organoids is difficult because experimental validation is laborious and time consuming. Although clustering analysis of the global transcriptome provides insight into the lineage markers and molecular similarity of differentiated cells or organoids with their counterpart human tissue,

this method does not provide a quantitative and standardized assessment of the similarity between these structures.

Previously, we developed a quantitative prediction system to assess the similarity to liver (LiGEP; Liver-specific Gene Expression Panel) of hPSC-derived hepatocytes and liver organoids to improve on qualitative quality assessment. Using the LiGEP algorithm, we can calculate the similarity between liver organoids or differentiated hepatocytes and human liver and provide valuable information for generating high-quality liver organoids and hepatocytes to researchers [24, 25]. However, because the LiGEP algorithm can only calculate similarity to liver, it is necessary to expand our study to other human organs. Additionally, we need to build a web-based analytics platform that is readily available to researchers.

Here, we developed quantitative calculation systems to assess similar to organs based on organ-specific gene panels using the GTEx public database (total of 8,555 samples, 53 tissues), including LuGEP for lung, StGEP for stomach and HtGEP for heart, and an analytical algorithm for direct comparison to human organs. We validated this analytical system with in-house RNA-seq data of 20 total RNA samples from different tissues. Moreover, we obtained the similarity (%) of hPSC-derived lung organoids, gastric organoids and CMs compared to the corresponding human organs. Finally, we developed a web-based user interface named Web-based Similarity Analytics System (W-SAS) to provide an analytical algorithm, similarity (percentage) and gene expression patterns for a direct comparison to human target organs (liver, lung, stomach, and heart). Thus, our quantitative calculation system and W-SAS can provide a useful platform for evaluating and improving the quality of differentiated organoids/cells.

Results

Development of a quantitative calculation system to assess similarity of hPSC-derived organoids and cells to organs

To assess the similarity of hPSC-derived organoids and cells to human organs at a quantitative level, in this study, we developed the W-SAS program for the quantitative assessment of similarity and quality of the hPSC-derived organoids and cells. First, the researcher will perform the RNA-seq analysis. Using raw RNA-seq data (TPM, FPKM/ RPKM values), the W-SAS program (organ-specific panels and algorithms) can calculate the similarity to the appropriate organ and provide a quantitative organ similarity score (%) and information on the gene expression patterns in organ-specific panels, directly comparing the target organs to the hPSC-derived organoids and cells. Using our system, researchers can receive important information for quality control of hPSC-derived organoids and cells (Fig. 1a).

Construction of Organ-specific Gene Expression Panels (Organ-GEP) for calculation of the similarity of hPSC-derived organoids and cells to organs

For selection organ-specific genes for each tissue (heart, lung, stomach), the analysis was performed in three steps (t-test, confidence interval, quantile comparison). Step 1: Gene selection was performed by comparing the mean and variance between heart, lung or stomach tissues and the remaining tissues. We performed paired t-tests to identify differentially expressed genes between two tissues (the heart, lung, or stomach vs. one of 42 tissues) for all possible cases. We defined a set of tissue-specific genes through the intersection of genes that showed a significant difference (p-value < 0.05) in the t-test results. We defined 2,843 heart-specific genes, 1,049 lung-specific genes and 466 stomach-specific genes. Step 2: Since the

genes acquired in the first step were chosen based on the difference in the mean and the variance between tissues, these genes were not only expressed uniquely in the particular tissue but also showed large variances in other tissues. Therefore, we used the confidence interval (CI) to overcome this problem. The CI is an estimate calculated from the statistics of the observed data, suggesting a range of possible values for the parameter; thus, it is used to identify genes that are specifically highly expressed in particular tissues. We calculated the lower bound of the 99% confidence interval (LCI) O_{Li} (ith gene's LCI) of the genes obtained in the first step for each tissue (heart, lung and stomach) and calculated the upper bound of the 99% confidence interval (UCI) $T_{1Ui}, \cdots, T_{42Ui}$ for the remaining 42 tissues. Then, we extracted the genes that had a higher LCI of the tissue (heart, lung and stomach) than the maximum UCI among the 42 tissues ($O_{Li} > \max(T_{1Ui}, \cdots, T_{42Ui})$). As a result of CI filtering, we identified candidate genes that are highly expressed in particular tissues (153 genes in heart, 189 genes in lung and 73 genes in stomach). Step 3: Although highly expressed tissue-specific genes of the three organs were previously identified, among the rest of the tissues, some genes had expression values in the top 25% of the samples that were higher than that of each tissue (heart, lung, stomach) because of a large variation in the expression values. To eliminate these false positive results, we performed quantile comparison between one of 3 tissues (heart, lung, stomach) and the remaining 42 tissues. First, we set the top 25% RPKM value of each tissue (heart, lung, stomach) as O_{qi} and set the top 25% RPKM values of the remaining 42 tissues as $T_{1qi}, \cdots, T_{42qi}$. Then, we selected the genes that met the following conditions: $O_{qi} > t \times \max(T_{1qi}, \cdots, T_{42qi}), t = 1.05$. Through quantile comparison, we defined the final organ-specific expressed genes (143 genes of heart, 145 genes of lung and 73 genes of stomach). To construct a gene panel that can reflect the characteristics and functions of each tissue, we added not only organ-specific

expressed genes but also genes related to tissue functions. As a result, three organ-specific gene panels (heart-specific gene expression panel with 144 genes, lung-specific gene expression panel with 149 genes and stomach-specific gene expression panel with 73 genes) were constructed (Supplementary Table S3).

Validation and characterization of Organ-GEPs

In the heatmap, the expression of each panel presents the specificity of its own tissue compared with other tissues. Moreover, in multigroup discriminant analysis (MDA), the expression panels for lung, heart, and stomach are clearly divided from those of other tissues, suggesting that we constructed organ-GEPs for calculation of similarity to the specific organ (Fig. 1b and 1c). To confirm that each organ-GEP reflected the organ-specific functions and characterizations, we performed Ingenuity Pathway Analysis (IPA) using each organ-GEP. The LuGEP was associated with "inflammatory response, formation lung, inflammation of respiratory system component, permeability of lung, fibrosis of lung and cellular infiltration of lung". These results suggested that our lung-specific gene expression panel supports lung function. The HtGEP was related to "contraction of heart, heart rate, contraction of cardiac muscle, cardiogenesis, organization of muscle cells, familial cardiomyopathy, muscle contraction, and enlargement of heart." Finally, StGEP was associated with "secretion of gastric acid, morphology of gastric mucosa, gastric lesion, secretion molecule, morphology of digestive system, and peptic disorder" (Fig. 1d). Thus, to calculate the similarity of hPSC-derived organoids to organs, we constructed organ-specific gene expression panels, and each panel reflected the functions of each organ.

Validation of Organ-GEP algorithms for direct comparison between target organs and organoids

We established an organ-specific calculation algorithm for determining cut-off values according to organ-specific genes via calculation of the distance based on the standard gene expression vectors between organoids using organ-specific expression gene panels (see Materials and Methods). To assess the algorithm, we used GTEx public data. In Figure 2a, LuGEP was used to calculate 100% similarity to lung of 320 lung tissues ($P > 0.001$); other tissues showed less than 20% similarity to lung, implying that the LuGEP algorithm could specifically distinguish the lung tissues among all human tissues. In addition, similar to LuGEP, HtGEP presented 100% similarity to 412 heart tissues ($P > 0.001$), but the HtGEP score showed a high percentage in muscle. In the StGEP analysis, the similarity to stomach represented 100% ($P > 0.001$) with 193 gastric samples. The results of other tissue samples presented under 10% similarity after calculation with the StGEP algorithm. Thus, we decided on the available organ-specific panel and algorithm for the prediction of target organ similarity with hPSC-derived organoids and cell types. Additionally, to confirm the organ panel and algorithm, we produced RNA-seq data with total RNA from 20 tissues purchased from Clontech (Human Total RNA Master Panel) and calculated the similarity using organ panels and algorithms. The similarity results showed 100% similarity to the lung, stomach, and heart (Fig. 2b). Thus, we suggest the possibility of a quantitative calculation method between human organs and hPSC-derived organoids and cells.

LuGEP and the algorithm can predict the similarity of the lung with hPSC-derived lung organoids

To validate the tissue similarity of human lung organoids by LuGEP and the algorithm, we generated a LBO by a stepwise differentiation method that mimics the development process of the human lung, as previously described (see Materials and Methods) [23]. First, undifferentiated hESCs were differentiated into DE expressing the surface markers CXCR4 and C-kit (Fig. 3a). DE cells were further differentiated into ventralized anterior-foregut endoderm-like cells and subsequently into LBO by embedding in Matrigel (Fig. 3a. Differentiated LBO showed branching morphologies similar to those observed in the developing lung, and it was composed of EPCAM-positive epithelium, which coexpressed the pulmonary endoderm marker NKX2.1 and distal pulmonary endoderm progenitor marker SOX9. Some cells expressed pulmonary cell type-specific markers, including alveolar type 2 cells (SFTPC), club cells (CC10), and ciliated cells (acTUB) (Fig. 3b). Before scoring the lung differentiation status from hESCs using the LuGEP algorithm, we verified the expression of representative stage-specific markers at each stage by quantitative RT-PCR. The expression of the pluripotency marker *Nanog* gradually decreased during differentiation, and the DE markers *GATA4* and *SOX17* were transiently upregulated on day 4. After patterning into the ventralized anterior foregut, we observed the induction of pulmonary cell type-specific markers, including markers of pulmonary endoderm (*NKX2.1*), proximal airway progenitors (*SOX2*), distal airway progenitors (*SOX9*), ciliated cells (*FOXJ1*), club cells (*SCGB1A*) and type 2 alveolar cells (*SFTA3, SFTPA1, SFTPB, LYZ*). However, D56 LBO did not show induction of basal cell marker (p63) and type 1 alveolar cell marker (PDPN), demonstrating insufficient differentiation compared to *in vivo* lung tissue (Fig. 3c).

To calculate the similarity of lung organoids to lung using the LuGEP algorithm, we performed LuGEP analysis with human ES, DE, lung organoid (days 21, 56), and human lung

tissues. Figure 3D shows that human ES cells and lung tissues presented 2.7%~8.5% and 100% similarity to lung, implying that LuGEP analysis could accurately distinguish lung tissue and other cells. However, day 4 DE had 17.9% similarity, and 21- and 56-day-old lung organoids presented 27.5% and 33.4% similarity, respectively, compared with the human lung (Fig. 3d). Additionally, compared to human lung tissues, each lung organoid exhibited different gene expression patterns of LuGEP. In PCA analysis with each sample (1 dot, 1 gene; color, gene expression level), we observed that the colors of each dot became similar to those of the human lung. The expression levels of type 2 alveolar cell markers (SFTP1 and SFTPB) gradually increased depending on the day of LBO. Endothelial PAS domain-containing protein 1 (EPAS1 or HIF-2 alpha) and NAPSA (aspartic proteinase), which are related to the inflammatory response, lung formation, and lung cancer, were gradually induced via IPA (Fig. 3e). Thus, we could identify specific gene expression patterns of LuGEP during the formation of LBO derived from hPSCs.

The calculation of similarity of human antrum-like gastric organoids (hGOs) to stomach with the StGEP and algorithm

To analyze the tissue similarity between the hGOs using StGEP, we generated hESC-derived antrum-like GOs (see Materials and Methods). First, we differentiated hESCs into DE and hESCs exposed to Activin A (Activin/nodal signaling activator) for 3 days, and we observed the expression of endoderm-specific markers such as SOX17 and FOXA2 (Fig. 4a). Additionally, in the presence of FGF4, CHIR99021 (WNT activator), Noggin (BMP inhibitor), and treatment with retinoic acid (RA) for the last day, DE cells were differentiated into posterior-foregut (PF) cells, and PF cells generated spherical shapes and expressed SOX2,

PDX1, HNF1β and HNF6 via qRT-PCR and immunocytochemical analysis (Fig. 4a and b). After 30~34 days, we observed 3D-cultured single spheroids, which transformed into spherical organoids, including stomach antrum-like glandular morphologies in Matrigel (Fig. 4c). In qPCR analysis, hGO contained gastric antrum-type cells such as mucus gland neck cells (*MUC6, TFF2*), surface mucus pit cells (*MUC5AC, TFF1, GKN1*), endocrine cells (*GHRL, SST, GAST*) and progenitor cells (*SOX9*), suggesting that hGO represented gastric antrum-like characteristics (Fig. 4d). Using hGO, we performed gastric similarity analysis using the StGEP and the algorithm. Figure 4d shows that hESCs and PF cells presented 2% and 10.3% similarity, respectively. However, hGO showed 51.7% and 33.9% similarity compared with the human stomach. In addition, the expression patterns of the StGEP in hGOs and PF cells showed different patterns compared with the human stomach (Fig. 4e). In PCA analysis, the expression of TFF1/2 and GKN1 gradually increased between PF and GO. Moreover, as GO_1 and GO_2 showed 51.7% and 33.9% similarity to stomach, respectively, we could identify a gene expression difference between GO_1 and GO_2. Thus, using the StGEP analysis results, we quantitatively evaluated the quality of hGOs between each sample directly compared to the human stomach.

Generation of CMs from hESCs and calculation of similarity to heart

Next, to verify the utilization of HtGEP and the algorithm, we differentiated CMs derived from hESCs by modulating Wnt signaling (see Materials and Methods). In qPCR analysis, the differentiated CMs with BMP4 exhibited high expression of cardiac transcription factors (*HAND2, GATA4, NKX2.5*), cardiac muscle markers (*cTnT*), ventricular CM markers (*MYL2, MYH7*) and atrial CM markers (*MYL7, MYH6*) and CM differentiation without BMP4 (Fig. 5a).

Additionally, in immunocytochemical analysis, we detected high expression of cardiac transcription factors (NKX2.5), cardiac muscle marker (cTnT), ventricular CM markers (MYL2), and atrial CM markers (MLC2a) in the differentiated CMs with BMP4 compared to the control CMs (Fig. 5b). Moreover, in FACS analysis with cardiac muscle-specific markers (cardiac troponin T, cTnT), hESCs and control CM showed 0.28% and 29% differentiated yields. However, differentiated CMs with BMP4 presented a 90% differentiated yield (Fig. 5c).

Using hESCs and differentiated CMs, we performed HtGEP analysis. In Figure 4C, undifferentiated samples showed values of 3.5% and 7.2%, and differentiated CM without BMP4 was 34.3%. Moreover, differentiated CM with BMP4 presented 83.4% similarity to heart (Fig. 5d). In an analysis of gene expression patterns, PCA results showed the difference in expression patterns between the CM control and CM+BMP4 groups (Fig. 5e). With treatment of BMP4 during CM differentiation, CM markers (MYH6, MYL3/7, TNNT2) were dramatically similar to the expression pattern of the human heart. Thus, using HtGEP analysis, we could identify the quantitative effects of CM maturation factors by comparing each step of CM differentiation.

User interface of the W-SAS

In this study, we also established a similarity calculation system for hPSC-derived lung and gastric organoids and CMs. To directly provide the calculation algorithm to the researcher, we developed a user-friendly interface (W-SAS, http://kobic.re.kr/wsas) for the calculation of similarity of organs to hPSC-derived organoids and cells (Fig. 6a). W-SAS is easy to use for the calculation of similarity. Figure 6b shows the RNA-seq results of each hPSC-derived organoid and cell uploaded to the W-SAS interface and the calculated and presented results of

similarity to the appropriate organ for each sample.

The W-SAS was constructed as a two-tiered server-client architecture. The server side was developed using Java Spring Framework with Java Development Kit (JDK) 1.7, and the client side was implemented using Bootstrap CSS Framework and JavaScript libraries, including Jquery, Ajax and D3.js. W-SAS is divided into two sections. The first section was the data upload. In this section, users uploaded the RNA-seq result (FPKM/RPKM, TPM values) formatted with Excel (csv, excel, txt files), selected the algorithm for each sample, and finally submitted the data for calculation of similarity. In the second section, after the calculation, the there were 5 types of results. (1) Similarity to an organ as a “percentage” and “bar graph”. (2) In the heatmap, the expression levels of each gene in the organ-specific gene expression panel were provided to a user for each sample. (3) The PCA result was also supplied to the researcher to determine the relationship between the target organ and organoids. (4) and (5) W-SAS suggested information on the expression profile for each gene in each sample (Fig. 6c). Thus, using the W-SAS interface, the researcher can rapidly receive important information about the similarity and expression gene set of the hPSC-derived organoids and cells compared to human target organs. Using W-SAS information, researchers can develop high-quality organoids via regulation of each panel of genes or alteration of culture conditions to improve similarity to organs.

Discussion

We constructed a system for prediction of similarity to three organs and a user interface (W-SAS) for hPSC-derived organoids and cells based on RNA-seq analysis. After calculation of the organoid data, researchers can receive information to improve the quality of organoids, such as percent similarity (%) and gene expression patterns, via a direct comparison between the target human organs and organoids.

The first advantage of W-SAS is a direct comparison between human target organs and differentiated organoids/cells. After the generation of hPSC-derived organoids/cells, W-SAS was used to calculate similarity to the target organ, showing a high percentage of similarity of the differentiated organoids/cells with the corresponding target organs. Second, W-SAS provides a researcher with an adequate number and names of target organ-specific genes between target human organs and organoids. Using this information can help to develop high-quality organoids via regulation of insufficient target organ-specific genes, such as genes with overexpression or depletion. The third advantage is the infinite expandability. For calculation of the similarity to a target organ, a target-specific gene expression panel and algorithm are needed. Here, we developed calculation algorithms and organ-specific gene-screening pipelines using RNA-seq data and applied them to all human organs to calculate the similarity with hPSC-derived organoids. Finally, the W-SAS interface is a user-friendly interface. We constructed a W-SAS interface for the calculation of similarity to a target organ with RNA-seq data of hPSC-derived organoids. To develop W-SAS, we focused on user comfort and organ expandability and designed a simple web page for researchers (“Data uploading”-“RUN”-“Result”). Interpretation of the results was also clearly provided to researchers about the similarity percentage. Moreover, if we construct an organ-specific gene expression panel and

analytical algorithm, this interface could serve researchers worldwide for the calculation of similarities to various organs.

However, W-SAS has some disadvantages. W-SAS was constructed based on transcriptome analysis using the GTEx database. Transcription data or computational network analysis are useful tools for the evaluation of lineage specifiers or differentiation stages during stem cell differentiation [26, 27, 28]. Moreover, transcriptome analysis showed higher sensitivity and reproducibility than proteome analysis. However, although transcriptome analysis can predict tissue-specific features, RNA-seq data do not reflect 100% of cell functions because total RNA is not translated as proteins [28]. W-SAS is an algorithm to analyze whether organ-specific genes are acquired at each stage of differentiation from hPSCs. Thus, in this study, organ development and function-associated genes were manually collected by keyword searches of the PubMed database to compensate for the disadvantage of the W-SAS calculation system. In addition, for organ-specific genes in Organ-GEPs, we included organ function-associated genes that show significant temporal expression in the stomach, heart, and lung and that show core signatures enriched with functional genes. However, to verify the cell function regrading similarity percentage, researchers must perform biochemical analyses to compensate for transcriptome analysis shortcomings.

Researchers generally performed FACS analysis with CM-specific markers (cTnT) to evaluate the CM differentiation rate derived from hPSCs [29]. In this study, the CM control and CM+BMP4 groups showed 29% and 92% in FACS analysis, but HtGEP analysis showed 34.3% and 83.4% (Fig. 5d). The heart is composed of various cell types, such as CMs, smooth muscle cells, and endothelial cells, and mainly contains 70~85% CMs in the heart [30]. Therefore, we could expect that the HtGEP score will not exceed 85% with CMs only. Although cTnT analysis

is an important method for the evaluation of CM differentiation derived from hPSCs, we considered that a more accurate differentiation and similarity assessment will be needed for a comprehensive analysis of the gene panel rather than measuring the differentiation with a single gene. Because HtGEP is a gene panel made from heart-specific genes, we suggest that HtGEP may reflect various features of heart-specific cell types. Therefore, the study of the remaining 17% similarity to heart will allow the generation of more heart-like heart organoids.

Conclusions

In summary, to generate high-quality organoids derived from hPSCs, researchers can visit the W-SAS interface to assess the organ similarity of their own hPSC-derived organoids and will receive information (similarity, heat map, PCA plot, expression patterns of panel) that can help generate high-quality organoids by engineering of gene expression. Thus, the feedback between W-SAS and researchers can become a vital troubleshooting step in the organoid research field.

Methods

Cell cultures

Human embryonic stem cells (hESCs) were cultured through a previously reported method [22, 31] and purchased from the WiCell Research Institute (Madison, WI, USA). The hESCs were maintained in hESC medium including Dulbecco's modified Eagle's medium (DMEM)/F-12 (Invitrogen, Carlsbad, CA, USA), 20% knockout serum replacement (Thermo Fisher), 1% penicillin/streptomycin (Invitrogen), 1% GlutaMAX (Invitrogen), 0.1 mM β-mercaptoethanol (Invitrogen), 1% nonessential amino acids (NEAA, Invitrogen) and 8 ng/ml basic fibroblast growth factor (bFGF, R&D Systems, Minneapolis, MN, USA). The cells were passaged every 7 days.

Transcriptome data acquisition and data preprocessing

We developed organ-specific gene expression panels that can reflect the characteristics and functions of each organ to develop a model that quantitatively predicts the differentiation of heart, lung, and stomach tissues. A total of 8,555 RNA-seq datasets (transcript RPKM of GTEx version 6) of 53 tissues were obtained through the publicly available GTEx database to select genes that could reflect the characteristics and functions of each tissue. While the RNA-seq data of the lung and stomach provide only one RPKM dataset each, the heart is divided into two subtissues (atrial appendage and left ventricle). We used both heart subtissues as a single heart sample, and all other subtissues of the remaining tissues were used as independent data. For the classification of 52 tissues, the MDS plot was calculated with 56,238 RPKM values of all genes, and the testis was excluded because it is clearly separated from other tissues. Genes

specifically expressed in testis can lead to false positive results when identifying organ-specific genes of three tissues. In addition, gender-specific tissues such as the ovary, uterus, vagina, fallopian tube and cervix were excluded. We also excluded whole blood and blood cell tissues because they include peripheral blood leukocytes. Finally, we prepared a total of 43 tissues (7,579 samples) to construct a panel of genes to reflect the characteristics and function of the heart, lung and stomach, excluding 7 gender-specific tissues and 2 blood tissues. Data preprocessing was performed before extracting the tissue-specific expression genes. In the data preprocessing, the genes were limited to protein-coding genes, and nonexpressed genes and/or genes with extremely low expression were filtered out. The gene sets of 43 tissues were matched to compare the gene expression differences between tissues. First, for identification of genes that could play a major role in tissue differentiation, 18,818 protein-coding genes were extracted through Ensembl gene ID information provided by GTEx RNA-seq data. Then, we removed 1,343 genes whose maximum value was less than 1 or third quantile value was less than 0 from all samples of the 43 tissues. Genes whose expression values are rarely measured in all tissues were not of interest, and by eliminating genes with low expression values, we can estimate the mean-variation relationship of data with high statistical reliability. Finally, we transformed the expression values into the log2 scale to perform normalization of a total of 17,475 gene sets.

Construction of the analytical algorithm

To construct an analytical algorithm, we used the standard gene expression vectors that can distinguish between organs and undifferentiated organoids from the expression values of the organ-specific gene expression panels that reflect the characteristics and functions of each

tissue. The standard gene expression vectors were defined in terms of the upper boundary of the organ (minimum expression value of the organ) and the lower boundary of the undifferentiated organoid (minimum expression value of the undifferentiated organoid). This strategy was employed when adjusting the expression boundary of differentiated organoids; in other words, the expression value of the differentiated organoids was replaced with the standard gene expression vector when it passed the boundaries. Based on the score boundaries, the status of differentiation was measured quantitatively by calculating the distance between the organ and the undifferentiated tissues. Using 73 heart samples and two undifferentiated organoid samples, we calculated the standard gene expression vector that can determine the degree of tissue development. First, each data point was normalized using log2 transformation. Using undifferentiated organoid and heart samples, we calculated the Manhattan distance close to the differentiated organoid. To calculate this, we performed the following steps. For n genes, the RPKM value of the differentiated organoid sample is $X = (x_1, \cdots, x_n)$, the jth RPKM value of the undifferentiated organoid sample is $U_j = (u_{j1}, \cdots, u_{jn})$, *the* $j = 1,2$ kth RPKM value of the heart organ sample is $H_k = (h_{k1}, \cdots, h_{kn})$, and $k = 1, \cdots, K$, with $K = 73$. The minimum RPKM of the two undifferentiated organoid samples is $U_m = (\min(u_{11}, u_{21}), \cdots, \min(u_{1n}, u_{2n})) = (u_{m1}, \cdots, u_{mn})$. Similarly, the minimum RPKM value for each gene of the heart samples is set at$H_m = (\min(h_{11}, \cdots, h_{K1}), \cdots, \min(h_{1n}, \cdots, h_{Kn})) = (h_{m1}, \cdots, h_{mn})$. U_m is the boundary of the undifferentiated organoid of the RPKM, and H_m is the minimum RPKM expression boundary of the heart. Using these parameters, we estimated the status of differentiation by calculating the Manhattan distance among organs and undifferentiated and differentiated organoids after adjusting the expression values to the score boundaries. For quantitative assessment, we found that the score of the unknown sample

approached 100 when the distance of the differentiated organoid was similar to that of the organ and approached zero if it was closer to the undifferentiated organoid. We compared differentiated organoid X with H_m for each gene and set the minimum value of $X_m = (\mathrm{median}(u_{m1}, x_1, h_{m1}), \cdots \mathrm{median}(u_{mn}, x_n, h_{mn}))$ as the RPKM value for the differentiated organoid. If $X_m = H_m$, the score value is set at 100%, and the closer the U_m, the more the score value decreases and the farther away it is from the heart. The score is as follows:

$$\mathrm{Score}_H = \frac{\|U_m - X_m\|_1}{\|U_m - H_m\|_1} \times 100(\%)$$

where $\|U_m - X_m\|_1 = \sum_{i=1}^{n} |u_{mi} - x_{mi}|$.

The same expression value is used for the lung and stomach. For the lung, the RPKM value of the k th lung organ sample is $L_k = (l_{k1}, \cdots, l_{kn})$, $k = 1, \cdots, K$, $K = 8$ and $X_m = (\min(x_1, l_{m1}), \cdots \min(x_n, l_{mn}))$. The score is as follows:

$$\mathrm{Score}_L = \frac{\|U_m - X_m\|_1}{\|U_m - L_m\|_1} \times 100(\%)$$

For the stomach, the RPKM value of the kth stomach sample is $S_k = (s_{k1}, \cdots, s_{kn})$, $k = 1, \cdots, K$, $K = 21$ and $X_m = (\min(x_1, s_{m1}), \cdots \min(x_n, s_{mn}))$. The score is as follows:

$$\mathrm{Score}_S = \frac{\|U_m - X_m\|_1}{\|U_m - S_m\|_1} \times 100(\%)$$

Differentiation of lung bud organoids from hESCs

Lung bud organoids were differentiated from hESCs as previously described (PMID: 28436965). Briefly, hESCs were maintained on feeder cells with PSC culture media:

DMEM/F12 supplemented with 20% knockout serum replacement (Gibco), 1% GlutaMAX (Gibco), 1% NEAA (Gibco), 1% penicillin-streptomycin (Gibco), 0.1% beta-mercaptoethanol (Gibco) and 20 ng/ml FGF-basic (R&D Systems). The day before differentiation into definitive endoderm (DE), hESCs were dissociated into single cells using Accutase (Millipore) and formed embryonic bodies (EBs) with serum-free differentiation media (SFM) supplemented with 3 ng/ml BMP4 (R&D Systems) and 10 mM Y27632 (Stem Cell Technologies). The next day, the EBs were differentiated into DE for 3 days with endoderm induction media: SFM supplemented with 100 ng/ml Activin A (R&D Systems), 2.5 ng/ml FGF-basic, 0.5 ng/ml BMP4 and 10 mM Y27632 under hypoxic conditions (5% O_2). For patterning the anterior foregut, cells were dissociated, seeded on fibronectin-coated plates and cultured in SFM supplemented with 200 ng/ml Noggin (R&D Systems) and 10 μM SB431542 (Tocris) for 24 hours. Subsequently, the medium was replaced with SFM supplemented with 1 μM IWP2 (R&D Systems) and 10 μM SB431542 for another 24 hours. For further ventral patterning of the anterior foregut, the cells were cultured in ventralizing/branching media: SFM supplemented with 50 nM all-trans retinoic acid (Sigma-Aldrich), 3 μM CHIR99021 (Tocris), 10 ng/ml BMP4, 10 ng/ml FGF10 (R&D Systems) and 10 ng/ml FGF7 (R&D Systems). After 48 hours, the cells were detached by pipetting and transferred to ultralow attachment 6-wll plates (Corning) to form spheroids and cultured until 21–25 days after differentiation. Ventralized anterior foregut spheroids were embedded on Matrigel droplets for branching morphogenesis of the lung bud organoids and continued culture with ventralizing/branching media. In this study, D56 LBO was used for analysis of lung specification.

Generation of human gastric organoids (hGOs)

hGOs were differentiated as described previously [14]. hESCs were seeded onto Matrigel-coated dishes and cultured in mTeSR1 (Stemcell Technologies, Vancouver, Canada) media. After 3 days, the cells were cultured in DE differentiation media containing RPMI-1640 and 100 ng/ml Activin A (R&D Systems) and gradually increasing FBS (Invitrogen) concentrations (0~2%) for 3 days. BMP4 (50 ng/ml, R&D Systems) was added to DE medium on the first day only. For differentiation into the posterior foregut, differentiated cells were cultured in RPMI-1640 media supplemented with 2% FBS (Invitrogen), 2 μM CHIR99021 (Torcris, Ballwin, MO, USA), 500 ng/ml FGF4 (Peprotech, Rocky Hill, NJ, USA), and 200 ng/ml Noggin (R&D Systems). The last day, 2 μM retinoic acid (Stemgent, Houston, Texas, USA) was added to the posterior foregut media. The 3D spheroids were collected and embedded in the Matrigel dome. After Matrigel polymerization, embedded spheroids were cultured with hGO media including advanced DMEM/F-12 (Invitrogen), 1X N2 (Invitrogen), 1X B27 (Invitrogen), 1% penicillin/streptomycin (Invitrogen), 15 mM HEPES (Invitrogen), 2 mM L-glutamine (Invitrogen), and 100 ng/ml EGF (R&D Systems). In addition, Noggin and retinoic acid were added for the first 3 days.

CM differentiation

For differentiation of hESCs into CMs, hESCs were seeded into 1% Geltrex (Thermo Fisher Scientific, Waltham, MA, USA)-coated tissue culture plates and grown to 70% cell confluency. For mesodermal induction, 6 μM CHIR 99021 (Tocris, Bristol, UK) was treated for 2 days in CM differentiation medium consisting of RPMI-1640, 212 μg/ml L-ascorbic acids (Sigma-Aldrich, St. Louis, MO, USA), and 500 μg/ml recombinant human albumin (Sigma-Aldrich).

After mesodermal induction, 2 μM Wnt-C59 was administered for another 2 days. The differentiated CMs were maintained in RPMI-1640 supplemented with 212 μg/ml L-ascorbic acids and 1X B27 supplement (Thermo Fisher Scientific). The antibodies used in this study are listed in Supplementary Table S1.

Quantitative real-time RT-PCR (qPCR)

Total RNA was obtained from harvested cells using an RNeasy Kit (Qiagen, Hilden, Germany), and a Superscript IV First-Strand Synthesis System Kit (Invitrogen, Carlsbad, CA) was used for reverse transcription. qPCR was performed on a 7500 Fast Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). Human heart RNA was purchased and used as a positive control. The sequences of the primers used in this study are presented in Supplementary Table S2.

Immunofluorescence

Cultured cells were fixed with 4% paraformaldehyde (PFA, Sigma-Aldrich, St. Louis, MO, USA) for 15 min at room temperature (RT) and permeabilized with 0.1% Triton X-100 in PBS. After the samples were blocked with 4% bovine serum albumin (BSA), primary antibodies were incubated at 4°C overnight. The cells were washed and incubated with secondary antibodies at RT for 1 hour. DAPI (1 mg/ml; Thermo Fisher Scientific) was used to stain nuclei [32]. The images were captured with fluorescence microscopy (IX51; Olympus, Tokyo, Japan).

Flow cytometry (FACS)

FACS analysis was performed to verify the differentiation efficiency of DE and CMs. The differentiated DE was dissociated into single cells and incubated with antibody in DPBS containing 2% FBS and 2 mM EDTA at RT for 30 min. After the cells were washed with dPBS, they were analyzed with Accuri C6 flow cytometry (BD Biosciences). For the differentiated CMs, cells were dissociated as single cells and fixed and permeabilized using the Transcription Factor Buffer Set (BD Biosciences, San Jose, CA) according to the manufacturer's instructions. Antibodies were diluted 1:40 and incubated at 4°C for 40 min. After incubation, the samples were washed twice with Perm/Wash Buffer (BD Biosciences). FACS analysis was performed with an Accuri C6 flow cytometer (BD Biosciences), and data were analyzed using FlowJo V10 software (TreeStar, USA). The antibodies used in this study are listed in Supplementary Table S1.

Acknowledgements

This work was supported by a grant from the Technology Innovation Program (No. 20008777) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea), the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science, ICT and Future Planning (NRF-2018M3A9H3023077), and the KRIBB Research Initiative Program. The funders had no role in the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

Conception and design: D.-S. Kim, C.-R. Jung, Mi.-O. Lee, M.-Y. Son, H.-S. Cho; Web page development: J.-W. Ryu, I. Jang, J. Choi, B. Lee, J. Jung; Development of methodology: Y. S. Son, K. B, Jung, Mi.-O. Lee, C.-R. Jung; Analysis and interpretation of data: J.-W. Ryu, J.-H. Oh, J.-H, Ahn, M.-Y. Son, J. Jung, K. Park, D. -S. Kim; Writing and review of manuscript: D.-S. Kim, C.-R. Jung, Mi.-O. Lee, M.-Y. Son, H.-S. Cho; Study supervision: D.-S. Kim, M.-Y. Son, H.-S. Cho

Additional information

Supplementary information is available for this paper

Competing interests

The authors declare that they have no competing interests.

Refference

1. Speakman J, Hambly C, Mitchell S, Krol E. The contribution of animal models to the study of obesity. *Lab Anim* **42**, 413-432 (2008).

2. Vandamme TF. Use of rodents as models of human diseases. *J Pharm Bioallied Sci* **6**, 2-9 (2014).

3. Awatade NT*, et al.* Human Primary Epithelial Cell Models: Promising Tools in the Era of Cystic Fibrosis Personalized Medicine. *Front Pharmacol* **9**, 1429 (2018).

4. Zeilinger K, Freyer N, Damm G, Seehofer D, Knospel F. Cell sources for in vitro human liver cell culture models. *Exp Biol Med (Maywood)* **241**, 1684-1698 (2016).

5. Cohen DE, Melton D. Turning straw into gold: directing cell fate for regenerative medicine. *Nat Rev Genet* **12**, 243-252 (2011).

6. Zhao T, Li Y, Deng H. Cell fate conversion-from the viewpoint of small molecules and lineage specifiers. *Diabetes Obes Metab* **18 Suppl 1**, 3-9 (2016).

7. Maury Y*, et al.* Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. *Nat Biotechnol* **33**, 89-96 (2015).

8. Burridge PW*, et al.* Chemically defined generation of human cardiomyocytes. *Nat Methods* **11**, 855-860 (2014).

9. Hannan NR, Segeritz CP, Touboul T, Vallier L. Production of hepatocyte-like cells from human pluripotent stem cells. *Nat Protoc* **8**, 430-437 (2013).

10. Huang SX*, et al.* Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nat Biotechnol* **32**, 84-91 (2014).

11. Morris SA, Daley GQ. A blueprint for engineering cell fate: current technologies to reprogram

cell identity. *Cell Res* **23**, 33-48 (2013).

12. Dye BR*, et al.* In vitro generation of human pluripotent stem cell derived lung organoids. *Elife* **4**, (2015).

13. Lancaster MA*, et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373-379 (2013).

14. McCracken KW*, et al.* Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. *Nature* **516**, 400-404 (2014).

15. Dutta D, Heo I, Clevers H. Disease Modeling in Stem Cell-Derived 3D Organoid Systems. *Trends Mol Med* **23**, 393-410 (2017).

16. Lancaster MA, Knoblich JA. Organogenesis in a dish: modeling development and disease using organoid technologies. *Science* **345**, 1247125 (2014).

17. Nam SA*, et al.* Graft immaturity and safety concerns in transplanted human kidney organoids. *Exp Mol Med* **51**, 1-13 (2019).

18. Nascimento JM*, et al.* Human Cerebral Organoids and Fetal Brain Tissue Share Proteomic Similarities. *Front Cell Dev Biol* **7**, 303 (2019).

19. Koo B, Choi B, Park H, Yoon KJ. Past, Present, and Future of Brain Organoid Technology. *Mol Cells* **42**, 617-627 (2019).

20. Phipson B*, et al.* Evaluation of variability in human kidney organoids. *Nat Methods* **16**, 79-87 (2019).

21. Camp JG*, et al.* Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc Natl Acad Sci U S A* **112**, 15672-15677 (2015).

22. Jung KB*, et al.* Interleukin-2 induces the in vitro maturation of human pluripotent stem cell-derived intestinal organoids. *Nat Commun* **9**, 3039 (2018).

23. Chen YW*, et al.* A three-dimensional model of human lung development and disease from pluripotent stem cells. *Nat Cell Biol* **19**, 542-549 (2017).

24. Kim DS*, et al.* A liver-specific gene expression panel predicts the differentiation status of in vitro hepatocyte models. *Hepatology* **66**, 1662-1674 (2017).

25. Mun SJ*, et al.* Generation of expandable human pluripotent stem cell-derived hepatocyte-like liver organoids. *J Hepatol*, (2019).

26. Au KF, Sebastiano V. The transcriptome of human pluripotent stem cells. *Curr Opin Genet Dev* **28**, 71-77 (2014).

27. Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ. CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903-915 (2014).

28. Wu JQ*, et al.* Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. *Proc Natl Acad Sci U S A* **107**, 5254-5259 (2010).

29. Lee MO*, et al.* Modelling cardiac fibrosis using three-dimensional cardiac microtissues derived from human embryonic stem cells. *J Biol Eng* **13**, 15 (2019).

30. Zhou P, Pu WT. Recounting Cardiac Cellular Composition. *Circ Res* **118**, 368-370 (2016).

31. Jung KB*, et al.* Blockade of STAT3 Causes Severe In Vitro and In Vivo Maturation Defects in Intestinal Organoids Derived from Human Embryonic Stem Cells. *J Clin Med* **8**, (2019).

32. Lee H*, et al.* Low-dose interleukin-2 alleviates dextran sodium sulfate-induced colitis in mice by recovering intestinal integrity and inhibiting AKT-dependent pathways. *Theranostics* **10**, 5048-5063 (2020).

Figure legends

Fig. 1 Construction of Organ-GEPs

a Research schematic summary of the W-SAS system. **b** A heatmap representing the gene expression of organ-GEPs (149 LuGEP, 73 StGEP, 144 HtGEP) in % samples from 21 tissues derived from the GTEx portal. **c** MDA plot analysis with RNA-seq results of organ-GEP derived from the GTEx portal. **d** Disease and biofunctional analysis using IPA (Ingenuity Pathway Knowledge Base). The top regulated disease and biofunction are selected, and the significance is displayed with a heatmap.

Fig. 2 Development and evaluation of the Organ-GEP algorithm

a Distribution of the Organ-GEP algorithm (LuGEP, StGEP, HtGEP) score. RNA-seq data were downloaded from the GTEx portal. A box plot shows the interquartile (IQR) range of each algorithm scores in 21 tissue types. **b** Validation of the Organ-GEP algorithm using in-house data (20 tissues). Pooled RNA-seq data of each tissue were purchased from Clontech.

Fig. 3 LuGEP indicates the similarity of hPSC-derived lung organoids to lung

a Representative flow cytometry plot and bright field images of stepwise differentiation to LBO. Flow cytometry plot (left panel) presented the percentage of DE differentiated cells (C-Kit$^+$ CXCR4$^+$) on differentiation day 4. Blue dots indicate analysis in undifferentiated hESCs (D0), and red dots indicate FACS analysis of DE cells (D4). The number indicates the percentage of C-Kit$^+$ CXCR4$^+$ differentiated cells in DE cells (D4.) Bright field images displayed the ventralized AFE spheroids (middle panel, D21), budding spheroids after embedding in

Matrigel (lower panel, D23), and branching LBO (right panel, D56 after differentiation). **b** Immunofluorescence staining images of hLBO (D56) sections. LBOs were fixed with 4% paraformaldehyde and stained with anti-NKX2.1, anti-SOX9, anti-SFTPC, anti-acTUB (Cy3, Red), anti-EPCAM, anti-CC10 (Alexa Fluor 488, white or green) and DAPI (blue). Scale bar, 50 μm. **c** qRT-PCR analysis of cell type-specific markers was performed in hESCs (D0), DEs (D4), vAFEs (D21) and LBOs (D56). **d** The result of the LuGEP algorithm with hESC-derived LBO and human lung. **e** PCA analysis with LuGEP results in hESC-derived LBO and human lung. One dot indicates one gene, and the exchange of color indicates the change in gene expression. (green; 0, red; 10).

Fig. 4 Application of StGEP analysis with hESC-derived gastric organoids

a Representative immunofluorescence staining images of gastric organoid sections. Anti-SOX17, anti-HNF1β, anti-PDX1, anti-ECAD (Alexa Fluor 488; green), anti-SOX9, anti-SST, anti-FOXA2, anti-SOX2, anti-KLF5, anti-MUC5AC, and anti-MUC6 antibodies (Alexa Fluor 633; red) were used. DAPI (blue). Scale bar, 200 μm and 100 μm. **b** qRT-PCR analysis using cell type-specific markers in the developmental stage of hESC-derived organoids. **c** Gastric organoids. Scale bar, 100 μm. **d** The result of the StGEP algorithm with hESC-derived gastric organoids and human stomach. **e** PCA analysis with StGEP results. The pattern of colors and dots shows the StGEP results regarding each developmental stage of hESC-derived gastric organoids (green; 0, red; 15).

Fig. 5 Quantitative similarity of hESC-derived CMs to heart using HtGEP analysis

a qRT-PCR analysis using cardiac transcription, muscle, ventricular, and atrial markers in the developmental stage of hESC-derived CMs. **b** Immunofluorescence staining images of hESC-derived CMs and CMs+BMP4. Anti-cTnT and anti-NKX2.5 antibodies (Alexa Fluor 488; green), anti-MYL2 and anti-MLC2a antibodies (Alexa Fluor 633; red), and DAPI (blue) were used. Scale bar, 50 μm. **c** FACS analysis using cardiac muscle-specific markers (cardiac troponin T, cTnT). The number indicates the differentiation rate of the hESC-derived CMs and CMs+BMP4. **d** The result of the HtGEP algorithm with the hESC-derived CMs and human heart. **e** PCA analysis with the HtGEP results. The change in color indicates the gene expression level of HtGEP in hESC-derived CMs (green; 0, red; 10).

Fig. 6 W-SAS is a user-friendly interface of the quantitative calculation system for similarity of hPSC-derived organoids and cells to organs

a Main page of W-SAS. **b** A workflow for W-SAS. Using RNA-seq results of hPSC-derived organoids and cells, W-SAS can calculate the similarity of each organoid and cell compared with the human target organ. **c** Results page of similarity for the hPSC-derived organoids. The results were divided into five parts: similarity percentage (1), heat map (2), PCA plot (3), and gene expression panel of organ-GEPs or samples (4 and 5).

Figures

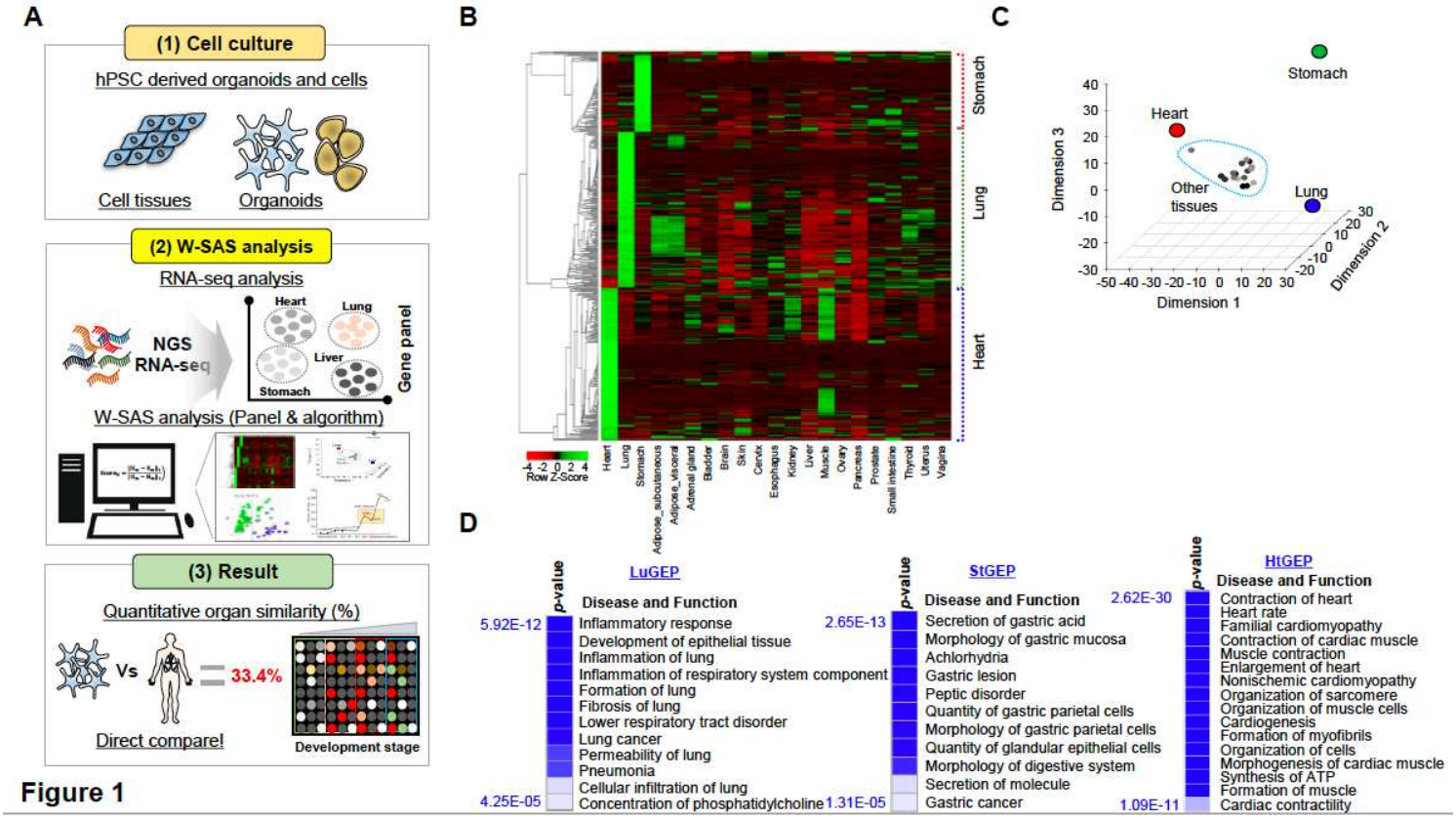

Figure 1

Figure 1

Construction of Organ-GEPs a Research schematic summary of the W-SAS system. b A heatmap representing the gene expression of organ-GEPs (149 LuGEP, 73 StGEP, 144 HtGEP) in % samples from 21 tissues derived from the GTEx portal. c MDA plot analysis with RNA-seq results of organ-GEP derived from the GTEx portal. d Disease and biofunctional analysis using IPA (Ingenuity Pathway Knowledge Base). The top regulated disease and biofunction are selected, and the significance is displayed with a heatmap.

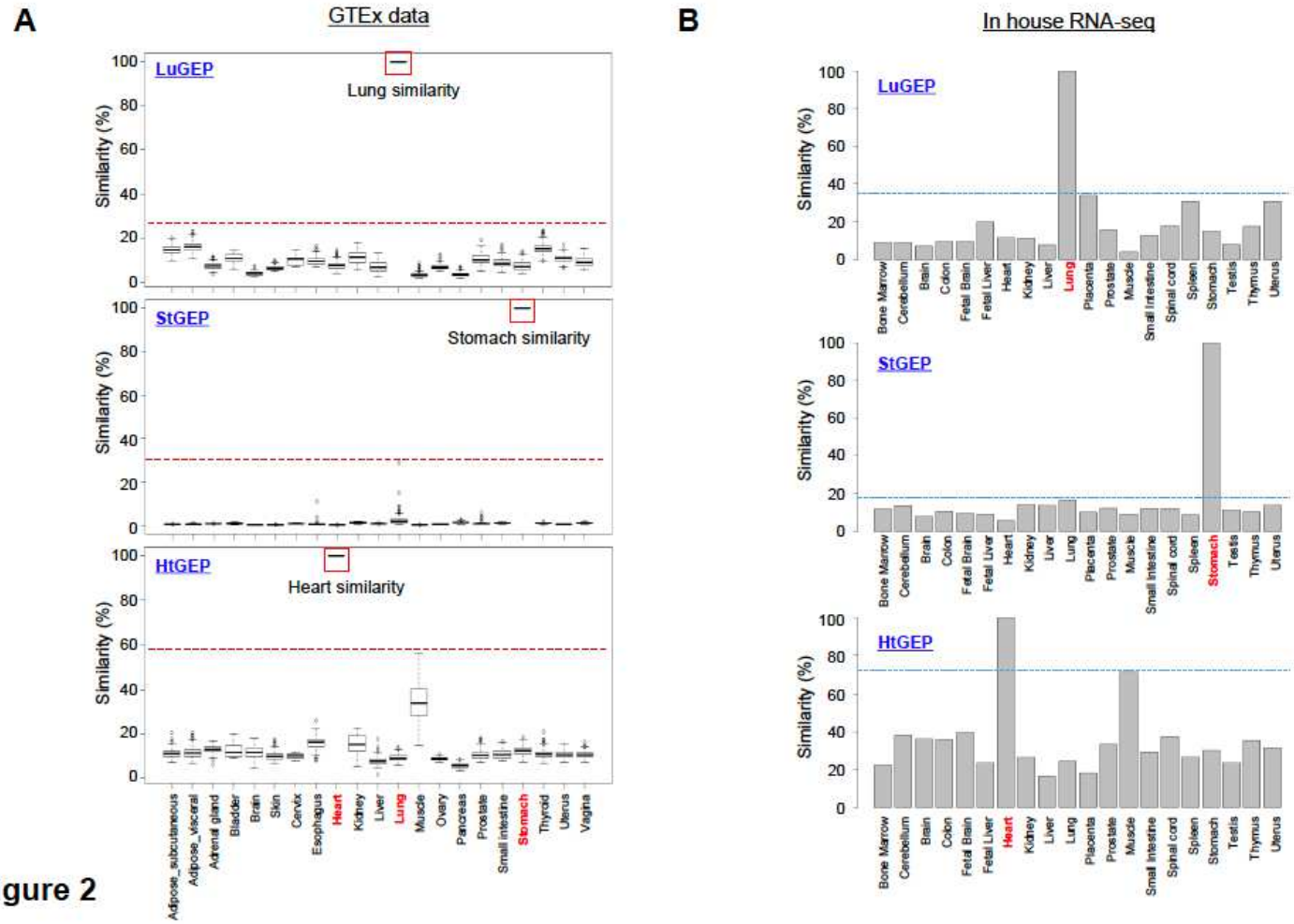

Figure 2

Figure 2

Development and evaluation of the Organ-GEP algorithm a Distribution of the Organ-GEP algorithm (LuGEP, StGEP, HtGEP) score. RNA-seq data were downloaded from the GTEx portal. A box plot shows the interquartile (IQR) range of each algorithm scores in 21 tissue types. b Validation of the Organ-GEP algorithm using in-house data (20 tissues). Pooled RNA-seq data of each tissue were purchased from Clontech.

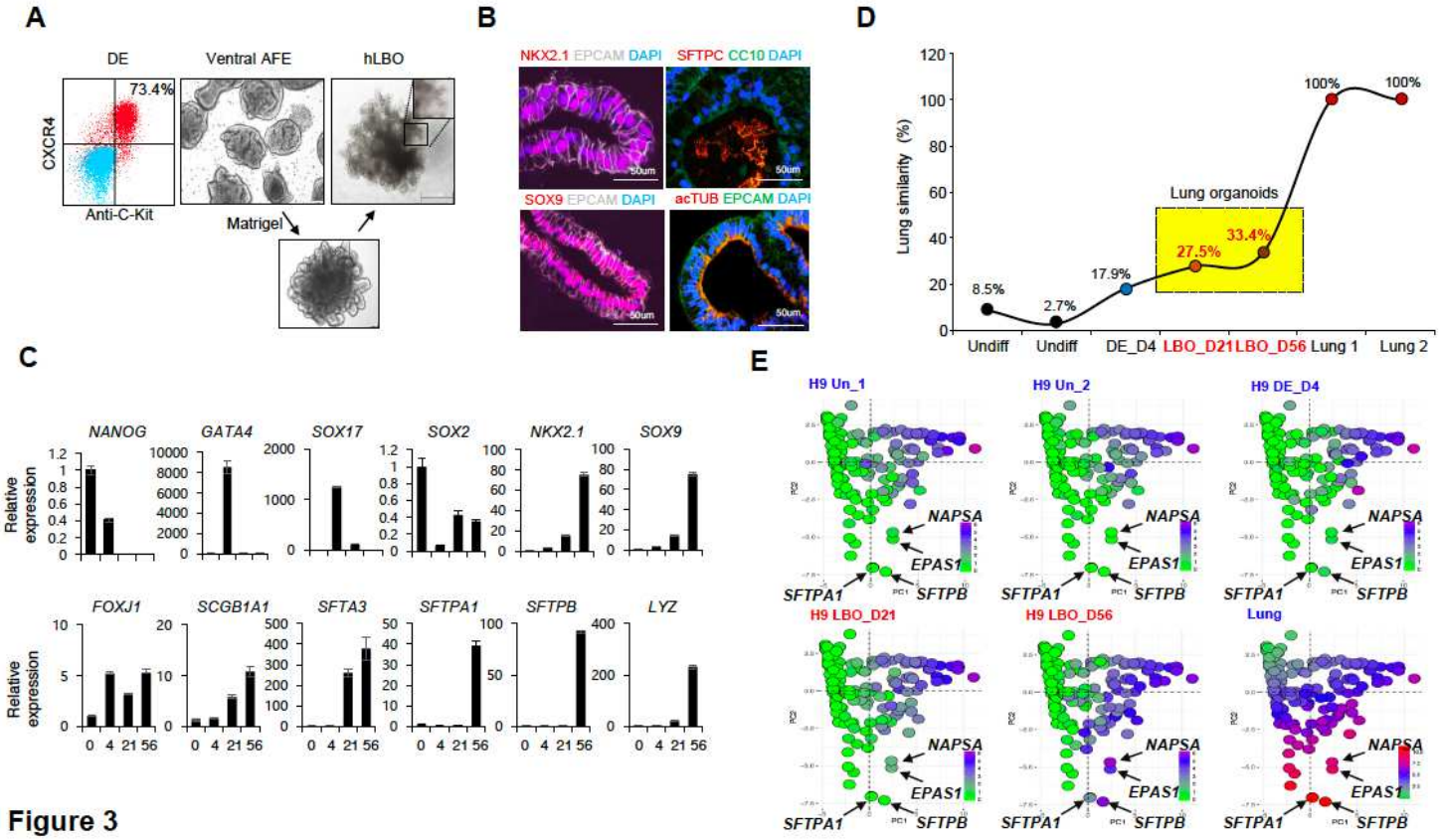

Figure 3

Figure 3

LuGEP indicates the similarity of hPSC-derived lung organoids to lung a Representative flow cytometry plot and bright field images of stepwise differentiation to LBO. Flow cytometry plot (left panel) presented the percentage of DE differentiated cells (C-Kit+ CXCR4+) on differentiation day 4. Blue dots indicate analysis in undifferentiated hESCs (D0), and red dots indicate FACS analysis of DE cells (D4). The number indicates the percentage of C-Kit+ CXCR4+ differentiated cells in DE cells (D4.) Bright field images displayed the ventralized AFE spheroids (middle panel, D21), budding spheroids after embedding in Matrigel (lower panel, D23), and branching LBO (right panel, D56 after differentiation). b Immunofluorescence staining images of hLBO (D56) sections. LBOs were fixed with 4% paraformaldehyde and stained with anti-NKX2.1, anti-SOX9, anti-SFTPC, anti-acTUB (Cy3, Red), anti-EPCAM, anti-CC10 (Alexa Fluor 488, white or green) and DAPI (blue). Scale bar, 50 μm. c qRT-PCR analysis of cell type-specific markers was performed in hESCs (D0), DEs (D4), vAFEs (D21) and LBOs (D56). d The result of the LuGEP algorithm with hESC-derived LBO and human lung. e PCA analysis with LuGEP results in hESC-derived LBO and human lung. One dot indicates one gene, and the exchange of color indicates the change in gene expression. (green; 0, red; 10).

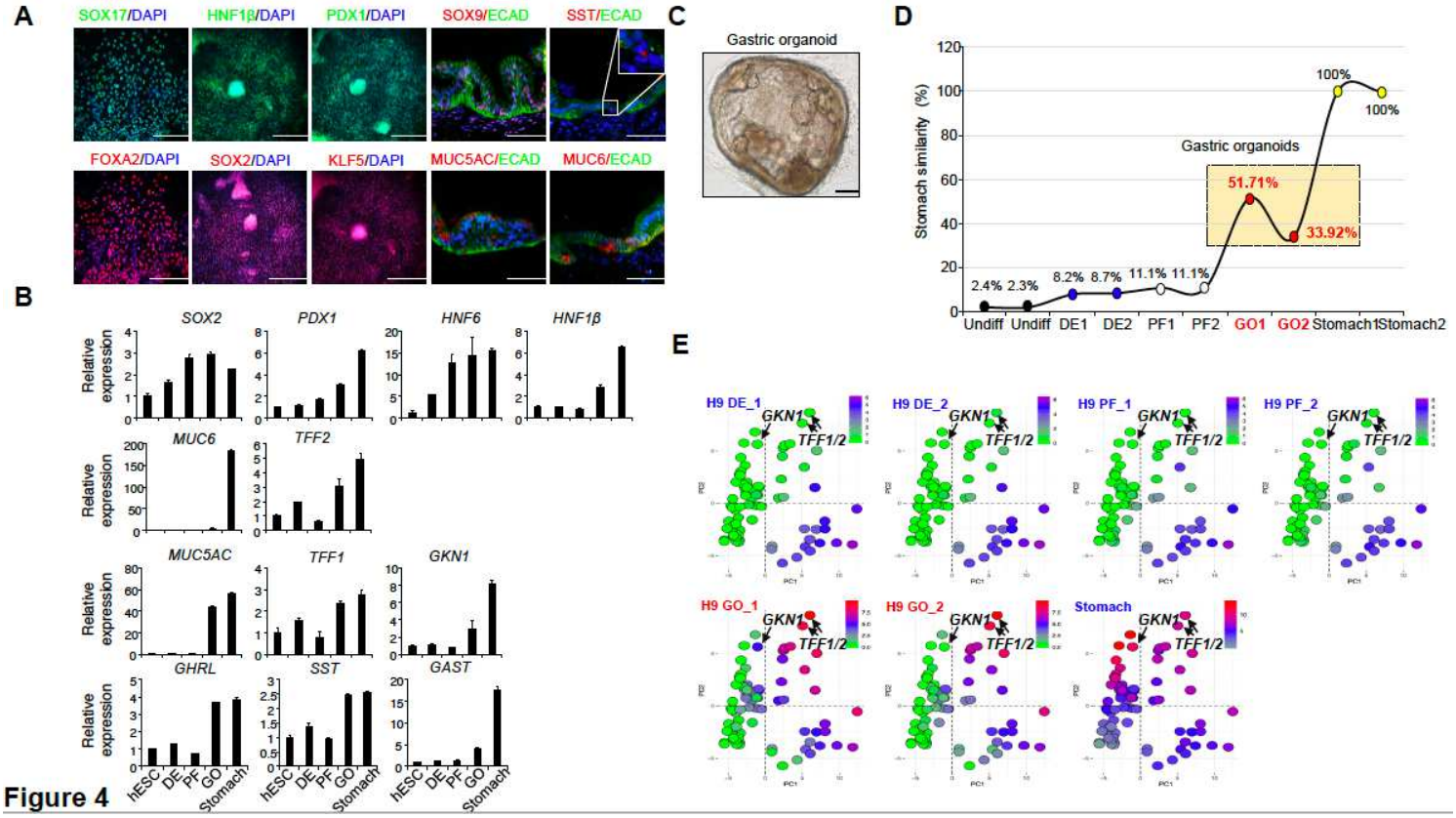

Figure 4

Figure 4

Application of StGEP analysis with hESC-derived gastric organoids a Representative immunofluorescence staining images of gastric organoid sections. Anti-SOX17, anti-HNF1β, anti-PDX1, anti-ECAD (Alexa Fluor 488; green), anti-SOX9, anti-SST, anti-FOXA2, anti-SOX2, anti-KLF5, anti-MUC5AC, and anti-MUC6 antibodies (Alexa Fluor 633; red) were used. DAPI (blue). Scale bar, 200 μm and 100 μm. b qRT-PCR analysis using cell type-specific markers in the developmental stage of hESC-derived organoids. c Gastric organoids. Scale bar, 100 μm. d The result of the StGEP algorithm with hESC-derived gastric organoids and human stomach. e PCA analysis with StGEP results. The pattern of colors and dots shows the StGEP results regarding each developmental stage of hESC-derived gastric organoids (green; 0, red; 15).

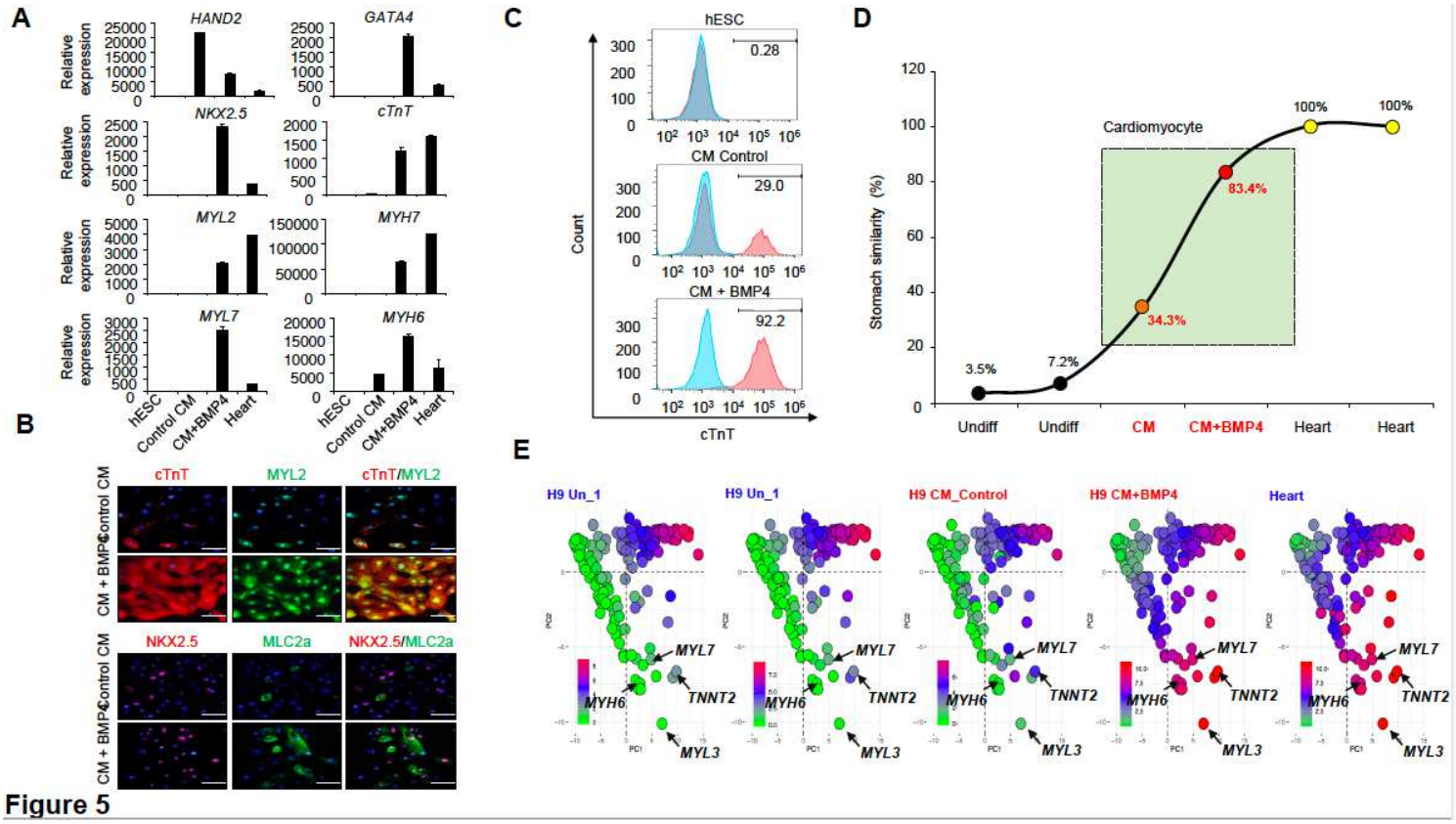

Figure 5

Figure 5

Quantitative similarity of hESC-derived CMs to heart using HtGEP analysis a qRT-PCR analysis using cardiac transcription, muscle, ventricular, and atrial markers in the developmental stage of hESC-derived CMs. b Immunofluorescence staining images of hESC-derived CMs and CMs+BMP4. Anti-cTnT and anti-NKX2.5 antibodies (Alexa Fluor 488; green), anti-MYL2 and anti-MLC2a antibodies (Alexa Fluor 633; red), and DAPI (blue) were used. Scale bar, 50 μm. c FACS analysis using cardiac muscle-specific markers (cardiac troponin T, cTnT). The number indicates the differentiation rate of the hESC-derived CMs and CMs+BMP4. d The result of the HtGEP algorithm with the hESC-derived CMs and human heart. e PCA analysis with the HtGEP results. The change in color indicates the gene expression level of HtGEP in hESC-derived CMs (green; 0, red; 10).

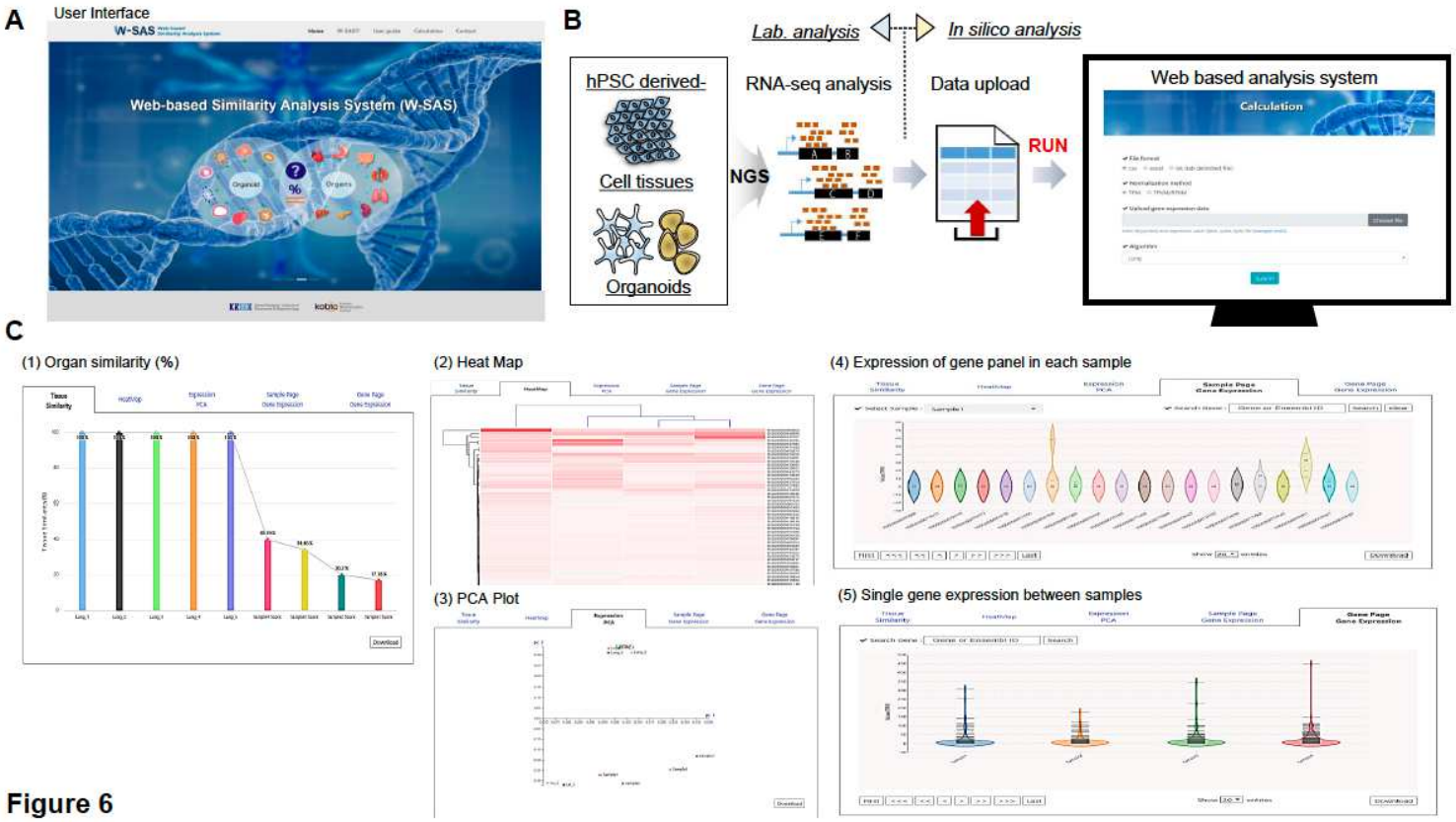

Figure 6

Figure 6

W-SAS is a user-friendly interface of the quantitative calculation system for similarity of hPSC-derived organoids and cells to organs a Main page of W-SAS. b A workflow for W-SAS. Using RNA-seq results of hPSC-derived organoids and cells, W-SAS can calculate the similarity of each organoid and cell compared with the human target organ. c Results page of similarity for the hPSC-derived organoids. The results were divided into five parts: similarity percentage (1), heat map (2), PCA plot (3), and gene expression panel of organ-GEPs or samples (4 and 5).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesfinal.pdf