The molecular basis of the cellular taxonomy of the human body
Alessandra Breschi, Carrie Davis, Sarah Djebali, Jesse Gillis, Dmitri D. Pervouchine, Anna Vlasova, Alex Dobin, Chris Zaleski, Jorg Drenkow, Cassidy Danyko, et al.

To cite this version:
Alessandra Breschi, Carrie Davis, Sarah Djebali, Jesse Gillis, Dmitri D. Pervouchine, et al.. The molecular basis of the cellular taxonomy of the human body. Human Genomics, 2018, 12 (1), Non paginé. hal-02629292

HAL Id: hal-02629292
https://hal.inrae.fr/hal-02629292
Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
A107

The molecular basis of the cellular taxonomy of the human body
Alessandra Breschi1,2, Carrie Davis1, Sarah Djebali3, Jesse Gillis1, Dmitri D. Pervouchine1, Anna Vlasova1, Alex Dobin1, Chris Zaleski2, Jorg Drenkow2, Cassidy Danyko1, Alexandria Scavelli1, Manuel Munoz1, Diego Gamito1,2, Ferran Reverter1, Thomas R. Gingeras3, Roderic Guigo1,2, Pervouchine5, Anna Vlasova1, Alex Dobin3, Chris Zaleski3, Jorg Drenkow3, U.S. National Cancer Institute (NCI) retired their NCI-60 panel of cell lines and xenografts are models of patient tumors. Since the first cell line and xenografts are models of patient tumors, and how these impact tissue phenotypes has not been well established. Here we have produced RNA sequencing data for a number of primary cells from multiple human body locations. The analysis of this data, together with additional epigenetic data also produced by the ENCODE project for a total of 146 primary cells, indicate that most cells in the human body belong to five major cell types: epithelial, endothelial, mesenchymal, neural and blood cells. These redefine, based on gene expression, the basic histological types in which tissues are usually classified. We identified genes specific to these cell types, including a core set of transcription factors (TFs). Cell type specific genes, particularly when lying in open chromatin domains, are enriched for motifs for these cell type specific TFs, suggesting that they are potential candidates to drive cell type specificity. We estimated the relative proportion in tissues of the different cell types based on the transcriptional profiles obtained from bulk tissue sections. This inferred cellular composition is a characteristic signature of tissues and reflects.

A108

Know thy cells: the classification of tumor models by tissue, disease, sex, and species
Heather Selby1, Mark Kon1,2, John Quackenbush3,4,5, and Benjamin Haibe-Kains2,6
1Boston University, Bioinformatics Program, Boston, MA, USA; 2Boston University, Mathematics and Statistics, Boston, MA, USA; 3Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA USA; 4Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA USA; 5Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada; 6Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada.

Correspondence: Heather Selby
Human Genomics 2018, 12(Suppl 1):A107

Cell lines and xenografts are models of patient tumors. Since the first patient-derived tumor cell line, HeLa, was created, new cell lines have been repeatedly established from patients. Large cell line collections now broadly capture the genomic diversity of patient tumors, and provide valuable insight into drug response. Patient-derived xenografts (PDXs) have begun to replace cell lines as more predictive experimental tumor models for drug development. PDXs are directly established from patient tumors in immune-deficient mice, and better resemble the actual patient tumor than cell lines. Recently, the U.S. National Cancer Institute (NCI) retired their NCI-60 panel of patient-derived tumor cell lines to refocus their anti-cancer drug screening on PDXs. Misidentified cell lines and xenografts are invalid models of patient tumors. The growth of cells in culture or xenografts makes tumor models susceptible to cross-contamination and/or mislabeling in spite of best laboratory practice. Although cell line misidentification is now widely-recognized, and authentication by short tandem repeat (STR) recommended, misidentified cell lines continue to be pervasive problems in cancer research. The International Cell Line Authentication Committee (ICLAC) has documented as many as 451 misidentified cell lines. Misidentified tumor models are often not from their original donor, and may even come from another tissue, disease, sex, or species.

To address the problem of misidentified tumor models, we developed a new classifier, Top Scoring-Binary Gene Pair (TS-BGP). The TS-BGP classifier, trained on tumor gene expression profiles, was used to predict the tissue, disease, sex, and species of both cell lines and PDXs. Binary gene pair signatures were extracted from the 31 tissues in the Genotype-Tissue Expression (GTEx) project, 32 diseases in The Cancer Genome Atlas (TCGA), 2 sexes in the GTEx project, and 2 species in the Encyclopedia of DNA Elements (ENCODE). The predictive accuracies of the leave-one-out-cross validations were 97.51% (GTEx; n=9115) for tissue, 90.89% (TCGA; n=10088) for disease, and 99.10% for sex (GTEx; n=9115), and 100% for species (ENCODE; n=622). The TS-BGP classifier identifies tumor models that accurately resemble patient tumors. With tumors models that resemble patient tumors, data generated in pre-clinical research can successfully identify translatable drugs and therapies that demonstrate clinical efficacy. The TS-BGP classification of tumor models by tissue, disease, sex, and species helps ensure the credibility, reproducibility, and translation of pre-clinical research.

A109

RIKEN Ageing Resource Data Project: Single cell transcriptome and epigenetic changes in mice
Tommy Terooate1,2, Tsukasa Kouno1, Matteo Guerrini3, Yasutaka Motomura4, Naoko Sato3, Takeshi Matsui4, Sidonia Fagarasan5, Hiroobu Fujivara1, Kazuyo Moro6, Hiroshi Ohno7, Ichiro Taniuchi7, Kosuke Hashimoto1, Piero Carninci8, Aki Minoda9
1RIKEN CLST-DGT, Yokohama, Japan; 2RIKEN IMS, Yokohama, Japan; 3RIKEN CDB, Kobe, Japan; 4Department of Internal Medicine, School of Medicine, Keio University, Tokyo, Japan.

Correspondence: Tommy Terooate; Aki Minoda

Bulk analyses of high-throughput genomic and proteomic technologies have produced invaluable publicly available data on ageing. However, these data are limited in terms of accurately defining cellular states, development and disease state. Single-cell omics on the other hand help overcome these obstacles and promises greater understanding of cellular processes. Consequently, we are currently generating a resource dataset for ageing studies with a focus on single cell RNA-seq. We present here our initial results produced from various tissues taken at different ages of the mouse. One aspect of our analysis will be to determine whether we observe dynamic evolution of cells in heterogeneity throughout ageing, which may reveal key regulatory pathways that are gradually deregulated during ageing. We will also present our plan for obtaining single cell ATAC-seq and multi-omic datasets (DNA methylation, CAGE, transl一流的me (ribosome profiling), proteomics and metabolomics) for the selected cell types of interest and integration analysis. Additionally, a unique aspect of our dataset is the utilization of both germ-free and SPF mouse models, which will enable us to introduce and access the effects of