Mapping human cell phenotypes to genotypes with single-cell genomics

Review Article

Author(s):
Camp, J. Gray; Platt, Randall; Treutlein, Barbara

Publication date:
2019-09-27

Permanent link:
https://doi.org/10.3929/ethz-b-000368526

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
Science 365(6460), https://doi.org/10.1126/science.aax6648

Funding acknowledgement:
758877 - Reconstructing human cortex development and malformation with single-cell transcriptomics (EC)
Mapping human cell phenotypes to genotypes with single-cell genomics

J. Gray Camp¹, Randall Platt², Barbara Treutlein²

Affiliations:
¹ Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland
² Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

ABSTRACT

The cumulative activity of all of our cells, with their myriad of interactions, life histories, and environmental experiences, gives rise to a condition that is uniquely human, and specific to each individual. It is an enduring goal to catalog our human cell types, to understand how they develop, how they vary between individuals, and how they fail in disease. Single-cell genomics has revolutionized this endeavor as sequencing-based methods provide a means to quantitatively annotate cell states based on high-information content and high-throughput measurements. Together with advances in stem cell biology and gene editing, we are in the midst of a fascinating journey to understand the cellular phenotypes that compose human bodies and how the human genome is used to build and maintain each cell. Here we will review recent advances into how single-cell genomics is being used to develop personalized phenotyping strategies that cross subcellular, cellular, and tissue scales to link our genome to our cumulative cellular phenotypes.

Phenotyping in the single-cell sequencing era

Phenotype can mean many things, but in general it is a way to classify a set of properties that arise from the interaction of an individual’s genotype with its environment. It is a reductionist, yet powerful, approach to take human phenotyping down to the level of single cells and use molecular states within cells to establish phenotypes at the molecular, cellular, and tissue/system level. Single-cell sequencing technologies can measure thousands of individual features per cell for thousands of cells at a time, providing a quantitative and ultra-high resolution snapshot of cell and molecular states composing a human tissue, organ, or other biosystem. Currently, there are protocols available to measure the RNA content, DNA sequence and methylation status, chromatin structure and accessibility, and protein composition in single cells(¹). In addition to molecular features, cell histories can be measured in certain scenarios where mutations have arisen in nuclear(2-4) or mitochondrial(5) DNA that distinguish lineages. Furthermore, in genetically tractable model systems (such as mice, zebrafish, and organoids), it is possible to record cell fate histories and infer lineage trees using reporter barcodes and genetic scarring based on RNA-guided CRISPR (clustered regularly interspaced short palindromic repeats)-associated (Cas) nucleases(6-9). So far, generally one or two single-cell measures have been used to phenotype individual cells within tissues in each experiment. However, the field is moving towards multimodal measurements from the same cell that capture
transcriptomic, genomic, epigenomic, and lineage states to enhance feature quantifications and give a richer picture of a cell’s phenotype at any given moment(10-12). These technologies are forcing scientists to grapple with previous notions of “cell type”, as cell properties can vary discretely (type or subtype), as well as continuously (state), and classification may not always follow rigid hierarchies(13). An additional layer of complexity arises in developing systems where cell states and tissue morphologies are changing rapidly with time. Single-cell sequencing captures transient states, and computational approaches enable trajectory(14) and lineage(9) reconstructions.

The microenvironment of each cell can also be critical to a individual cell’s phenotype. Spatial transcriptomic and proteomic approaches based on multiplexed RNA hybridizations(15, 16) or protein immunohistochemistry(17), in situ sequencing(18-20), mass cytometry(21) or other strategies(22) can be used to measure cell types and states in situ. In addition, methods are being developed to analyze spatial locations of molecules within a cell and cells within a tissue using barcoded oligonucleotides that can couple together when the molecules are close enough to physically interact(23, 24). These pairwise interactions are encoded in DNA and can be measured by high-throughput sequencing, and computational analyses enable spatial reconstruction of cell interactions based on the molecular proximitities.

Single-cell phenotyping has provided extraordinary atlases of model organisms that span organ systems from the same animal(25-27), link cell morphologies with molecular features(28), resolve cell type classifications(18, 29) and gradients of cell states (30, 31) spatially, and assemble cell fates maps using lineage recording(7). This wave is extending to developing, mature, and aged human organs(32, 33), identifying previously undocumented human cell types(34-36) and providing a quantitative framework to classify human cell subtypes and other states(37, 38). A major technical goal is to integrate(39, 40) all possible feature detection methods (RNA, DNA, chromatin state, lineage, etc.) performed independently or in combination in single cells, and render these measurements into three-dimensional spatial volumes with subcellular and temporal resolution (Figure 1). By measuring cell heterogeneity with such high information content techniques, emergent phenotypes at the tissue or biosystem level (system phenotype) such as composition, regulatory states, cell interactions, spatial constructions, and differentiation trajectories can be compared across environmental and genetic conditions. Because human development is not deterministic, it will be interesting to see how variable organ reconstructions are across different humans. Similar to the Human Genome Project, the Human Cell Atlas (HCA) consortium will confront the challenge of how to generate a reference organ map when the organ may look different in every person. The process of assessing such variations using integrated multi-modal measurements, as is being proposed in HCA projects over the next few years, will help to create common coordinated frameworks for sampling tissues, processing samples and data, and computational comparisons between methods and individuals.

Disease associated cell phenotypes

As each human and mouse organ is mapped at single-cell resolution across space and time, a next phase of inquiry is to understand how genetic changes impact human phenotypes. Disease association studies have used naturally occuring mutations to identify disease-causing
genes; these genes are Mendelian, have strong-effect coding mutations, or have been experimentally validated. Disease-causing genes can be mapped to cell types by identifying the cell population where the gene is expressed (multi-organ(41), kidney(42), cortex(43); Figure 2a).

A notable example was shown recently when researchers mapped out the cell types within the adult human lung, and identified a new human cell type (“ionocyte”), which was the only type that expressed CFTR at high levels and therefore likely mediates the lung pathology observed in cystic fibrosis (35, 36). However, most disease associated mutations identified through GWAS are likely regulatory, and the genes they regulate are currently unknown. Multimodel measures combining single-cell DNA, RNA, and chromatin can locate portions of the genome that are active in a certain cell type or state providing a link from the regulatory genome to the cell and even tissue phenotype. Future exploration into population measures of cell and tissue phenotypes at the single-cell level(44, 45), similar to what has been done using bulk RNAseq of adult human organs from the GTEx project(46), will be critical to identify the mechanism of action of most GWAS variants.

In addition, researchers have started to phenotype human diseased tissues (Figure 2b), thereby elucidating spatial glial neuron interactions in Amyotrophic Lateral Sclerosis(47), identifying disease associated microglia in Alzheimers disease(48) and fibroblasts in Arthritis (49), profiling pancreatic islets in health and type 2 diabetes(50), and identifying cellular rewiring during colitis(51). These types of analyses promise to bring about a new phase of molecular disease classification and diagnosis, and guide the development of therapies that can target the specific cell types impacted by a given disease. This is already happening for various cancers(52), however, there are major challenges requiring innovation to bring single-cell sequencing technologies to patients with other disorders on a clinical scale. Protocols to minimize cell loss during experiments could make it feasible to work with minute amounts of input material (53). Apart from cancer, it is difficult to acquire tissue from most diseased microenvironments in patients, and miniaturized biopsies from healthy and diseased regions of a tissue could open up disorders that lack clear molecular phenotypes. Industrialization of the single-cell –omics pipeline from sample preparation and sequencing, to data analysis is also required. This could include protocol optimizations to reduce cost of cell throughput while retaining sensitivity(25, 54, 55), to increase sample multiplexing based on reference polymorphisms(45) or tagging(56, 57), and to compress phenotyping by experimentally enriching for diagnostic features or select against non-diagnostic features(58). In situ sequencing approaches may be another route to increased throughput once methods have been industrialized. Finally, robust software pipelines will be needed that can rapidly analyze the high-dimensional data and output perturbation landscapes that are able to diagnose disease.

Comparing human with mouse and other species can reveal the power and limitations of model organisms for understanding human genotype to phenotype relationships. In many cases, the same broad cell classes are found in mammalian tissues, and cell states can be integrated across species(39). For example, recent single-cell analysis of the mouse kidney suggested that most known kidney disease-associated genes in humans map to the mouse cell counterparts(59). However there are cases in which the human and mouse tissues have diverged significantly in terms of cell composition and gene expression. For example, humans and other primates have a specialized area of the retina called the fovea, which can be distinguished based on the particular proportion and types of retinal neurons that have distinct
expression signatures from other areas of the retina (60). Mice on the other hand lack a fovea altogether, making it a poor model system for many blindness disorders. It may be possible to leverage such differences to understand the underlying genetic mechanisms that control human cell phenotypes. Altogether, comparisons between species, between human individuals, and between healthy and disease conditions can illuminate variety of human cell phenotypes, and help to link the genome to specific cell states. Now we need creative strategies to move past correlation and establish the functional relevance of the observations from these highly-resolved maps.

**Stem cell and genetic manipulation tools**

There has been an exciting revolution in the stem cell biology field, which has made it possible to generate diverse human cell types in controlled 2D cultures (61), and to generate complex 3D tissues that resemble the primary tissue/organ counterparts (termed organoids) (62, 63). Organoid protocols have been developed for various parts of the brain, liver, intestine, lung, kidney, stomach, etc., and the protocols are being optimized for stereotyped morphology (64) and inter-organoid reproducibility (65, 66). Many of these systems have been analyzed by single-cell genomics and the data compared with their primary tissue counterparts to quantitate how accurately cells states are recapitulated (67). The power of these *in vitro* systems is that they are specific to an individual, enable replicate measurements over a time course, are amenable to genetic manipulation and lineage recording in diverse environmental conditions, could be used for high-throughput screening, and can recapitulate certain disease phenotypes (Figure 3). For example, recent work showed that cerebral organoids recapitulate neuronal migration defects observed in patients with periventricular nodular heterotopia, and single-cell transcriptomics identified a perturbed differentiation trajectory (68). In addition, cerebral organoids were used to identify specific cell states that were sensitive to hypoxia conditions that occur in premature births, and organoids were used to screen for small molecules that prevented loss of of these cell states (69). As detailed above, integrating multiple single-cell genomic measurements together with cell history recorders and 3D spatial reconstructions will soon provide very exciting high-resolution phenotyping strategies in these personalized models of disease.

Culture systems that can recapitulate human development and physiology *in vitro* enable researchers to use RNA-guided CRISPR-associated (Cas) nucleases to interrogate these systems and link genotype to phenotype. CRISPR-Cas nucleases come in various natural as well as synthetically engineered flavors enabling diverse genome and epigenome modifications (70). These tools were originally established based on the Cas9 protein for gene editing in immortalized mammalian cell lines (71-73) and then harnessed to link phenotype to genotype using both a forward and reverse genetics approach (Figure 4). Reverse genetic approaches based on CRISPR, generally termed pooled CRISPR screens, involve the: i) production of genome-scale or sub-genomic gRNA libraries; ii) low multiplicity of infection delivery to cells such that single cells receive single perturbations; iii) enrichment or depletion of cells based on a cellular phenotype of interest (e.g. proliferation, death, or presence/absence of selectable marker or reporter); and iv) identification and analysis of genes corresponding to enriched and/or depleted gRNAs. Pooled CRISPR screens were originally demonstrated in...
human cancer cell lines (74, 75) but were recently extended and optimized in human iPSC culture systems to identify genes that regulate pluripotency(76). Moving beyond gene knockout, catalytically inactivated Cas proteins fused to effector domains enable diverse perturbations, including transcriptional activation (CRISPRa) or inhibition (CRISPRi)(77), DNA methylation or demethylation(78), histone acetylation(79), DNA (80) or RNA (81) base editing, as well as others(82). Many of these RNA-guided Cas effector proteins have successfully been used in pooled CRISPR screens further expanding our capacity to link diverse genetic and epigenetic features to phenotypes(77, 83-85). Researchers have started to bring these technologies to human iPSC cells by establishing stable cell lines that have inducible expression of the different Cas effectors, and these lines can then be used to explore phenotype-genotype relationships in a diversity of human cellular and tissue contexts.(5, 86, 87)

More recently, exciting work has merged CRISPR screening with single-cell genomic readouts(88-91). In these methods pools of gRNAs are introduced into cells along with a Cas protein such that cells express different gRNAs. Transcriptomes can be sequenced in single-cells and the gRNA that is expressed per cell can be determined based on an associated barcode, or through direct sequencing of the gRNA. In this way, the effect of many different gene perturbations can be examined in the same experiment with single-cell resolution. There are several considerations when designing single-cell perturbation screens in organoids, and optimizations on the initial protocols to reduce barcode recombination will enable more sensitive and accurate readouts(92). One needs to determine how many genes can be targeted based on a power analysis taking into account cell heterogeneity of the system (# cells), proportion of mutant and wild-type cells, depth of sequencing, # reads per cell, effect size of perturbation, and cost. Clonal selection within a stem cell culture or within an organoid can have an impact on the results, especially if the organoid system is initiated from a composite of many different stem cell clones (e.g. cerebral organoids). Cas protein can be constitutively or transiently expressed, or induced through multiple strategies (e.g. Tet/On, Cre) and gRNAs can be introduced into the cells or organoid through different delivery methods (AAV, lentivirus, transposon). Currently the gRNA or barcode needs to be the read-out rather than the genomic lesion making the readout correlative. If the proportion of mutant cells is too high, then the organoid may not develop properly and the presence of a sufficient proportion of wild type cells could buffer mutant effects. It is important to incorporate a selection feature of the cells receiving the gRNA (e.g. Fluorescence). Finally, genetic perturbations might be cell autonomous or non-autonomous and this can be difficult to distinguish in pooled screens with mosaic organoids, making arrayed screening important alternatives. CRISPR/Cas9 screening based on single-cell sequencing in iPSC-derived organoids has a rich future in the exploration of human cell phenotype to genotype relationships. This will be made possible through innovations to increase throughput to perform combinatorial genetic interaction and massively multiplexed screens, new technologies to combine screening with molecular recording(7, 93) and lineage tracing, and in situ readouts(94) based on in situ sequencing technologies to open up exploration into spatial effects of genetic perturbation.

**Conclusions and outlook:** In this review, we aimed to provide a foundational overview of the current state of single-cell genomic-based phenotyping of human organs and organoids, and how CRISPR-Cas technologies will enable phenotypes to be functionally linked to regions of the
human genome. We envision that the descriptive phase of single-cell genomics, where cell
phenotypes are catalogued for each healthy human tissue, will culminate in 4D resolved in silico
simulations that enable researchers to walk into the tissue, point to a cell at a location within the
tissue at a particular time point, and know its molecular features and its interaction with other
cells within the microenvironment. In the short term, the goal is to integrate and render different
feature (RNA, DNA, chromatin state, lineage) measurements into three-dimensional spatial
volumes with subcellular and temporal resolution. It will be exciting to incorporate single-cell
molecular measurements with in toto imaging of developing or cleared human organs/organoids
with cellular and subcellular resolution (95-98). Innovations in virtual-reality microscopy are
starting to bring together immersive visualization and simulation of imaged-based data with
hand gesture control, and the first virtual reality platform for the visualisation and analysis of
single-cell gene expression data has been developed (99).

Perturbation screens in human cells could be integrated with 4D spatiotemporal models
to weigh the functional relevance of genes and regulatory regions for establishing molecular,
cellular and systems-level phenotypes. However, there are still many limitations that require
innovations in the stem cell and organoid field that will enhance the biological insight that can be
attained from these efforts. Specifically, organoid morphology may not be stereotyped, there are
missing or off-target lineages, organoids are not integrated with other relevant organ systems,
and iPSC-derived organoids follow development and may not reflect processes in adults. It will
be important to continue to use single-cell genomics to assess the precision of novel organoid
protocols, and compare lineage and fate maps with other mammalian counterparts.

Establishment of stem cell resources from different populations (100) or the same individuals
from whom there are reference atlases will be useful for establishing a foundation for
quantitative comparisons across protocols. Furthermore, well characterized iPSC lines that
contain various flavors of CRISPR/Cas systems for genetic perturbation screens and cell fate
recording, together with a suite of diverse fluorescent reporters (101), will push the field forward.

Finally, a drive toward industrialization of single-cell sequencing pipelines on diseased
tissues and organoid models could bring exciting prospects for disease classification and
personalized medicine. This will require close collaborations between basic and clinical
researchers, as well as industry partners, to identify the unmet medical needs where single-cell
sequencing could have the most immediate effect. There are many disorders where multiple
genetic drivers are known, however it is increasingly clear that the different causative genes are
expressed in very different cell types (102). This is a major conundrum where molecular
dissection of the disorder in primary tissues and/or organoid models could enable a refined
classification of the disorder, and also identify mechanisms that underly particular disease
presentations. Optimization of organoid protocols together with increased single-cell genomic
throughput would also enable testing disease-associated environmental conditions and potential
pharmaceutical-, gene- and cell-based therapies. There are many obstacles that remain,
however the field is moving forward at an extraordinary pace and it will be exciting to see where
it goes from here.
Figure 1: Human phenotyping in the single-cell genomics era. (a) Many different single-cell genomic methods have been developed to profile cell heterogeneity in human organs across space and time. (b-c) These measurements can be integrated to build phenotypic maps that cross sub-cellular, cellular, and tissue/system scales.

Figure 2: Human organ maps can resolve disease phenotypes. (a) Single-cell resolved human organ phenotypic maps can be used to identify cell states that are likely most impacted by human disease. (b) In the future, diseases can be grouped into molecularly defined subclasses based on single-cell genomic (SCG) phenotyping. Many obstacles remain for bringing SCG phenotyping directly to patients in a clinical setting.
Figure 3: Human organoids to recapitulate human phenotypes in vitro. (a) Human organoids derived from induced pluripotent or adult stem cells are genetically tractable and can recapitulate complex tissue level phenotypes in controlled culture environments.

Figure 4: Genetic manipulation toolkit to link phenotype to genotype using stem cells. A diverse array of genetic manipulation tools based on the CRISPR/Cas system can be deployed in organoids to test predictions from single-cell genomic surveys of organs and organoids and link genotype to phenotype. Shown are examples of potential CRISPR/Cas toolkit applications in human organoids.

References:


**ACKNOWLEDGEMENTS**

We thank Treutlein and Camp labs, and B. Roska for helpful discussions. **Funding:** This work was supported by the Chan Zuckerberg Initiative (BT/JGC), European Research Council (ANTHROPOID, JGC; ORGANOMICS; BT), Swiss National Science Foundation (JGC, RJP), ETH domain Personalized Health and Related Technologies (RJP), Brain and Behavior Research Foundation (RJP), National Centre of Competence in Research—Molecular Systems Engineering (RJP, BT), ETH Zürich (RJP, BT), and Institute of Molecular and Clinical Ophthalmology Basel (JGC). **Author Contributions:** All authors contributed to researching and writing the study. **Competing interests:** The authors have no competing interests. **Data and materials availability:** There is no new data generated in this manuscript.