Single Cell Genomics and Disease

Geoff Stanley

BIOC 258, Final Project, Fall 2013

Introduction

Single-cell genomics is the set of techniques that al.lows researchers to characterize the genomes of individual cells. It has only been very recently that the amplification and sequencing techniques have become accurate and cheap enough to analyze the extremely small amount of genetic material contained in one cell. Analysis at the single-cell level is leading to advances in developmental biology, cancer biology, immunology, and neuroscience.

Early Attempts at Single-Cell Genomics

Genetic analysis (excluding gene expression techniques like histochemical staining, RNA fluorescence, and fluorescent-tagged protein transfection, which are outside the purview of this report) at the single cell level was first accomplished at the Arnheim lab in 1988 (Li et al., 1988). They first isolated single diploid cells from individuals and amplified a portion of the beta globin gene, which is mutated in sickle-cell disease. They were able to produce between 5 and 500 fmol of amplified globin gene from the two initial genes at a PCR efficiency of 65%, al.though it is not clear how pure the resulting product was. They al.so amplified sequences from single sperm, and were able to type 70 sperm on two different genes. However, their amplification capabilities were limited to one or two genes.

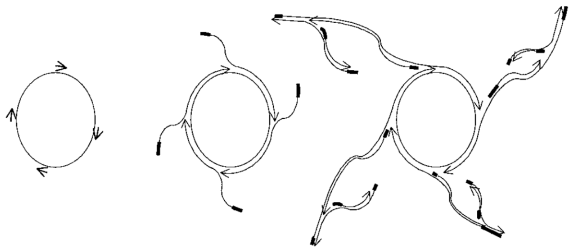
Critical Technology: Whole-Genome Amplification (WGA)

For single-cell genomics to work, researchers must be able to obtain a large and comprehensive amount of genetic material from a single cell. The DNA in a single cell is very

small: a rule of thumb is that a 1 Mbp length of dsDNA weighs 1 pg, so a human sperm cell contains ~3 ng of DNA. The method to do so was first developed in the lab of Norman Arnheim in 1992 (Zhang et al., 1992), which they termed primer-extension preamplication (PEP). The method employed a random mixture of 15-bp primers, representing up to 4^{15} starting sequences (although they did not actually measure how comprehensive their mixture of primers was). This ensured wide coverage of the genome at the cost of an extremely low copy number per gene. They then took 30 al.iquots from a single sperm's PCR product, amplified one of 12 specific sequences, and measured for its presence using a blot. They showed (to a 95% confidence level) that they could produce at least 30 copies of at least 78% of the genome, results that paved the way for future single-cell genomics and human whole-genome analysis. The PEP technique was further improved in 1999 (Dietmaier et al., 1999) by adding additional elongation steps.

A similar technique, called multiple displacement amplification (MDA) was developed in 2001 and used to sequence whole bacterial genomes. Bacterial genomes are much smaller than human genomes, and the human body contains a very large, heterogeneous population of bacteria, so bacterial genomic sequencing could provide significant insights into the relationship between bacterial populations and human health. In developing the MDA technique, researchers were able to show a 10 000-fold amplification of whole bacterial genomes which is currently used in microfluidic devices to amplify the genomes of single bacteria (Kalinsky et al., 2011). MDA is now commercially available (e.g., the REPLI-g Kit from QIAGEN GmbH).

Fig. 1. Amplifying a whole bacteria genome.



Critical Technology: Microfluidics

A problem that single-cell genetics soon faced was the laborious manner in which single cells had to be collected and individually placed into wells, lysed, and analyzed. This was typically done by hand using micropipettes or by laser microdissection of prepared slides (Dietmaier et al., 1999, Schutz et al., 1998). The advent of microfluidics brought a method of doing high-throughput single-cell genomics, critical for increasing the number of genes a researcher can investigate and the accuracy and sensitivity to which those genes can be typed (Whitesides 2006). Microfluidic chips used for sequencing are made from PDMS, a flexible material that can be manufactured using photolithography to form a high density of pressure-actuated small channels and valves whose dimensions are on the order of 10 to 100μm, similar to that of single cells. At these volumes, the Reynolds number (the ratio of inertial to viscous forces) is vanishingly low, meaning that fluid turbulence is nonexistant and cells can be directed to areas of the chip quickly and accurately (Whitesides et al., 2006). One of the earliest such devices used for single-cell studies was devised at Stanford (Wheeler et al., 2003). This device isolated individual cells by flowing them in a stream at a T-junction (Fig. 2). Cells are trapped in the flow stagnation area where the stream splits into two, and the small size of the trapping area means that only one cell gets "sucked in" to it.

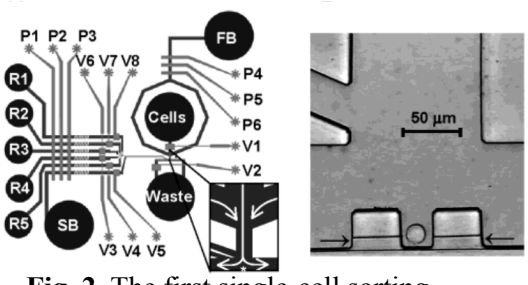


Fig. 2. The first single-cell sorting microfluidic chip. The cell is trapped by flow stagnation, but does not experience high forces, unlike in filter traps.

The first microfluidic PCR reactions were published the same year (Liu et al., 2003). The power of microfluidics for high-throughput sequencing was first demonstrated by this paper, wherein 400 different

PCR reactions were performed using only 41 pipetting steps, a 300 fold improvement over the 1200 pipetting steps that would have been needed using typical PCR techniques. Each reaction chamber had a volume of just 3 nL, meaning that small quantities of reagents could be used for many reactions.

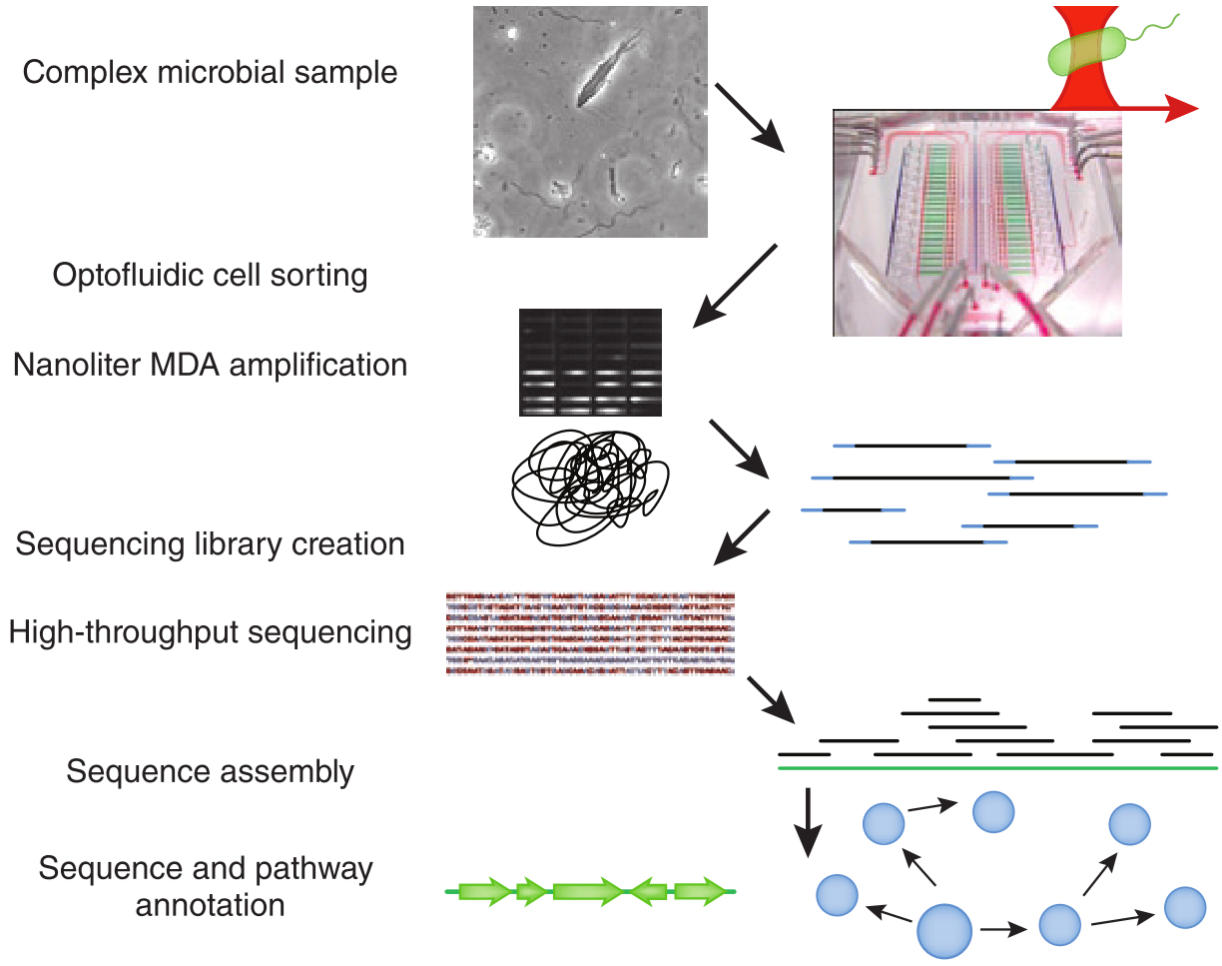


Fig. 3. The process of high-throughput, single cell sequencing, shown here for a diverse microbial colony as might be found in a human. Cell sorting is done via optical trap and are flown into chambers to be lyzed and have their genomes amplified. Reconstruction of the genomes are then carried out using bioinformatic techniques. From Kalisnky et al., 2011.

The two processes of genome amplification and microfluidic cell sorting have been married to produce high-throughput, single-cell genomic assays, with workflows as pictured in Fig. 3.

Lineage Mapping

The power of single-cell sequencing lies in its ability to detect genetic differences between cells in heterogeneous populations. Thus single-cell analysis is readily applicable to the creation of lineage maps of human cells, which use genetic variations among cells to map the binary trees that describe their lineage. As cells divide and differentiate, point mutations accumulate and are passed down. Theoretically, if researchers can determine which mutations were developed at which points al.ong each lineage tree, any given cell can be placed within that tree. Using computation modeling of mouse and human genomes, they surmised that roughly 1.5 million microsatellite mutations should exist in after 40 division, and that they can reconstruct large portions of this tree by sequencing just several hundred microsatellites (Frumkin et al., 2005). This technique was then applied to develop lineage maps of mouse cells lines (Wasserton et al., 2008).

Cancer Cell Genomics

Single-cell genomics is critical to understanding tumor cell development because tumors are characterized by very high genomic instability and mutation rate, and thus significant heterogeneity between cells (Klein et al., 1999). One of the first applications of the above technique of lineage mapping was to understand how tumors develop (Frumkin et al., 2008). The classical theory of tumorogenesis states that tumor lines are monoclonal, starting with a single mutated cell that divides uncontrollably. As the tumor grows, the more aggressive cells are selected for and the tumor eventually becomes metastatic. To assess this model, tumor cells and

adjacent normal cells were laser microdissected from stained tissue and sequenced using whole-genome amplification (Frumkin et al.. 2008). By comparing the lineage maps of the tumor cells to adjacent normal cells, they found that the tumor originated from a single cell. Using average rats of tumor cell proliferation, they calculated that the cell originated ~5 months before the biopsy (although this was not confirmed using other methods, so it is not clear how accurate this is). They al.so were able to type the first of the tumor cells to metastasize, a technique that could lead to better understanding of the genetic mutations that lead metastasis and means of specifically targeting the most dangerous cancer cells.

Single-cell techniques have recently been used to better understand acute myeloid lymphoma (AML), caused aggressive cancer cells with multiple mutations that originate in bone marrow hematopoetic stem cells (HSCs) (Jan M et al. 2012). This system presents a genetic mystery because bone marrow cells are typically short-lived and have a low rate of mutations, so it is unclear how such aggressive tumor cells could arise from them. They isolated HSCs, using FACS, that they determined to have normal function. Sequencing determined that some HSCs carried preleukemic mutations, i.e., they had one or more of the mutations found in leukemic cells, just not enough to affect their phenotype. They created clonal populations from those cells, and did single-cell sequencing on hundreds of cells from each colony. They found specific mutations that led to the development of leukemic cells (Fig. 4), thus demonstrating how leukemic cells are derived from HSCs and indicating potential targets, the preleukemic

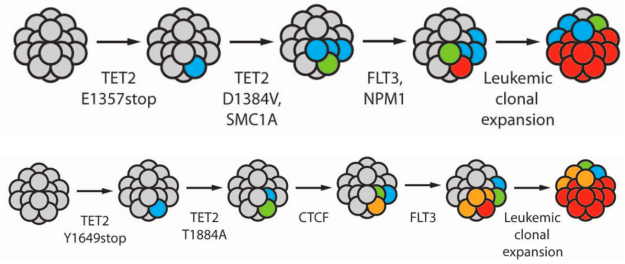


Fig. 4. A schematic of the lineage of AML cells. TET2 is a commonly-seen early mutation, and FLT3 is one of the last to develop.

HSCs, for leukemia detection and treatment. Following this paper, similar characterizations were made for diseases like breast cancer (Powell et al. 2012) and kidney tumors (Xu X et al. 2012). In the latter study, they used exome sequencing of single cells to demonstrate an unexpectedly high degree of genetic diversity among renal cancer cells. They found that, while some characteristic renal cancer mutations were found in many cancer cells, most of the mutations in an individual renal cancer cell were common to only a few other cells, a phenomenon that was undetectable in ensemble assays.

Whole-Genome Haplotyping

Although much sequencing of the human genome can be done on an ensemble level, there is a fundamental loss of information due to the diploid nature of human cells. Each cell contains two copies of every gene, each with potentially different mutations, and the haplotype is the combination of al.leles on each of the two genome copies. This information is potentially critical for disease; for example, a patient with two mutations on a gene could have one mutation on each chromosome, resulting in no functional protein production, or both mutations on only one chromosome, possibly resulting in sufficient production of functional protein (Fan, H.C. et al., 2011). The single-chromosome typing invented by Fan et al. is able to reconstitute this information at the single cell level. To do this, they identified single cells in a microfluidic chip that were in the metaphase and lysed them to release the chromosomes. The individual chromosomes are then flowed past chambers, where they were shown to be randomly deposited. The number of chambers was such that Poisson statistics made it highly unlikely that more than one chromosome was in a chamber (and even more unlikely that two homologous chromosomes shared a chamber, the only condition that would necessarily invalidate the method). Each

chromosome was then amplified using multiple displacement amplification (described earlier). They were able to phase, using direct sequencing, 96% of 1.2 million SNPs available on the Illumina HumanOmni1-Quad chip. This al.lowed them to measure the permutations of recombinations from father and mother that led to the unique child genome, and could potentially use the technique to measure recombination events in the absence of family information.

Human Bacterial Genomics

A typical rule of thumb is that humans carry ~100 trillion bacterial cells, outnumbering human cells 10:1. It is thus widely believed that the makeup of human flora can be both a cause and an indicator of health and disease (Qin et al. 2010). However, ensemble genetic analyses on the human microbiome, which is highly diverse, misses much of the potential information available. Lab techniques that rely on growing colonies of microbes to analyze single species inevitably select for those that grow well under laboratory conditions, and it is estimated that <1% of bacteria and less than half of bacterial phyla have been successfully cultured (Marcy, Y et al. 2007). Using microfluidic isolation of rod-shaped bacteria from the human mouth, and MDA for WGA, they found many species in the TM7 phylum that had not been characterized. Using BLAST mapping, they found that most of the genetic material of their isolated bacteria was unlike that of any other bacterial organism sequenced.

Immunology

Immune cells are a highly diverse group of cells with significant cell-to-cell variation in gene expression. Particularly, B cells must bind to specific antigens, of which there are millions of different types. Recently, Joshua Weinstein and coworkers developed a method to

simultaneously measure the expression of antibodies and mutations in the antibody-producing genes of B cells (Weinstein et al. 2013). They characterized the feedback loops of antibody expression and gene mutation, which controls which antibodies are expressed. Currently, researchers in the Quake laboratory are trying to build microfluidic assays for the detection and characterization of the diversity of antibody heavy chains from large groups of B cells.

Conclusion/Future directions

Significant advances have been made that have al.lowed high-throughput genetic analysis of single cells. This has driven discoveries in cancer cell genomics, human bacterial phyla, and immunology, and has made techniques like whole-genome haplotyping possible. However, there remains much science to be done using single-cell genomics techniques. The diversity of human bacteria needs to be more thoroughly characterized. Changes in the distribution of bacterial species with respect to diet, disease, exertion, stress, etc. can be mapped using single-cell techniques, and microfluidic chips with orders of magnitude higher throughput must be developed to provide even small fractions of coverage of the whole human bacteriome. Characterizations of single cancer cells have only just begun, and future researchers will develop methods for the detection and elimination of metastatic and near-metastatic cells. Patients could even be given cocktails of chemotherapy drugs tailored to the genetic profiles of their cancer. Personally, I will be researching the diversity of neuronal cell populations using single cell techniques in my rotation with the Quake lab quarter, and I am excited to be at the forefront of this burgeoning field.

References

1. Zhang, L. et al.. Whole genome amplification from a single cell: implications for genetic analysis. Proc. Natl. Acad. Sci. 89, 5847–5851 (1992).
2. Dolezel J, Bartos J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. Cytometry A. 2003 Feb 51(2):127-8
3. Jeffreys, A. J., Wilson, V., Neumann, R. & Keyte, J. Hypervariable minisatellltes can be amplified from human DNA by the polymerase chain reaction , using primers from DNA flanking the minlsatelllte. 10953–10971 (1988).
4. Li, H. H. et al.. Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335, 414–7 (1988).
5. Klein, C. a et al.. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. Proc. Natl. Acad. Sci. U. S. A. 96, 4494–9 (1999).
6. Schütze, K. & Lahr, G. Identification of expressed genes by laser-mediated manipulation of single cells. Nat. Biotechnol. 16, 737–42 (1998).
7. Dietmaier, W. et al.. Multiple mutation analyses in single tumor cells with improved whole genome amplification. Am. J. Pathol. 154, 83–95 (1999).
8. Whitesides, G. M. The origins and the future of microfluidics. Nature 442, 368–73 (2006).
9. Wheeler, A. R. et al.. Microfluidic device for single-cell analysis. Anal. Chem. 75, 3581–6 (2003).
10. Liu, J., Hansen, C. & Quake, S. R. Solving the “world-to-chip” interface problem with a microfluidic matrix. Anal. Chem. 75, 4718–23 (2003).
11. Kalisky, T. & Quake, S. R. Single-cell genomics. Nat. Methods 8, 311–4 (2011).
12. Frumkin, D., Wasserstrom, A., Kaplan, S., Feige, U. & Shapiro, E. Genomic variability within an organism exposes its cell lineage tree. PLoS Comput. Biol. 1, e50 (2005).
13. Frumkin, D. et al.. Cell lineage analysis of a mouse tumor. Cancer Res. 68, 5924–31 (2008).
14. Wasserstrom A, Adar R, Shefer G, et al.. Reconstruction of cell lineage trees in mice. PloS ONE (2008) ; 3:e1939.
15. Fan, H. C., Wang, J., Potanina, A. & Quake, S. R. Whole-genome molecular haplotyping of single cells. Nat. Biotechnol. 29, 51–7 (2011).
16. Jan, M. et al.. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Sci. Transl. Med. 4, 149ra118 (2012).
17. Qin, J. et al.. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65 (2010).
18. Powell, A. a et al.. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. PLoS One 7, e33788 (2012).
19. Xu, X. et al.. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148, 886–95 (2012).
20. Marcy, Y. et al.. Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc. Natl. Acad. Sci. U. S. A. 104, 11889–94 (2007).