Single-Cell Research
An Overview of Recent Single-Cell Research Publications Featuring Illumina® Technology
<table>
<thead>
<tr>
<th></th>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Introduction</td>
</tr>
<tr>
<td>7</td>
<td>Applications</td>
</tr>
<tr>
<td></td>
<td>Cancer</td>
</tr>
<tr>
<td></td>
<td>Metagenomics</td>
</tr>
<tr>
<td></td>
<td>Stem Cells</td>
</tr>
<tr>
<td></td>
<td>Developmental Biology</td>
</tr>
<tr>
<td></td>
<td>Immunology</td>
</tr>
<tr>
<td></td>
<td>Neurobiology</td>
</tr>
<tr>
<td></td>
<td>Drug Discovery</td>
</tr>
<tr>
<td></td>
<td>Reproductive Health</td>
</tr>
<tr>
<td></td>
<td>Microbial Ecology and Evolution</td>
</tr>
<tr>
<td></td>
<td>Plant Biology</td>
</tr>
<tr>
<td></td>
<td>Forensics</td>
</tr>
<tr>
<td></td>
<td>Allele-Specific Gene Expression</td>
</tr>
<tr>
<td>50</td>
<td>Sample Preparation</td>
</tr>
<tr>
<td>54</td>
<td>Data Analysis</td>
</tr>
<tr>
<td>60</td>
<td>DNA Methods</td>
</tr>
<tr>
<td></td>
<td>Multiple-Strand Displacement Amplification</td>
</tr>
<tr>
<td></td>
<td>Genome &amp; Transcriptome Sequencing</td>
</tr>
<tr>
<td>68</td>
<td>Epigenomics Methods</td>
</tr>
<tr>
<td></td>
<td>Single-Cell Assay for Transposase-Accessible Chromatin Using Sequencing</td>
</tr>
<tr>
<td></td>
<td>Single-Cell Bisulfite Sequencing/Single-Cell Whole-Genome Bisulfite Sequencing</td>
</tr>
<tr>
<td></td>
<td>Single-Cell Methylome &amp; Transcriptome Sequencing</td>
</tr>
<tr>
<td></td>
<td>Single-Cell Reduced-Representation Bisulfite Sequencing</td>
</tr>
<tr>
<td></td>
<td>Single-Cell Chromatin Immunoprecipitation Sequencing</td>
</tr>
<tr>
<td></td>
<td>Chromatin Conformation Capture Sequencing</td>
</tr>
<tr>
<td></td>
<td>Droplet-Based Chromatin Immunoprecipitation Sequencing</td>
</tr>
<tr>
<td>78</td>
<td>RNA Methods</td>
</tr>
<tr>
<td></td>
<td>Designed Primer-Based RNA Sequencing</td>
</tr>
<tr>
<td></td>
<td>Single-Cell Universal Poly(A)-Independent RNA Sequencing</td>
</tr>
</tbody>
</table>
Quartz-Seq
Smart-Seq
Smart-Seq2
Single-Cell Methylome & Transcriptome Sequencing
Genome & Transcriptome Sequencing
Genomic DNA and mRNA Sequencing
T Cell-Receptor Chain Pairing
Unique Molecular Identifiers
Cell Expression by Linear Amplification Sequencing
Flow Cell-Surface Reverse-Transcription Sequencing
Single-Cell Tagged Reverse-Transcription Sequencing
Fixed and Recovered Intact Single-Cell RNA Sequencing
Cell Labeling via Photobleaching
Indexing Droplets
Drop-Seq
CytoSeq
Single-Cell RNA Barcoding and Sequencing
High-Throughput Single-Cell Labeling
INTRODUCTION

Living tissues are composed of a variety of cell types. Each cell type has a distinct lineage and unique function that contribute to tissue and organ biology and, ultimately, define the biology of the organism as a whole. The lineage and development stage of each cell determine how they respond to other cells and to their native environment. In addition, subpopulations of cells of the same type are often genetically heterogeneous from each other as well as from other cell types. Recently, scientists have launched the ambitious Human Cell Atlas Project, an international collaborative effort to map all cell types in the human body by using single-cell sequencing.2

“Single-cell sequencing has emerged as a revolutionary method that reveals biological processes with unprecedented resolution and scale, and has already greatly impacted biology and medicine.” - Benitez et al. 2017

Much of the initial impetus for single-cell tissue sequencing has come from cancer research, where cell lineage and detection of residual disease is of paramount importance. Currently, single-cell approaches are also used to improve our understanding of other complex biological systems, including the central nervous system (CNS), immune system, and mammalian development.4

Single-cell sequencing is also an effective approach to characterize organisms that are difficult to culture in vitro. Advances in single-cell sequencing have improved the detection and analysis of infectious diseases, food-borne pathogens, and microbial diversities in the environment or the gut.6,7,8,9

The high accuracy and specificity of next-generation sequencing (NGS) makes it ideal for single-cell and low-level DNA/RNA sequencing. The growing collection of published single-cell techniques includes detection of DNA mutations, copy-number variants (CNVs), DNA-protein binding, RNA splicing, and the measurement of mRNA expression.10 More recently, microfluidics platforms and droplet-based methods have enabled massively parallel sequencing of mRNA in large numbers of individual cells.11,12 The function of an individual cell is largely governed by interactions with its neighbors. This spatial context is typically lost in single-cell sequencing experiments, but new methods13,14 and analysis algorithms15 are combining measurements of single-cell gene expression with spatial localization within tissues.

This review highlights recent publications demonstrating how Illumina technology is being used in single-cell sequencing applications and techniques. To learn more about Illumina sequencing and microarray technologies, visit www.illumina.com.

The same gene can be expressed at different levels, and influenced by different control mechanisms, in different cell types within the same tissue.

Reviews

Baslan T and Hicks J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. Nat Rev Cancer. 2017;17:557-569


Tanay A and Regev A. Scaling single-cell genomics from phenomenology to mechanism. Nature. 2017;541:331-338


Wang Y and Navin NE. Advances and applications of single-cell sequencing technologies. Mol Cell. 2015;58:598-609
Applications

Cancer

Tumor progression occurs through driver mutations that undergo Darwinian selection for successive clonal expansion of tumor subclones. As a result, advanced tumors may contain a number of unique subclones with different sets of mutations, different histopathology, and different responses to therapy. Molecular profiling of all subclones at diagnosis is important, because a subclone that makes up only 0.3% of a primary tumor can become the predominant clone following relapse.

Deep sequencing can detect subclone abundance as low as 1% of the total tumor cell population, but single-cell sequencing approaches are required to fully characterize therapeutic efficacy on rare cell populations.

“Understanding single cancer cells at their individual level and as an ensemble is bound to advance our understanding of not only therapeutic resistance but all facets of tumor biology.” - Baslan and Hicks 2017

Circulating tumor cells (CTC) can also be used to detect cancer. Cell-free detection of cancer nucleic acid markers—so-called liquid biopsies—may prove more sensitive and reproducible. Single-cell approaches for the molecular profiling of cancer stem cells (CSCs) and disseminated cancer cells also add to our understanding of tumor development, metastasis, and therapeutic response.

Recent clinical data have demonstrated that therapeutic enhancement of immune system function can improve cancer outcomes. Antibodies that block cytotoxic T-lymphocyte–associated protein 4 (CTLA-4) as well as programmed death 1 (PD-1) induce clinical responses in a number of cancers, including melanoma, lung cancer, renal cancer, bladder cancer, and Hodgkin’s lymphoma. Single-cell sequencing approaches offer the possibility for a deeper understanding of the complex interactions among immune cells and tumor cells, as well as a more thorough characterization of the cellular ecosystem of tumors.


Intratumor heterogeneity: The progressive accumulation of somatic mutations results in a heterogeneous polyclonal tumor, in which different clones may respond differently to treatment.
Reviews


References

Single-cell approaches have demonstrated that tumors are genetically heterogeneous. In this study, the authors used the HiSeq™ 2500 System to perform RNA-Seq of 515 individual cells isolated from breast cancer tumors. Single-cell transcriptomic analyses identified heterogeneous as well as conserved gene expression signatures for subtype-specific breast cancer cells. In addition to breast cancer cells, the authors identified gene expression signatures for T cells, B cells, and macrophages. Individual T cells could be classified as activated or exhausted, suggesting that immune cell interactions within tumors are dynamic and distinct. These data suggest that individual breast cancer tumor microenvironments contain cells with unique patterns of gene expression, and that tumor subtypes are shaped by tumor cells as well as immune cells within the tumor microenvironment.

Illumina Technology: Nextera™ XT DNA Sample Prep Kit, TruSeq™ Rapid SBS Kit, HiSeq 2500 System


Cancer stem cells are rare cell types present in some tumors that possess the characteristics of normal stem cells and can give rise to tumors. In addition, cancer stem cells can be resistant to chemotherapy and can persist during remission. Because cancer stem cells are vastly outnumbered by tissue cells and are difficult to isolate, genetically characterizing them has been challenging. In this study, the authors used single-cell transcriptomic analyses to characterize cancer stem cells in chronic myeloid leukemia (CML). CML is typically characterized by the presence of the BCR-ABL fusion gene and is less genetically complex than other cancers. The authors developed a single-cell protocol for improved detection of BCR-ABL combined with RNA sequencing on the HiSeq 2000/4000 systems. They subsequently performed RNA-Seq on over 2000 individual stem cells from CML samples. Their data revealed the heterogeneity of CML cancer stem cells and identified gene expression signatures that correlated with poor treatment response in CML. Finally, their data suggest a gene expression signature for cancer stem cells, already present at CML diagnosis, that confers resistance to chemotherapy.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2000/4000 System


Checkpoint blockade cancer immunotherapies act by altering the function of host T cells, and they have made a dramatic impact on the treatment of advanced non-small cell lung cancer (NSCLC). Tumor-infiltrating myeloid cells are host immune cells that present tumor-associated antigens to host T cells, and they contribute to tumor progression and treatment response. However, a detailed understanding of how the immune landscape develops in response to tumor cues is currently lacking. To address this knowledge gap, the authors performed single-cell analysis of immune cells from patient NSCLC tumors, noncancerous lung tissues, and blood. They found a higher number of immune cells in tumor than in nontumor lung tissue. Further, T and B cells were enriched in early-stage lung tumors, whereas natural killer (NK) cells were significantly reduced. In addition, the authors used single-cell RNA sequencing to characterize the genetic diversity of tumor-infiltrating monocytes from early-stage NSCLC and identified gene expression signatures for these monocyte cell types across NSCLC stages.

Illumina Technology: Unidentified Illumina System

The centrosome is the main microtubule-organizing center and regulator of cell-cycle progression in animal cells. In a normal cell, a centrosome is copied only once per cell cycle, with one copy passed on to the daughter cells during mitosis. In cancer cells, however, extra centrosomes are common; they correlate with poor prognosis and abnormal chromosome number, or aneuploidy. It is currently unclear whether extra centrosomes cause cancer or simply arise due to tumorigenesis. To address this question, the authors created transgenic mice overexpressing Plk4, a gene that regulates centrosome duplication. These transgenic mice have a persistent increase in centrosome number in cells from numerous tissues. The authors then karyotyped 99 individual epidermal cells by single-cell genomic copy-number analysis using the HiSeq 2500 System. Their data demonstrated that centrosome amplification was sufficient to promote chromosome segregation errors and aneuploidy in transgenic mouse cells. Further, starting at just 36 weeks, the Plk4 transgenic mice developed spontaneous aneuploid tumors. Taken together, the data suggest that centrosome amplification is sufficient to drive aneuploidy and cancer formation in mice.

Illumina Technology: TruSeq DNA Sample Preparation Kit v2, HiSeq 2500 System


Single-cell sequencing approaches can be useful for interrogating the genetic diversity of tumor cells, as well as characterizing the tumor microenvironment. However, many bulk datasets from large cohorts already exist, such as The Cancer Genome Atlas (TCGA). In this study, the authors combined single-cell RNA-Seq of 14,266 individual cells from 16 isocitrate dehydrogenase (IDH)-mutant gliomas, with 165 TCGA bulk RNA profiles. The TCGA RNA data demonstrated differences in gene expression in the 2 different subtypes of IDH-mutant gliomas, astrocytoma and oligodendroglioma. Single-cell RNA-Seq demonstrated that differences in gene expression were due to signature tumor events and tumor microenvironment composition. Notably, glial cell lineages were similar in both glioma subtypes. The study demonstrates the power of combining single-cell and bulk RNA-Seq approaches, and the data shed new light on the genetic composition of IDH-mutant glioma.

Illumina Technology: Nextera XT Library Prep Kit, NextSeq™ 500 System


Hepatocellular carcinoma (HCC) is a leading cause of cancer deaths worldwide, with the highest incidence in Asia and Africa. Despite the promise for other tumor types, cancer immunotherapies have not been successful in treating HCC, even though HCC tumors contain significant levels of tumor-infiltrating monocytes. In this study, the authors use single-cell RNA-Seq to understand and characterize tumor-infiltrating monocytes in HCC. Specifically, they isolated more than 5000 individual T cells from peripheral blood, tumor, and adjacent normal tissues from 6 HCC samples. They performed deep RNA-Seq on single T cells as well as T-cell receptor (TCR) sequencing using HiSeq 2500/4000 Systems. The combination of single-cell transcriptional profiles and TCR sequences identified 11 different T-cell subtypes. Notably, the authors found that exhausted CD8+ T cells and Treg were enriched and clonally expanded in HCC. This study demonstrates the utility of single-cell sequencing approaches in characterizing the tumor microenvironment and characteristics of tumor-infiltrating cells in HCC.

Illumina Technology: HiSeq 2500/4000 Systems

Single-cell sequencing has the potential to inform treatment response and drug resistance by assessing malignant, microenvironmental, and immunologic states within tumors. In this study, the authors applied scRNA-Seq to 4645 single cells (malignant, stromal, immune, and endothelial) isolated from 19 patients with metastatic melanoma. They found that malignant cells within the same tumor displayed transcriptional heterogeneity associated with cell cycle, spatial context, and drug resistance. The same tumor had cells with high expression levels of microphthalmia-associated transcription factor (MITF), as well as cells with low MITF levels and elevated levels of AXL kinase (cells prone to early drug resistance). Infiltrating T-cell analysis revealed exhaustion programs, connection to T-cell activation/expansion, and patient variability. This study demonstrates how single-cell genomics can unravel the cellular ecosystem of tumors, with implications for targeted and immune therapies.

**Illumina Technology: Nextera XT Sample Preparation Kit, NextSeq 500 System**


Glioblastoma is one of the deadliest forms of cancer. Glioblastoma tumors have mutations in a number of druggable pathways, but current targeted therapies have proven ineffective due to rapid and universal drug resistance. Specifically, the mechanistic target of rapamycin (mTOR) pathway is a key driver in 90% of glioblastomas, yet tumor cells develop rapid resistance to mTOR-targeted therapies. In this study, the authors used the NextSeq 500 system to obtain single-cell genomic data, which they correlated with single-cell proteomic data in tumor cells treated with mTOR inhibitor. Their data showed that resistance to mTOR inhibitor in glioblastoma tumor cells occurred within days of drug therapy. Surprisingly, the correlation with single-cell sequencing data demonstrated that this drug resistance proceeds via nongenetic mechanisms, through upregulation of specific signaling phosphoproteins. This study suggests a novel approach for designing drug combination therapy in treating glioblastoma.

**Illumina Technology: NextSeq 500 System**


Metagenomics

Although microorganisms represent the most abundant and diverse life forms on earth, in some environments only 0.1%–1% of the members have been cultivated. Single-cell genomics approaches have led to a renewed understanding of microbial ecology, and they have greatly expanded our view of the tree of life. This understanding has also revealed the extent and importance of viruses in the environment and their role in shaping bacterial populations. When combined with function- or phenotype-based screens, single-cell sequencing can enable the identification of microbial genomes that encode the function or phenotype. Surprisingly, bacterial cell colonies—the paragon of microbial homogeneity—also display complex collective dynamics that allow for adaptation to their local environment.

“Single-cell genome sequencing of individual archaeal and bacterial cells is a vital approach to decipher the genetic makeup of uncultured microorganisms.”

- Bowers et al. 2017

Reviews


Single-cell genomics has led to a number of individual draft genomes for uncultivated microbes; however, multiple-strand displacement amplification (MDA) artifacts during the amplification step lead to incomplete and uneven coverage. Metagenomic data sets do not suffer the same sequence bias, but the genomic complexity of microbial communities precludes the recovery of draft genomes. In this study, the authors developed a new method for generating population genome assemblies from metagenomic-guided, single-cell amplified genome assembly data. They validated the approach by completing single-cell amplified genomes for Marine Group 1 Thaumarchaeota and SAR324 clade bacterioplankton. The improved method assembly of the SAR324 clade genome revealed the presence of many genes not present in the single-cell amplified genome.

Illumina Technology: MiSeq™ and HiSeq 2000 System


In microbial ecology studies, 16S rRNA sequencing can identify microbial community members, whereas shotgun metagenomics can determine the functional diversity of the community. However, combining the 2 approaches is technically challenging. In this study, the authors developed emulsion, paired isolation, and concatenation PCR (epicPCR), a technique that links functional genes and phylogenetic markers. They applied the technique to millions of uncultured individual cells from the freshwater Upper Mystic Lake in Massachusetts. Specifically, they profiled the sulfate-reducing community within the freshwater lake community and were able to identify new putative sulfate reducers. The method is suitable for identifying functional community members, tracing gene transfer, and mapping ecological interactions in microbial cells.

Illumina Technology: MiSeq System

References


Prochlorococcus is a photosynthetic cyanobacterium that is highly abundant in the ocean surface and contributes to marine productivity. Previous studies have shown that Prochlorococcus populations from different sites in the Atlantic Ocean can be highly diverse genomically, suggesting that ocean habitats can affect genomic structure and diversity. In this study, the authors sampled wild populations of cells from habitats in the Pacific and Atlantic Oceans. They used the Illumina GAIIx System to perform single-cell sequencing of Prochlorococcus from the different sites. They found that Prochlorococcus from the 2 different ocean sites were composed of nonoverlapping and distinct subpopulations with different genomic backbones. These data suggest that environmental selection pressures, as well as geographic separation, can shape Prochlorococcus populations.

Illumina Technology: GAIIx System


Marine sediments are the largest carbon sink on the planet, with half of chemosynthetic oceanic carbon fixation occurring in coastal sediments. However, the microbes responsible for this activity are unknown. By surveying bacterial 16S rDNA gene diversity from 13 coastal sediments across Europe and Australia, the authors identified groups of Gammaproteobacteria that were affiliated with sulfur-oxidizing bacteria. 13C-carbon assimilation studies showed that these uncultured Gammaproteobacteria accounted for 80% of carbon fixation in coastal sediments. Finally, the authors isolated individual cells from the environmental sample and performed single-cell whole-genome sequencing (WGS) to identify genes that linked hydrogen-oxidizing activity with sulfur-oxidizing Gammaproteobacteria.

Illumina Technology: MiSeq™ and HiSeq 2000 System


Molecular environmental surveys using 16S rRNA sequencing have greatly expanded our knowledge of microbial phylogenetic diversity. However, some bacterial and archaeal clades can be systematically underrepresented in current surveys or missed altogether. In this study, the authors analyzed 5.2 Tb of metagenomic data and discovered a novel bacterium (Candidatus Kryptonia) found exclusively in geothermal springs. The lineage had been missed in classical metagenomic surveys, because of mismatches in commonly used 16S rRNA primers. The authors combined metagenomic data with single-cell sequencing to generate high-quality genomes that represented 4 unique genera within this phylum.

Illumina Technology: MiSeq System

Mende DR, Aylward FO, Eppley JM, Nielsen TN and DeLong EF. Improved Environmental Genomes via Integration of Metagenomic and Single-Cell Assemblies. *Front Microbiol.* 2016;7:143

Single-cell genomics has led to a number of individual draft genomes for uncultivated microbes; however, multiple-strand displacement amplification (MDA) artifacts during the amplification step lead to incomplete and uneven coverage. Metagenomic data sets do not suffer the same sequence bias, but the genomic complexity of microbial communities precludes the recovery of draft genomes. In this study, the authors developed a new method for generating population genome assemblies from metagenomic-guided, single-cell amplified genome assembly data. They validated the approach by completing single-cell amplified genomes for Marine Group 1 Thaumarchaeota and SAR324 clade bacterioplankton. The improved method assembly of the SAR324 clade genome revealed the presence of many genes not present in the single-cell amplified genome.

Illumina Technology: TruSeq LT Nano Kit, MiSeq System


In microbial ecology studies, 16S rRNA sequencing can identify microbial community members, whereas shotgun metagenomics can determine the functional diversity of the community. However, combining the 2 approaches is technically challenging. In this study, the authors developed emulsion, paired isolation, and concatenation PCR (epicPCR), a technique that links functional genes and phylogenetic markers. They applied the technique to millions of uncultured individual cells from the freshwater Upper Mystic Lake in Massachusetts. Specifically, they profiled the sulfate-reducing community within the freshwater lake community and were able to identify new putative sulfate reducers. The method is suitable for identifying functional community members, tracing gene transfer, and mapping ecological interactions in microbial cells.

Illumina Technology: MiSeq System

Oceanic plankton is the world's largest ecosystem and is composed of viruses, prokaryotes, microbial eukaryotes, phytoplankton, and zooplankton. This ecosystem structure and composition are influenced by environmental conditions and nutrient availability. In this study, the authors analyzed 313 plankton samples from the *Tara* Oceans expedition and obtained viral, eukaryotic, and prokaryotic abundance profiles from Illumina-sequenced metagenomes and 18S rDNA V9 sequences. They used network inference and machine-learning methods to construct an interactome among plankton groups. In particular, the authors confirmed predicted virus-host interactions by comparing putative host contigs with viral data from single-cell genomes.

**Illumina Technology:** Illumina-sequenced metagenomes (+tags) and 18S rDNA V9 sequences

---


Mende DR, Aylward FO, Eppley JM, Nielsen TN and DeLong EF. Improved Environmental Genomes via Integration of Metagenomic and Single-Cell Assemblies. *Front Microbiol.* 2016;7:143
Stem Cells

Human life begins from a single oocyte, which undergoes mitotic divisions to generate a population of cells that make up the human embryo. Embryonic stem cells (ESCs) are pluripotent stem cells derived from the inner cell mass of a blastocyst, an early-stage preimplantation embryo. Each stem cell undergoes a series of cell divisions that results in a specific lineage, which determines its genetic code and response to local environmental factors. This process gives rise to an array of unique, genetically heterogeneous cells. Sequencing these single stem cells during differentiation has helped elucidate the underlying mechanisms.

“Single-cell sequencing provides powerful tools for characterizing the omic-scale features of heterogeneous cell populations including, those of stem cells.” - Wen and Tang 2016

Hematopoietic stem cells (HSCs), and the mechanisms regulating their differentiation into erythroid, myeloid, or lymphoid lineages, are a unique example of cellular development. Single-cell sequencing has helped to identify a population of neural stem cells that become activated in response to brain injury, suggesting a possible approach for treating traumatic brain injury. Induced pluripotent stem cells (iPSCs), a type of pluripotent stem cell that can be generated directly from adult cells, also have potential use in cell-replacement therapies. Single-cell sequencing has helped to characterize the genetic heterogeneity of individual iPSCs, mechanisms regulating their differentiation and pluripotency, and self-organization into organoid structures.


Embryonic stem cells.
An overview of recent publications featuring Illumina technology

Reviews


References


HSCs are pluripotent stem cells found in the bone marrow, and they give rise to all myeloid and lymphoid cell lineages in the blood. In mice, a small subset of HSCs are dormant, dividing only 5 times per lifetime and maintaining a high long-term reconstitution potential. Dormant HSCs can transition into activated HSCs, which are still quiescent but primed for cell-cycle entry. The mechanisms regulating this transition are not well understood. In this study, the authors used single-cell RNA-Seq to compare gene expression in individual dormant HSCs, activated HSCs, and multipotent progenitor cells. They isolated HSCs using fluorescence-activated cell sorting (FACS) with cells from transgenic mice overexpressing an HSC-specific green fluorescent protein (GFP) marker. The single-cell transcriptome data revealed that dormant HSCs were characterized by low expression levels of the Myc gene and high expression levels of retinoic acid genes. Further, dormant HSCs expressed low levels of genes required for essential biosynthetic pathways, and transition to activated HSCs was characterized by a continuous increase in biosynthetic gene expression. Overall, their data shed new light on the biological pathways that regulate activation of HSCs and suggest that retinoic acid-vitamin A may regulate HSC plasticity in mice.

Illumina Technology: Nextera XT Sample Preparation Kit, HiSeq 2000/2500 Systems


Bulk RNA-Seq approaches average gene expression data from individual cells together, destroying crucial information on cell-state transition. Consequently, bulk RNA-Seq is not suitable for characterizing cellular states or their role in development. Single-cell RNA-Seq approaches allow for the measurement of gene expression from large numbers of individual cells, allowing the cataloging and definition of cellular states. In the case of hematopoiesis, current models suggest that HSCs differentiate stepwise into discrete multipotent progenitor cell lineages. In the current study, the authors combined single-cell RNA-Seq with functional analyses to challenge this model. They used the HiSeq 2500 System to perform single-cell sequencing of 1,413 single HSCs isolated from human bone marrow. They also individually cultured ex vivo 2038 single HSCs to characterize lineage potential. These 2 data sets were then integrated to quantitatively link transcriptomic and functional single-cell data. Their data suggest that individual HSCs gradually acquire lineage biases across multiple directions, without transitioning into discrete, hierarchically organized progenitor cells.

Illumina Technology: HiSeq 2500 System


Bone marrow stem cell transplants are commonly used to treat conditions such as leukemia and lymphoma. Following transplantation, mesenchymal stem cells can fuse with parenchymal cells of the brain, liver, small intestine, and heart. The resultant effects of this hybridization are not known. In this study, the authors used the MiSeq system to perform scRNA-Seq of individual mesenchymal stem cell–cardiomyocyte hybrids to characterize global gene expression. The expression of cell-cycle genes was generally decreased in hybrids. However, for most other gene groups, individual hybrids were genetically distinct. Moreover, 2 hybrids were genetically similar to breast cancer cells, suggesting that monitoring stem cell transplantation for tumor emergence is warranted.

Illumina Technology: Nextera XT Sample Preparation Kit, MiSeq System


iPSCs offer the potential for autologous cell-replacement therapies. Maintaining the genetic integrity of the cultured iPSCs is an important consideration in potential therapy. The authors characterized the frequency of somatic mitochondrial DNA (mtDNA) mutations in cells derived from young or elderly subjects. The data show that the mtDNA mutation frequency in iPSCs increases with the age of the individual. The results highlight the importance of genetically monitoring mtDNA mutations in iPSCs, especially those that are generated from older patients.

**Illumina Technology: Nextera XT Sample Preparation Kit, MiSeq System**


The scarcity of tissue-specific stem cells, and the complexity of their surrounding environment, make single-cell sequencing methods imperative for characterizing these cell types. In this study, the authors used single-cell sequencing and weighted gene coexpression network analysis to identify CD133+ ependymal cells from the adult mouse forebrain neurogenic zone as NSCs. These subpopulations of cells were enriched for immune-responsive genes, as well as genes encoding angiogenic factors. Administration of vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (bFGF) enhanced migration and elicited differentiation into neurons and glia.

**Illumina Technology: HiSeq 2500 System**


Hematopoiesis is the developmental process by which bone marrow-derived HSCs differentiate into erythroid, myeloid, or lymphoid lineages. In this study, the authors combined massively parallel scRNA-Seq with fluorescence-activated cell sorting (FACS), chromatin profiling, genetic perturbation, and computational modeling to characterize the transcriptome of myeloid progenitor populations. The single-cell transcription data grouped HSCs into multiple progenitor subgroups with 7 differentiation states. These data provide a new reference model for studying hematopoiesis at the single-cell level.

**Illumina Technology: NextSeq 500 and HiSeq 1500 System**


“Assaying the genomes of multiple cells during development reveals developmental lineage trees whereas cellular transcriptomes inform us about the regulatory state of cells and their gradual restriction in potency.” - Marioni and Arendt 2017

The study of human preimplantation development has been based typically on small numbers of samples, often pooled.63 These studies fail to capture a detailed view of the first days of human preimplantation development. Recent single-cell sequencing studies are beginning to capture a more comprehensive and detailed view.64

Single-cell approaches are also providing a window into embryonic development in other model organisms (such as mouse,65 zebrafish,66 and Caenorhabditis elegans67), and they are leading to a greater understanding of the regulatory mechanisms underlying embryonic development in humans.68,69 Comparative single-cell studies are even beginning to reveal the evolutionary history of germ layers—a fundamental concept in developmental biology for the past 150 years.70

During human embryogenesis, the single-cell zygote divides several times to form a morula.
Reviews


References


Somatic mutations are permanent changes in the DNA sequence of a genome. They arise because of mistakes in DNA replication in cells or from environmental factors. Previous studies using single-cell sequencing approaches have demonstrated that individual neurons contain unique single-nucleotide variant (SNV) somatic mutations, a phenomenon referred to as genomic mosaicism. Although genomic mosaicism is common among multiple cell types, neurons are postmitotic, and thus the SNVs have the potential to persist and alter neuronal function. In this study, the authors isolated individual neurons from the forebrains of embryonic samples and grew them into small clones in culture. They used the HiSeq platform to analyze the genomes of these samples, and they identified 200–400 SNVs per cell. The SNV patterns in these embryonic neurons resembled those found in cancer cell genomes. By analyzing the genomes of neurons at different stages of embryonic development, they determined that the mutation rate speeds up during neurogenesis. These data suggest that embryonic neurons are protected from mutations, compared to neurons at later stages of neurodevelopment.

Illumina Technology: HiSeq 2500 and HiSeq X Systems


Single-cell sequencing approaches have been instrumental in highlighting the genetic and cellular diversity of biological tissues. In this study, the authors extended this concept in a tour de force single-cell sequencing study of an entire organism. They used the NextSeq 500 System to perform single-cell RNA-Seq of all 50,000 cells of the Caenorhabditis elegans L2 larval stage. To accomplish this, they developed a combinatorial indexing method to uniquely label the transcriptomes of large numbers of single cells, a method termed sci-RNA-seq. Their single-cell transcriptomic data indicated that the C. elegans L2 larva is composed of 27 different cell types, including some that are extremely rare within the organism. Further, they integrated the sci-RNA-seq data with chromatin immunoprecipitation sequencing (ChIP-Seq) data to determine the effects of transcription factors on specific cell types. Their work demonstrates the ability of single-cell sequencing approaches to reconstruct the cell complement for an entire organism, and it provides a valuable resource for nematode biologists.

Illumina Technology: Nextera XT Sample Preparation Kit, NextSeq 500 System


For Research Use Only. Not for use in diagnostic procedures.

Single-cell Research

The liver is an organ that is critically important for detoxifying metabolites, producing carrier proteins, and producing biochemicals necessary for digestion. Anatomically, it is composed of hepatic lobules, hexagonal-shaped structures vital for carrying blood out from the liver. The hepatic lobules are composed of 15 radial layers of hepatocytes, the primary cell type within the liver, and each radial layer specializes in a distinct process. Given the highly specialized microenvironment within each hepatic lobule, bulk sequencing approaches lack the resolution to characterize zonal gene expression within each radial layer. In this study, the authors combined single-cell RNA sequencing of dissociated individual hepatocytes with single-molecule fluorescence in situ hybridization (smFISH) of liver tissue sections. Single-cell RNA-Seq enabled measurement of gene expression patterns of thousands of cells, while smFISH enabled measurement of in situ expression levels for numerous landmark genes. The combined data revealed the zonal expression of all expressed genes within the hepatic lobule, with 50% of all liver genes expressed zonally. This approach is suitable for reconstructing spatial gene expression organization of other organs.

Illumina Technology: HiSeq 2500 System


Human preimplantation embryos often display chromosome mosaicism, commonly euploid-aneuploid mosaicism with complements of both normal and abnormal cells. This mosaicism occurs early in development, within the first few cell divisions following fertilization. Although preimplantation mosaicism is common, and results in high rates of early human pregnancy failures, the fate of aneuploid cells in the embryo is still unclear. In this study, the authors developed a mouse model for preimplantation chromosome mosaicism by treating developing mouse embryos with reversine during the 4- to 8-cell stage. The developing mosaic embryos were then characterized by using a combination of live-cell imaging and single-cell sequencing. The data show that aneuploid cells were eliminated from the embryo through apoptosis, starting just before implantation. Mosaic euploid-aneuploid embryos had comparable developmental potential to normal embryos, as long as they contained a sufficient proportion of euploid cells.

Illumina Technology: Nextera XT Sample Preparation Kit, HiSeq 2000/2500 Systems


Plant stomata facilitate plant gas exchange with the atmosphere. In Arabidopsis, the production and pattern of stomata proceeds from a discrete lineage that can be parsed into intermediate steps. Despite the biological significance of ribonuclease L (RNase L), the RNAs cleaved by this enzyme are poorly defined. In this study, the authors used Illumina sequencing to reveal the frequency and location of RNase L cleavage sites within host and viral RNAs. The method was optimized and validated using viral RNAs cleaved with RNase L and RNase A, and RNA from infected and noninfected HeLa cells. The authors identified discrete genomic regions susceptible to RNase L and other single-strand–specific endoribonucleases. Monitoring the frequency and location of these cleavage sites within host and viral RNAs may reveal how these enzymes contribute to health and disease.

Illumina Technology: TruSeq SBS Kit v3–HS, HiSeq 2000 System


Cochlear and vestibular sensory epithelia in the inner ear use similar cell types to transduce 2 types of stimuli: sound and acceleration. However, each individual sensory epithelium is composed of anatomically and physiologically heterogeneous cell types, which have eluded transcriptional characterization due to the limited numbers of each cell type. In this study, the authors performed RNA-Seq on 301 individual cells from the utricular and cochlear sensory epithelia of newborn mice. Cluster analysis determined distinct transcriptional profiles for each cell type. Comparison of expression data from cell types within utricles and cochleae demonstrated divergence between auditory and vestibular cells.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, Nextera XT DNA Sample Preparation Index Kit, HiSeq 1000 System

Germ layers give rise to all of the tissues and organs in an animal, and serve as an organizing principle of developmental biology. The mesoderm is present in complex bilaterian animals but not in phyla Cnidaria and Ctenophora (comb jellies), suggesting that the mesoderm was the final germ layer to evolve. The authors used the HiSeq 2000 system to analyze the transcriptome of individual *C. elegans* blastomeres (AB, MS, E, C, and P3) that collectively account for the entire embryo. They also generated a whole-embryo time course using cell expression by linear amplification sequencing (CEL-Seq), spanning the single-cell stage to the free-living larva, at 10-minute resolution. They found that the gene expression program of *C. elegans* mesoderm was induced after those of the ectoderm and endoderm. Further, the endoderm expression program activated earlier than the endoderm program. This result was also observed for expression of endoderm orthologs in frog (*Xenopus tropicalis*), sea anemone (*Nematostella vactensis*), and sponge (*Amphimedon queenslandica*). Taken together, these observations suggest that the endoderm program dates back to the origin of multicellularity, whereas the ectoderm originated as a secondary germ layer.

**Illumina Technology: HiSeq 2000 System**


Neurons are postmitotic cells, so their genomes are particularly susceptible to DNA damage. In this study, the authors surveyed the landscape of somatic single-nucleotide variants (SNVs) in the human brain by performing single-cell WGS of 36 individual cortical neurons. The most abundant SNVs included noncoding, noncoding RNA, intronic, and intergenic SNVs. Coding, truncating, splice, and silent SNVs were much less abundant. Moreover, the data showed that each cortical neuron had a distinctive genome that harbored up to 1580 somatic SNVs. Finally, the somatic SNVs created nested linkage trees, demonstrating that somatic mutations could be used to reconstruct the developmental lineage of neurons.

**Illumina Technology: TruSeq Nano LT Sample Preparation Kit, MiSeq, HiSeq 2000 and HiSeq X Ten Systems**


scRNA-Seq is an established method for discovering novel cell types, understanding regulatory networks, and reconstructing developmental processes. However, scRNA-Seq typically involves dissociating cells from tissues and thus disrupting their native spatial context. To capture spatial context in scRNA-Seq data, the authors developed Seurat, a computational strategy that combines scRNA-Seq with complementary in situ hybridization data for a smaller set of “landmark” genes that guides spatial assignment. They validated Seurat by spatially mapping 851 individual cells from dissociated zebrafish embryos and creating a transcriptome-wide map of spatial patterning. Seurat was able to localize rare subpopulations of cells correctly, and it could map spatially restricted cells as well as those with a more scattered pattern of expression.

**Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2500 System**
An overview of recent publications featuring Illumina technology


Human preimplantation development requires embryonic genome activation and maternal transcript degradation during the first 3 days after fertilization. Studies of the process at this stage are challenging technically, given the sparse samples and lack of suitable methods. To overcome these hurdles, the authors performed scRNA-Seq of 348 oocytes, zygotes, and individual blastomeres from 2- to 3-day old human embryos. They showed that 32 and 129 genes are transcribed during the transition from oocyte to 4-cell stage and from 4- to 8-cell stage, respectively. Several of the transcribed genes were unannotated PRD-like homeobox genes, including ARGFX, CPHX1, CPHX2, DPRX, DUXA, DUXB, and LEUTX.

Illumina Technology: GAIIx, HiSeq 1000, HiSeq 2000 and MiSeq System


Cereal endosperm is a main source of food, feed, and raw material worldwide, yet genetic control of endosperm cell differentiation is not well defined. In this study, the authors coupled laser-capture microdissection (LCM) and Illumina sequencing to profile mRNAs in 5 major cell types of differentiating endosperms and 4 compartments of maize (Zea mays) kernels. They identified mRNAs that specifically accumulate in each compartment, as well as genes predominantly expressed in 1 or multiple compartments. Their results demonstrate that the MRP-1 transcription factor can activate gene regulatory networks within the basal endosperm transfer layer. These data provide a high-resolution gene activity atlas of the compartments of the maize kernel. The study also uncovers the regulatory modules associated with differentiation of the major endosperm cell types.

Illumina Technology: TruSeq DNA Sample Preparation v2 Kit, HiSeq 2000 System


Wurtzel O, Oderberg IM and Reddien PW. Planarian Epidermal Stem Cells Respond to Positional Cues to Promote Cell-Type Diversity. Dev Cell. 2017;40:491-504 e495

Chan SS, Chan HH and Kyba M. Heterogeneity of Mesp1+ mesoderm revealed by single-cell RNA-seq. Biochem Biophys Res Commun. 2016;474:469-475


Immunology

The immune system consists of a number of specialized cell types that play unique roles in the host immune response. In the adaptive immune system, T and B lymphocytes (T and B cells) express specific surface receptors (T-cell receptors [TCRs] and B-cell receptors [BCRs]) that recognize and engage specific antigens presented on the surface of antigen-presenting cells via the major histocompatibility complex (MHC). Individual immune cell types are typically isolated by FACS, based on specific surface molecular markers.73 Due to the technical limitations of FACS, FACS-isolated cells can still consist of mixed populations at various stages of development or activation, and FACS approaches are limited to available markers.74 In contrast, single-cell sequencing approaches are not limited by specific molecular markers, and they can identify unique gene expression patterns and splice variants in T cells and B cells.75, 76

“Advances in computer science and single-cell sequencing technology have created a data-driven revolution in immunology.” - Neu et al. 2017

Single-cell sequencing approaches are refining our understanding of the role played by allergen-specific B cells in food allergies,77, 78 as well as the B cell–mediated neutralizing antibody response to infectious agents.79 Single-cell sequencing approaches are also elucidating novel mechanisms that regulate T cell differentiation and biology in human autoimmune diseases, including rheumatoid arthritis,80 systemic lupus erythematosus (SLE),81 multiple sclerosis,82 type 1 diabetes mellitus,83 and autoimmune encephalomyelitis.84

Antibodies that block CTLA-4 as well as PD-1 induce clinical responses in a number of cancers, including melanoma, lung cancer, renal cancer, bladder cancer, and Hodgkin’s lymphoma.85, 86 Single-cell approaches offer the possibility for a deeper understanding of the complex interactions among immune cells and tumor cells, as well as a more thorough characterization of the cellular ecosystem of tumors.87,88,89

B-Cell Repertoire

Antibodies are produced by B cells in a developmentally ordered series of somatic gene rearrangements that continue throughout the life of an organism. Antibodies are composed of disulfide-linked heavy (V<sub>H</sub>) and light (V<sub>L</sub>) chains, which determine antigen-binding specificity. Each B cell contains a unique pair of V<sub>H</sub> and V<sub>L</sub> that are encoded by several distinct gene loci. To predict antibody-antigen binding specificity accurately, V<sub>H</sub> and V<sub>L</sub> genes should be analyzed from the same single B cell.\(^{90}\)

The antibody V<sub>H</sub> repertoire is generated through somatic recombination of variable (V), diversity (D), and joining (J) gene segments. VJ recombination of V<sub>L</sub>, V<sub>H</sub>, and V<sub>L</sub> heterodimeric pairing completes the antibody.

T-Cell Repertoire

Every T cell expresses unique TCRs, which are heterodimeric proteins composed of a unique combination of \(\alpha\) and \(\beta\) chains. TCRs engage with peptide antigens presented by MHCs on the surface of antigen-presenting cells.\(^{91,92}\) Several single-cell sequencing methods are available to sequence TCRs without disrupting \(\alpha\) and \(\beta\) chain pairing through cell lysis.\(^{93,94}\)

TCR-antigen-MHC interaction and TCR gene recombination. An antigen-presenting cell presents peptide antigen bound to MHC (blue). Similar to mechanisms generating antibody diversity, the TCR (orange) repertoire is generated through V(D)J recombination. TCRs bind to MHC-presented antigens. The complementarity-determining region 3 (CDR3) domain is shown in purple.\(^{95}\)

For Research Use Only. Not for use in diagnostic procedures.

Reviews


Proserpio V and Mahata B. Single-cell technologies to study the immune system. Immunology. 2016;147:133-140

Schober K and Busch DH. A synergistic combination: using RNA-seq to decipher both T-cell receptor sequence and transcriptional profile of individual T cells. Immunol Cell Biol. 2016;94:529-530


References


Single-cell approaches have demonstrated that tumors are genetically heterogeneous. In this study, the authors used the HiSeq 2500 System to perform RNA-Seq of 515 individual cells isolated from breast cancer tumors. Single-cell transcriptomic analyses identified heterogeneous as well as conserved gene expression signatures for subtype-specific breast cancer cells. In addition, the authors identified gene expression signatures for T cells, B cells, and macrophages. Individual T cells could be classified as activated or exhausted, suggesting that immune cell interactions within tumors are dynamic and distinct. These data suggest that individual breast cancer tumor microenvironments contain cells with unique patterns of gene expression. Further, tumor subtypes are shaped by tumor cells, as well as immune cells, within the tumor microenvironment.

Illumina Technology: Nextera XT DNA Sample Prep Kit, TruSeq Rapid SBS Kit, HiSeq 2500 System


As cells age, they undergo a progressive decline in physiology and function. However, the molecular mechanisms underlying this age-dependent decline are not currently understood. In this study, the authors used single-cell RNA-Seq to uncover the effects of aging on the immune system. They used the HiSeq 2500 System to sequence the RNA of naïve and memory CD4+ T cells in young and old mice, in both stimulated and unstimulated states. They found that older T cells exhibited much more variability in gene expression compared with younger T cells. This heightened variability was also associated with a lack of coordination of the immune response in older mice. These data suggest that age-related transcriptional variability among individual cells may be a biomarker for aging cells.

Illumina Technology: Nextera XT DNA Sample Prep Kit, HiSeq 2500 System


All cellular blood components are derived from HSCs in a biological process called hematopoiesis. Among the numerous cell types that are formed, monocytes and dendritic cells are key components of the immune system. Dendritic cells process antigens and present these to T cells, thus activating the adaptive immune response. Both cell types are composed of several subtypes that have been determined through a combination of morphology, localization, and expression of surface markers. In this study, the authors performed single-cell sequencing of 2400 individual cells isolated from human blood and enriched for dendritic cells and monocytes by FACS. This unbiased strategy resulted in a reclassification of 6 dendritic cell and 4 monocyte subtypes. These data provide a more refined understanding of dendritic cells and monocytes, and also suggest a new approach to map the entire taxonomy of the immune system.

Illumina Technology: Nextera XT DNA Sample Kit, HiSeq 2500 System

HCC is a leading cause of cancer deaths worldwide, with the highest incidence in Asia and Africa. Despite the promise for other tumor types, cancer immunotherapies have not been successful in treating HCC, even though HCC tumors contain significant levels of tumor-infiltrating monocytes. In this study, the authors use single-cell RNA-Seq to understand and characterize tumor-infiltrating monocytes in HCC. Specifically, they isolated more than 5000 individual T cells from peripheral blood, tumor, and adjacent normal tissues from 6 HCC samples. They performed deep RNA-Seq on single T cells as well as TCR sequencing using HiSeq 2500/4000 Systems. The combination of single-cell transcriptional profiles and TCR sequences identified 11 different T-cell subtypes. Notably, the authors found that exhausted CD8⁺ T cells and Tregs were enriched and clonally expanded in HCC. This study demonstrates the utility of single-cell sequencing approaches in characterizing the tumor microenvironment and characteristics of tumor-infiltrating cells in HCC.

Illumina Technology: HiSeq 2500/4000 Systems


Single-cell sequencing has led to an enhanced resolution in the characterization of transcriptional heterogeneity in cancer, the immune system, and in pluripotent stem cells. scRNA-Seq measurements typically rely on single time-point measurements. They provide a snapshot of tissue heterogeneity, but extending these measurements to serial time points could potentially enhance our understanding of the mechanisms for generating tissue heterogeneity. In this study, the authors developed a microfluidic platform that enabled off-chip scRNA-Seq after multigenerational lineage tracking. They used the platform to collect single-cell transcriptional data for lineages of mouse lymphocytic leukemia cells, as well as primary murine CD8+ T cells. Their results reveal transcriptional signatures for each cell type that depend on lineage and cell cycle.

Illumina Technology: NextSeq 500 System


TCRs are able to recognize antigen with a high degree of specificity. This is due, in part, to the high degree of genetic heterogeneity of TCRs produced by the recombination of V(D)J loci during T-cell development. In this study, the authors developed TraCeR, a computational method that allowed them to reconstruct full-length paired TCR sequences from scRNA-Seq data. The ability to interrogate recombined TCR sequences in the context of scRNA-Seq data allowed the authors to link T-cell specificity with functional response, by revealing clonal relationships between cells and their transcriptional profiles.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2500 System

For Research Use Only. Not for use in diagnostic procedures.

Single-cell sequencing enables the detailed evaluation of genetic and transcriptional features present in individual cells within tumors. The technology has the potential to inform treatment response and drug resistance by assessing malignant, microenvironmental, and immunologic states within tumors. In this study, the authors applied scRNA-Seq to 4645 single cells (malignant, stromal, immune, and endothelial) isolated from 19 patients with metastatic melanoma. They found that malignant cells within the same tumor displayed transcriptional heterogeneity associated with cell cycle, spatial context, and drug resistance. The same tumor had cells with high expression levels of MITF, as well as cells with low MITF levels and elevated levels of AXL kinase (cells prone to early drug resistance). Infiltrating T-cell analysis revealed exhaustion programs, connection to T-cell activation/expansion, and patient variability. This study demonstrates how single-cell genomics can unravel the cellular ecosystem of tumors, with implications for targeted and immune therapies.

**Illumina Technology:** Nextera XT Sample Preparation Kit, NextSeq 500 System


The interactions between immune cells and invading pathogens determine the course of infection. Bulk cell-sequencing approaches can mask the heterogeneous, stochastic, and dynamic nature of the cell-pathogen interactions. In this study, the authors used the HiSeq 2500 system to perform scRNA-Seq with fluorescent markers to probe the responses of 150 individual macrophages to invading strains of *Salmonella* bacteria. Their data showed that variable PhoPQ activity, which upregulates *Salmonella* virulence factors in infecting *Salmonella*, drove variable host cell type I interferon (IFN) responses by modifying lipopolysaccharides (LPS) in a subset of bacteria. The results suggest that functional heterogeneity in the host cell response to infection is linked to cell-to-cell variations in the population of infecting pathogens.

**Illumina Technology:** HiSeq 2500 System


Combinatorial labeling of single cells is rapid and relatively inexpensive, and it can boost the throughput of massively parallel single-cell sequencing approaches dramatically. In this study, the authors developed CytoSeq, a method to label large numbers of individual cells combinatorially. Individual cells are placed in single wells, along with combinatorial libraries of beads containing cell- and transcript-barcoding probes. The authors performed CytoSeq on human peripheral blood mononuclear cells (PBMCs) and used the MiSeq system to sequence amplified cDNAs. They analyzed several genes and were able to identify major subsets of PBMCs. In addition, by comparing cellular heterogeneity in naïve and cytomegalovirus (CMV)-activated CD8+ T cells, they identified rare cells specific to the CMV antigen. CytoSeq can be applied to complex mixtures of cells of varying size and shape, as well as to other biomolecules.

**Illumina Technology:** MiSeq System


Autoimmune encephalomyelitis (EAE) in mice is widely studied as an animal model of human CNS demyelinating diseases, including multiple sclerosis and acute disseminated encephalomyelitis. IL-17-producing Th17 cells are a critically important part of the adaptive immune system but are also implicated in the pathogenesis of autoimmunity. In this study, the authors used the HiSeq 2000/2500 system to perform RNA-Seq of 976 individual Th17 cells isolated from CNS and lymph nodes of mice with EAE. Computational analysis of scRNA-Seq data revealed the marked genetic heterogeneity of Th17 cells and related them to a spectrum of Th17 cells spanning regulatory to pathogenic functional states. The authors identified and validated 46 four genes (Grp65, Ptp2, Toso, and Cd5l) that regulate Th17 pathogenicity, suggesting possible new drug targets in autoimmunity.

**Illumina Technology:** Nextera XT DNA Sample Preparation Kit, HiSeq 2000/2500 Systems
An overview of recent publications featuring Illumina technology

For Research Use Only. Not for use in diagnostic procedures.


Acute SLE is a recurrent autoimmune disease that attacks various tissues and has no cure. Autoimmune activity is associated with surges in B cells. The only SLE therapy approved by the US Food and Drug Administration (FDA), belimumab, targets B-cell activating factor. In this study, the authors isolated B cells from patients experiencing SLE flares and used deep sequencing and proteomic approaches to analyze the diversity of B cells. They demonstrated that B cells from patients with SLE flares were polyclonal. By sequencing single B cells, they also identified a subpopulation of newly activated naïve B cells that provided an important source of autoantibodies during SLE flares, suggesting that SLE autoreactivities occur during polyclonal activation. These results may guide patient treatment options and facilitate the design of future SLE therapies.

**Illumina Technology: MiSeq System**


TCRs interact with MHC class II antigen complexes.
Neurobiology

Among cells comprising specific brain regions, single-cell sequencing approaches have discovered genomic mosaicism in individual neurons, including CNVs and somatic SNVs. The genetic variations underlying this genomic mosaicism in the brain arise during fetal development, and while their functional relevance is not yet fully clear, recent data suggest that aging and neurodegeneration are associated with somatic mutations in neurons. Recent single-cell sequencing studies have also identified high rates of somatic LINE-1 element (L1) retrotransposition in the hippocampus and cerebral cortex that could have major implications for normal brain function; however, other studies have determined that rates of L1 retrotransposition in the brain are lower than first reported.

“Single-cell RNA sequencing is an exciting new technology allowing the analysis of transcriptomes from individual cells, and is ideally suited to address the inherent complexity and dynamics of the central nervous system.” - Ofengeim et al. 2017

In addition to CNS genomic diversity, recent studies have extended our understanding of CNS transcriptome diversity at the single-cell level from both adult brain as well as 3-dimensional brain organoids. Single-cell transcriptomics has identified mechanisms regulating neurodevelopment, and scRNA-Seq studies have recently begun to unravel new biological details of sensory neurons, interneurons, and other cell types in the brain. New technical achievements in single-cell sequencing combine scRNA-Seq with electrophysiological recording of individual neurons, as well as characterizing gene expression patterns associated with experience-driven induction of activity in individual hippocampal neurons. Finally, single-cell analysis of 6000 methylomes from individual neuronal nuclei recently identified neuronal subtypes in the human frontal cortex.

Single-cell sequencing approaches uncover genetic mosaicism in neurons.
Reviews
Harbom LJ, Chronister WD and McConnell MJ. Single neuron transcriptome analysis can reveal more than cell type classification: Does it matter if every neuron is unique? Bioessays. 2016;38:157-161
Bae BI, Jayaraman D and Walsh CA. Genetic changes shaping the human brain. Dev Cell. 2015;32:423-434

References

Somatic mutations are permanent changes in the DNA sequence of a genome. They arise because of mistakes in DNA replication in cells or from environmental factors. Previous studies using single-cell sequencing approaches have demonstrated that individual neurons contain unique SNV somatic mutations, a phenomenon referred to as genomic mosaicism. Although genomic mosaicism is common among multiple cell types, neurons are postmitotic, and thus the SNVs have the potential to persist and alter neuronal function. In this study, the authors isolated individual neurons from the forebrains of embryonic samples and grew them into small clones in culture. They used the HiSeq platform to analyze the genomes of these samples, and they identified 200–400 SNVs per cell. The SNV patterns in these embryonic neurons resembled those found in cancer cell genomes. By analyzing the genomes of neurons at different stages of embryonic development, they determined that the mutation rate speeds up during neurogenesis. These data suggest that embryonic neurons are protected from mutations, compared to neurons at later stages of neurodevelopment.

Illumina Technology: HiSeq 2500 and HiSeq X Systems


Single-cell RNA-Seq approaches have been useful in defining diverse neuronal subtypes within the brain. In this study, the authors developed a technique to characterize the epigenomic diversity within individual neuronal cells. This technique, single-nucleotide methylcytosine sequencing (snmC-seq), takes advantage of the significant amount of 5-methylcytosine found in neurons. The authors isolated 3400 individual neurons from mouse cortex and collected nuclei by FACS. They treated the nuclear DNA treated with bisulfite to convert unmethylated cytosine into uracil, while 5-methyl- and 5-hydroxymethylcytosines remained unaffected. Finally, they pooled, amplified, and sequenced the samples using the HiSeq 4000 System. The authors used snmC-seq to generate 6000 methylomes from the isolated neuronal nuclei, and their data identified 16 mouse neuronal subtypes in mouse brain. They also performed snmC-seq on isolated nuclei from human cortex samples, and they identified 21 human neuronal subtypes. The single-cell methylome technique thus expanded the taxonomy of brain cell diversity. Moreover, it identified regulatory elements associated with neuronal diversity.

Illumina Technology: TruSeq Methylation Kit, HiSeq 4000 System


Aging is associated with increased somatic mutations in cells, as errors in replication accrue. It is unclear whether this is true for postmitotic neurons that do not divide. In this study the authors isolated 161 individual postmitotic neurons from individuals aged 4 months to 82 years. They performed single-cell WGS of each individual neuron to identify SNVs. Their data demonstrated that SNVs increased linearly with increasing age. They also isolated individual neurons from 8 individuals with early-onset neurodegenerative diseases. Single-cell WGS of these neurons indicated SNVs were more abundant than in normal neurons. Overall, these data suggest that age-related accumulation of SNVs in neurons may be important in neurodegenerative diseases.

Illumina Technology: TruSeq Nano LT Sample Preparation Kit, HiSeq 2000 and HiSeq X10 Systems

---

For Research Use Only. Not for use in diagnostic procedures.

An overview of recent publications featuring Illumina technology 29

The brain is an enormously complex organ, and 2-dimensional tissue cultures have limited utility in understanding brain function. Recently, 3-dimensional tissue culture brain organoids have been developed in the hope of understanding brain biology, disease, and development. In this study, the authors attempted to characterize the extent to which brain organoids mimicked the complexity and cellular diversity of whole brains. They performed single-cell RNA-Seq of over 80,000 individual cells isolated from 31 human brain organoids. The single-cell gene expression signatures demonstrated that human brain organoids developed and generated a wide diversity of cell types, including subtypes related to adult cortical neurons as well as retinal cells. Moreover, the neuronal connectivity and synapses were present within cells of the organoid. Finally, the organoids developed neuronal networks spontaneously, and photosensitive cells were even sensitive to light stimulation. Overall, the data suggest that human brain organoids have enormous potential as *in vitro* models of the human brain.

**Illumina Technology:** Nextera XT DNA Library Preparation Kit, NextSeq 500 System

---


In this study, the authors describe Patch-Seq, a method that combines whole-cell electrophysiological characterization, morphological characterization, and scRNA-Seq. They performed Patch-Seq on 58 neuronal cells from layer 1 of the mouse neocortex. After they characterized individual neurons electrophysiologically, they aspirated the cell contents through the patch-clamp pipette and prepared them for RNA-Seq. The authors classified cells based on electrophysiology and morphology, as well as by patterns of gene expression. Their data show that gene expression patterns could infer axonal arborization and the action potential amplitude of individual neurons.

**Illumina Technology:** HiSeq 2000 System

---


scRNA-Seq has been a key method for the subclassification of cells that would otherwise be indistinguishable based solely on morphology or anatomy. Profiling the transcriptome of individual neurons in response to activation is important for characterizing brain function. In this study, the authors performed RNA-Seq on isolated nuclei (snRNA-Seq) from individual mouse neurons taken from the dentate gyrus of the hippocampus and stimulated by pentylentetrazole. There were large-scale changes in the activated neuronal transcriptome, including induction of mitogen-activated protein kinase (MAPK) pathway genes, after brief, novel environment exposure. The data show that snRNA-Seq of activated neurons allows for analysis of gene expression beyond immediate early genes.

**Illumina Technology:** HiSeq 2500 System

---

For Research Use Only. Not for use in diagnostic procedures.

Single-cell Research
An overview of recent publications featuring Illumina technology

For Research Use Only. Not for use in diagnostic procedures.


Given the complexity of the human brain, understanding the genetic and functional diversity of individual cells is of paramount importance. In this study, the authors performed scRNA-Seq on more than 1600 individual cells from the mouse primary visual cortex. Their data analysis identified 49 unique transcriptomic cell types. A subset of these transcriptomic cell types displayed specific and differential electrophysiological and axon projection properties, which confirms that single-cell transcriptomic signatures can be linked to specific cellular phenotypes.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, MiSeq and HiSeq 2000/2500 Systems


Postmitotic, somatic mutations are known to cause cancer, but these mutations may also lead to diverse neurological diseases, including cortical malformations, epilepsy, intellectual disability, and neurodegeneration. Studying pathogenic somatic mutations is challenging due to the variety of ways that their effects are shaped by normal development. The overall pattern of somatic mutation distribution in the human brain is not well characterized. In this study, the authors used the HiSeq 2000 system to perform high-coverage WGS of individual neurons. Somatic mutation analyses of individual neurons from several CNS locations identified multiple cell lineages in the brain, marked by different L1 retrotransposition events. The patterns of somatic mutations mirrored known somatic mutation disorders of brain development, suggesting that focally distributed somatic mutations are also present in normal brains.

Illumina Technology: HiSeq 2000 System


In mammals, odor detection is mediated by G protein–coupled olfactory receptors on neurons in the nasal olfactory epithelium. Mature neurons typically express a single olfactory receptor per neuron. In this study, the authors used the HiSeq 2500 system to perform scRNA-Seq of single epithelial neurons during mouse development, with multiple cells from each stage of development sequenced. The single-cell data confirmed that most neurons expressed high levels of only 1 olfactory receptor. However, many immature neurons expressed multiple olfactory receptors at low levels, with a single neuron capable of expressing olfactory receptors from up to 7 different chromosomes. The data show that developmental pathways ultimately restrict olfactory receptor expression in mature neurons.

Illumina Technology: TruSeq DNA Sample Preparation Kit, HiSeq 2500 System


Within the human brain, pools of adult NSCs participate in brain maintenance and regeneration following injury. The balance of activation and quiescence of NSCs depends on the induction of specific transcription factors. In this study, the authors used the HiSeq 2000 system to perform scRNA-Seq of NSCs isolated from the brain subventricular zone. They identified the expression of lineage-specific transcription factors in a specific subpopulation of dormant NSCs. They also discovered that brain ischemic injury induced interferon signaling in dormant NSCs, promoting their entry into a primed-quiescent state. This study unveils the general molecular principles underlying NSC activation and suggests potential avenues for brain regenerative medicine.

Illumina Technology: Nextera XT Sample Preparation Kit, HiSeq 2000 System


Neurons are postmitotic cells, so their genomes are particularly susceptible to DNA damage. In this study, the authors surveyed the landscape of somatic SNVs in the human brain by performing single-cell WGS of 36 individual cortical neurons. The most abundant SNVs included noncoding, noncoding RNA, intronic, and intergenic SNVs. Coding, truncating, splice, and silent SNVs were much less abundant. Moreover, the data showed that each cortical neuron had a distinctive genome that harbored up to 1580 somatic SNVs. Finally, the somatic SNVs created nested linkage trees, demonstrating that somatic mutations could be used to reconstruct the developmental lineage of neurons.

Illumina Technology: TruSeq Nano LT Sample Preparation Kit, MiSeq, HiSeq 2000 and HiSeq X Ten Systems
The scarcity of tissue-specific stem cells, and the complexity of their surrounding environment, make single-cell sequencing methods imperative for characterizing these cell types. In this study, the authors used single-cell sequencing and weighted gene coexpression network analysis to identify CD133+ ependymal cells from the adult mouse forebrain neurogenic zone as NSCs. These subpopulations of cells were enriched for immune-responsive genes, as well as genes encoding angiogenic factors. Administration of VEGF and bFGF enhanced migration and elicited differentiation into neurons and glia.

Illumina Technology: HiSeq 2500 System

Somatic L1 retrotransposition occurs during neurogenesis and serves as a potential source of genotypic variation among neurons. Pronounced L1 activity is present in the hippocampus, but its biological consequence is unclear. In this study, the authors used the MiSeq system to perform single-cell retrotransposon capture sequencing (RC-Seq) on individual hippocampal neurons, hippocampal glial cells, and cortical neurons. These experiments established that L1-driven mosaicism is abundant in the hippocampus with 13.7 somatic L1 insertions per hippocampal neuron, compared to previous estimates of < 0.1 L1 insertions per cortical neuron. Given the abundance of L1 insertions, it is reasonable to assume that L1-driven mosaicism may alter brain biology and function.

Illumina Technology: MiSeq System

The brain is built from a large number of cell types, of which neurons are a subset. To understand brain function, it is important to characterize this complex collection of highly specialized cells. In this study, the authors performed quantitative scRNA-Seq based on 3005 single-cell transcriptomes from the primary somatosensory cortex and hippocampal CA1 regions. RNA molecules were counted using unique molecular identifiers (UMIs). Biclustering methods revealed 9 major classes of cells, including pyramidal neurons, interneurons, oligodendrocytes, astrocytes, microglia, vascular endothelial cells, mural cells, and ependymal cells. The researchers identified 47 molecularly distinct subclasses within the 9 cell types. They also identified unique marker genes that aligned with cell type, morphology, and location.

Illumina Technology: HiSeq 2000 System

Diversity of brain cells.


Drug Discovery

Modern drug discovery involves the identification of novel drug targets that are highly relevant to a disease process, followed by the development of highly specific, target-based therapeutic interventions. In a given tissue, cells may express up to 400 druggable genes, suggesting a relative abundance of drug targets. However, tissues are composed of a variety of genetically diverse cell types, and the resultant biological noise in bulk sequencing approaches can drown out the specific genes of interest. Single-cell sequencing approaches show great promise in discovering and characterizing drug targets, and in improving pharmaceutical screens.

"The emergence of single-cell omics tools over the past 5 years has happened at a lightning pace, and the potential for their use in discovery and development of broad classes of therapies and therapeutic strategies is high." - Health et al. 2016

Single-cell sequencing of tumor cells is enhancing our understanding of tumor heterogeneity, drug response, and resistance mechanisms in cancer. Similarly, single-cell analysis of circulating tumor cells highlights their genetic heterogeneity, therapeutic sensitivity, and metastatic potential.

Recent clinical data have demonstrated that therapeutic enhancement of immune system function can benefit patients with cancer. Antibodies that block CTLA-4 as well as PD-1 induce clinical responses in a number of cancers, including melanoma, lung cancer, renal cancer, bladder cancer, and Hodgkin’s lymphoma. Single-cell sequencing approaches offer the possibility for a deeper understanding of the complex interactions among immune cells and tumor cells, as well as a more thorough characterization of the cellular ecosystem of tumors.

In addition to uncovering mechanisms of tumor-host immune interactions, single-cell approaches also uncover host immune responses to infectious, pathogenic microorganisms. Our understanding and characterization of microbial diversity continues to expand, and single-cell sequencing can help in characterizing the abundance of infectious pathogens in our environment. Finally, genomic characterization of uncultivable microbes has the potential for the discovery of novel natural products that could be used as new chemical scaffolds in pharmaceutical development.

Single-cell sequencing approaches are refining our understanding of allergen-specific B cells in food allergies, as well as the B cell-mediated neutralizing antibody response to infectious agents. Single-cell sequencing approaches are also elucidating novel mechanisms regulating T-cell differentiation and biology in human autoimmune diseases, including rheumatoid arthritis, SLE, multiple sclerosis, type 1 diabetes mellitus, and neuroinflammation.

An overview of recent publications featuring Illumina technology

Reviews


References


Alzheimer's disease is a chronic neurodegenerative disease with no known cure. Its incidence increases with age, with approximately 6% of people age 65 and older affected by the disease.1-16 To better understand the cellular etiology of Alzheimer's disease, the authors performed single-cell RNA-Seq of approximately 8000 individual immune cells from healthy transgenic mice as well as those with Alzheimer's disease. The data revealed a novel microglial cell that was associated with neurodegenerative disease. These microglia were found to contain Aβ particles, a molecular signature of Alzheimer's disease. The data suggest that this novel immune cell type has the potential to restrict neurodegeneration, with potential implications for treatment of Alzheimer's disease.

Illumina Technology: NextSeq 500 System


Many tumors can be successfully treated using chemotherapy, but they eventually develop resistance to cancer drugs. Consequently, understanding the mechanisms of drug resistance is an active area of research. About 60% of melanoma tumors carry a V600E mutation in the BRAF kinase gene, rendering them sensitive...
to the BRAF inhibitor, vemurafenib; however, a subset of tumor cells will ultimately develop resistance. In this study, the authors isolated single tumor cells from melanoma patient samples, which they then cultured in the presence of vemurafenib. Targeted DNA sequencing of 119 cancer-related genes in parental cells as well as vemurafenib-resistant subclones demonstrated that no new mutations developed in resistant subclones. They then performed population-based RNA-Seq of parental cells and vemurafenib-resistant cells, identifying increased expression of several known resistance marker genes in resistant cells. Moreover, drug-exposed resistant cells showed increased expression of many of the resistance marker genes. Overall, these data suggest a novel mechanism for drug resistance in single tumor cells.

**Illumina Technology: HiSeq 2500 System**


Tumor cell genetic heterogeneity plays a role in drug resistance; however, in this study, the authors focused on understanding how cancer phenotypic heterogeneity might contribute to cancer progression. They isolated isogenic clonal cancer cells displaying high morphological variation. These highly variable subpopulations displayed increased survival and metastatic potential in response to numerous anticancer drugs. The authors performed scRNA-Seq of these clones, and they found that morphological variants expressed enhanced transcriptomic variability, despite their genetic stability. Finally, they correlated high variability in expression of the spliceosome-machinery genes with transcriptomic variability and promotion of metastasis. These data demonstrate a novel aspect of intraclonal tumor heterogeneity, and further suggest a mechanism for increased drug resistance, survival, and metastasis.

**Illumina Technology: Nextera Extended Exome Sequencing Kit, HiSeq 2500 System**


Glioblastoma is one of the deadliest forms of cancer. Glioblastomas tumors have mutations in several druggable pathways, but current targeted therapies have proven ineffective due to rapid and universal drug resistance. Specifically, the mTOR pathway is a key driver in 90% of glioblastoma, yet tumor cells develop rapid resistance to mTOR-targeted therapies. In this study, the authors used the NextSeq 500 system to obtain single-cell genomic data, which they correlated with single-cell proteomic data in tumor cells treated with mTOR inhibitor. Their data showed that resistance to mTOR inhibitor in glioblastoma tumor cells occurred within days of drug therapy. Surprisingly, the correlation with single-cell sequencing data demonstrated that this drug resistance proceeds via nongenetic mechanisms, through upregulation of specific signaling phosphoproteins. This study suggests a novel approach for designing drug combination therapy in treating glioblastoma.

**Illumina Technology: NextSeq 500 System**


In the adaptive immune system, proinflammatory IL-17–producing Th17 cells contribute to pathogen clearance, but they have also been implicated in autoimmune and inflammatory diseases. Although Th17 cells display extensive cellular heterogeneity, the genetic basis of this heterogeneity is not well characterized. In this study, the authors used the HiSeq 2000/2500 system to perform scRNA-Seq of Th17 cells isolated from patients with EAE. They compared this data set with scRNA-Seq data from normal Th17 cells differentiated in vitro. Their data showed genetic heterogeneity in Th17 cells from autoimmune patients. Further, they used transgenic mice to validate 4 genes (Gpr65, Pltp, Toso, and Cd5L) involved in developing EAE. These findings may lead to the discovery of novel drug targets in autoimmune disease.

**Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2000/2500 Systems**


Intratumor heterogeneity correlates with poor cancer clinical prognosis, but the mechanism for this correlation is not entirely understood. In this study, the authors isolated 34 PDX lung adenocarcinoma tumor cells and performed scRNA-Seq using the HiSeq 2000/2500 system. By clustering 69 lung adenocarcinoma–prognostic genes, including KRAS, they could classify the PDX cells into 4 distinct subgroups. scRNA-Seq of the PDX cells that survived anticancer drug treatment demonstrated that tumor cells with activated KRAS variants were targeted by anticancer drugs, even though KRAS itself was not the target. Their scRNA-Seq data also suggest that the tumor cells responsible for drug resistance can be masked by the genomics of the bulk tumor.

**Illumina Technology: Nextera XT DNA Sample Prep Kit, HiSeq 2000/2500 Systems**
Hepatitis C virus (HCV) is an RNA virus of the Flaviviridae family and is the leading cause of liver disease worldwide. HCV requires the liver-specific tumor suppressor microRNA(miR)-122 for replication, but the regulation of endogenous miRNAs by HCV infection has not been characterized. In this study, the authors generated global miRNA-target interaction maps during HCV infection and demonstrated that HCV viral RNA sequesters miR-122 to derepress normal host targets. They then validated and quantified the RNA sequestration at the single-cell level by analyzing expression data from reporters containing miR-122 sites, confirming significant derepression during HCV infection. The findings describe a model for HCV-induced miR-122 sequestration and suggest that this mechanism may enhance the long-term oncogenic potential of HCV.

Illumina Technology: MiSeq and HiSeq 2000 Systems

Single-cell sequencing of individual tumor cells holds the potential of better correlating genetic heterogeneity with the mechanism of drug response and resistance. In this study, the authors characterized the heterogeneity in single-cell gene expression across 336 lung adenocarcinoma cells derived from cell lines. They also analyzed lung adenocarcinoma cells before and after treatment with the multi-tyrosine kinase inhibitor, vandetanib. They found that relative expression diversity of cellular housekeeping genes was reduced in cancer cells exposed to vandetanib. In contrast, the expression diversity of genes targeted by vandetanib (including EGFR and RET) remained constant. Their data demonstrate that patterns in gene expression divergence play important roles in tumor cells acquiring drug resistance; further, this genetic diversity is not revealed by RNA-Seq of bulk tumors.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2500 System

Acute SLE is a recurrent autoimmune disease that attacks various tissues and has no cure. Autoimmune activity is associated with surges in B cells. The only SLE therapy approved by the US FDA, belimumab, targets B-cell activating factor. In this study, the authors isolated B cells from patients experiencing SLE flares and used deep sequencing and proteomic approaches to analyze the diversity of B cells. They demonstrated that B cells from patients with SLE flares were polyclonal. By sequencing single B cells, they also identified a subpopulation of newly activated naïve B cells that provided an important source of autoantibodies during SLE flares, suggesting that SLE autoreactivities occur during polyclonal activation. These results may guide patient treatment options and facilitate the design of future SLE therapies.

Illumina Technology: MiSeq System

B lymphocytes (B cells) secrete antibodies.


Reproductive Health

Prenatal diagnosis refers to testing for diseases or genetic conditions in a fetus or embryo before birth. Birth defects can be diagnosed by detecting detrimental chromosomal or DNA aberrations. Traditionally, this diagnosis has used invasive procedures, such as amniocentesis. Noninvasive prenatal testing (NIPT) offers greater confidence over traditional testing methods and only requires a blood draw. NIPT through cell-free DNA (cfDNA) sequencing is used clinically and is improving prenatal diagnoses.151

"Single-cell sequencing enables global sequencing of the genome, exome, transcriptome, or epigenome for a comprehensive analysis of the early embryo."

- Zhu et al. 2017

Single-cell sequencing can also be useful in the realm of PGS/PGD, where embryos can be screened for chromosomal abnormalities prior to implantation.152,153 Research shows how sequencing aids genetic testing of in vitro–generated gametes (IVG) either for research or for implantation.154 Single-cell sequencing diagnostics can accelerate research, improve the early detection of embryo status, and assure transplantation of a healthy in vitro fertilized (IVF) embryo.155,156

Reviews


Genetic testing of an embryo can provide early insights into chromosomal abnormalities.
References


Single-cell transcriptome analysis has been performed previously for a number of biopsy tissues, but those for endometrium have not been published. In this study, the authors present a complete pipeline for endometrial single-cell gene expression profiling, from sampling to data analysis. They dissociated individual stromal and epithelial cells from endometrium and isolated the relevant cell types using FACS followed by scRNA-Seq. In parallel, they cultured endometrial stromal cells and compared global expression profiles with uncultured cells. The authors found that 2661 genes (out of 8622 detected genes) were more active in cultured stromal cells than in biopsy cells. Further, 241 genes were differentially expressed in biopsy vs cultured individual cells, at significant levels.

Illumina Technology: HiSeq System


During human development, the zygote undergoes cellular division to establish 3 cell types of the mature blastocyst: trophectoderm (TE), primitive endoderm (PE), and epiblast (EPI). In this study, the authors used the HiSeq 2000 system to sequence the transcriptomes of 1529 individual cells from 88 human preimplantation embryos. They developed a comprehensive transcriptional map of human preimplantation development, demonstrating that lineages of TE, PE, and EPI all occur simultaneously and coincide with blastocyst formation. Their data also identified novel genes, such as ARGFX, FOXA2, and LINC00261, that may be important for preimplantation development. Finally, their transcriptional map identified unique features of X-chromosome dosage compensation in humans.

Illumina Technology: HiSeq 2000 System


PGD allows for genetic profiling of human embryos to screen for genetic disorders prior to implantation. Although recent technical advances in single-cell sequencing highlight the promise of this approach in PGD, DNA amplification bias and relatively long turnaround time limit this approach. In this study, the authors combined haplotype-resolved parental genome sequencing with rapid embryo genotyping to predict the whole-genome sequence of a 5-day human embryo. The authors predicted inheritance at 3 million paternal/maternal heterozygous sites with > 99% accuracy, as well as the transmission of a parental HBA1/HBA2 deletion. The data suggest that PGD may facilitate the diagnosis of genetic diseases in embryos.

Illumina Technology: HiSeq 2000 System, CytoSNP-12 Chip


NGS methods have improved the precision of PGS/PGD. Although the precision has been limited by false-positive and false-negative SNVs, linkage analysis can overcome this challenge. In this study, the authors developed MARSALA, a method that combines NGS using the HiSeq platform with single-cell WGA. The method allows for embryo diagnosis with a single-molecule precision and significantly reduces false-positive and false-negative errors. This is the first integrated NGS-based PGD procedure that simultaneously detects disease-causing mutations and chromosome abnormalities, and performs linkage analyses.

Illumina Technology: HiSeq 2500 System


Before analyzing single-cell DNA-Seq data, DNA copy-number aberrations must be differentiated from WGA artifacts. This requirement makes DNA copy-number profiling and haplotyping of single-cell sequencing data challenging. In this study, the authors developed a single-cell genome analysis method that determined haplotypes and copy number across the genome of a single cell—a process called, haplarithmisis. The method deciphered SNP allele fractions of single cells and integrated these data into a computational workflow for imputation of linked disease variants (siCHILD). The authors validated the method by determination of haplotypes carrying disease alleles in single-cell genomes from individual lymphocytes and human blastomeres derived from human IVF embryos.

Illumina Technology: TruSeq DNA LT Sample Preparation Kit, HumanCytoSNP-12v2.1 BeadChips, HiSeq 2000/2500 Systems

For Research Use Only. Not for use in diagnostic procedures.

Single-cell Research
Microbial Ecology and Evolution

Not only do we discover new species and microorganisms continually, current NGS technology can help us to understand the dynamics of microbial ecology and evolution. These discoveries include host-species interactions that generate selection pressures, which can lead to the evolution of a species. Such insight is critical to understanding complex ecosystems and the many unique microbes that comprise them. Sequencing can also facilitate the detection of these new species and their evolved roles in their respective environments.

“New genomic information from metagenomics and single cell genomics has provided insights into microbial metabolic cooperation and dependence, generating new avenues for cultivation efforts.”

- Solden et al. 2016

Single-cell analysis provides a better assessment of how different organisms pressure selection and the evolution of cohabitants, as well as host-pathogen interactions.
Reviews


References


Marine sediments are the largest carbon sink on the planet, with half of chemosynthetic oceanic carbon fixation occurring in coastal sediments. However, the microbes responsible for this activity are unknown. By surveying bacterial 16S rRNA gene diversity from 13 coastal sediments across Europe and Australia, the authors identified groups of Gammaproteobacteria that were affiliated with sulfur-oxidizing bacteria. 14C-carbon assimilation studies showed that these uncultured Gammaproteobacteria accounted for 80% of carbon fixation in coastal sediments. Finally, the authors isolated individual cells from the environmental sample and performed single-cell WGS to identify genes that linked hydrogen-oxidizing activity with sulfur-oxidizing Gammaproteobacteria.

**Illumina Technology: MiSeq and HiSeq 2000 Systems**


In microbial ecology studies, 16S rRNA sequencing can identify microbial community members, whereas shotgun metagenomics can determine the functional diversity of the community. However, combining the 2 approaches is technically challenging. In this study, the authors developed emulsion, paired isolation, and concatenation PCR (epiPCR), a technique that links functional genes and phylogenetic markers. They applied the technique to millions of uncultured individual cells from the freshwater Upper Mystic Lake in Massachusetts. Specifically, they profiled the sulfate-reducing community within the freshwater lake community and were able to identify new putative sulfate reducers. The method is suitable for identifying functional community members, tracing gene transfer, and mapping ecological interactions in microbial cells.

**Illumina Technology: MiSeq System**


SAR11 bacteria are the most abundant microbes in the earth’s oceans, constituting half of all microbial cells in the oxygen-rich surface ocean. Although considered aerobic, SAR11 are also abundant in marine environments where oxygen levels are low. In this study, the authors used the MiSeq system to sequence 19 single-cell amplified genomes from a subpopulation of SAR11 bacteria isolated from ocean oxygen-minimum zones. They found that the SAR11 bacteria that had adapted to their low-oxygen environment encoded abundant respiratory nitrate reductases. These enzymes perform the first step in denitrification, a microbially facilitated process of nitrate reduction that may ultimately produce molecular nitrogen (N2). These data redefine the ecological niche of earth’s most abundant organismal group and suggest that SAR11 bacteria contribute to nitrogen loss in oxygen-minimum zones.

**Illumina Technology: Nextera XT DNA Sample Preparation Kit, MiSeq and HiSeq Systems**


Genetic diversity is a key determinant in the ability of viruses to escape immunity and vaccination, develop drug resistance, and cause disease. It is assumed that single virions constitute viral infectious units. However, the authors performed single-cell sequencing of 881 VSV plaques derived from 90 individual infected cells and showed that individual virus infectious units were comprised of multiple genetically diverse viral genomes. They also found that several genome viral variants could be delivered simultaneously to the same individual cells, and the rate of spontaneous virus mutation varied across individual cells, with implications for viral yield. This study at the single-cell level has implications for our understanding of viral diversity and evolution.

**Illumina Technology: MiSeq System**
An overview of recent publications featuring Illumina technology


Viral infections can alter the composition and metabolic potential of marine communities, as well as the evolution of host populations. All oceanic microbes are potentially impacted by viral infections; however, our understanding of host-virus interactions is limited. In this study, the authors used single-cell WGS of 58 isolated oceanic microbes to identify genomic blueprints of viruses inside or attached to individual bacterial and archaeal cells. The data include the first known viruses of Thaumarchaeota, Marinimicrobia, Verrucomicrobia, and Gammaproteobacteria. They demonstrate that single-cell genomics approaches can provide insight into host-virus interactions in complex environments.

**Illumina Technology:** NextSeq 500 System


Oceanic plankton is the world’s largest ecosystem and is composed of viruses, prokaryotes, microbial eukaryotes, phytoplankton, and zooplankton. This ecosystem structure and composition are influenced by environmental conditions and nutrient availability. In this study, the authors analyzed 313 plankton samples from the Tara Oceans expedition and obtained viral, eukaryotic, and prokaryotic abundance profiles from Illumina-sequenced metagenomes and 18S rDNA V9 sequences. They used network inference and machine-learning methods to construct an interactome among plankton groups. In particular, the authors confirmed predicted virus-host interactions by comparing putative host contigs with viral data from single-cell genomes.

**Illumina Technology:** Illumina-sequenced metagenomes (=tags) and 18S rRNA V9 sequences

Phytoplankton accounts for half of all photosynthetic activity on Earth.


The bacterial family Rickettsiaceae includes the epidemic typhus-causing pathogen *Rickettsia prowazekii,* and thus Rickettsiaceae host-pathogen interactions are of great interest. In this study, the authors discovered Candidatus Arcanobacter lacustris, a Rickettsiaceae sister lineage alphaproteobacterium isolated from Damariscotta Lake. They used the HiSeq 2000 system to perform single-cell WGS of Candidatus Arcanobacter lacustris. Phylogenetic and comparative analysis of its genome revealed the presence of chemotaxis and flagellar genes. These genes are unique in the Rickettsiaceae family and suggest that the ancestor of Rickettsiaceae may have had a facultative lifestyle.

**Illumina Technology:** HiSeq 2000 System


For Research Use Only. Not for use in diagnostic procedures.
Plant Biology

Although transcriptomics has advanced our understanding of plant development, single-cell transcriptomics has not yet been employed as widely in plants.\textsuperscript{166,167} Single-cell sequencing approaches have great potential to further our understanding of plant biology.\textsuperscript{168,169} However, plant cells are enclosed in a rigid cell-wall matrix, and isolating individual plant cells remains challenging technically. In Arabidopsis, a number of techniques have been used to isolate single cells, including protoplasting to remove plant cell walls followed by FACs,\textsuperscript{170,171} and cell-wall digestion followed by glass micropipetting.\textsuperscript{172} In maize kernels, RNA-Seq of LCM compartments has

“The continued application of single-cell plant system biology is likely to transform our view of the relationships, development, and evolution of distinct cell-types in plants.” - Libault et al. 2017

Single-cell analysis will allow a better assessment of the nature of plant stem cells, plant cell plasticity, and local cellular response to environmental changes.
revealed details of plant endosperm cell differentiation, and WGS of individual maize microspores has advanced our understanding of plant meiotic recombination. While technical challenges remain, scRNA-Seq methods are poised to revolutionize our knowledge of plant biology.

Reviews
Efroni I and Birnbaum KD. The potential of single-cell profiling in plants. Genome Biol. 2016;17:65

References
Like mammalian systems, functional biology in plants is driven by spatially oriented gene expression programs that are differentially regulated in different tissues. Resolving this spatial organization in gene expression using single-cell RNA-Seq of individual plant cells is feasible, and it has been achieved by using laser-capture microdissection (LCM, see below) and FACS of plant tissues. In this study, the authors present a novel method for generating spatial transcriptome data in plants. They imaged Arabidopsis thaliana tissue cryosections on barcoded arrays containing spot-specific barcodes coupled to oligo(dT) oligonucleotides. After cDNA synthesis, they used a NextSeq System to characterize gene expression within each array spot. By integrating the spatial information with gene expression data, the authors identified gene expression differences in 8 tissue domains of Arabidopsis thaliana.

Illumina Technology: NextSeq System

Plant stomata facilitate plant gas exchange with the atmosphere. In Arabidopsis, the production and pattern of stomata proceeds from a discrete lineage that can be parsed into intermediate steps. Despite the biological significance of RNase L, the RNAs cleaved by this enzyme are poorly defined. In this study, the authors used illumina sequencing to reveal the frequency and location of RNase L cleavage sites within host and viral RNAs. The method was optimized and validated using viral RNAs cleaved with RNase L and RNase A, and RNA from infected and noninfected HeLa cells. The authors identified discrete genomic regions susceptible to RNase L and other single-strand–specific endoribonucleases. Monitoring the frequency and location of these cleavage sites within host and viral RNAs may reveal how these enzymes contribute to health and disease.

Illumina Technology: TruSeq SBS Kit v3–HS, HiSeq 2000 System
Li X, Li L and Yan J. Dissecting meiotic recombination based on tetrad analysis by single-microspore sequencing in maize. Nat Commun. 2015;6:6648

Meiotic recombination plays an important role in genetic diversity by contributing to allele assortment, creating a substrate for natural selection, and evolving eukaryotic genomes. Maize has been used successfully as a genetic model for the dissection of recombination variation, but understanding single-meiotic events at nucleotide-level resolution has been impossible previously, due to the difficulty in sequencing single plant cells and gametes. In this study, the authors developed a method for isolation and WGS of the 4 microspores of a single maize tetrad. They used the HiSeq 2000 system to sequence 96 individual microspores, identifying 600,000 high-quality SNPs, which allowed them to characterize recombination patterns at very high resolution. Their high-resolution recombination map revealed that crossovers were more likely to occur in genic rather than intergenic regions; further, they were especially common in the 5'- and 3'-end regions of annotated genes.

**Illumina Technology:** TruSeq DNA Sample Preparation Kit, HiSeq 2000 System


Cereal endosperm is a main source of food, feed, and raw material worldwide, yet genetic control of endosperm cell differentiation is not well defined. In this study, the authors coupled LCM and Illumina sequencing to profile mRNAs in 5 major cell types of differentiating endosperms and 4 compartments of maize (Zea mays) kernels. They identified mRNAs that specifically accumulate in each compartment, as well as genes predominantly expressed in 1 or multiple compartments. Their results demonstrate that the MRP-1 transcription factor can activate gene regulatory networks within the basal endosperm transfer layer. These data provide a high-resolution gene activity atlas of the compartments of the maize kernel. The study also uncovers the regulatory modules associated with differentiation of the major endosperm cell types.

**Illumina Technology:** TruSeq DNA Sample Preparation Kit, HiSeq 2000 System


Forensics

NGS has revolutionized many aspects of modern forensics, including short tandem repeat (STR) analysis, monozygotic twin discrimination, Y chromosome analysis, mitochondrial whole-genome studies, age estimation, cause-of-death determination, bodily fluid identification, forensic microbiological analysis, species identification, and ancestry inference. Single-cell forensic analysis was first reported in 1997, when van Oorschot et al. performed STR analysis for 226 individual buccal cells isolated by micromanipulation. Single-cell sequencing methods are suited to criminal investigations that are hampered by very small amounts of biological material. In sexual assault crimes, sperm cells can degrade or can be contaminated by the victim’s epithelial cells, but mtDNA typing of individual sperm cells can resolve this issue.

Single-cell mtDNA analysis has also been applied to individual cells of other human tissues and blood.

Reviews


Yao YG, Kajigaya S and Young NS. Mitochondrial DNA mutations in single human blood cells. Mutat Res. 2015;779:68-77


References

Jayaprakash AD, Benson EK, Gone S, et al. Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA. Nucleic Acids Res. 2015;43:2177-2187

In addition to the nuclear genome, eukaryotic cells also carry a mitochondrial genome, and mtDNA profiling is a useful tool in forensic analysis. Heteroplasmy, the occurrence of multiple mtDNA haplotypes in a cell, can increase the strength of DNA evidence in cases of historical significance. However, since mtDNA makes up less than 1% of total cell DNA, characterizing mtDNA diversity has proved challenging. In this study, the authors developed Mseq, a method for purifying and sequencing mtDNA. They used the MiSeq system to sequence mtDNA from human PBMCs to show that heteroplasmy is maintained stably in individual daughter cells over multiple cell divisions.

Illumina Technology: MiSeq System

For Research Use Only. Not for use in diagnostic procedures.

Forensic casework often focuses on microscopic or trace amounts of biological material left behind at a crime scene. Forensic profiles from these samples are demonstrated with “touch-DNA” evidence, which is understood to be DNA from skin cells transferred to an object through physical contact. In this study, the authors used the Genome Analyzer IIx to obtain transcriptome data from bulk human tissues, as well as from very low amounts of RNA (5-25 pg) from a few cells. By comparing the expression data across samples, they identified 5 mRNA markers highly specific to human skin that could be detected in almost all touch-DNA samples.

Illumina Technology: GAIIx System


Allele-Specific Gene Expression

Diploid organisms have 2 sets of chromosomes, 1 from each parent. Genes can be transcribed from 1 allele (monoallelic expression) or from 2 alleles (biallelic expression). Population sequencing provides a global representation of gene expression, but the expression levels of rare isoforms may be lost. Single-cell sequencing approaches can detect these rare isoforms, as well as changes between monoallelic and biallelic expression.

Compared to established methods such as RNA fluorescence in situ hybridization (FISH), RNA sequencing, PCR, and live-cell imaging, single-cell RNA sequencing provides the most accurate representation of monoallelic or biallelic expression in individual cells within a population.\(^\text{185}\) Combining RNA expression data with SNP data can define specific SNPs that lead to preferential allele expression or silencing. Additionally, it can elucidate their subsequent role in cancer or tissue-specific differentiation.\(^\text{186,187}\)

Mitotic cells accumulate SNPs that can play a role in determining random monoallelic gene expression.\(^\text{186}\)


Reviews


References


Mammalian cells have 2 alleles from which gene transcription can occur, but whether mRNAs are actively transcribed from 1 or both alleles is a subject of intense research. In this study, the authors used the HiSeq 2000 system to perform RNA-Seq of 203 single human primary fibroblasts, to determine allele-specific expression levels. Their data showed that, for the majority of genes in a cell, transcripts were derived from only 1 of 2 alleles. Moreover, genes expressing both alleles in a given cell were rare, and allele-specific expression correlated with cellular transcript levels.

Illumina Technology: Nextera XT DNA Kit, TruSeq RNA Kit, TruSeq DNA Kit, HiSeq 2000 System


In single-cell DNA-Seq, sequence artifacts are introduced by requisite DNA amplification methods, such as MDA and multiple annealing and looping–based amplification cycles (MALBAC). In this study, the authors developed a new statistical method for quantitative assessment of single-cell DNA amplification bias due to WGA. By comparing MDA and MALBAC DNA libraries, they provided a benchmark comparison of single-cell libraries generated by MDA and MALBAC and also identified universal features of genomic coverage bias at the amplicon level. Their statistical models allowed for calibration of allelic bias in single-cell WGA data.

Illumina Technology: MiSeq and HiSeq 2500 Systems


SAMPLE PREPARATION

Isolating individual cells is the first step in single-cell sequencing workflows, and many techniques are available. In addition to well-established methods (including FACS, serial dilution, micropipetting, and LCM), microfluidics and drop-based techniques have increased the throughput of single-cell sequencing workflows, enabling greater accuracy and specificity in single-cell data analysis. This section highlights some techniques used for isolating single cells from suspensions or tissues (Table 1).

190. Wang Y and Navin NE. Advances and applications of single-cell sequencing technologies. Mol Cell. 2015;58:598-609
<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACS</td>
<td>Microdroplets with single cells isolated using electric charge</td>
<td>• Specific immunotagging of cell-surface markers improves accuracy</td>
<td>• Requires specific antibodies/markers</td>
<td>$</td>
</tr>
<tr>
<td>Serial dilution</td>
<td>Serial dilution to 1 cell per well melanogaster</td>
<td>• Simple approach</td>
<td>• Probability of isolating multiple cells</td>
<td>$</td>
</tr>
<tr>
<td>Mouth pipetting</td>
<td>Isolate single cells with glass pipettes</td>
<td>• Simple approach</td>
<td>• Technically difficult</td>
<td>$</td>
</tr>
<tr>
<td>Robotic micromanipulation</td>
<td>Robotic micropipettes isolate single cells</td>
<td>• High accuracy</td>
<td>• Low throughput</td>
<td>$$$</td>
</tr>
<tr>
<td>Microfluidics platforms</td>
<td>Microfluidic chips isolate cells in flow channels</td>
<td>• Isolate cells from small volumes</td>
<td>• Requires uniform cell size</td>
<td>$$$</td>
</tr>
<tr>
<td>Optical tweezers</td>
<td>Dissociated cell suspension</td>
<td>• Focused and controlled cell isolation</td>
<td>• Technically challenging</td>
<td>$$$</td>
</tr>
<tr>
<td>Single nuclei</td>
<td>Isolate nuclei from tissue homogenates and sort by FACS</td>
<td>• Gentle treatment avoids gene expression artifacts</td>
<td>• Cytoplasmic transcripts and small RNAs are not detectable</td>
<td>$</td>
</tr>
<tr>
<td>Nanofilters</td>
<td>Size-exclusion filtration on filters</td>
<td>• Cells selected by size</td>
<td>• Cells can adhere to filters</td>
<td>$</td>
</tr>
<tr>
<td>Mag Sweeper</td>
<td>Rotating magnet with EpCAM antibodies</td>
<td>• Enrichment of rare cells</td>
<td>• Requires markers for isolation</td>
<td>$</td>
</tr>
<tr>
<td>Micro-manipulation</td>
<td>Dissociated cell suspension</td>
<td>• Can isolate diverse cell types from mixed population</td>
<td>• Low throughput</td>
<td>$</td>
</tr>
<tr>
<td>TIVA</td>
<td>Photo-activatable mRNA capture molecule from live single cells</td>
<td>• Compatible with live tissues, retaining single-cell microenvironment</td>
<td>• Low throughput</td>
<td>$$$</td>
</tr>
<tr>
<td>CellSearch</td>
<td>Magnets with antibody-conjugated nanoparticles</td>
<td>• High throughput</td>
<td>• Bias toward isolation markers</td>
<td>$$$</td>
</tr>
<tr>
<td>CellCleeector</td>
<td>Robotic capillary micromanipulator</td>
<td>• High throughput</td>
<td>• Expensive</td>
<td>$$$</td>
</tr>
<tr>
<td>DEP-Array</td>
<td>Microchip with dielectric cages</td>
<td>• High sensitivity enables isolation of rare cells</td>
<td>• Low throughput</td>
<td>$$$</td>
</tr>
<tr>
<td>LCM</td>
<td>Cells are cut from tissue section with laser under microscope</td>
<td>• Preserves spatial context</td>
<td>• Technically challenging</td>
<td>$$$</td>
</tr>
</tbody>
</table>

For Research Use Only. Not for use in diagnostic procedures.
Reviews
Wang Y and Navin NE. Advances and applications of single-cell sequencing technologies. Mol Cell. 2015;58:598-609

References
Since single-cell sequencing methods often involve dissociation of cells and loss of spatial information, methods that retain spatial information in single-cell genomic analysis are critically important. The authors developed a cell-labeling via photobleaching (CLaP) method that combines cellular labeling with single-cell genomics. Individual cells are labeled in culture by laser photobleaching, followed by isolation based on a wide variety of distinguishing characteristics. In this study, the authors used CLaP to tag a number of different cells from lines grown in monolayers. They isolated individual cells using drop-based microfluidics and performed RNA-Seq using the HiSeq 2500 system. The ability to combine spatial information with single-cell genomics makes this method well suited for studying tissue heterogeneity.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2500 System

Determining the genotype and phenotype of individual microbial cells is fundamentally important in understanding microbial evolution. Single-cell sequencing techniques, including WGS, currently allow detection of mutants at high resolution. However, similar approaches for phenotypic analysis are lacking. In this study, the authors present a drop-based microfluidics system that allows the genetic detection of heritable phenotypes in evolving bacterial populations. At various time intervals, they sampled cells and isolated them in 100 nL drops, then monitored growth monitored using a fluorescent protein reporter. The authors used this approach to follow E. coli populations during 30 days of starvation. The data showed that the phenotypic diversity of the E. coli increased with starvation, and single-cell sequencing was able to identify mutations corresponding to each phenotypic class.

Illumina Technology: HiSeq 2500 System

FACS, followed by MDA, is the current standard for single-cell sample processing. Processing cells in individual wells can increase the cost of single-cell sequencing, due to increased costs for reagents, consumables, and equipment for high-throughput liquid handling. To reduce the cost of parallel single-cell sequencing, the authors developed an approach for isolating single cells and preparing DNA libraries in bulk, followed by sorting afterward. They embedded Rhodobacter sphaeroides cells in alginate microspheres and subjected them to MDA. They extracted DNA from individual microspheres and sequenced it using the MiSeq system. This approach has the potential to improve the process for generating sequencing-ready DNA from many individually isolated cells.

Illumina Technology: MiSeq System

In this study, the authors present a new scalable high-density microfluidic platform for solid-phase capture of RNA on glass coverslips or on polymer beads. They trapped single-cell lysates in sealed picoliter microwells capable of printing RNA on glass or capturing RNA on beads. They combined this sample preparation approach with a scalable technology for scRNA-Seq based on CEL-Seq. The technology is relatively inexpensive, with consumable costs of $0.10–$0.20 per cell and is capable of processing hundreds of individual cells in parallel.

**Illumina Technology:** TruSeq RNA-Seq Library Preparation Kit, NextSeq 500 and HiSeq 2500 Systems


scRNA-Seq can profile gene expression over the entire cell transcriptome, but cell isolation typically results in loss of spatial context. *In situ* hybridization is an excellent technique for identifying the location of gene expression, but it is restricted to a fixed number of genes. In this study, the authors present a protocol for *in situ* profiling of gene expression in cells and tissues. In this approach, RNA is converted into crosslinked cDNA amplicons and sequenced manually on a confocal microscope. The approach has the added benefit of enriching for context-specific transcripts over housekeeping/structural genes, while preserving the tissue architecture for transcript localization.

**Illumina Technology:** Nextera XT DNA Sample Preparation Kit, MiSeq System


WGA is a critical component of single-cell sequencing pipelines, and MDA is the most common WGA method in single-cell sequencing. Despite its widespread use, MDA typically produces uneven genome coverage due to amplification bias and the formation of DNA chimeras. To overcome this limitation, the authors developed droplet MDA that minimizes these technical artifacts. They used microfluidics to compartmentalize extracted DNA fragments into 67 pl droplets, where the individual fragments were then amplified using MDA. This approach was validated by sequencing the droplet MDA products of *E. coli* cells, with genome recovery improving to 89%, compared to 59% using traditional MDA.

**Illumina Technology:** Nextera XT DNA Sample Preparation Kit, MiSeq System


DATA ANALYSIS

Single-cell sequencing poses unique challenges for data analysis. Individual mammalian cells contain 50,000–300,000 transcripts, and gene expression values among individual cells can vary significantly. Although several hundred thousand transcripts may be expressed per individual cell, up to 85% of these are present at only 1–100 copies. Therefore, it is critically important in scRNA-Seq to capture low-abundance mRNA transcripts and amplify the synthesized cDNA to ensure that all transcripts are ultimately represented uniformly in the library. Spike-in quantification standards of known abundance can help distinguish technical variability/noise from biologically meaningful gene expression changes. Molecular indexing can also correct for sequencing biases and recent improvements in automated sample handling can reduce technical variability even more.

“Single-cell analysis provides a new venue for bioinformatics, as bulk-cell data analysis methods may not be directly applicable to single-cell data.” - Ylicin et al. 2016

DNA amplification and single-cell DNA-Seq technical artifacts can be reduced by using computational algorithms specifically designed for this purpose. This section highlights some analysis methods used for single-cell sequencing data (Table 2).

Table 2. Summary of Data Analysis Methods for Single-Cell Sequencing

<table>
<thead>
<tr>
<th>Name</th>
<th>Algorithm</th>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daley &amp; Smith[211]</td>
<td>Coverage</td>
<td>DNA-Seq</td>
<td>Estimates gain in coverage with increased sequencing depth from initial shallow sequencing using Bayes Poisson models.</td>
</tr>
<tr>
<td>Varbin[212]</td>
<td>CNV</td>
<td>DNA-Seq</td>
<td>Uses variable bin sizes to call CNVs.</td>
</tr>
<tr>
<td>SNS[213]</td>
<td>CNV</td>
<td>DNA-Seq</td>
<td>Uses variable bin sizes to call copy numbers.</td>
</tr>
<tr>
<td>Xu et al.[214]</td>
<td>CNV</td>
<td>DNA-Seq</td>
<td>Uses a simplified negative binomial distribution to call CNVs.</td>
</tr>
<tr>
<td>siCHILD[215]</td>
<td>Haplotyping &amp; CNV</td>
<td>DNA-Seq</td>
<td>Determines haplotypes, CNV, and segregation origin haplotypes across the genome of a single cell via haplithrism.</td>
</tr>
<tr>
<td>Velvet-SC[216]</td>
<td>Assembly</td>
<td>DNA-Seq</td>
<td>Addresses low-coverage regions by using de Bruijn graphs with a dynamic cut-off.</td>
</tr>
<tr>
<td>SmashCell[218]</td>
<td>Assembly annotation</td>
<td>DNA-Seq</td>
<td>Uses a tree with branches representing different choice of algorithm or parameters, mostly used in metagenomics.</td>
</tr>
<tr>
<td>Kim &amp; Simon[219]</td>
<td>Evolutionary tree</td>
<td>DNA-Seq</td>
<td>Likelihood function for allel dropouts, Bayesian approach for mutation ordering, temporal relationships among mutation sites.</td>
</tr>
<tr>
<td>PyClone[220]</td>
<td>Clonal population</td>
<td>DNA-Seq</td>
<td>A statistical model for inference of clonal population structures in cancers.</td>
</tr>
<tr>
<td>Subramanian &amp; Schwartz[221]</td>
<td>Clonal population</td>
<td>DNA-Seq</td>
<td>Computational approach for learning tumor progression from single-cell sequencing data using k-mer counts.</td>
</tr>
</tbody>
</table>

---


---

For Research Use Only. Not for use in diagnostic procedures.

Single-cell Research
An overview of recent publications featuring Illumina technology

1. Ji et al. [2016] - Allele dropout DNA-Seq
   - Method to control for false negatives from single-cell amplification data due to allele dropout in mutation calling.

2. GRM [2017] - Normalization RNA-Seq
   - Fits polynomial gamma regression model to fragments per kilobase of transcript per million mapped reads (FPKM) data from spike-ins.

3. SAMstr [2018] - Normalization RNA-Seq
   - Uses spike-in controls to normalize and estimate transcript numbers per cell; tolerates variations in sequencing depth.

4. BASICS [2019] - Identifying variable genes RNA-Seq
   - Fully Bayesian approach that jointly models extrinsic spike-in molecules with genes from cells of interest.

5. Brennecke et al. [2020] - Identifying variable genes RNA-Seq
   - Statistical method that allows the user to assess whether observed gene variation provides evidence of high biological variability.

6. Kim et al. [2021] - Identifying variable genes RNA-Seq
   - Uses spike-ins to estimate parameters related to technical noise, allowing for differences in variability across cells.

7. scLVM [2022] - Noise reduction RNA-Seq
   - Single-cell latent variable model estimates proportion of variation associated with hidden factors to identify subpopulations.

8. OEtfinder [2023] - Noise reduction RNA-Seq
   - Uses orthogonal polynomial regression to identify genes with significantly increased expression artifacts in specific capture sites on the Fluidigm C1 platform.

9. PCA/t-SNE [2024] - Subpopulation ID RNA-Seq
   - Linear/nonlinear dimension-reduction approach for unsupervised clustering of cells.

10. ZIFA [2025] - Subpopulation ID RNA-Seq
    - Dimensionality reduction method that models dropout characteristics to improve simulated and biological data sets.

11. Destiny [2026] - Subpopulation ID RNA-Seq
    - Extends diffusion maps to handle zeros and sampling data heterogeneities in single-cell data.

12. SNN-Cliq [2027] - Subpopulation ID RNA-Seq
    - Uses shared nearest-neighbor-based similarity graphs. Partitioning of the graphs automatically identifies subgroups of cells.

13. RaceID [2028] - Subpopulation ID RNA-Seq
    - Two technical noise sources: random sampling (Poissonian) noise and variability due to sequencing efficiency characterization.

14. SCUBA [2029] - Subpopulation ID RNA-Seq
    - Uses k-means to cluster data along a binary tree detailing bifurcation events for time-course data.

15. BackSPIN [2030] - Subpopulation ID RNA-Seq
    - A divisive biclustering method based on sorting points into neighborhoods.

16. PAGODA [2031] - Subpopulation ID RNA-Seq
    - Principal component analysis (PCA) for gene sets to identify those where first PCA exceeds significantly exceeds genome-wide background expectation.

17. MAST [2032] - Differential detection RNA-Seq
    - Two-part generalized linear model characterizing expression heterogeneity by parameterizing stochastic dropout and bimodal expression distributions.

18. SCDE [2033] - Differential detection RNA-Seq
    - Single-cell differential expression uses a separate model for dropouts and a Bayesian model for differential expression.

19. scDD [2034] - Differential detection RNA-Seq
    - Bayesian modeling framework characterizing expression within a biological condition and with differential distributions across conditions.

20. Monocle [2035] - Pseudotemporal ordering RNA-Seq
    - Uses independent component analysis for dimension reduction and minimum spanning tree for cell ordering.

21. Waterfall [2036] - Pseudotemporal ordering RNA-Seq
    - Clustering method for determining the temporal ordering of the expression profiles of individual cells assayed by RNA-Seq.
For Research Use Only. Not for use in diagnostic procedures.

Illumina Technology: HiSeq 2500 and HiSeq X Ten Systems

Authors validated their procedure for accurately analyzing SNVs in single-cell genomics.

- whole-genome sequencing using the HiSeq 2500 and HiSeq X Ten Systems on SCMDA-amplified and nonamplified single cells from these clones and used SCMDA to amplify the single-cell genomic DNA. They performed SCaller. In this study, the authors isolated nonamplified genomic DNA from fibroblast clones. They also isolated single-cell multiple displacement amplification (SCMDA) and an associated single-cell variant-calling algorithm, prone to artifacts that are associated with WGA. To overcome this technical challenge, the authors developed a computational strategy to infer cellular localization by integrating scRNA-Seq data with in situ RNA patterns.

- A computational tool to identify groups of putative oscillating genes and cyclic order of samples for each group.

- A method to assess cell-state hierarchies from single-cell data using a metric to assess cell-to-cell similarities and a graph-building algorithm.

NGS is routinely used to detect SNVs in genomic DNA from tissues, but analyzing SNVs in single cells is prone to artifacts that are associated with WGA. To overcome this technical challenge, the authors developed single-cell multiple displacement amplification (SCMDA) and an associated single-cell variant-calling algorithm, SCaller. In this study, the authors isolated nonamplified genomic DNA from fibroblast clones. They also isolated individual cells from these clones and used SCMDA to amplify the single-cell genomic DNA. They performed whole-genome sequencing using the HiSeq 2500 and HiSeq X Ten Systems on SCMDA-amplified and nonamplified samples and identified SNVs using SCaller. By comparing SNVs from single cells and parent clones, the authors validated their procedure for accurately analyzing SNVs in single-cell genomics.

Illumina Technology: HiSeq 2500 and HiSeq X Ten Systems

References


NGS is routinely used to detect SNVs in genomic DNA from tissues, but analyzing SNVs in single cells is prone to artifacts that are associated with WGA. To overcome this technical challenge, the authors developed single-cell multiple displacement amplification (SCMDA) and an associated single-cell variant-calling algorithm, SCaller. In this study, the authors isolated nonamplified genomic DNA from fibroblast clones. They also isolated individual cells from these clones and used SCMDA to amplify the single-cell genomic DNA. They performed whole-genome sequencing using the HiSeq 2500 and HiSeq X Ten Systems on SCMDA-amplified and nonamplified samples and identified SNVs using SCaller. By comparing SNVs from single cells and parent clones, the authors validated their procedure for accurately analyzing SNVs in single-cell genomics.

Illumina Technology: HiSeq 2500 and HiSeq X Ten Systems

Reviews


Single-cell genomics has led to a number of individual draft genomes for uncultivated microbes; however, MDA artifacts during the amplification step lead to incomplete and uneven coverage. Metagenomic data sets do not suffer the same sequence bias, but the genomic complexity of microbial communities precludes the recovery of draft genomes. In this study, the authors developed a new method for generating population genome assemblies from metagenomic-guided, single-cell amplified genome assembly data. They validated the approach by completing single-cell amplified genomes for Marine Group 1 Thaumarchaeota and SAR324 clade bacterioplankton. The improved method assembly of the SAR324 clade genome revealed the presence of many genes not present in the single-cell amplified genome.

Illumina Technology: TruSeq LT Nano Kit, MiSeq System


scRNA-Seq methods present an unbiased approach for studying complex tissues and diseases. However, the data suffer from high levels of technical noise and strong dependence on expression magnitude. Cell-to-cell differences can prove challenging when clustering cells based on important biological differences. For example, partitioning methods including k-means clustering and a BackSPIN algorithm may classify cells based on cell cycle rather than tissue-specific signaling. The authors introduce pathway and gene set overdispersion analysis (PAGODA) that overcomes this challenge by detecting all significant and potentially overlapping pathways in which measured cells can be classified.

Illumina Technology: HiSeq 2000 System


Modern single-cell sequencing techniques, particularly those involving massively parallel approaches, often result in the isolation of cells that are stressed, broken, or killed. These low-quality cells can lead to data artifacts, and they must be excluded from analysis. In this study, the authors present the first tool for scRNA-Seq that can process and remove low-quality cells in a simple and rigorous way. The analysis pipeline uses a highly-curated set of 20 biologic and technical features that are incorporated into a machine-learning algorithm. The authors validated the approach on CD4+ T cells, bone marrow dendritic cells, and mouse ESCs. The method also defined a new type of low-quality cell that was not detectable visually.

Illumina Technology: HiSeq 2000 System


scRNA-Seq data sets suffer from inherent technical noise that can challenge the identification of cell subpopulations. To overcome this challenge, as well as unknown hidden factors affecting gene expression heterogeneity, the authors developed a model (scLVM) to account for unobserved factors in RNA-Seq data sets, and validated their model using individual mouse ESCs. They also used the HiSeq 2000 system to perform RNA-Seq of individual T cells over the course of naïve T cells differentiating into T-2 cells. They applied the scLVM model to differentiating T-cell RNA-Seq data sets and corrected for cell cycle gene expression. They were able to identify 2 subpopulations of differentiating T cells that were not revealed by using nonlinear PCA or k-means clustering alone.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2000 System


Characterizing constituent cell types is critical for understanding the function of a given organ or tissue. Existing methods for characterizing cell types involve imaging and isolating cells based on specific markers, but this approach is challenging if the cell types are rare, such as CSCs or CTCs. In this study, the authors used the HiSeq 2500 system to perform RNA-Seq on hundreds of randomly selected cells from mouse intestinal organoids. To characterize cell subpopulations within the organoids, they developed RaceCellD, a computational method for identifying rare cell types in complex populations of cells. They validated this algorithm by identifying a single hormone-producing cell type in a population of sampled organoid cells, and they identified Reg4 as a novel marker for these rare enteroendocrine cells. Finally, they used Reg4 to capture these rare cells to investigate their genetic heterogeneity, identifying a number of enteroendocrine lineages.

Illumina Technology: HiSeq 2500 System

scRNA-Seq has the potential to capture oscillation dynamics in populations of individual cells and to discover oscillations missed in bulk sequencing experiments. However, continuous RNA-Seq time-series experiments are not feasible, and synchronization may not be possible for most oscillatory systems. The Monocle computational algorithm was developed previously to address this challenge in scRNA-Seq data by pseudo-temporal ordering of the data from a few different time points. In this study, the authors developed Oscope, a computational method that identifies and characterizes the transcriptional dynamics of oscillating genes using scRNA-Seq data from unsynchronized cells. They validated Oscope by applying the model to various scRNA-Seq Illumina data sets, including human ESCs, and they discovered an oscillatory pattern related to capture-site and output-well positions on the Fluidigm C1 chip.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2500 System

Statistical approaches can help identify and characterize oscillating genes, such as cell cycle genes, in scRNA-Seq data sets.


scRNA-Seq is an established method for discovering novel cell types, understanding regulatory networks, and reconstructing developmental processes. However, scRNA-Seq typically involves dissociating cells from tissues and thus disrupting their native spatial context. To capture spatial context in scRNA-Seq data, the authors developed Seurat, a computational strategy that combines scRNA-Seq with complementary in situ hybridization data for a smaller set of “landmark” genes that guides spatial assignment. They validated Seurat by spatially mapping 851 individual cells from dissociated zebrafish embryos and creating a transcriptome-wide map of spatial patterning. Seurat was able to localize rare subpopulations of cells correctly, and it could map spatially restricted cells as well as those with a more scattered pattern of expression.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2500 System


Somatic stem cells contribute to tissue development and regeneration, but a systematic molecular analysis of stem-cell behavior has proved elusive due to challenges in cellular heterogeneity. In this study, the authors used scRNA-Seq to characterize the developmental dynamics of adult hippocampal qNSCs. They also developed a bioinformatic pipeline, called Waterfall that quantified single-cell expression data along a reconstructed developmental trajectory. The combination of scRNA-Seq and Waterfall identified molecular signatures of adult qNSCs, and defined molecular cascades underlying qNSC activation and neurogenesis.

Illumina Technology: HiSeq 2500 System
An overview of recent publications featuring Illumina technology

For Research Use Only. Not for use in diagnostic procedures.

During the processing of single-cell DNA-Seq data, DNA copy-number aberrations must be differentiated from WGA artifacts. This requirement makes DNA copy-number profiling and haplotyping of single-cell sequencing data challenging. In this study, the authors developed a single-cell genome analysis method that determined haplotypes and copy number across the genome of a single cell—a process called, haplarithmisis. The method deciphered SNP allele fractions of single cells and integrated these data into a computational workflow for imputation of linked disease variants (siCHILD). The authors validated the method by determination of haplotypes carrying disease alleles in single-cell genomes from individual lymphocytes and human blastomeres derived from human IVF embryos.

**Illumina Technology:** TruSeq DNA LT Sample Preparation Kit, HumanCytoSNP-12v2.1 BeadChips, HiSeq 2000/2500 Systems


Before analyzing single-cell DNA-Seq data, DNA copy-number aberrations must be differentiated from WGA artifacts. This requirement makes DNA copy-number profiling and haplotyping of single-cell sequencing data challenging. In this study, the authors developed a single-cell genome analysis method that determined haplotypes and copy number across the genome of a single cell—a process called, haplarithmisis. The method deciphered SNP allele fractions of single cells and integrated these data into a computational workflow for imputation of linked disease variants (siCHILD). The authors validated the method by determination of haplotypes carrying disease alleles in single-cell genomes from individual lymphocytes and human blastomeres derived from human IVF embryos.

**Illumina Technology:** TruSeq DNA LT Sample Preparation Kit, HumanCytoSNP-12v2.1 BeadChips, HiSeq 2000/2500 Systems


In single-cell DNA-Seq, sequence artifacts are introduced by requisite DNA amplification methods, such as MDA and MALBAC. In this study, the authors developed a new statistical method for quantitative assessment of single-cell DNA amplification bias due to WGA. By comparing MDA and MALBAC DNA libraries, they provided a benchmark comparison of single-cell libraries generated by MDA and MALBAC and also identified universal features of genomic coverage bias at the amplicon level. Their statistical models allowed for calibration of allelic bias in single-cell WGA data.

**Illumina Technology:** MiSeq and HiSeq 2500 Systems


Mende DR, Aylward FO, Eppley JM, Nielsen TN and DeLong EF. Improved Environmental Genomes via Integration of Metagenomic and Single-Cell Assemblies. *Front Microbiol.* 2016;7:143


DNA METHODS

DNA replication during mitosis is not perfect, and progressive generations of cells accumulate somatic mutations. Consequently, each cell in our body has a unique genomic signature, which allows the reconstruction of cell-lineage trees with very high precision. These cell-lineage trees can predict the existence of small subpopulations of stem cells. This information is instructive in cancer development as well as in preimplantation and genetic diagnoses.

Single-cell DNA-Seq can identify acquired somatic mutations and CNVs, allowing researchers to trace back lineages of differentiated cells. Single-cell genomics is also an effective approach for characterizing microorganisms that are difficult or impossible to culture in vitro. Advances in single-cell genomics have led to improvements in diagnosing infectious disease outbreaks, understanding antibiotic-resistant strains and food-borne pathogens, and classifying microbial diversity in the environment or in the gut. Newer techniques are using multiplexing and microfluidics platforms to improve the throughput of single-cell DNA-Seq and to lower costs.

This section highlights some single-cell DNA-Seq methods and recent publications demonstrating how Illumina technology is being used in single-cell DNA-Seq techniques. To learn more about Illumina sequencing methods, visit www.illumina.com/techniques/sequencing.html.

“Single-cell sequencing has uncovered the breadth of genomic heterogeneity between cells in a variety of contexts, including somatic aneuploidy in the mammalian brain and intratumor heterogeneity.”

- Vitak et al. 2017


For Research Use Only. Not for use in diagnostic procedures.
Single-cell Research
Reviews


Wang Y and Navin NE. Advances and applications of single-cell sequencing technologies. Mol Cell. 2015;58:598-609

References


One of the challenges in single-cell genomics is accurate CNV analysis. Single-cell genomes must be amplified prior to copy-number analysis, and although there are several different WGA techniques, all suffer from a degree of sequence bias and low amplification fidelity. In this study, the authors developed Linear Amplification via Transposon Insertion (LIANTI), a superior WGA method. In this method, genomic DNA from single cells is fragmented and tagged using a LIANTI transposon, followed by addition of double-strand T7 promoters at both ends of the fragmented strands. These DNAs are then transcribed into genomic RNAs, followed by RNase digestion and second-strand synthesis to generate barcoded and double-stranded LIANTI amplicons. These amplicons are then ready for library prep and subsequent NGS. The authors demonstrated that LIANTI provided accurate CNV detection in individual human BJ cells, with the smallest false-positive rate of any single-cell WGA method.

Illumina Technology: HiSeq 2500/4000 Systems


NGS is routinely used to detect SNVs in genomic DNA from tissues, but analyzing SNVs in single cells is prone to artifacts that are associated with WGA. To overcome this technical challenge, the authors developed single-cell multiple displacement amplification (SCMDA) and an associated single-cell variant-calling algorithm, SCaller. In this study, the authors isolated nonamplified genomic DNA from fibroblast clones. They also isolated individual cells from these clones and used SCMDA to amplify the single-cell genomic DNA. They performed whole-genome sequencing using the HiSeq 2500 and HiSeq X Ten Systems on SCMDA-amplified and nonamplified samples and identified SNVs using SCaller. By comparing SNVs from single cells and parent clones, the authors validated their procedure for accurately analyzing SNVs in single-cell genomics.

Illumina Technology: HiSeq 2500 and HiSeq X Ten Systems


Single-cell sequencing studies have revealed that tumors consist of diverse populations of genetically distinct cells. Single-cell sequencing can also trace cell lineages, as well as identify genetic mutations that might be otherwise hidden in bulk tissue samples. Most human tumor samples are formalin-fixed and paraffin-embedded (FFPE) for clinical diagnostic purposes. While this technique does preserve the sample, formalin fixation chemically alters the genomic DNA and makes it unsuitable for single-cell sequencing. To address this challenge, the authors developed a method for single-cell, whole-genome copy-number profiling of DNA extracted from FFPE tissues. They observed matching copy-number profiles of nuclei isolated from matched FFPE, frozen cancer cell lines, and fresh neoplastic tissue. This approach enables single-cell genomics applications for research or clinical analysis of FFPE samples.

Illumina Technology: HiSeq 2500 System

To fully understand the mechanisms by which the genome, transcriptome, and DNA methylome interact at the single-cell level, these 3 separate methods ideally should be applied to the same individual cell. In this study, the authors report scTrio-Seq, a method that can analyze genomic CNVs, the DNA methylome, and the transcriptome of an individual mammalian cell simultaneously. They used scTrio-Seq in 25 individual hepatocellular carcinoma primary cells to identify 2 subpopulations of cells. They also found that large-scale CNVs can cause proportional changes in RNA expression in subsets of genes, but the CNVs did not affect DNA methylation in the relevant genomic regions.

Illumina Technology: HiSeq 2000/2500 Systems


Tumor cell heterogeneity is known to play a role in disease progression, therapeutic resistance, and metastasis. However, our understanding of tumor heterogeneity is limited, due to a lack of sensitive approaches for interrogating genetic heterogeneity at a genome-wide scale. In this study, the authors developed a DNA amplification method that combined bioinformatic and molecular approaches to enable highly multiplexed single-cell sequencing. They applied this technique to produce genome-wide CNV profiles of up to 100 individual human cancer cells as well as biopsied tissues on a single lane of a HiSeq system. The method enables rapid profiling of thousands of single-cell genomes.

Illumina Technology: HiSeq System


Current WGA methods can be limited by fluctuations in amplification yield, as well as false-positive and false-negative SNV errors. The authors developed an emulsion-based amplification method (eWGA) that can overcome amplification bias and detect SNVs with high accuracy. Single-cell DNA is divided into aqueous droplets in oil where DNA fragments can be amplified to saturation, minimizing the differences in amplification gain among the emulsified fragments. The method is compatible with MDA and can detect CNVs and SNVs in single cells with improved amplification evenness and accuracy.

Illumina Technology: MiSeq and HiSeq 2500 Systems


Multiple-Strand Displacement Amplification

MDA is commonly used for sequencing microbial genomes due to its ability to amplify templates larger than 0.5 Mbp, but it can also be used to study genomes of other sizes. In this method, 3’-blocked random hexamer primers are hybridized to the template, followed by synthesis with Phi 29 polymerase. Phi 29 performs strand-displacement DNA synthesis, allowing for efficient and rapid DNA amplification. Deep sequencing of the amplified DNA allows for accurate representation of reads, while sequencing depth provides better alignment and consensus for sequences (Table 3).

A schematic overview of MDA.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Template can be circular DNA (plasmids, bacterial DNA).</td>
<td></td>
</tr>
<tr>
<td>- Can sequence large templates.</td>
<td></td>
</tr>
<tr>
<td>- Can perform single-cell sequencing or sequencing for samples with very limited starting material.</td>
<td></td>
</tr>
<tr>
<td>- Strong amplification bias. Genome coverage as low as ~6%.</td>
<td></td>
</tr>
<tr>
<td>- PCR biases can underrepresent GC-rich templates.</td>
<td></td>
</tr>
<tr>
<td>- Contaminated reagents can impact results.</td>
<td></td>
</tr>
</tbody>
</table>

References


NGS is routinely used to detect SNVs in genomic DNA from tissues, but analyzing SNVs in single cells is prone to artifacts that are associated with WGA. To overcome this technical challenge, the authors developed single-cell multiple displacement amplification (SCMDA) and an associated single-cell variant-calling algorithm, SCaller. In this study, the authors isolated nonamplified genomic DNA from fibroblast clones. They also isolated individual cells from these clones and used SCMDA to amplify the single-cell genomic DNA. They performed whole-genome sequencing using the HiSeq 2500 and HiSeq X Ten Systems on SCMDA-amplified and nonamplified samples and identified SNVs using SCaller. By comparing SNVs from single cells and parent clones, the authors validated their procedure for accurately analyzing SNVs in single-cell genomics.

Illumina Technology: HiSeq 2500 and HiSeq X Ten Systems


This method is a refinement to SNES and includes the addition of DNA barcoding to allow multiplexing of 48–96 individual cells into single sequencing reactions. Compared to SNES, this new technique has higher throughput and reduced cost. The authors suggest that SNES is more suitable for detecting point mutations and indels at base-pair resolution.

Illumina Technology: HiSeq 2000 System


Despite the great potential of single-cell sequencing methods to advance the understanding of tissue heterogeneity, current single-cell DNA-Seq methods are challenged by technical errors and poor physical coverage data. In this study, the authors developed single-nucleus exome sequencing (SNES), a single-cell DNA-Seq method that combines flow-sorting of G1/0- or G2/M nuclei, time-limited MDA, exome capture using the TruSeq Exome Enrichment Kit, and sequencing on the HiSeq 2000 system. They validated SNES by sorting and sequencing single nuclei from a fibroblast cell line. The method generated 96% coverage of individual cells and demonstrated 92% detection efficiency for SNVs and 85% for indels in single cells.

Illumina Technology: TruSeq Exome Enrichment Kit, HiSeq 2000 System
In this study, the authors compared MDA, MALBAC, and GenomePlex amplification methods in sequencing of individual hippocampal neurons. They amplified genomic DNA from individual hippocampal neurons using 3 different amplification methods, followed by sequencing at shallow depth on a HiSeq 2000 system. Their results showed that single-cell sequencing results from MALBAC and GenomePlex methods were highly reproducible and had high success rates. MALBAC did display significant GC bias, but it was overcome by using bioinformatics tools. Overall, they determined that MALBAC and GenomePlex performed better for detecting CNVs.

Illumina Technology: HiSeq 2000 System

WGA is a critical component of single-cell sequencing pipelines, and MDA is the most common WGA method in single-cell sequencing. Despite its widespread use, MDA typically produces uneven genome coverage due to amplification bias and the formation of DNA chimeras. To overcome this limitation, the authors developed droplet MDA that minimizes these technical artifacts. They used microfluidics to compartmentalize extracted DNA fragments into 67 pl droplets, where the individual fragments were then amplified using MDA. This approach was validated by sequencing the droplet MDA products of E. coli cells, with genome recovery improving to 89%, compared to 59% using traditional MDA.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, MiSeq System

Genome & Transcriptome Sequencing

Genome & transcriptome sequencing (G&T-Seq) is a protocol that can separate and sequence genomic DNA and full-length mRNA from single cells. In this method, single cells are isolated and lysed. RNA is captured using biotinylated oligo(dT) capture primers and separated from DNA using streptavidin-coated magnetic beads. Smart-Seq2 is used to amplify captured RNA on the bead, while MDA is used to amplify DNA. After sequencing, integrating DNA and RNA sequences provides insights into the gene-expression profile of single cells (Table 4).

A schematic overview of G&T-Seq.

Table 4. Advantages and Disadvantages of G&T-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Compatible with any WGA method.</td>
<td>• Physical separation of DNA and RNA can increase the risk of sample loss or contamination.</td>
</tr>
<tr>
<td>• No 3’-end bias in sequence reads because full-length transcripts are captured.</td>
<td>• Physical separation of DNA and RNA increases handling time.</td>
</tr>
<tr>
<td>• Because DNA and RNA are physically separated and amplified independently, there is no need to mask coding sequences during analysis.</td>
<td></td>
</tr>
</tbody>
</table>
References


Multiparameter single-cell sequencing is a powerful tool that has uncovered relationships among genomic, transcriptional, and epigenetic heterogeneity. In this study, the authors developed single-cell methylome & transcriptome sequencing (scM&T-Seq), a multiparameter sequencing method that allows methylome and transcriptome profiling in the same cell. They used the G&T-Seq protocol to purify single-cell DNA that was then subjected to single-cell bisulfite conversion (scBS-Seq). The authors performed scM&T-Seq on 61 mouse ESCs. They found that gene expression levels of many pluripotency factors were negatively associated with DNA methylation. These data demonstrate that epigenetic heterogeneity is an important mechanism of fluctuating pluripotency in ESCs. They also demonstrate that scM&T-Seq can illuminate the poorly understood relationship between transcriptional and DNA-methylation heterogeneity in single cells.

Illumina Technology: Nextera XT Kit, HiSeq 2000 System


Single-cell genomic sequencing has provided insights into cellular heterogeneity, as well as cellular lineage and development. Single-cell transcriptomic sequencing has refined our understanding of cell types and states. In this study, the authors developed G&T-Seq, a method that allows for the separation and subsequent sequencing of genomic DNA and full-length mRNA from single cells. It complements the genomic DNA and mRNA sequencing (DR-Seq) method, but it can be used with any WGA method and also provides full-length transcripts from the same cell. The authors performed G&T-Seq-enabled transcriptome analysis by using a modified Smart-Seq2 protocol and automated the method on a robotic liquid-handling platform. They used the HiSeq platform to sequence numerous single-cell types, including human cancer cells, reversine-treated mouse embryo blastomeres, and iPSC-derived neurons. Notably, G&T-Seq analysis of aneuploid blastomeres demonstrated that chromosomal gains/losses led to increases/losses in chromosome-wide relative gene expression during a single cell division.

Illumina Technology: Nextera XT Kit, MiSeq, HiSeq 2500 and HiSeq X Ten Systems

Multiple Annealing and Looping–Based Amplification Cycles

MALBAC is intended to address some of the shortcomings of MDA. In this method, MALBAC primers randomly anneal to the DNA template. A polymerase with displacement activity at elevated levels amplifies the template, generating semiamplicons. As the amplification and annealing process is repeated, the semiamplicons are amplified into full amplicons that have a 3’ end complementary to the 5’ end. As a result, full-amplicon ends hybridize to form a looped structure that inhibits further amplification of the looped amplicon, while only the semiamplicons and genomic DNA undergo amplification. Deep sequencing full-amplicon sequences allows for accurate representation of reads, while sequencing depth provides improved alignment for consensus sequences (Table 5).

A schematic overview of MALBAC.
Table 5. Advantages and Disadvantages of MALBAC.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Can sequence large templates.</td>
<td>• Polymerase is relatively error-prone compared to Phi 29</td>
</tr>
<tr>
<td>• Can perform single-cell sequencing or sequencing for samples with very limited starting material.</td>
<td>• Temperature-sensitive protocol.</td>
</tr>
<tr>
<td>• Full-amplicon looping inhibits overrepresentation of templates, reducing PCR bias.</td>
<td>• Genome coverage up to ~90%, but some regions of the genome are consistently underrepresented.</td>
</tr>
<tr>
<td>• Can amplify GC-rich regions.</td>
<td></td>
</tr>
<tr>
<td>• Uniform genome coverage.</td>
<td></td>
</tr>
<tr>
<td>• Lower allele drop-out rates compared to MDA.</td>
<td></td>
</tr>
</tbody>
</table>

References


In this study, the authors compared MDA, MALBAC, and GenomePlex amplification methods in sequencing of individual hippocampal neurons. They amplified genomic DNA from individual hippocampal neurons using 3 different amplification methods, followed by sequencing at shallow depth on a HiSeq 2000 system. Their results showed that single-cell sequencing results from MALBAC and GenomePlex methods were highly reproducible and had high success rates. MALBAC did display significant GC bias, but it was overcome by using bioinformatics tools. Overall, they determined that MALBAC and GenomePlex performed better for detecting CNVs.

Illumina Technology: HiSeq 2000


NGS methods have improved the precision of PGS/PGD. Although the precision has been limited by false-positive and false-negative SNVs, linkage analysis can overcome this challenge. In this study, the authors developed MARSALA, a method that combines NGS using the HiSeq platform with single-cell WGA. The method allows for embryo diagnosis with a single-molecule precision and significantly reduces false-positive and false-negative errors. This is the first integrated NGS-based PGD procedure that simultaneously detects disease-causing mutations and chromosome abnormalities and performs linkage analyses.

Illumina Technology: HiSeq 2500 System


Genomic DNA and mRNA Sequencing

DR-Seq studies the genomic and transcriptomic relationship of single cells via sequencing. Nucleic acid amplification prior to physical separation reduces sample loss and the risk of contamination. DR-Seq involves multiple amplification steps, including the quasilinear amplification technique similar to MALBAC.

First, mRNAs are reverse-transcribed from lysed single cells using poly(dT) primers with Ad-1x adapters, producing single-stranded cDNA (sscDNA). The Ad-1x adapter sequence contains cell-identifying barcodes, 5’ Illumina adapters, and a T7 promoter. Next, both gDNA and sscDNA are amplified simultaneously via quasilinear WGA with Ad-2 primers. These primers are similar to MALBAC adapters, containing 8 random nucleotides for random priming followed by a constant 27-nucleotide
tag at the 5’ end. Products of this amplification step are split in halves. One half is prepared for genome sequencing, in which gDNA are PCR-amplified and “liberated” of their Ad-2 adapters before DNA library prep and sequencing. The other half is prepared for transcriptome sequencing, whereby second strands are synthesized for the cDNAs and amplified by in vitro transcription. The resulting RNA products are produced only from cDNA fragments flanked with Ad-1x and Ad-2, omitting amplification of the gDNA fragments. The RNA library is prepared for sequencing following the Illumina small-RNA protocol. Sequencing gDNA and mRNA from the same cell preserves information between the genome and its expression levels (Table 6).

A schematic overview of DR-Seq.

![Diagram of DR-Seq process]

Table 6. Advantages and Disadvantages of DR-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Interrogates genomic and transcriptomic behavior from a single cell.</td>
<td>• Manual single-cell isolation prevents high-throughput adaptation.</td>
</tr>
<tr>
<td>• Amplification prior to separation reduces sample loss and contamination.</td>
<td>• Quasilinear amplification is temperature-sensitive.</td>
</tr>
<tr>
<td>• Length-based identifier used to remove duplicate reads.</td>
<td>• RNA reads are 3’-end-biased.</td>
</tr>
<tr>
<td>• Quasilinear amplification reduces PCR bias.</td>
<td></td>
</tr>
</tbody>
</table>

Reference


Single-cell genomics and transcriptomics are promising tools for quantifying genetic and expression variability among individual cells. In this study, the authors describe DR-Seq, a method to quantify the genome and transcriptome of the same cell simultaneously. DR-Seq does not require physical separation of nucleic acids before amplification, which helps to minimize the chances for sample loss or contamination. The authors amplified gDNA and cDNA from mouse ESCs, subsequently divided the nucleic acids for further amplification and library construction, and sequenced both libraries using a HiSeq 2500 system. They demonstrated that genes with high cell-to-cell variability in transcript numbers have low CNVs, and vice versa.

Illumina Technology: HiSeq 2500 System

EPIGENOMICS METHODS

Epigenetics—the mechanisms of temporal and spatial control of gene activity, independent of DNA sequence—plays a crucial role in embryogenesis, differentiation, lineage specification, and cancer evolution.\textsuperscript{285,286} During embryogenesis, differentiating cells acquire epigenetic alterations\textsuperscript{287} that subsequently influence RNA expression and cellular phenotype.\textsuperscript{288,289} In differentiated somatic cells, as well as in stem cells, epigenetic markers can be regulated by lifestyle, environmental factors, chemical exposure, stress, and other factors.\textsuperscript{290} Epigenetic changes play a role in a number of diseases, including cancer, neurodegenerative diseases, cardiovascular disease, and respiratory disease.\textsuperscript{291}

“There are today we can probe the majority of epigenetic dimensions with single-cell resolution.”

- Kelsey, Stegle, and Reik 2017

Bulk sequencing of tissues lacks the resolution required to understand how genotypically identical individual cells develop unique phenotypes as a result of unique spatial localization and temporal order. Single-cell epigenomics techniques, including DNA methylation and chromatin immunoprecipitation sequencing (ChIP-Seq), can be combined with RNA expression and SNP data to identify the mechanistic role of epigenetics in gene regulation precisely.\textsuperscript{292} Recently, massively parallel sequencing techniques have been developed to analyze epigenomics in thousands of individual cells, allowing us to understand epigenomic heterogeneity at unprecedented resolution.\textsuperscript{293,294} Further, the development of single-cell multiparameter methods has enabled simultaneous profiling of genomic and transcriptomic changes in individual cells,\textsuperscript{295,296} with a recent study even demonstrating simultaneous single-cell profiling of genomic, transcriptomic, and epigenomic (triple-omics) changes within individual cells.\textsuperscript{297}

This section highlights some single-cell epigenomics sequencing methods and recent publications demonstrating how Illumina technology is being used in single-cell epigenomics techniques. To learn more about Illumina sequencing methods, visit www.illumina.com/techniques/sequencing.html.

\begin{thebibliography}{99}
\item Alegria-Torres JA, Baccarelli A and Boltlat V. Epigenetics and lifestyle. Epigenomics. 2011;3:267-277
\item Hyun BR, McElwee JL and Soloway PD. Single molecule and single cell epigenomics. Methods. 2015;72:41-50
\item Greeneleaf WJ. Assaying the epigenome in limited numbers of cells. Methods. 2015;72:51-56
\end{thebibliography}
The dynamic composition of chromatin during different stages of the cell cycle, or from one cell type to another, is regulated through multiple epigenetic mechanisms.

Reviews
Greenleaf WJ. Assaying the epigenome in limited numbers of cells. Methods. 2015;72:51-56

References
Stevens TJ, Lando D, Basu S, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature. 2017;544:59-64
Understanding cell structure has typically been facilitated through cell biology studies involving microscopy imaging techniques. In the case of the cell nucleus, these cell biology data have been enhanced using chromosome conformation capture epigenetic methods, like 3C and Hi-C. In these epigenomic methods, DNA sequences that are in close physical proximity within the nucleus are digested, ligated, and analyzed by NGS. To more fully resolve the 3D structure of the mammalian genome, the authors combined imaging with improved single-cell Hi-C. Further, they show that integrating the epigenomic data with transcriptomic (RNA-Seq) and ChIP-Seq data provides finer detail regarding the organization of regulated genes. Their data show that structures of loops and topological-associated domains (TADs) are highly variable from cell to cell. Conversely, active enhancer and promoter structure is consistent genome-wide in every cell.

Illumina Technology: Nextera XT DNA Library Preparation Kit, MiSeq and HiSeq Systems

To fully understand the mechanisms by which the genome, transcriptome, and DNA methylome interact at the single-cell level, these 3 separate methods ideally should be applied to the same individual cell. In this study, the authors report scTrio-Seq, a method that can analyze genomic CNVs, the DNA methylome, and the transcriptome of an individual mammalian cell simultaneously. They used scTrio-Seq in 25 individual hepatocellular carcinoma primary cells to identify 2 subpopulations of cells. They also found that large-scale CNVs can cause proportional changes in RNA expression in subsets of genes, but the CNVs did not affect DNA methylation in the relevant genomic regions.

Illumina Technology: HiSeq 2000/2500 Systems


Chromothripsis is a new mutational phenomenon in cancer and congenital disorders. In this process, extensive DNA rearrangements and oscillating patterns of DNA copy number are restricted to one or a few chromosomes. The mechanism underlying chromothripsis is not known, but it has been proposed to involve physical isolation of chromosomes in micronuclei. In this study, the authors combined single-cell genome sequencing with live cell imaging to demonstrate that micronucleus formation can lead to a spectrum of genomic rearrangements, including chromothripsis. Specifically, the mechanism for chromothripsis appears to involve the fragmentation and subsequent reassembly of single chromatids within single micronuclei.

Illumina Technology: HiSeq and MiSeq Systems


Wu X, Houe A, Suzuki T and Zhang Y. Simultaneous mapping of active DNA demethylation and sister chromatid exchange in single cells. Genes Dev. 2017;31:511-523


Single-Cell Assay for Transposase-Accessible Chromatin Using Sequencing

The single-cell assay for transposase-accessible chromatin using sequencing (scATAC-Seq) is a protocol for mapping accessible regions in the genome of single cells by combining microfluidics and Tn5 tagmentation. In scATAC-Seq, cell suspensions are loaded into a microfluidics system and sorted individually. Here, cells undergo lysis, and Tn5 transposase tags open chromatin regions with sequencing barcodes. Tagged DNA fragments are purified and amplified with cell-specific barcodes. Libraries from all single cells are then pooled, and deep sequencing provides base-pair resolution of nucleosome-free regions in the genome (Table 7).

References


Methods to investigate genome-wide DNA accessibility have revealed substantial variation in regulatory regions across a wide diversity of cells. In order to test whether this heterogeneity exists within individual cells, the authors developed scATAC-Seq. In this method, individual cells are captured and assayed using a microfluidics platform. After PCR amplification and barcoding, the authors sequenced these single-cell libraries using HiSeq and NextSeq systems. They generated DNA accessibility maps from 254 GM12878 lymphoblastoid cells, as well as other cell lines. Their data demonstrate single-cell epigenetic heterogeneity.

Illumina Technology: Nextera DNA Sample Prep Kit, HiSeq and NextSeq Systems


This study used combinatorial indexing to measure chromatin accessibility in thousands of single cells. The authors isolated nuclei and tagged them in bulk with Tn5 transposases, in each of many wells. Next, they pooled these barcoded nuclei, diluted them, and redistributed them to a second set of wells where a second barcode was introduced using PCR. The authors integrated this combinatorial indexing with scATAC-Seq298 to measure chromatin accessibility for more than 15,000 human and mouse single cells. They sequenced the scATAC-Seq libraries on the MiSeq system, and their data identified relevant differences in chromatin accessibility between cell types.

Illumina Technology: MiSeq and NextSeq Systems

Table 7. Advantages and Disadvantages of scATAC-Seq

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Deep sequencing of open chromatin regions in single cells.</td>
<td>• Lower throughput than combinatorial indexing scATAC-Seq (maximum of 96 cells in parallel).298</td>
</tr>
<tr>
<td>• High average reads per cell (70,000 reads) compared to combinatorial indexing scATAC-Seq.</td>
<td></td>
</tr>
<tr>
<td>• Capture of each viable cell individually confirmed through microscopy in the microfluidics device.</td>
<td></td>
</tr>
</tbody>
</table>

For Research Use Only. Not for use in diagnostic procedures.
Single-Cell Bisulfite Sequencing/Single-Cell Whole-Genome Bisulfite Sequencing

Single-cell bisulfite sequencing (scBS-Seq) or single-cell whole-genome bisulfite sequencing (scWGBS) are versions of the well-established bisulfite sequencing (BS-Seq) and whole-genome bisulfite sequencing (WGBS) post-bisulfite adapter-tagging (PBAT) protocols, modified to detect methylated cytosines in genomic DNA from single cells. In this method, after single cells are isolated, genomic DNA is treated with sodium bisulfite, which fragments the DNA. The converted DNA then undergoes random priming several times and is PCR-amplified for sequencing. Deep sequencing provides single-nucleotide resolution of methylated cytosines from single cells (Table 8).

A schematic overview of scBS/scWGBS.

Table 8. Advantages and Disadvantages of scBS/scWGBS.

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Covers CpG and non-CpG methylation throughout the genome at single-base resolution.</td>
<td>• Bisulfite converts unmethylated cytosines to thymidines, reducing sequence complexity, which can make it difficult to create alignments.</td>
</tr>
<tr>
<td>• Covers 5mC in dense, less dense, and repeat regions</td>
<td>• SNPs where a cytosine is converted to thymidine will be missed upon bisulfite conversion.</td>
</tr>
<tr>
<td></td>
<td>• Bisulfite conversion does not distinguish between 5mC and 5hmC.</td>
</tr>
</tbody>
</table>

References


In this study, the authors developed a technique to characterize the epigenomic diversity within individual neuronal cells. This technique, single-nucleotide methylcytosine sequencing (snmC-seq), takes advantage of the significant amount of 5-methylcytosine found in neurons. The authors isolated 3400 individual neurons from mouse cortex and collected nuclei by FACS. They treated the nuclear DNA treated with bisulfite to convert unmethylated cytosine into uracil, while 5-methyl- and 5-hydroxymethylcytosines remained unaffected. Finally, they pooled, amplified, and sequenced the samples using the HiSeq 4000 System. The authors used snmC-seq to generate 6000 methylomes from the isolated neuronal nuclei, and their data identified 16 mouse neuronal subtypes in mouse brain. They also performed snmC-seq on isolated nuclei from human cortex samples, and they identified 21 human neuronal subtypes. The single-cell methylome technique thus expanded the taxonomy of brain cell diversity. Moreover, it identified regulatory elements associated with neuronal diversity.

Illumina Technology: TruSeq Methylation kit, HiSeq 4000 System


WGBS is currently the most popular method for methylation mapping. In this study, the authors describe a WGBS method modified for single cells (scWGBS). They sorted cells using FACS and bisulfite-converted the DNA directly in lysed cells. Next, they prepared single-strand libraries and sequenced them using the HiSeq 2000/2500 system. They validated the method using more than 250 samples in 3 in vitro models of cellular differentiation, including the K562 erythroleukemia-derived cell line, the HL60 cell line, and induced mouse ESCs. In all 3 models, scWGBS detailed characteristic patterns of epigenome remodeling and cell-to-cell heterogeneity.

Illumina Technology: HiSeq 2000/2500 Systems
An overview of recent publications featuring Illumina technology

Single-Cell Methylome & Transcriptome Sequencing

scM&T-Seq allows parallel analysis of both epigenetic and gene expression patterns from single cells using Smart-Seq2 and scBS-Seq. scM&T-Seq is built upon G&T-Seq, but instead of using MDA for DNA sequencing, it uses scBS-Seq to interrogate DNA methylation patterns.

First, single cells are isolated and individually lysed. Then, mRNAs are isolated using streptavidin-coupled mRNA capture primers, physically separating them from DNA strands. Smart-Seq2 is used to generate cDNA libraries from the mRNA, which involves reverse transcription with template switching and tagmentation. DNA libraries are prepared via scBS-Seq, which involves bisulfite conversion of DNA strands to identify methylated cytosines. Both libraries are now ready for sequencing (Table 9).

A schematic overview of scM&T-Seq.

Table 9. Advantages and Disadvantages of scM&T-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Investigates links between epigenetic and transcriptional heterogeneity in single cells.</td>
<td>• Smart-Seq2 is not strand-specific and applicable to only poly(A)+ RNA.</td>
</tr>
<tr>
<td>• Because DNA and RNA are physically separated and amplified independently, there is no need to mask coding sequences during analysis.</td>
<td>• Does not distinguish between 5mC and 5hmC.</td>
</tr>
</tbody>
</table>

References


Multiparameter single-cell sequencing is a powerful tool that has uncovered relationships among genomic, transcriptional, and epigenetic heterogeneity. In this study, the authors developed scM&T-Seq, a multiparameter sequencing method that allows methylome and transcriptome profiling in the same cell. They used the G&T-Seq protocol to purify single-cell DNA that was then subjected to single-cell bisulfite conversion (scBS-Seq). Using the HiSeq 2000 system, the authors performed scM&T-Seq on 61 mouse ESCs. They found that gene expression levels of many pluripotency factors were negatively associated with DNA methylation. These data demonstrate that epigenetic heterogeneity is an important mechanism of fluctuating pluripotency in ESCs. They also demonstrate that scM&T-Seq can illuminate the poorly understood relationship between transcriptional and DNA-methylation heterogeneity in single cells.

Illumina Technology: Nextera XT Kit, HiSeq 2000 System


In this study, the authors developed a method that simultaneously profiles the methylome and the transcriptome of the same individual cell (scMT-Seq). The method is very similar to scM&T-Seq, except that scMT-Seq uses single-cell reduced-representation bisulfite sequencing (scRRBS) for methylome analysis. The authors validated this method by simultaneously profiling the transcriptome and DNA methylome in individual sensory neurons of the dorsal root ganglion (DRG). Their data identified transcriptome and DNA methylome heterogeneity in DRG neurons. They also found that gene methylation and expression are positively correlated, but only for those genes that contain CpG island promoters.

Illumina Technology: MiSeq and HiSeq 2500 Systems

For Research Use Only. Not for use in diagnostic procedures.
Single-Cell Reduced-Representation Bisulfite Sequencing

scRRBS uses one or multiple restriction enzymes on genomic DNA to produce sequence-specific fragmentation. The fragmented genomic DNA is then treated with bisulfite and sequenced. It is the method of choice to study specific regions of interest. It is particularly effective where methylation is high, such as in promoters and repeat regions (Table 10).

Table 10. Advantages and Disadvantages of scRRBS.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Provides genome-wide coverage of CpGs in islands at single-base resolution.</td>
<td>• Restriction enzymes cut at specific sites, providing biased sequence selection.</td>
</tr>
<tr>
<td>• Covers areas dense in CpG methylation</td>
<td>• Measures 10%-15% of all CpGs in the genome.</td>
</tr>
<tr>
<td></td>
<td>• Cannot distinguish between 5mC and 5hmC.</td>
</tr>
<tr>
<td></td>
<td>• Does not cover non-CpG areas, genome-wide CpGs, and CpGs in areas without the enzyme restriction site.</td>
</tr>
</tbody>
</table>

Reference

Single-Cell Chromatin Immunoprecipitation Sequencing

scChIP-Seq is a well-established method to map specific protein-binding sites. In this method, DNA-protein complexes are crosslinked in vivo. Samples are then fragmented and treated with an exonuclease to trim unbound oligonucleotides. Protein-specific antibodies are used to immunoprecipitate the DNA-protein complex. The DNA is extracted and sequenced, giving high-resolution sequences of the protein-binding sites (Table 11).

A schematic overview of scChIP-Seq.

Table 11. Advantages and Disadvantages of scChIP-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Provides base-pair resolution of protein-binding sites.</td>
<td>• Nonspecific antibodies can dilute the pool of DNA-protein complexes of interest.</td>
</tr>
<tr>
<td>• Can map specific regulatory factors or proteins.</td>
<td>• The target protein must be known and be able to raise an antibody.</td>
</tr>
<tr>
<td>• The use of exonuclease eliminates contamination by unbound DNA.</td>
<td></td>
</tr>
</tbody>
</table>

Reference


In this study, the authors combined microfluidics, DNA barcoding, and sequencing to analyze chromatin data at single-cell resolution. They validated the technology by assaying thousands of individual cells, followed by deconvolution of a mixture of ESCs, fibroblasts, and hematopoietic progenitors into chromatin state maps for each cell type. Although the data from each single cell covered only 1000 reads, the ability to assay thousands of individual cells allowed them to identify a spectrum of subpopulations of ESCs, defined by differences in chromatin signatures of pluripotency and differentiation timing. The method revealed aspects of epigenetic heterogeneity not captured by scRNA-Seq alone.

Illumina Technology: HiSeq 2500 System
Chromatin Conformation Capture Sequencing

Chromatin conformation capture sequencing (Hi-C\textsuperscript{307} or 3C-Seq\textsuperscript{308}) is used to analyze chromatin interactions. In this method, DNA-protein complexes are crosslinked using formaldehyde. The sample is fragmented, and the DNA is ligated and digested. The resulting DNA fragments are PCR-amplified and sequenced. Deep sequencing provides base-pair resolution of ligated fragments (Table 12).

![Chromatin conformation capture (3C and Hi-C)](image)

A schematic overview of Hi-C/3C-Seq.

Table 12. Advantages and Disadvantages of Hi-C/3C-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Allows detection of long-range DNA interactions.</td>
<td>• Detection may result from random chromosomal collisions.</td>
</tr>
<tr>
<td>• High-throughput method.</td>
<td>• 3C PCR is difficult, and it requires careful controls and experimental design.</td>
</tr>
<tr>
<td></td>
<td>• Needs further confirmation of interaction.</td>
</tr>
<tr>
<td></td>
<td>• Requires large amounts of starting material due to multiple steps.</td>
</tr>
</tbody>
</table>

References


Table 12. Advantages and Disadvantages of Hi-C/3C-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Allows detection of long-range DNA interactions.</td>
<td>• Detection may result from random chromosomal collisions.</td>
</tr>
<tr>
<td>• High-throughput method.</td>
<td>• 3C PCR is difficult, and it requires careful controls and experimental design.</td>
</tr>
<tr>
<td></td>
<td>• Needs further confirmation of interaction.</td>
</tr>
<tr>
<td></td>
<td>• Requires large amounts of starting material due to multiple steps.</td>
</tr>
</tbody>
</table>

References


During interphase, mammalian chromosomes interact with the nuclear lamina (NL) through structures termed lamina-associated domains (LADs). In this study, the authors developed a modified DNA adenine methyltransferase identification (DamID) method to characterize genome-wide mapping of chromosome-NL interactions in 118 individual KBM7 cells. The data showed that 15% of the genome contacted the NL in most of the individual cells analyzed, and that this contact frequency was locus-specific. Chromosome-LN contact sites that were stable across cells were poor in genes, compared to those sites that were more variable across cells, suggesting that these sites may serve a structural rather than epigenetic role. Hi-C analysis also showed that loci with intrachromosomally coordinated NL contacts were in close proximity in the nuclear space.

\textit{Illumina Technology: HiSeq 2000/2500 Systems}


Hi-C provides pairwise information on genomic regions that are within spatial proximity of each other in the nucleus. In this study, the authors modified single-cell Hi-C with in-nucleus ligation, in order to characterize the thousands of chromatin interactions that occur in individual cells. This modification allows for magnetic-bead capture of labeled, crosslinked ligation junctions and PCR amplification of single-cell Hi-C libraries. The authors validated this approach by performing single-cell Hi-C in individual mouse T helper 1 (T\textsubscript{H}1) cells. The resulting T\textsubscript{H}1 interactome maps provided information on nuclear genome organization and chromosome structure.

\textit{Illumina Technology: GA\textsubscript{x} System}
Droplet-Based Chromatin Immunoprecipitation Sequencing

Single-cell droplet-based chromatin immunoprecipitation sequencing (Drop-ChIP-Seq) analyzes the chromatin states of single cells by utilizing microfluidics, unique molecular barcodes, and NGS. First, single cells are isolated into droplets containing lysis buffer and MNase, and then fused with another droplet carrying distinct oligonucleotides. These oligonucleotides hold the sequences for cell-specific barcodes, sequencing adapter, and restriction sites. DNA ligase is also fused with the droplet to complete the tagging process. Next, carrier chromatin is introduced into the pooled droplets, followed by standard ChIP-Seq procedures (Table 13).

A schematic overview of Drop-ChIP-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Analyzes chromatin states from single cells in a highly parallel manner.</td>
<td>• Requires a large number of sample cells.</td>
</tr>
<tr>
<td>• Unique molecular barcoding reduces the risk posed by nonspecific antibodies.</td>
<td></td>
</tr>
</tbody>
</table>

Reference


In this study, the authors combined microfluidics, DNA barcoding, and sequencing to analyze chromatin data at single-cell resolution. They validated the technology by assaying thousands of individual cells, followed by deconvolution of a mixture of ESCs, fibroblasts, and hematopoietic progenitors into chromatin state maps for each cell type. Although the data from each single cell covered only 1000 reads, the ability to assay thousands of individual cells allowed them to identify a spectrum of subpopulations of ESCs, defined by differences in chromatin signatures of pluripotency and differentiation timing. The method revealed aspects of epigenetic heterogeneity not captured by scRNA-Seq alone.

Illumina Technology: HiSeq 2500 System
RNA METHODS

Low-level RNA detection refers to both detection of rare RNA molecules in a cell-free environment (such as circulating tumor RNA) and the expression patterns of single cells. Tissues consist of a multitude of different cell types, each with a distinctly different set of functions. Even within a single cell type, the transcriptomes are highly dynamic and reflect temporal, spatial, and cell cycle–dependent changes. Cell harvesting, handling, and technical issues with sensitivity and bias during amplification add additional levels of complexity. To resolve this multitiered complexity would require analyzing many thousands of cells. The use of unique barcodes has greatly increased the number of samples that can be multiplexed and pooled at little to no decrease in reads associated with each sample. Recent improvements in cell capture and sample preparation will provide more information, faster, and at lower cost. This development promises to expand our understanding of cell function fundamentally, with significant implications for research and human health.

“Single-cell RNA-seq has become instrumental for interrogating cell types, dynamic states, and functional processes in complex tissues.” - Habib et al. 2017

Recently, massively parallel sequencing techniques have been developed to analyze gene expression levels in thousands of individual cells, allowing us to understand transcriptional heterogeneity at unprecedented resolution. Further, the development of single-cell multiparameter methods has enabled simultaneous profiling of transcriptomic and epigenomic changes in individual cells. A recent study even demonstrated simultaneous single-cell profiling of genomic, transcriptomic, and epigenomic (triple-omics) changes within individual cells.

This section highlights some scRNA-Seq methods and recent publications demonstrating how Illumina technology is being used in scRNA-Seq. To learn more about Illumina sequencing methods, visit www.illumina.com/techniques/sequencing.html.

A

B

Single-cell transcriptomics approaches can characterize gene expression in individual cells of a tissue or organ.
Reviews


Faridani OR and Sandberg R. Putting cells in their place. Nat Biotechnol. 2015;33:490-491


Trapnell C. Defining cell types and states with single-cell genomics. Genome Res. 2015;25:1491-1498


Wang Y and Navin NE. Advances and applications of single-cell sequencing technologies. Mol Cell. 2015;58:598-609

References


Pooled CRISPR screening has become a powerful tool in functional genomics, and it has been used to identify novel biological mechanisms, drug targets, and biomarkers. In this approach, cells are transduced with a library of guide RNA (gRNA) vectors, and gRNA distribution is measured before and after selective challenge. These screening approaches are well suited for relatively simple biological outputs like cell survival, cell proliferation, or single gene expression. In this study, the authors developed CRISPR droplet screening (CROP-seq) to enable more complex CRISPR-based gene expression screens. They transduced a gRNA plasmid library into Jurkat cells to express 1 detectable gRNA per cell, then challenged the cells with anti-CD3 and anti-CD28 antibodies to activate T cells. They performed Drop-seq on the Jurkat cells to identify single-cell transcriptomes associated with individual gRNAs and identified genes known to be important for T cell activation. This approach links CRISPR screening with transcriptome responses in thousands of individual cells and should enable CRISPR screens with single-cell transcriptome resolution.

Illumina Technology: BaseSpace™, HiSeq 2500 System


Single-cell RNA-Seq has become an indispensable technique for characterizing cell types, states, and function in complex tissues. Sequencing individual nuclei has similarly proved useful in cell-type discrimination, particularly for tissues where single cells are difficult to dissociate. Current approaches for single-nucleus sequencing rely on the use of FACS or C1 microfluidics to sort individual nuclei into multwell plates for downstream analysis. In this study, the authors developed DroNc-seq, a method that combines single-nucleus sequencing with the microfluidics-based Drop-seq method. The authors modified Drop-seq to incorporate single nuclei within microfluidic droplets, as opposed to single individual cells. They performed massively parallel RNA-Seq for approximate 39,000 individual nuclei from mouse and human brain, at low cost and with similar sensitivity to Drop-seq.

Illumina Technology: NextSeq 500 System


To fully understand the mechanisms by which the genome, transcriptome, and DNA methylome interact at the single-cell level, these 3 separate methods ideally should be applied to the same individual cell. In this study, the authors report scTrio-Seq, a method that can analyze genomic CNVs, the DNA methylome, and the transcriptome of an individual mamalian cell simultaneously. They used scTrio-Seq in 25 individual hepatocellular carcinoma primary cells to identify 2 subpopulations of cells. They also found that large-scale CNVs can cause proportional changes in RNA expression in subsets of genes, but the CNVs did not affect DNA methylation in the relevant genomic regions.

Illumina Technology: HiSeq 2000/2500 Systems

For Research Use Only. Not for use in diagnostic procedures.

An overview of recent publications featuring Illumina technology.

In this study, the authors developed CRISP-seq, an NGS method that integrates massively parallel single-cell RNA-Seq with pooled CRISPR screens. They used this technique to characterize immune development and signaling pathways in thousands of individual mouse myeloid cells. By profiling the genomic perturbations and transcriptome in individual cells, they identified multiple interacting factors that act in concert to regulate immune signaling pathways. In particular, they identified specific sets of genes whose expression was regulated by Stat1 and Rela in myeloid cells.

Illumina Technology: MiSeq and NextSeq 500 Systems


scRNA-Seq can profile gene expression over the entire cell transcriptome, but cell isolation typically results in loss of spatial context. In situ hybridization is an excellent technique for identifying the location of gene expression, but it is restricted to a fixed number of genes. In this study, the authors present a protocol for in situ profiling of gene expression in cells. In this approach, RNA is converted into crosslinked cDNA amplicons and sequenced manually on a confocal microscope. The approach has the added benefit of enriching for context-specific transcripts over housekeeping/structural genes, while preserving the tissue architecture for transcript localization.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, MiSeq System


Individual mammalian cells display a wide diversity in cellular size and volume, yet they maintain the same absolute DNA content. Therefore, it is likely that compensatory mechanisms exist to maintain a constant concentration of gene expression products, despite differences in DNA concentration. In this study, the authors used single-molecule counting and single-cell image analysis to demonstrate that individual human primary foreskin fibroblasts globally control transcription to compensate for variability in the ratio of DNA to cellular content. They performed scRNA-Seq using the NextSeq 500 system and found that ubiquitously expressed “housekeeping” genes exhibited lower levels of expression noise than other genes.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, NextSeq 500 System


Designed Primer–Based RNA Sequencing

Designed primer–based RNA sequencing (DP-Seq) is a method that amplifies mRNA from limited starting material, as low as 50 pg. In this method, a specific set of heptamer primers is designed. Enriched poly(A)-selected mRNA undergoes first-strand cDNA synthesis. Designed primers are then hybridized to first-strand cDNA, followed by second-strand synthesis and PCR. Deep sequencing of amplified DNA allows for accurate detection of specific mRNA expression at the single-cell level (Table 14).

A schematic overview of DP-Seq.

Table 14. Advantages and Disadvantages of DP-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Uses as little as 50 pg of starting material.</td>
<td>• The sequences of the target areas must be known to design the heptamers.</td>
</tr>
<tr>
<td>• Low transcript-length bias.</td>
<td>• Exponential amplification during PCR can lead to primer-dimers and spurious PCR products.</td>
</tr>
<tr>
<td></td>
<td>• Some read-length bias.</td>
</tr>
</tbody>
</table>

Reference

Single-Cell Universal Poly(A)-Independent RNA Sequencing

Single-cell universal poly(A)-independent RNA sequencing (SUPeR-Seq) sequences non-poly(A) and poly(A)⁺ RNAs from single cells. It is designed particularly for mapping circular RNA (circRNA) species. RNA samples from lysed single cells are annealed to random primers with universal anchor sequences (AnchorX-T15N6) and reverse-transcribed to generate the first strand of cDNA. Unreacted primers are digested to avoid primer dimers, prior to the addition of a poly(A) tract to the 3’ end of the cDNA. This is done by introducing dATPs and ddATPs in a 100:1 ratio, respectively. A second set of random primers, also with a universal anchor sequence (AnchorY-T24) anneals to the newly synthesized poly(A) tract. A second cDNA strand is generated by reverse transcription, and the cDNA molecules are purified by gel electrophoresis. The purified cDNA molecules are PCR-amplified using 5’-amine-terminated primers and prepared for sequencing by the TruSeq DNA library preparation protocol. After sequencing the cDNA library, circRNAs are identified from the dataset by finding 2 exonic reads that are distal in the reference genome but adjacent to each other, with 1 inverted over the other in the dataset. The inversion of 1 adjacent exon signifies the circularization of the RNA (Table 15).

A schematic overview of SUPeR-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifies circular RNA from single cells.</td>
<td>Relies on dataset analysis to identify circRNAs.</td>
</tr>
<tr>
<td>Avoids 3’ bias by using random primers with anchor sequences.</td>
<td></td>
</tr>
<tr>
<td>Able to identify novel circRNAs due to random primers.</td>
<td></td>
</tr>
</tbody>
</table>

Reference


Although numerous scRNA-Seq methods have been developed, all of them specifically detect polyadenylated RNAs. A substantial amount of RNA expressed in mammalian cells lacks a poly(A) tail. In this study, the authors describe SUPeR-Seq, a poly(A)-independent method for scRNA-Seq. By performing SUPeR-Seq on mouse preimplantation embryos, they discovered 2891 circRNAs and 913 novel linear transcripts. This discovery allowed them to analyze the abundance of circRNAs in mammalian embryonic development and to identify sequence features of circRNAs.

Illumina Technology: TruSeq DNA Sample Preparation Kit, TruSeq RNA Sample Preparation Kit, HiSeq 2000/2500 Systems
Quartz-Seq

The Quartz-Seq method optimizes whole-transcript amplification (WTA) of single cells. In this method, a reverse-transcription (RT) primer with a T7 promoter and PCR target is first added to extracted mRNA. Reverse transcription synthesizes first-strand cDNA, after which the RT primer is digested by exonuclease I. A poly(A) tail is then added to the 3' ends of first-strand cDNA, along with a dT primer containing a PCR target. After second-strand generation, a blocking primer is added to ensure PCR enrichment in sufficient quantity for sequencing. Deep sequencing allows for accurate, high-resolution representation of the whole transcriptome of a single cell (Table 16).

A schematic overview of Quartz-Seq.

Table 16. Advantages and Disadvantages of Quartz-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Single-tube reaction suitable for automation.</td>
<td>• PCR biases can underrepresent GC-rich templates.</td>
</tr>
<tr>
<td>• Digestion of RT primers by exonuclease I</td>
<td>• Amplification errors caused by polymerases will be represented and sequenced incorrectly.</td>
</tr>
<tr>
<td>• Short fragments and byproducts are suppressed during enrichment.</td>
<td>• Targets smaller than 500 bp are preferentially amplified by polymerases during PCR.</td>
</tr>
</tbody>
</table>

Reference


RNA-Seq and transcriptional profiling of single cells have expanded our understanding of cellular heterogeneity at levels not achievable using bulk sequencing. Importantly, the cell cycle can be a major driver of transcriptional heterogeneity in scRNA-Seq. In this study, the authors analyzed 6 supervised computational methods to predict G1, S, or G2M phase using ESC transcriptome data. These data were generated, using the Quartz-Seq method, from libraries constructed with the TruSeq Stranded RNA Sample Preparation Kit. By comparing the performance of each algorithm on various scRNA-Seq datasets from various organisms, the authors conclude that a principal component analysis–based approach provides the best results.

Illumina Technology: TruSeq Stranded RNA Sample Preparation Kit
Smart-Seq

Smart-Seq was developed as a single-cell sequencing protocol with improved read coverage across transcripts. Complete coverage across the genome allows the detection of alternative transcript isoforms and SNPs. In this protocol, cells are lysed, and the RNA is hybridized to an oligo(dT)-containing primer. The first strand is then created with the addition of a few untemplated C nucleotides. This poly(C) overhang is added exclusively to full-length transcripts. An oligonucleotide primer is then hybridized to the poly(C) overhang and used to synthesize the second strand. Full-length cDNAs are PCR-amplified to obtain nanogram amounts of DNA. The PCR products are purified for sequencing (Table 17).

A schematic overview of Smart-Seq.

Table 17. Advantages and Disadvantages of Smart-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses as little as 50 pg of starting material.</td>
<td>Not strand-specific.</td>
</tr>
<tr>
<td>Can be used with unknown mRNA sequences.</td>
<td>No early multiplexing.</td>
</tr>
<tr>
<td>Provides improved coverage across transcripts.</td>
<td>Transcript length bias with inefficient transcription of reads over 4 kb.</td>
</tr>
<tr>
<td>Results in high levels of mappable reads.</td>
<td>Preferential amplification of high-abundance transcripts.</td>
</tr>
<tr>
<td></td>
<td>The purification step may lead to loss of material.</td>
</tr>
<tr>
<td></td>
<td>Could be subject to strand-invasion bias.</td>
</tr>
</tbody>
</table>

Reference

Smart-Seq2

Smart-Seq2 incorporates several improvements over the original Smart-Seq protocol. The new protocol includes a locked nucleic acid (LNA), an increased MgCl2 concentration, betaine, and elimination of the purification step to improve the yield significantly. In this protocol, single cells are lysed in a buffer that contains free dNTPs and tailed oligo(dT) oligonucleotides with a universal 5’ anchor sequence. Reverse transcription is performed, which also adds 2–5 untemplated nucleotides to the cDNA 3’ end. A template-switching oligo (TSO) is added, which carries 2 riboguanosines and a modified guanosine to produce an LNA as the last base at the 3’ end. After the first-strand reaction, the cDNA is amplified using a limited number of cycles. Tagmentation is then used to construct sequencing libraries quickly and efficiently from the amplified cDNA.

A schematic overview of Smart-Seq2.

<table>
<thead>
<tr>
<th>mRNA fragment</th>
<th>mRNA template (Smart)</th>
<th>First-strand synthesis with MMLV reverse transcriptase</th>
<th>PCR</th>
<th>Tegmentation</th>
<th>Gap repair, enrichment PCR and PCR purification</th>
<th>Enrichment-ready fragment</th>
</tr>
</thead>
</table>

Table 18. Advantages and Disadvantages of Smart-Seq2.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Uses as little as 50 pg of starting material.</td>
<td>• Not strand-specific.</td>
</tr>
<tr>
<td>• Can be used with unknown mRNA sequences.</td>
<td>• No early multiplexing.</td>
</tr>
<tr>
<td>• Provides improved coverage across transcripts.</td>
<td>• Only suitable for poly(A)+ RNA.</td>
</tr>
</tbody>
</table>

References


To perform RNA-Seq of single cells, protease treatment has been used to dissociate individual cells from tissues. In this study, the authors showed that this protease-digestion approach altered the transcriptome of individual neurons. To overcome this challenge, they isolated nuclei from postmortem human brain homogenates and sorted them by FACS. They also used Smart-Seq2 to perform cDNA synthesis from nuclear mRNAs and the MiSeq system for sequencing of Nextera XT barcoded libraries. This approach is amenable to any tissue in which single-cell dissociation requires harsh treatment.

Illumina Technology: Nextera XT DNA Library Preparation Kit, MiSeq System

Single-Cell Methylome & Transcriptome Sequencing

scM&T-Seq allows parallel analysis of both epigenetic and gene expression patterns from single cells using Smart-Seq2 and scBS-Seq. scM&T-Seq is built upon G&T-Seq, but instead of using MDA for DNA sequencing, it uses scBS-Seq to interrogate DNA methylation patterns.

First, single cells are isolated and individually lysed. Then, mRNAs are isolated using streptavidin-coupled mRNA capture primers, physically separating them from DNA strands. Smart-Seq2 is used to generate cDNA libraries from the mRNA, which involves reverse transcription with template switching and tagmentation. DNA libraries are prepared via scBS-Seq, which involves bisulfite conversion of DNA strands to identify methylated cytosines. Both libraries are now ready for sequencing (Table 19).

A schematic overview of scM&T-Seq.

Table 19. Advantages and Disadvantages of scM&T-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Investigates links between epigenetic and transcriptional heterogeneity in single cells.</td>
<td>• Smart-Seq2 is not strand-specific and applicable to only poly(A)+ RNA.</td>
</tr>
<tr>
<td>• Because DNA and RNA are physically separated and amplified independently, there is no need to mask coding sequences during analysis.</td>
<td>• Does not distinguish between 5mC and 5hmC.</td>
</tr>
</tbody>
</table>

References


Multiparameter single-cell sequencing is a powerful tool that has uncovered relationships among genomic, transcriptional, and epigenetic heterogeneity. In this study, the authors developed scM&T-Seq, a multiparameter sequencing method that allows methylome and transcriptome profiling in the same cell. They used the G&T-Seq protocol to purify single-cell DNA that was then subjected to scBS-Seq. Using the HiSeq 2000 system, the authors performed scM&T-Seq on 61 mouse ESCs. They found that gene expression levels of many pluripotency factors were negatively associated with DNA methylation. These data demonstrate that epigenetic heterogeneity is an important mechanism of fluctuating pluripotency in ESCs. They also demonstrate that scM&T-Seq can illuminate the poorly understood relationship between transcriptional and DNA-methylation heterogeneity in single cells.

Illumina Technology: Nextera XT Kit, HiSeq 2000 System


In this study, the authors developed a method that simultaneously profiles the methylome and the transcriptome of the same individual cell (scMT-Seq). The method is very similar to scM&T-Seq, except that scMT-Seq uses scRRBS for methylome analysis. The authors validated this method by simultaneously profiling the transcriptome and DNA methylome in individual sensory neurons of the DRG. Their data identified transcriptome and DNA methylome heterogeneity in DRG neurons. They also found that gene methylation and expression are positively correlated, but only for those genes that contain CpG island promoters.

Illumina Technology: MiSeq and HiSeq 2500 Systems
**Genome & Transcriptome Sequencing**

G&T-Seq is a protocol that can separate and sequence genomic DNA and full-length mRNA from single cells. In this method, single cells are isolated and lysed. RNA is captured using biotinylated oligo(dT) capture primers and separated from DNA using streptavidin-coated magnetic beads. Smart-Seq2 is used to amplify captured RNA on the bead, while MDA is used to amplify DNA. After sequencing, integrating DNA and RNA sequences provides insights into the gene-expression profile of single cells.

A schematic overview of G&T-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Compatible with any WGA method.</td>
<td>• Physical separation of DNA and RNA can increase the risk of sample loss or contamination.</td>
</tr>
<tr>
<td>• No 3'-end bias in sequence reads because full-length transcripts are captured.</td>
<td>• Physical separation of DNA and RNA increases handling time.</td>
</tr>
<tr>
<td>• Because DNA and RNA are physically separated and amplified independently, there is no need to mask coding sequences during analysis.</td>
<td></td>
</tr>
</tbody>
</table>

**References**


Multiparameter single-cell sequencing is a powerful tool that has uncovered relationships among genomic, transcriptional, and epigenetic heterogeneity. In this study, the authors developed scM&T-Seq, a multiparameter sequencing method that allows methylome and transcriptome profiling in the same cell. They used the G&T-Seq protocol to purify single-cell DNA that was then subjected to scBS-Seq. Using the HiSeq 2000 system, the authors performed scM&T-Seq on 61 mouse ESCs. They found that gene expression levels of many pluripotency factors were negatively associated with DNA methylation. These data demonstrate that epigenetic heterogeneity is an important mechanism of fluctuating pluripotency in ESCs. They also demonstrate that scM&T-Seq can illuminate the poorly understood relationship between transcriptional and DNA-methylation heterogeneity in single cells.

**Illumina Technology: Nextera XT Kit, HiSeq 2000 System**


Single-cell genomic sequencing has provided insights into cellular heterogeneity, as well as cellular lineage and development. Single-cell transcriptomic sequencing has refined our understanding of cell types and states. In this study, the authors developed G&T-Seq, a method that allows for the separation and subsequent sequencing of genomic DNA and full-length mRNA from single cells. It complements the DR-Seq method, but it can be used with any WGA method and also provides full-length transcripts from the same cell. The authors performed G&T-Seq-enabled transcriptome analysis by using a modified Smart-Seq2 protocol and automated the method on a robotic liquid-handling platform. They used the HiSeq platform to sequence numerous single-cell types, including human cancer cells, reverse-treat mouse embryo blastomeres, and iPSC-derived neurons. Notably, G&T-Seq analysis of aneuploid blastomeres demonstrated that chromosomal gains/losses led to increases/losses in chromosome-wide relative gene expression during a single cell division.

**Illumina Technology: Nextera XT Kit, MiSeq, HiSeq 2500 and HiSeq X Ten Systems**


Single-cell genomic sequencing has provided insights into cellular heterogeneity, as well as cellular lineage and development. Single-cell transcriptomic sequencing has refined our understanding of cell types and states. In this study, the authors developed G&T-Seq, a method that allows for the separation and subsequent sequencing of genomic DNA and full-length mRNA from single cells. It complements the DR-Seq method, but it can be used with any WGA method and also provides full-length transcripts from the same cell. The authors performed G&T-Seq-enabled transcriptome analysis by using a modified Smart-Seq2 protocol and automated the method on a robotic liquid-handling platform. They used the HiSeq platform to sequence numerous single-cell types, including human cancer cells, reverse-treat mouse embryo blastomeres, and iPSC-derived neurons. Notably, G&T-Seq analysis of aneuploid blastomeres demonstrated that chromosomal gains/losses led to increases/losses in chromosome-wide relative gene expression during a single cell division.

**Illumina Technology: Nextera XT Kit, MiSeq, HiSeq 2500 and HiSeq X Ten Systems**
Genomic DNA and mRNA Sequencing

DR-Seq studies the genomic and transcriptomic relationship of single cells via sequencing. Nucleic acid amplification prior to physical separation reduces sample loss and the risk of contamination. DR-Seq involves multiple amplification steps, including the quasilinear amplification technique similar to MALBAC.

First, mRNAs are reverse-transcribed from lysed single cells using poly(dT) primers with Ad-1x adapters, producing sscDNA. The Ad-1x adapter sequence contains cell-identifying barcodes, 5' Illumina adapters, and a T7 promoter. Next, both gDNA and sscDNA are amplified simultaneously via quasilinear WGA with Ad-2 primers. These primers are similar to MALBAC adapters, containing 8 random nucleotides for random priming followed by a constant 27-nucleotide tag at the 5' end. Products of this amplification step are split in halves. One half is prepared for genome sequencing, in which gDNA are PCR-amplified and "liberated" of their Ad-2 adapters before DNA library prep and sequencing. The other half is prepared for transcriptome sequencing, whereby second strands are synthesized for the cDNAs and amplified by in vitro transcription. The resulting RNA products are produced only from cDNA fragments flanked with Ad-1x and Ad-2, omitting amplification of the gDNA fragments. The RNA library is prepared for sequencing following the Illumina small-RNA protocol. Sequencing gDNA and mRNA from the same cell preserves information between the genome and its expression levels (Table 21).

A schematic overview of DR-Seq.

Table 21. Advantages and Disadvantages of DR-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Interrogates genomic and transcriptomic behavior from a single cell.</td>
<td>• Manual single-cell isolation prevents high-throughput adaptation.</td>
</tr>
<tr>
<td>• Amplification prior to separation reduces sample loss and contamination.</td>
<td>• Quasilinear amplification is temperature-sensitive.</td>
</tr>
<tr>
<td>• Length-based identifier used to remove duplicate reads.</td>
<td>• RNA reads are 3'-end-biased.</td>
</tr>
<tr>
<td>• Quasilinear amplification reduces PCR bias.</td>
<td></td>
</tr>
</tbody>
</table>

Reference


Single-cell genomics and transcriptomics are promising tools for quantifying genetic and expression variability among individual cells. In this study, the authors describe DR-Seq, a method to quantify the genome and transcriptome of the same cell simultaneously. DR-Seq does not require physical separation of nucleic acids before amplification, which helps to minimize the chances for sample loss or contamination. The authors amplified gDNA and cDNA from mouse ESCs, subsequently divided the nucleic acids for further amplification and library construction, and sequenced both libraries using a HiSeq 2500 system. They demonstrated that genes with high cell-to-cell variability in transcript numbers have low CNVs, and vice versa.

Illumina Technology: HiSeq 2500 System
T Cell–Receptor Chain Pairing

Functional TCRs are heterodimeric proteins composed of unique combinations of α and β chains. For an accurate functional analysis, both subunits must be sequenced together to avoid disrupting the α- and β-chain pairing during the cell lysis step.333

Cell-based emulsion RT-PCR technique for identifying TCR α-β chain pairing. Released TCR-α and TCR-β mRNAs are reverse-transcribed, amplified, and overlap-extended within each droplet. Products are extracted from the emulsion and fused molecules of interest are selectively amplified. Nonfused molecules are suppressed with blocking primers.334

References


Unique Molecular Identifiers

UMIs are molecular tags that can be used to detect and quantify unique mRNA transcripts.335 In this method, mRNA libraries are generated by fragmentation and then reverse-transcribed to cDNA. Oligo(dT) primers with specific sequencing linkers are added to the cDNA. Another sequencing linker with a 10 bp random label and an index sequence is also added to the 5’ end of the template, which is amplified and sequenced. Sequencing allows for high-resolution reads, enabling accurate detection of true variants (Table 22).

A schematic overview of UMIs.

Table 22. Advantages and Disadvantages of UMIs.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Can sequence unique mRNA transcripts.</td>
<td>• Targets smaller than 500 bp are preferentially amplified by polymerases during PCR.</td>
</tr>
<tr>
<td>• Can be used to detect transcripts occurring at low frequencies.</td>
<td></td>
</tr>
<tr>
<td>• Transcripts can be quantified based on sequencing reads specific to each barcode.</td>
<td></td>
</tr>
<tr>
<td>• Can be applied to multiple platforms to karyotype chromosomes as well.</td>
<td></td>
</tr>
</tbody>
</table>
Cell Expression by Linear Amplification Sequencing

CEL-Seq utilizes barcoding and pooling of RNA to overcome challenges from low input.336 In this method, each cell undergoes reverse transcription with a unique barcoded primer in its individual tube. After second-strand synthesis, cDNA from all reaction tubes are pooled and PCR-amplified. Paired-end deep sequencing of the PCR products allows for accurate detection of sequence derived from sequencing both strands (Table 23).

A schematic overview of CEL-Seq.

Table 23. Advantages and Disadvantages of CEL-Seq.

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Barcoding and pooling allows for multiplexing and studying many different single cells at a time.</td>
<td>• Strongly 3' biased. 335</td>
</tr>
<tr>
<td>• Contamination between samples is greatly reduced due to processing a single tube per cell.</td>
<td>• Abundant transcripts are preferentially amplified.</td>
</tr>
<tr>
<td>• Uses fewer steps than single-cell tagged reverse-transcription sequencing (STRT-Seq).</td>
<td>• Requires at least 400 pg of total RNA.</td>
</tr>
<tr>
<td>• Shows very little read-length bias. 337</td>
<td></td>
</tr>
<tr>
<td>• Strand-specific.</td>
<td></td>
</tr>
</tbody>
</table>

Reference

Cooper DA, Jha BK, Silverman RH, Hesselberth JR and Barton DJ. Ribonuclease L and metal-ion-independent endoribonuclease cleavage sites in host and viral RNAs. Nucleic Acids Res. 2014;42:5202-5216


In this study, the authors present a new scalable high-density microfluidic platform for solid-phase capture of RNA on glass coverslips or on polymer beads. They trapped single-cell lysates in sealed picoliter microwells capable of printing RNA on glass or capturing RNA on beads. They combined this sample preparation approach with a scalable technology for scRNA-Seq based on CEL-Seq. The technology is relatively inexpensive, with consumable costs of $0.10–$0.20 per cell and is capable of processing hundreds of individual cells in parallel.

Illumina Technology: TruSeq RNA-Seq Library Preparation Kit, NextSeq 500 and HiSeq 2500 Systems
Flow Cell–Surface Reverse-Transcription Sequencing

Flow cell–surface reverse-transcription sequencing (FRT-Seq) is a transcriptome-sequencing technique developed in 2010.\textsuperscript{339} It is strand-specific, free of amplification, and is compatible with paired-end sequencing. To begin with, poly(A\textsuperscript{+}) RNA samples are fragmented by metal-ion hydrolysis and dephosphorylated. Next, P7 primers are ligated to the 3’ end of the fragments. The adapter sequence starts at the 5’ terminus with 20 nucleotides of RNA, followed by DNA nucleotides. The primers are also 5’ phosphorylated and blocked with dideoxycytosine (ddC) at the 3’ end. Following 3’-adapter ligation, fragments are size-selected for nucleotide fragments longer than the adapter. The 5’ ends of the fragments are phosphorylated and ligated to P5 adapters. These adapters are blocked with an amino-C6 linker at the 5’ end. Now that the fragments are flanked with adapters, they are hybridized to the flow cell and reverse-transcribed before cluster generation and sequencing (Table 24). FRT-Seq has the potential to overcome problems associated with RNA amplification in single-cell sequencing, but it has not yet been used in scRNA-Seq applications.\textsuperscript{340}

A schematic overview of FRT-Seq.

Table 24. Advantages and Disadvantages of FRT-Seq.

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Strand-specific poly(A\textsuperscript{+}) mRNA sequencing for transcriptome analysis.</td>
<td>• Requires a large amount of input RNA material (250 ng).</td>
</tr>
<tr>
<td>• No amplification step—gives more accurate representation of the total mRNA population, preventing amplification bias.</td>
<td>• Selects only poly(A\textsuperscript{+}) mRNA samples.</td>
</tr>
</tbody>
</table>

Reference


Single-Cell Tagged Reverse-Transcription Sequencing

STRT-Seq is a method similar to CEL-Seq that involves unique barcoding and sample pooling to overcome the challenges of samples with limited material.\textsuperscript{341} In this method, single cells are first picked in individual tubes, where first-strand cDNA synthesis occurs using an oligo(dT) primer with the addition of 3–6 cytosines. A helper oligo promotes template switching, which introduces the barcode in the cDNA. Barcoded cDNA is then amplified by single-primer PCR. Deep sequencing allows for accurate transcriptome sequencing of individual cells (Table 25).
A schematic overview of STRT-Seq.

Table 25. Advantages and Disadvantages of STRT-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Barcoding and pooling allows for multiplexing and studying many different single cells at a time.</td>
<td>• PCR biases can underrepresent GC-rich templates.</td>
</tr>
<tr>
<td>• Contamination between samples is greatly reduced due to processing a single tube per cell.</td>
<td>• Nonlinear PCR amplification can lead to biases affecting reproducibility and accuracy.</td>
</tr>
<tr>
<td></td>
<td>• Amplification errors caused by polymerases will be represented and sequenced incorrectly.</td>
</tr>
<tr>
<td></td>
<td>• Targets smaller than 500 bp are preferentially amplified by polymerases during PCR.</td>
</tr>
</tbody>
</table>

Reference


Fixed and Recovered Intact Single-Cell RNA Sequencing

FRISCR sequencing characterizes transcriptome profiles from fixed and stained single cells. First, the cell suspension is fixed with paraformaldehyde, permeabilized, and immunostained. Individual cells are then sorted into tubes using FACS. These cells are lysed and crosslinking is reversed crosslink by incubation at 56°C for 1 hour. mRNA from the cells is isolated by dT25 magnetic bead pull-down. The mRNA sequencing library is prepared by following Smart-Seq2 procedures: 1) template-switching reverse transcription using Moloney murine leukemia virus reverse transcriptase; 2) PCR-amplifying the resulting cDNAs; and 3) preparing a cDNA library using the Nextera XT Library Preparation Kit. The fragments are now flanked with adapters and are ready for sequencing (Table 26).

A schematic overview of FRISCR sequencing.
Table 26. Advantages and Disadvantages of FRISCR Sequencing.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Full-length mRNA transcriptome profiling from fixed and stained single cells.</td>
<td>• 3’ to 5’ bias.</td>
</tr>
<tr>
<td>• Immunostaining enables targeting of rare cell populations.</td>
<td></td>
</tr>
<tr>
<td>• Generates full-length mRNA reads.</td>
<td></td>
</tr>
<tr>
<td>• Significantly more mRNA recovered compared to fixed cells from Triton-X100 lysis.</td>
<td></td>
</tr>
</tbody>
</table>

Reference


The human neocortex develops from rare progenitor cells, especially RG. These cells have been difficult to characterize, since they are rare and are defined by a combination of position, morphology, and intracellular markers. The authors developed a method that allows RNA-Seq of individual fixed, stained, and sorted cells, known as FRISCR sequencing. They sorted individual RG cells by FACS and prepared single-cell mRNA libraries using Smart-Seq2 followed by sequencing using the MiSeq system. They demonstrated that expression data from fixed and purified single cells were similar to that obtained from live cells. Their data also identified subpopulations of ventricular zone–enriched RG and subventricular zone–localized RG, as well as new molecular markers for each subtype.

Illumina Technology: Nextera XT Library Preparation Kit, MiSeq System

Cell Labeling via Photobleaching

CLaP is a noninvasive, laser-based labeling technique for single cells. CLaP uses lasers to crosslink specific cells with fluorescent tags before isolating the single cells for sequencing.

In CLaP, cells of interest are tagged by crosslinking biotin-4-fluorescein (B4F) with the cell membrane using laser irradiation. Streptavidin-conjugate fluorescent labels are then bound to biotinylated cells. These steps can be repeated to tag multiple cell types with a variety of fluorescent tags. Tagged cells are subsequently isolated and processed to generate cDNA libraries before sequencing (Table 27).

A schematic overview of CLaP.

Table 27. Advantages and Disadvantages of CLaP.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Noninvasive, targeted laser-based single-cell labeling.</td>
<td>• Image-based selection limits the potential for high-throughput applications.</td>
</tr>
<tr>
<td>• Automated image-based cell selection is possible.</td>
<td>• Diffusion of reagents through the extracellular matrix and continuous laser illumination limit the procedure for 3-dimensional environments/tissues.</td>
</tr>
<tr>
<td>• Fluorescence-based tags can be substituted with other labels, such as electron-dense molecules.</td>
<td>• Cellular specificity may be decreased slightly in primary cell cultures.</td>
</tr>
<tr>
<td>• Multicolored fluorescent stains can be used.</td>
<td></td>
</tr>
</tbody>
</table>
Since single-cell sequencing methods often involve dissociation of cells and loss of spatial information, methods that retain spatial information in single-cell genomic analysis are critically important. The authors developed CLaP, a method that combines cellular labeling with single-cell genomics. Individual cells are labeled in culture by laser photobleaching, followed by isolation based on a wide variety of distinguishing characteristics. In this study, the authors used CLaP to tag a number of different cells from lines grown in monolayers. They isolated individual cells using drop-based microfluidics and performed RNA-Seq using the HiSeq 2500 system. The ability to combine spatial information with single-cell genomics makes this method well suited for studying tissue heterogeneity.

Illumina Technology: Nextera XT DNA Sample Preparation Kit, HiSeq 2500 System

Indexing Droplets

Indexing droplets (inDrop) are used for high-throughput single-cell labeling. This approach is similar to Drop-Seq, but it uses hydrogel microspheres to introduce the oligos.

Single cells from a cell suspension are isolated into droplets containing lysis buffer. After cell lysis, cell droplets are fused with a hydrogel microsphere containing cell-specific barcodes and another droplet with enzymes for reverse transcription. Droplets from all the wells are pooled and subjected to isothermal reactions for reverse transcription. The barcode-oligos anneal to poly(A) mRNAs and act as primers for reverse transcriptase. Each mRNA strand now has cell-specific barcodes. The droplets are pooled, broken, and the mRNAs are purified. The 3' ends of the cDNA strands are ligated to adapters, amplified, annealed to indexed primers, and amplified further before sequencing (Table 28). The sequencing method is similar to CEL-Seq.

Table 28, Advantages and Disadvantages of inDrop.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• High throughput single-cell transcriptome profiling using microfluidics.</td>
<td>• Droplets may contain 2 cells or 2 different types of barcodes.</td>
</tr>
<tr>
<td>• Low cost: $0.1 per cell (experiments require 100 cells).</td>
<td></td>
</tr>
<tr>
<td>• Highly scalable to larger cell quantities.</td>
<td></td>
</tr>
<tr>
<td>• No fragmentation step.</td>
<td></td>
</tr>
</tbody>
</table>
An overview of recent publications featuring Illumina technology

For Research Use Only. Not for use in diagnostic procedures.
CytoSeq

The gene expression cytometry protocol known as CytoSeq enables gene expression profiling of thousands of single cells. In this method, single cells are first randomly deposited into wells. A combinatorial library of beads with specific capture probes is added to each well. After cell lysis, mRNAs hybridize to the beads, which are then pooled for reverse transcription, amplification, and sequencing. Deep sequencing provides accurate, high-coverage gene expression profiles of several single cells (Table 30).

References


Single-cell RNA-Seq has become an indispensable technique for characterizing cell types, states, and function in complex tissues. Sequencing individual nuclei has similarly proved useful in cell-type discrimination, particularly for tissues where single cells are difficult to dissociate. Current approaches for single-nucleus sequencing rely on the use of FACS or C1 microfluidics to sort individual nuclei into multiwell plates for downstream analysis. In this study, the authors developed DroNc-seq, a method that combines single-nucleus sequencing with the microfluidics-based Drop-seq method. The authors modified Drop-seq to incorporate single nuclei within microfluidic droplets, as opposed to single individual cells. They performed massively parallel RNA-Seq for approximate 39,000 individual nuclei from mouse and human brain, at low cost and with similar sensitivity to Drop-seq.

Illumina Technology: NextSeq 500 System


One of the bottlenecks in scRNA-Seq is the limitation in the number of individual cells that can be separated and analyzed. In this study, the authors developed Drop-seq, a massively parallel scRNA-Seq method that uses uniquely barcoded primer beads together with captured single cells in droplets. This encapsulation method allows for processing of thousands of individual cells by RNA-Seq on a microfluidics platform. The authors validated this technique by applying it to the mouse retina. After Drop-seq, they used the NextSeq system to perform RNA-Seq on approximately 45,000 cells, and they identified 39 distinct cell populations within mouse retina. Their results demonstrate that Drop-seq can be used to understand the biology of complex tissues with diverse cell types.

Illumina Technology: Nextera XT DNA Sample Prep Kit, NextSeq 500 System

Table 29. Advantages and Disadvantages of Drop-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Analyzes sequences of single cells in a highly parallel manner.</td>
<td>• Requires custom microfluidics device to perform droplet separation.</td>
</tr>
<tr>
<td>• Unique molecular and cell barcodes enable cell- and gene-specific identification of mRNA strands.</td>
<td>• Low gene-per-cell sensitivity compared to other scRNA-Seq methods.345</td>
</tr>
<tr>
<td>• Reverse transcription with template-switching PCR produces high-yield reads from single cells.</td>
<td>• Limited to mRNA transcripts.</td>
</tr>
<tr>
<td>• Low cost—$0.07 per cell ($653 per 10,000 cells)—and fast library prep (10,000 cells per day).</td>
<td></td>
</tr>
</tbody>
</table>


An overview of recent publications featuring Illumina technology

For Research Use Only. Not for use in diagnostic procedures.

Single-Cell RNA Barcoding and Sequencing

Single-cell RNA barcoding and sequencing (SCRB-Seq) is a cost-efficient, multiplexed scRNA-Seq technique. SCRB-Seq isolates single cells into wells using FACS. After cell lysis, poly(A)+ mRNAs are annealed to a custom primer containing a poly(T) tract, UMI, well barcode, and biotin. Template-switching reverse transcription and PCR amplification are carried out on the mRNA, generating barcoded full-length cDNA. cDNA strands from all wells are pooled together to be purified. They are amplified by PCR and purified further. cDNA libraries are prepared using the Nextera XT kit with modified i5 primers. The resultant cDNA fragments are size-selected for 300–800 bp and sequenced (Table 31).

Table 30. Advantages and Disadvantages of CytoSeq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can readily scale to tens/hundreds of thousands of cells.</td>
<td>Sequencing depth requires large number of reads (eg, 200,000 transcripts per cell require 2 million reads for 10x coverage: 2 billion reads for 1000 cells).</td>
</tr>
<tr>
<td>Complements and expands the capabilities of fluorescence or mass spectrometry–based cytometry.</td>
<td>A single run can be relatively expensive and time consuming.</td>
</tr>
<tr>
<td>Detects any transcribed mRNA without the limitations of antibody availability.</td>
<td>Involves a trade-off between depth of sequencing and differential gene expression.</td>
</tr>
<tr>
<td>Enables rare cell characterization on small samples with insufficient cells for traditional flow cytometry.</td>
<td></td>
</tr>
<tr>
<td>Allows direct analysis of complex samples of heterogeneous cell size and shape.</td>
<td></td>
</tr>
</tbody>
</table>

References


This study used combinatorial indexing to measure chromatin accessibility in thousands of single cells. The authors isolated nuclei and tagged them in bulk with Tn5 transposases, in each of many wells. Next, they pooled these barcoded nuclei, diluted them, and redistributed them to a second set of wells where a second barcode was introduced using PCR. The authors integrated this combinatorial indexing with scATAC-Seq347 to measure chromatin accessibility for more than 15,000 human and mouse single cells. They sequenced the scATAC-Seq libraries on the MiSeq system, and their data identified relevant differences in chromatin accessibility between cell types.

Illumina Technology: MiSeq and NextSeq Systems


Combinatorial labeling of single cells is rapid and relatively inexpensive, and it can boost the throughput of massively parallel single-cell sequencing approaches dramatically. In this study, the authors developed CytoSeq, a method to label large numbers of individual cells combinatorially. Individual cells are placed in single wells, along with combinatorial libraries of beads containing cell- and transcript-barcoding probes. The authors performed CytoSeq on human PBMCs and used the MiSeq system to sequence amplified cDNAs. They analyzed several genes and were able to identify major subsets of PBMCs. In addition, by comparing cellular heterogeneity in naïve and CMV-activated CD8+ T cells, they identified rare cells specific to the CMV antigen. CytoSeq can be applied to complex mixtures of cells of varying size and shape, as well as to other biomolecules.

Illumina Technology: MiSeq System

Table 31. Advantages and Disadvantages of SCRB-Seq.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Cost-efficient, high-throughput single-cell transcriptome profiling.</td>
<td>• Template-switching reverse transcription is heavily biased to full-length mRNA. [349]</td>
</tr>
<tr>
<td>• Highly sensitive gene-detection results compared to popular scRNA-Seq</td>
<td></td>
</tr>
<tr>
<td>techniques. [348]</td>
<td></td>
</tr>
</tbody>
</table>

Reference

High-Throughput Single-Cell Labeling
Hi-SCL generates transcriptome profiles for thousands of single cells using a custom microfluidics system, similar to Drop-Seq [350] and inDrop. [351] Single cells from a cell suspension are isolated into droplets containing lysis buffer. After cell lysis, cell droplets are fused with a droplet containing cell-specific barcodes and another droplet with enzymes for reverse transcription. Droplets from all the wells are pooled and subjected to isothermal reactions for reverse transcription. The barcode-oligos anneal to poly(A)+ mRNAs and act as primers for reverse transcriptase. Now that each mRNA strand has cell-specific barcodes, droplets are broken and the mRNAs are purified. The 3' ends of the cDNA strands are ligated to adapters, amplified, annealed to indexed primers, and amplified further before sequencing (Table 32).

Table 32. Advantages and Disadvantages of Hi-SCL.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• High-throughput, single-cell transcriptome profiling using a microfluidics system.</td>
<td>• Lack of UMIs in oligos may create amplification noise.</td>
</tr>
<tr>
<td>• Low cost—$0.1 per cell (experiment requires 100 cells).</td>
<td>• Droplets may contain 2 cells or 2 different types of barcodes.</td>
</tr>
<tr>
<td>• Highly scalable to larger cell quantities.</td>
<td></td>
</tr>
<tr>
<td>• No fragmentation step.</td>
<td></td>
</tr>
</tbody>
</table>
References


Given the importance of single-cell data, there is a great need to increase the throughput of sequencing pipelines. Methods that physically separate large numbers of individual cells into wells or chambers of microfluidics chips are vital to this effort. Hi-SCL uses drop-based libraries of oligonucleotide barcodes to index individual cells. The drops are used as containers on a microfluidics platform, and the tagged molecules from different cells can be mixed without losing cell-of-origin information. In this study, the authors used MiSeq and HiSeq systems to validate Hi-SCL by performing RNA-Seq on hundreds of mouse ESCs and MEFs. They demonstrated that single-cell data could recapitulate bulk expression data and that single-cell data could distinguish ESCs from fibroblasts. Compared to Fluidigm C1 and CEL-Seq,352 Hi-SCL proved to be a faster and cheaper method for massively parallel sequencing.

Illumina Technology: MiSeq and HiSeq Systems


Genomic Solutions for Cell Biology and Complex Disease Research

Complex diseases are the result of multiple genetic and environmental factors. They are distinguished from Mendelian traits (or simple traits) as they do not follow a specific model of inheritance and are usually more frequent in the population. Although some of these diseases are highly heritable, currently known genetic variants can explain only some of the estimated heritability. This review gives a general overview on how genomic technologies and NGS can help in the study of complex diseases.

Cancer and Immune System Research Review

Advances in high-throughput sequencing have dramatically improved our knowledge of the cancer genome and the intracellular mechanisms involved in tumor progression and response to treatment. While the primary focus to date has been on the cancer cell, this technology can also be used to understand the interaction of the tumor cells and the cells in the surrounding tumor microenvironment.

Expression analysis of the RNA levels can be used to determine the activation of pathways in the tumor microenvironment. Since common signaling pathways are involved in manifestation of several hallmarks of cancer, including cancer cell proliferation, survival, invasion, metastasis, and immunosuppression, targeting these shared signaling pathways in combination with immunotherapy may be a promising strategy for cancer treatment.