Single Cell Genomics: Advances and Future Perspectives

Iain C. Macaulay¹, Thierry Voet¹,²*

¹ Single Cell Genomics Centre, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom, ² Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, Belgium

Abstract: Advances in whole-genome and whole-transcriptome amplification have permitted the sequencing of the minute amounts of DNA and RNA present in a single cell, offering a window into the extent and nature of genomic and transcriptomic heterogeneity which occurs in both normal development and disease. Single-cell approaches stand poised to revolutionise our capacity to understand the scale of genomic, epigenomic, and transcriptomic diversity that occurs during the lifetime of an individual organism. Here, we review the major technological and biological breakthroughs achieved, describe the remaining challenges to overcome, and provide a glimpse into the promise of recent and future developments.

Introduction

The cell is a fundamental unit of biology, in which the blueprint of the genome is transcribed and translated into biological form and function. Almost all of our current understanding of the genome and its regulation has been derived from studies carried out at the population level—typically thousands or millions of cells analysed in bulk. The resulting analysis, although unquestionably informative, often neglects any heterogeneity that occurs within the population of cells.

The genome, despite being widely thought of as stable throughout normal development, has a small probability of acquiring genetic mutations with every cell division [1,2]. Over sufficient divisions, genomic heterogeneity within the organism—known as somatic variation—is a certainty. While such variation lies at the root of many disorders [3,4], including cancer [5], recent studies revealed unexpected levels of genomic variation in normal and diseased tissue, suggesting higher rates of genetic lesion than previously expected [6–12]. Still, little is known about the rate and nature of DNA mutation and how this is influenced by genetic background, lifestyle, and many other factors.

The transcriptome is naturally more dynamic than the genome, reflecting the function—or type—of the cell. There is considerable evidence indicating that cell-to-cell variability in gene expression is ubiquitous, even within a phenotypically “homogeneous” population of cells [13]. The extent of transcriptional heterogeneity and the diversity of cell types in tissues remain, however, largely unknown.

The genomic and transcriptomic composition of individual cells is lost in conventional sequencing studies, which analyse DNA and/or RNA extracted from large populations of cells; and de novo genome mutation and transcriptomic variations in cells will be largely concealed in the bulk signal. Clear insights into many biological processes—from normal development to tumour evolution—will thus only be gained from a detailed understanding of genomic, epigenomic, and transcriptional variation at the single-cell level. Furthermore, some cell types are so rare that single-cell approaches become paramount to their identification and characterisation.

Advances in techniques for the isolation of single cells (Figure 1A), whole genome or transcriptome amplification, and genome-wide analysis platforms—primarily next-generation sequencing (NGS) devices—paved the way for high-resolution analysis of the genome or transcriptome from one cell, which reveals previously obscured biological complexity.

Single-Cell Whole-Genome Amplification: Methods and Limitations

A diploid human cell contains approximately 7 pg genomic DNA; necessitating amplification prior to microarray- or NGS-based analyses to detect various classes of genetic variation (Figure 1B–1F). Current whole-genome amplification (WGA) principles are based on Multiple Displacement Amplification (MDA), Polymerase Chain Reaction (PCR), or a combination of both (Figure 2A–2C).

Unfortunately, no WGA method is faultless, and their various imperfections can considerably affect the interpretation of the microarray or NGS readout [14]. The breadth of genomic coverage, the amplification bias due to local differences in richness for guanine and cytosine bases (%GC-bias), the prevalence of allelic drop outs (ADO), preferential allelic amplifications (PA), chimeric DNA-molecules, and nucleotide copy errors can vary significantly between different WGA approaches, making some methods better suited than others for detecting specific classes of genetic variation [14–17] (Figure 2D). A comparative analysis of all WGA methods, including the investigation of the advantageous effects of reducing the reaction volume to a nanoliter scale [18,19], against a benchmark case is acute.

Advances in NGS and Bioinformatics Permit High-Resolution Screening of a Single-Cell Genome

Single-cell WGA products have been analysed using a variety of high-throughput platforms, including DNA-microarrays, SNP-arrays, and NGS (Figure 2D). A key difficulty in the interpretation

Citation: Macaulay IC, Voet T (2014) Single Cell Genomics: Advances and Future Perspectives. PLoS Genet 10(1): e1004126. doi:10.1371/journal.pgen.1004126

Editor: Nancy Maizels, University of Washington, United States of America

Published January 30, 2014

Copyright: © 2014 Macaulay, Voet. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: We acknowledge the Wellcome Trust (UK; http://www.wellcome.ac.uk/ and http://www.sanger.ac.uk/), the Research Foundation Flanders (FWO; http://www.fwo.be/; Belgium) [FWO-G.093.11 to TV], and the KU Leuven (http://www.kuleuven.be/; Belgium; SymBioSys, PFV/10/016 to TV]. The funders had no role in the preparation of the article.

Competing Interests: I have read the journal’s policy and have the following conflicts: TV is a named inventor on submitted patent applications PCT/EP2011/006211; PCT/EP2013/070858; ZL913096.

* E-mail: Thierry.Voet@med.kuleuven.be
Figure 1. Detection of various classes of genetic variation using single-cell WGA-NGS approaches. A) The most prominent methods for (i–ii) isolating individual cells (including (i) creation of single-cell suspensions—usually by enzymatic tissue disaggregation—and subsequent cell isolation through manual micro-pipetting [37,38,57,105], fluorescence-activated cell sorting [106,107] or microfluidics devices [18,81,108], and (ii) laser-capture microdissection [109,110]) as well as (iii) isolating single nuclei [12,32,56,111] are indicated, accompanied with particular advantages and disadvantages. A comprehensive review of single-cell isolation methods is presented by Shapiro et al. [112]. B–D) Subsequently, the cell is lysed and its genome amplified. A standard sequencing library can be prepared from the WGA product for paired-end (or single-end) sequencing. The resulting
of single-cell WGA data on any platform is the separation of the numerous WGA artifacts from the genuine genetic variants present in the template genome.

Standard DNA-microarrays can detect copy number variations (CNVs) larger than 2.5 Mb from a single-cell genome [20–22], while targeted array comparative genomic hybridizations can discover approximately 1 Mb-sized DNA imbalances [23], although remarkably, CNVs as small as 56 kb in single-cell PCR-based WGA products have been detected [24]. Similarly, SNP-arrays can find copy number aberrations encompassing millions of bases in a cell [25–28], but have the advantage of enabling the discovery of copy neutral DNA anomalies and regions of loss-of-heterozygosity (LOH), and allow inferring genome-wide haplotypes [29–31].

NGS has a number of advantages over microarrays enabling improved resolution and accuracy in variant calling [14]. First, NGS can examine every nucleotide amplified from the cell and allows genome-wide discovery of the full spectrum of DNA mutations (Figure 1E), while microarrays only probe for certain CNV loci (Figure 2D). Second, sequencing provides digital precision, with one digital unit representing a mapped sequence read. Finally, paired-end sequencing and mapping discloses the linkage between both ends of each linear DNA-molecule in a sequencing library of a single-cell WGA product, allowing the identification of structural variations via read-pairs mapping discordantly to the reference genome (Figure 1Ei).

Analytical challenges remain in interpreting single-cell NGS data for the full spectrum of genetic variants. Although WGA imperfections due to genome base composition (e.g. %GC-bias) can be computationally corrected for [17,32], the potential for PA and ADO can still generate local distortion in copy number, requiring distinct analyses to distinguish genuine copy number changes from WGA artefacts. Allelic fractions of heterozygous SNPs [26,33,34] or aberrantly mapping read pairs following paired-end sequencing of the WGA product [17] can be used to increase confidence in CNV measurements (Figure 1B–1F). For instance, a real deletion of a diploid locus should show LOH and discordantly mapping read-pairs that explain the DNA loss. Furthermore, the cell cycle stage of the isolated cell must be considered, further complicating the analysis, as cells in S-phase generate a dynamic copy number profile, leading to false structural DNA-imbalance discoveries [35].

The identification of the full spectrum of intra- and inter-chromosomal (un)balanced structural variants in a single-cell WGA product is still in its infancy—the main difficulty being to filter true structural variants from chimeric DNA generated during WGA, as well as issues with genome coverage (Figure 1Ei, 1F). Although filters have been designed to permit the detection of the structural architecture of DNA copy number variation [17] and even to detect L1-retrotransposition [36], many structural variants are still missed in single-cell analyses. Base alterations, such as SNPs, can be detected in single-cell WGA products (Figure 1Ei). However, to call accurate and reliable base substitutions in non-haploid loci, one requires the data of at least three cells to discriminate the variant from a WGA or sequencing error [16,37,38], and as such, detailed characterisation of extremely rare cells or sub-clones within populations may not be possible. Despite these hurdles, several groups have proven the efficacy of single-cell NGS to detect multiple classes of mutation within a genome and even to detect sister chromatid exchanges following single-cell Strand-seq [39]. Step-by-step bioinformatics protocols for analyzing Strand-seq data [40] as well as for copy number profiling single cells through NGS [32] or microarray analysis [34] and commercial solutions (e.g. platforms used within [21,41]) are surfacing.

Single-Cell Genomics Reveals the Extent of Somatic Variation in Development and Disease

The study of multiple classes of mutations at the single-cell level revealed startling insights into the genomic variation that can occur during the human life cycle. Following single-cell genome-wide analysis, up to 7% [18,42] and up to 70% [43–45] of male and female gametes, respectively, contain numerical chromosomal anomalies due to meiotic mis-segregations. Furthermore, sequencing of haploid single sperm cells revealed a base mutation rate of $2 \times 10^{-6}$—which is severalfold higher than measurements from genome-sequenced pedigree data [46]. Single-cell analyses of human embryos following in vitro fertilization (IVF) demonstrated that the very first cell cycles of human life are prone to numerical and structural chromosome instability [17,25–27,44,47–51]. Various observations indirectly hint that an in vivo conceptus faces a similar period of increased genomic vulnerability [52–55], suggesting that the first cell divisions may represent a spring of DNA mutation, which does not necessarily undermine normal development [8], but can lead to a spectrum of conditions, including loss of conception, genetic disorders, and genetic variation development.

Several studies sequenced and dissected cancer genomes to single-cell resolution, with the aim of understanding tumour development and progression of the disease. Copy number landscapes of single nuclei from primary mammary ductal carcinomas and a paired metastatic liver tumour were generated following low-coverage sequencing. This revealed various chromosomal rearrangements, followed by distinct phases of clonal expansion during tumour evolution and metastasis [56]. Subsequent single-cell exome sequencing studies in bladder [57], kidney [38], and hematopoietic neoplasms [37] provided a detailed characterisation of base mutations in specific genes. Similarly, whole-genome sequencing of multiple MALBAC-amplified cells (Figure 2C) revealed a base mutation rate of a cancer cell line to be increased 10-fold when compared to estimated germ-line ciphers [16]. Furthermore, by sequencing daughter cells of a single mitotic division, the acquisition of new CNVs could be demonstrated for a breast cancer cell line [17].
Figure 2. Overview of single-cell WGA approaches. A) Multiple displacement amplification (MDA) initiates with random priming of denatured single-cell DNA template, followed by a 30°C isothermal amplification using a DNA-polymerase with strand-displacement activity, typically phi29 [113]. When the 3’-end of a newly synthesized fragment reaches the 5’-end of an adjoining primed string of nucleotides, it will displace the latter, liberating single-stranded DNA for new primer annealing and DNA-synthesis. B) Primer extension pre-amplification (PEP)-PCR [114], degenerate-oligonucleotide primed (DOP)-PCR [115], and linker-adaptor (LA)-PCR [116] use PCR for WGA. C) WGA methods like PicoPlex [117,118] and MALBAC [16] use displacement pre-amplification to generate PCR-amplifiable fragments (abbreviated as DA-PCR methods here). Specifically, MALBAC initiates with multiple rounds of displacement pre-amplification using a primer that anneals randomly throughout the genome, but contains a specific sequence allowing full amplicons to form looped pre-amplification products of a cell’s template DNA. This looping protects previously copied segments from further pre-amplification. Multiple rounds of the displacement pre-amplification reaction, interspersed by a denaturation step, increase the probability that random priming will occur across the genome. D) Classes of genetic variation reported in single cells following WGA and analysis. The proofreading capacity of phi29 improves sequence fidelity during WGA [18,37,38,119]. Furthermore, MDA amplifies the majority of a cell’s genome and appears a preferred method for SNP genotyping [15,17,18] or base mutation detection [18,37,38,120], but ADO and PA occurs. Following MDA, single-cell copy number profiles can be distorted [15,17] —although improvements are emerging [19] —and chimeric DNA amplification products are created [17,121]. In general, the (DA-)PCR-based WGA products more accurately preserve the copy number profile of the template genome [15–17] and can be used for SNP genotyping and base mutation detection [16].

doi:10.1371/journal.pgen.1004126.g002
**A**

95% of transcripts

Frequency

Log copy number of transcript/cell

- Quantifiable/detectable transcriptome (10+ copies/cell)
- Noisy/non-reliably detectable transcriptome (1-10 copies/cell)

**B**

Single cell transcriptome analysis

- Blue dots: Single cell
- Green dots: Co-expression
- Red dots: Co-expression

**C**

Bulk analysis

- Blue bars: Bulk analysis
- Green bars: Bulk analysis
- Red bars: Bulk analysis

**D**

Co-expression Matrix (single cell)

- Green circles: Co-expression
- Red circles: No co-expression

**E**

Co-expression Matrix (bulk analysis)

- Green circles: Co-expression
- Red circles: No co-expression

**F**

STRTR Amplification

First Strand Synthesis with Oligo dT + anchor sequence

6-mer barcode

Template Switching during First Strand Synthesis

Pooling of RT products and single-primer PCR amplification of ss cDNA

Capture of cDNA target with streptavidin beads and subsequent cDNA fragmentation

Generation of Illumina sequencing library

**G**

CEL-Seq Amplification

First Strand Synthesis with Barcoded Oligo dT with illumina adaptor sequence

Second Strand Synthesis

Pooling of RT-products (uniquely barcoded) and linear amplification by T7 polymerase

Fragmentation of amplified RNA

Ligation of 3’ adapter

Reverse transcription

PCR amplification of amplicons with both adaptors

**H**

Tang/Surani Method

First Strand Synthesis with Oligo dT + anchor sequence

Poly(A) tailing first strand cDNA

Oligo dT primed Second Strand Synthesis

PCR amplification of cDNA

Fragmentation, Library Prep, Sequencing

**I**

SMARTer Amplification

First Strand Synthesis with Oligo dT + anchor sequence

Template Switching during First Strand Synthesis

PCR amplification of ss cDNA

Generation of full length cDNA

Fragmentation, Library Prep, Sequencing

**J**

5’UTR Exon Exon Exon Exon 3’UTR polyA tail

**Coverage**

- Tang/Surani
- STRTR
- SMARTer
- CEL-Seq
- Ovation
- SMART

*Transcribed 3’ regions of transcripts have also been detected with STRTR

*Enable multiplexed amplification of single-cell RNA due to early barcoding (see panels F and G)*

*Can retain strand information

*Can include random primers for first strand cDNA synthesis, whereas all other WTA methods currently rely on oligo-dT priming
Figure 3. Single-cell transcriptomics. A) A single cell is thought to contain a few hundred thousand mRNA transcripts, present in a log-normal distribution of abundances, with as much as 85% speculated to be present between 1–100 copies. Current amplification methods are thought to reliably detect transcripts at >5–10 copies per cell. B) Single-cell transcriptomes reveal heterogeneity in gene (co)expression that bulk analysis would not permit. Six single cells are shown, with heterogeneous expression of three genes. C) Bulk analysis would detect uniform expression of all three genes. D–E) Single-cell analysis would reveal underlying heterogeneity but also indicate that two of these genes showed a pattern of co-expression. F) STRT-Seq [71,72] is initiated by reverse transcription using an oligo-dT adaptor-primer. At the 5′ cap of the transcript, non-template nucleotides are added by the reverse transcriptase, permitting hybridisation of a barcoded template-switching adaptor-primer. Following pooling of barcoded RT-products, PCR amplification is performed, after which the 5′-ends are captured and sequenced. G) CEL-Seq [73] is initiated using a barcoded oligo-dT primer, which also contains a 5′ adapter sequence and T7 RNA polymerase priming site. Complimentary RNA is generated from the cDNA by T7 RNA polymerase. The cRNA is then fragmented and prepared for (3′-end) paired-end sequencing. H) The Tang/Surani method [74,75], and improved derivatives [69], first generates, then 3′ polyadenylates, first strand cDNA. Priming with adaptor-conjugated oligo-dT generates double stranded cDNA, which is then amplified by PCR and sequenced. I) The SMARTer method [76,79] uses template-switching to generate a full-length transcript with adaptor sequences at both ends. These sequences are then used to prime PCR amplification of the transcriptome. The full-length cDNA is used as input for sequencing. J) Overview of the sequence coverage of a typical transcript which would be obtained by each of the currently available single-cell RNA-seq methods [71–80].

doi:10.1371/journal.pgen.1004126.g003

Single-cell genome sequencing continues to provide new insights into genomic instability in various cell types and developmental processes [12]. It will lead to a better understanding of the acquisition of genetic changes during induced pluripotent stem cell derivation and reprogramming, and to insights in the effects of mutagens, carcinogens, ageing, or germ-line genetic profile on general mutation burden. The methods will enable dissection of the genetic content of individual cells in normal organs, premalignant states, and established tumours, providing insights into the operation of genome maintenance in health and disease.

Single-Cell Genomics in the Clinics

Single-cell genomics is providing cutting-edge clinical applications, notably in the genetic diagnosis of preimplantation human embryos following IVF (reviewed by Van der Aa et al. [58]). Furthermore, developing embryos shed cells in the maternal blood stream following implantation; the potential to capture and analyse such circulating foetal cells [59,60] may broaden the scope and precision of current non-invasive prenatal testing of circulating foetal DNA in maternal plasma. Single-cell genomics is also applied for studying blood-borne circulating tumour cells (CTCs) [61–63], derived from a solid tumour, to investigate the value of CTCs, in addition to tumour-cell-free DNA [64], for guiding diagnosis, prognosis, and treatment of the cancer.

Single-Cell Whole-Transcriptome Amplification: Methods and Limitations

A human cell likely contains less than 1 pg of mRNA [65]. Transcripts are thought to be expressed over several orders of magnitude, where many transcripts have low level expression (5–20 transcript copies per cell) [66], with more than 85% having less than 100 copies per cell (Figure 3A). The transcriptome of an individual cell is not fixed, but reflects the functionality of the cell, as well as its responses to acute extrinsic and intrinsic stimuli. In addition to this “regulated” heterogeneity, there is also transcriptional “noise”—heterogeneity which emerges from the kinetics of transcription and mRNA decay between cells within a population. In gene expression analyses of bulk cell populations, it may be impossible to distinguish changes in expression from changes in the cellular composition of the population. Similarly, genes perceived to be co-expressed at the population level may in fact be mutually exclusively expressed when observed at the single-cell level (Figure 3B–3E).

Single-cell reverse transcription and whole-transcriptome amplification (WTA) methods have been developed to permit qPCR [67–69], microarray [70], and more recently, NGS analyses of the transcriptomes of single cells. Various single-cell RNA-seq methods now exist, each offering an overview of either the 5′-end [71,72], 3′-end [73] or even the full length [74–80] of transcripts from a single cell (Figure 3F–3J).

Reverse transcription, the initial step in each RNA amplification method, and subsequent conversion of cDNA into amplifiable molecules are likely key limiting factors in the detection and quantification of transcripts in single cells. It is estimated that on average only 5–25% of mRNA-molecules are converted to amplifiable cDNA [72]. Additionally, PCR-based amplification methods have the potential for non-linear amplification, resulting in the distortion of the relative abundance of transcripts. In vitro transcription based WTA-methods, such as CEL-Seq [73] (Figure 3G), may arguably avoid such complications through linear amplification of the transcriptome. Furthermore, nanoliter-scale reactions can demonstrate benefits over microliter-scale processes [81].

NGS and Bioinformatics Analyses of Single-Cell Transcriptomes

At the most basic level of analysis, a single-cell RNA-seq experiment gives a readout of the abundance of a transcript within a cell. For 3′- or 5′-end sequencing, this is calculated simply as the number of reads mapping to a particular transcript, normalised to the overall number of reads for that cell. If full length RNA is analysed, the number of reads mapping to each transcript is normalised both for the number of reads per cell and, additionally, for transcript length.

Comparative analyses can be applied to measure differences in normalised gene expression between cells. Genes with heterogeneous expression can be identified by their variability within the population; subsequent clustering of variable genes may allow identification of subsets of genes that are co-expressed within a sub-population of cells. Such approaches have been used to dissect specific “bimodal” gene expression patterns within a population of cells [82] and to distinguish co-expression modules in early embryogenesis [83]. While many of the analytical tools for “bulk” mRNA sequencing are also applied for single-cell data, necessary tools specific for single-cell transcriptomics are starting to emerge [72,84,85].

The broad range of transcript abundance in a single cell presents a particular challenge for any amplification method—transcripts present at extremely low levels can still have important biological consequences, and yet, they may be undetectable due to inefficiency of the amplification approach. Even if they are detectable, the influence of technical noise and stochastic effects at these low levels may result in unreliable measurements of relative...
abundance within or between individual cells. Thus, a major challenge in quantitative analysis of single-cell transcriptome data is understanding technical noise within or between the samples [86]. The inclusion of RNA spike-ins, such as those developed by the External RNA Controls Consortium (ERCC) [87], can give particular insights into the relative efficiency, detection limits, and technical noise of each amplification method [84,85]. Furthermore, single molecule counting approaches—which incorporate a unique identifier into every molecule prior to amplification—will indicate the extent to which individual RNA molecules are amplified [88].

**Insights from Single-Cell Transcriptomics**

Single-cell RNA-seq has already been applied to catalogue allele specific expression and expression of long non-coding RNAs in single blastomeres [74,89] as well as to dissect transcriptional programmes in single cells derived from human and mouse embryos [83,90], revealing insights into the transcriptional modules that are activated at critical points during development. SMARTer WTA [76,79] (Figure 3I) has been used to detect differential exon usage between single cells [82] and to demonstrate a bimodality in gene expression in a phenotypically homogeneous population of bone marrow dendritic cells upon treatment with lipopolysaccharide (LPS) [82]. Here, even genes which were highly expressed were restricted to only a subset of the population—an observation that would have been missed had the population been analysed at a bulk level.

Single-cell transcriptomics has the power to dissect mixed populations of cells; conversely, if only limited material is available, it may permit characterisation of the transcriptome of extremely rare cells, such as CTGs [76].

**The Future: Less Amplification, More Cells, More Types of Data**

Many undesirable consequences of WGA and WTA remain to be solved. The ability to reduce [19] or even eliminate amplification of DNA or RNA before sequencing could increase the accuracy and reliability of single-cell analysis. Input requirements for library preparation continue to reduce, and direct library preparation from single-cell genomes has been demonstrated [39,91]. The capacity to directly sequence unamplified DNA and RNA derived from single cells, however, requires further innovation, though direct sequencing of single molecules is already a possibility for DNA and RNA [92,93]. Furthermore, translation of molecular counting principles to single-cell DNA sequence analyses may allow more accurate measurements of CNVs and enable base-error correction [88,94], in addition to haplotyping approaches. Interpreting the full epigenomic status of a cell remains a challenge, but protocols for single-cell DNA-methylation [95–98] and chromosome conformation capture [99] assays are emerging. Excitingly, methodology to analyse both the (epi)genome and transcriptome of the same cell in parallel is in development and will offer a powerful platform to analyse the exact relationship between genomic variation, regulation, and gene expression.

Typical single-cell sequencing studies have focussed on small numbers of cells (10 s–100 s) but have already demonstrated the potential to distinguish complex heterogeneity at this level. The application of automated cell capture, amplification, and library preparation systems—particularly those utilising nanofluidics approaches—will dramatically increase the scale and affordability of single-cell analyses, such that much larger experiments will emerge.

**Towards a Phylogenetic Tree of a Human Lifetime, and the Discovery of New Cell Types**

Studies like the 1000 Genomes Project have contributed greatly to our understanding of genetic and phenotypic variation amongst individuals within a population. However, these studies are grounded on the assumption that the genome of the individual is “fixed” in tissues throughout life.

Considerable evidence is emerging that somatic genomic variation is both common and dynamic in a human being [12,100–104], although little is known about its scale, origin, rate, and nature. Dedicated bioinformatics analyses can extract only the most prevalent heterogeneities (>5% of cells) from populations of cells, representing likely just the tip of the iceberg. To truly understand the full extent of genomic heterogeneity, from conception to death, single-cell genomes must be investigated.

Large-scale single-cell sequencing projects, performed on cells from endodermal, mesodermal, and ectodermal tissues from an individual, will enable construction of a phylogenetic tree of a human lifetime and mapping of the contribution of genetic heterogeneity to the organism. Concurrent single-cell (epi)genomic/transcriptomic studies, on a large enough scale, will allow definitive sub-classification of cell types by gene expression profiles and (epi)genetic status, replacing or enhancing the current schema. Such studies will reveal, in ways that studies of bulk populations cannot address, the relationship between genome sequence, epigenetic status, and gene expression, determining the functional capacity of the cell.

**Conclusion**

The last few years have seen rapid development of technologies and methods that permit highly detailed analysis of the genome and transcriptome of a single cell. In parallel, various observations have been made that suggest that both genomic and transcriptomic heterogeneity within an organism may have been considerably underestimated. Single-cell approaches now stand poised to illuminate this new layer of biological complexity during normal development and disease.

**References**

1. Nussbaum RL, McInnes RR, Willard HF (2007) Thompson & Thompson Genetics in Medicine. Philadelphia: Saunders Elsevier. 600 p.
2. Strachan T, Andrew R (2010) Human Molecular Genetics. New York: Garland Science. 781 p.
3. Biesecker LG, Spinner NB (2013) A genomic view of mosaicism and human disease. Nat Rev Genet 14: 307–320.
4. Poduri A, Evrony GD, Cai X, Walsh CA (2013) Somatic mutation, genomic variation, and neurological disease. Science 341: 1237758.
5. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. Nature 458: 719–724.
6. De S (2011) Somatic mosaicism in healthy human tissues. Trends Genet 27: 217–223.
7. Kloosterman WP, Guryev V, van Roosmalen M, Duran KJ, de Bruijn E, et al. (2011) Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum Mol Genet 20: 1916–1924.
8. Voet T, Vanneste E, Vermeech JR (2011) The human cleavage stage embryo is a cradle of chromosomal rearrangements. Cytoenet Genome Res 133: 160–168.
9. Dumanski JP, Piotrowski A (2012) Structural genetic variation in the context of somatic mosaicism. Methods Mol Biol 838: 249–272.
10. Kloosterman WP, Cuppen E (2013) Chromothripsis in congenital disorders and cancer: similarities and differences. Curr Opin Cell Biol 25: 341–348.
11. Korbel JO, Campbell PJ (2013) Criteria for inference of chromothripsis in cancer genomes. Cell 152: 1226–1236.
12. McConnell MJ, Lindberg MR, Bremand KJ, Piper JC, Voet T, et al. (2013) Mosaic copy number variation in human brains. Science 342: 632–637.

13. Huang S (2009) Non-genetic heterogeneity of cells in development: more than just noise. Development 136: 3853–3862.

14. Kumar P, Zhang H, Esteek M, Van der As A, Voet T (2013) How to analyse a single blastomere? Application of whole-genome technologies: micro-arrays and next generation sequencing. In: Sermon K, Viville S, editors. Textbook of Human Reproductive Genetics. Cambridge: Cambridge University Press. In press.

15. Treff NR, Su J, Tao X, Northrop LE, Scott RT, Jr. (2011) Single-cell whole-genome amplification technique impacts the accuracy of SNP microarray-based genotyping and copy number analyses. Mol Hum Reprod. 17: 335–345.

16. Zong C, Liu S, Chapman AR, Nie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 335: 1622–1626.

17. Voet T, Kumar P, Van Loo P, Goeke SL, Marshall J, et al. (2013) Single-cell paired-end genome sequencing reveals variable selection procedure per cell cycle. Nucleic Acids Res 41: 6119–6138.

18. Wang J, Fan HC, Behr B, Qaake SR (2012) Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human cells. Cell 153: 40–51.

19. Gole J, Gore A, Richards A, Chiu YJ, Fung HL, et al. (2013) Massively parallel amplification on tiling oligo arrays. Nucleic Acids Res 37: e105.

20. Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, et al. (2009) Real-time combined genotyping and copy number analyses. Mol Hum Reprod. 15: 281–310.

21. Johnson DS, Gerelos G, Baner J, Ryan A, Cinnigou G, et al. (2009) Preclinical validation of a microarray method for full molecular karyotyping of blastomers in a 24-hour protocol. Hum Reprod. 25: 1066–1075.

22. Bi W, Brennan A, Shaw CA, Chapman AR, Nie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 335: 1622–1626.

23. Gele J, Gore A, Richards A, Chinn YJ, Fung HL, et al. (2013) Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microarrays. Nat Biotechnol 31: 1129–1132.

24. Greig JB, O’Hear AC, Waldhausen-Gill J, Hoffmann EM, Auer M, et al. (2009) Identification of small gains and losses in single cells after whole genome amplification on tiling oligo arrays. Nucleic Acids Res 37: e105.

25. Alfarawati S, Fragouli E, Colls P, Wells D (2011) First births after preimplantation genetic diagnosis of aneuploidy: the BRADY study. Hum Reprod 26: 1560–1574.

26. Fiorentino F, Spizzichino L, Bono S, Birlicik A, Kokkali G, et al. (2011) PGD of familial haplotyping and embryo analysis for Preimplantation Genetic Diagnosis (PGD) using DNA microarrays: a proof of principle study. J Assist Reprod Genet 30: 1595–1603.

27. Xu X, Hou Y, Yin X, Bao L, Tang A, et al. (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148: 885–896.

28. Falconer E, Hills M, Naumann U, Poos SS, Chavez EA, et al. (2012) DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. Nat Methods 9: 1107–1112.

29. Hills M, O’Neill K, Falconer E, Brinkman R, Landsorp PM (2013) BAIT: Organizing genomes and mapping rearrangements in single cells. Genome Med 5: 82.

30. Mertzanidou A, Spitz C, Nguyen HT, Van de Velde H, Sermon K (2013) Evolution of aneuploidy up to Day 4 of human preimplantation development. Hum Reprod 28: 1716–1724.

31. Treff NR, Su J, Tao X, Liu S, Wang Y, Yang M, Li J, et al. (2012) Probing mitotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. Science 338: 1627–1630.

32. Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, et al. (2012) Genome-wide copy number profiling in individual hematopoietic progenitors by digital RT-PCR. Proc Natl Acad Sci U S A 109: 21083–21088.

33. Ni X, Zhuo M, Su Z, Duan J, Gao Y, et al. (2013) Reproducible copy number analysis of single cells. Nat Protoc 7: 1024–1041.

34. Treff NR, Su J, Taylor D, Scott RT, Jr. (2011) Telomere DNA deficiency is associated with development of human embryonic aneuploidy. PLoS Genet 7: e1002161. doi:10.1371/journal.pgen.1002161

35. Alfarawati S, Fragouli E, Colls P, Wells D (2012) Embryos of robertsonian translocation carriers exhibit a mitotic interchromosomal effect that enhances genetic instability during early development. PLoS Genet 8: e1002952. doi:10.1371/journal.pgen.1002952

36. Navin N (1999) Cytogetics of Spontaneous Abortion. In: Greisen S, Keagle M, editors. The Principles of Clinical Cytogenetics. New Jersey: Humana Press. pp. 317–343.

37. Macklon NS, Gerards JP, Fauser BC (2002) Conception to ongoing pregnancy: the ‘black box’ of early pregnancy loss. Hum Reprod Update 8: 333–343.

38. Santos MA, Kuikj EW, Macklon NS (2010) The impact of ovarian stimulation for IVF on the developing embryo. Reproduction 139: 23–34.

39. Robberecht C, Voet T, Zamani Esteki M, Nowakowska BA, Vermes CH (2013) Nonallelic homologous recombination between retrotransposons elements is a driver of the novo unbalanced translocations. Genome Res 23: 411–418.

40. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, et al. (2011) Tumour evolution inferred by single-cell sequencing. Nature 472: 90–94.

41. Li Y, Xu X, Song I, Hou Y, Li Z, et al. (2012) Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. GigaScience 1: 12.

42. Van der As A, Zamani Esteki M, Vermes CH, Voet T (2013) Preimplantation genetic diagnosis guided by single-cell genomics. Genome Med 5: 71.

43. Elia S, Price J, Dockter M, Wachtel S, Thorapal A, et al. (1992) First trimester prenatal diagnosis of trisomy 21 in fetal cells from maternal blood. Lancet 340: 1033.

44. Simpson JL (2013) Cell-free fetal DNA and maternal serum analytes for monitoring embryonic and fetal status. Fetal Steril 99: 1124–1134.

45. Heitzer E, Auer M, Gauch C, Pichler M, Ulu P, et al. (2013) Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. Cancer Res 73: 2965–2975.

46. Ni X, Zhao M, Su Z, Duan J, Gao Y, et al. (2012) Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. Proc Natl Acad Sci U S A 110: 21803–21808.

47. Swennenhuis JF, Reumers J, Thys K, Aersens J, Terstappen LW (2013) Efficiency of whole genome amplification of Single Circulating Tumor Cells enriched by CellSearch and sorted by FACS. Genome Med 5: 106.

48. Navin N, Hicks J (2011) Future medical applications of single-cell sequencing in cancer. Genome Med 3: 31.

49. Nakazaki ES (2004) Microarrays and the gene expression profile of a single cell. Ann N Y Acad Sci 1020: 92–100.

50. Subkhankulova T, Gilchrist MJ, Livesey FJ (2008) Modelling and measuring the transcriptome among phenotypically identical cells. BMC Genomics 9: 22.

51. Alfarawati S, Fragouli E, Colls P, Wells D (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148: 885–896.

52. Pflueger S (1999) Cytogenetics of Spontaneous Abortion. In: Greisen S, Keagle M, editors. The Principles of Clinical Cytogenetics. New Jersey: Humana Press. pp. 317–343.
94. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, et al. (2012) Detection of
91. Falconer E, Lansdorp PM (2013) Strand-seq: A unifying tool for studies of
87. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, et al. (2011) Synthetic spike-
86. Marinov GK, Williams BA, McCue K, Schroth GP, Gertz J, et al. (2013) From
85. Katayama S, Tohonen V, Linnarsson S, Kere J (2013) SAMstrt: statistical test
83. Xue Z, Huang K, Cai C, Cai L, Jiang CY, et al. (2013) Genetic programs in
79. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, et al. (2013)
78. Pan X, Durrett RE, Zhu H, Tanaka Y, Li Y, et al. (2013) Two methods for full-
77. Kohn AB, Moroz TP, Barnes JP, Netherton M, Moroz LL (2013) Single-cell
75. Tang F, Barbacioru C, Nordman E, Lee C, et al. (2012) Full-length mRNA-Seq
74. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. (2009) mRNA-Seq
72. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, et al. (2012) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-Seq. Genome Res 21: 1160–1167.
71. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, et al. (2012) Highly multiplexed and strand-specific single-cell RNA 5’ end sequencing. Nat Protoc 7: 813–828.
69. Huang H, Goto M, Tsunoda H, Sun L, Taniguchi K, et al. (2013) Non-biased and efficient global amplification of a single-cell cDNA library. Nucleic Acids Res. In press.
68. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2012) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
67. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
66. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
65. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
64. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
63. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
62. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
61. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
60. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
59. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.
58. Guo G, Kuang M, Kuang H, Wang L, Li X, et al. (2011) Novel amplification strategy for full-length RNA-seq. Nucleic Acids Res 40: 2112–2124.