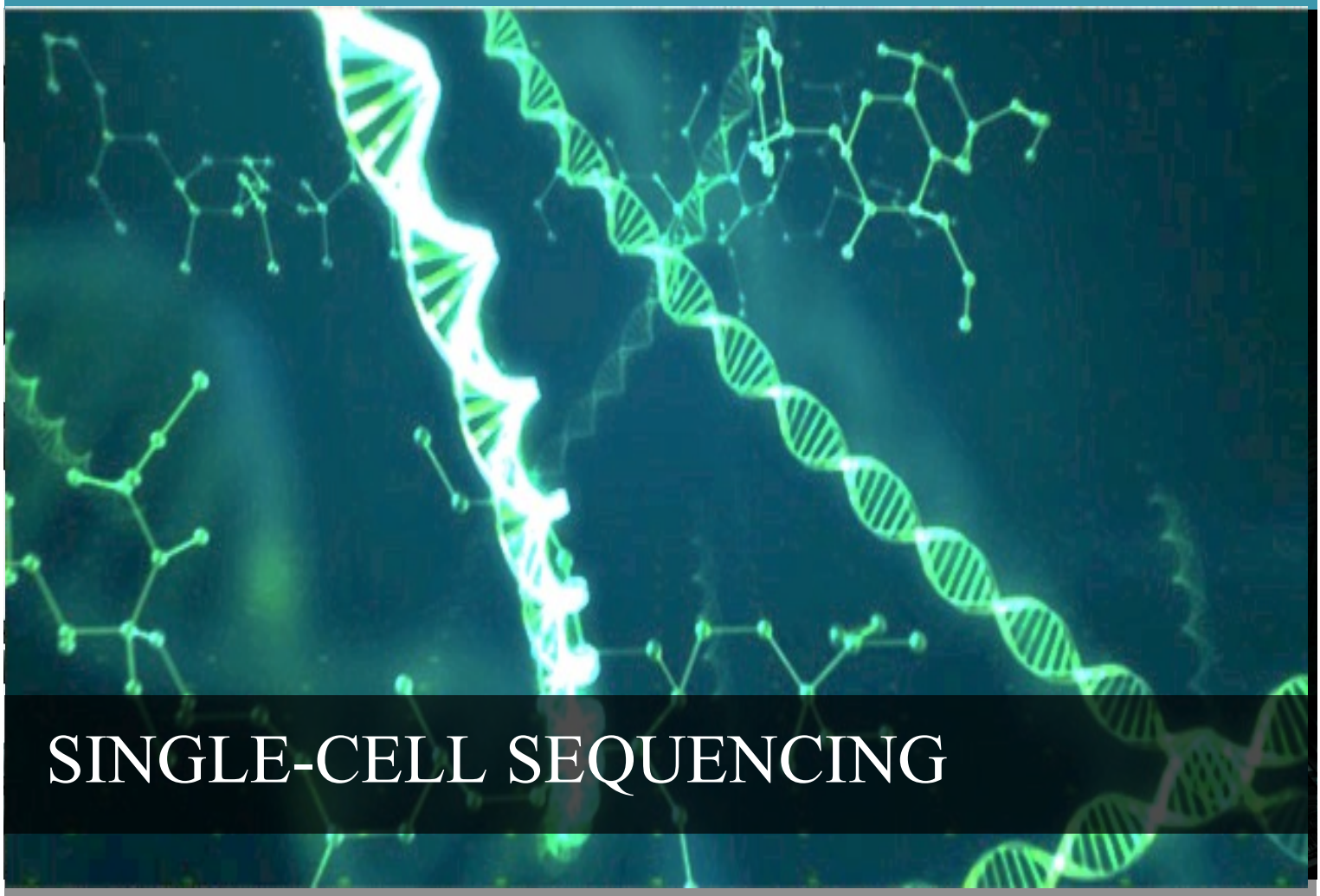

SINGLE-CELL SEQUENCING

Jaena Han
Professor: Doug Brutlag, PhD

Biochem 158: Genomics, Bioinformatics, and Medicine

Jaena Han
Biochem 158

Introduction

Once considered a technical feat only possible for a few specialized labs, single-cell sequencing is rapidly becoming more robust and broadly accessible.[12] Like typical next generation sequencing experiments, the protocols of a single-cell sequencing generally adhere to the following steps: isolation of single cell, nucleic acid extraction and amplification, sequencing library preparation, sequencing and bioinformatic data analysis (Fig. 1).[15] Sequencing using one cell is much more difficult than sequencing from cells in bulk, as the minimal amount of starting materials make degradation, sample loss and contamination exert pronounced effects on quality of sequencing data.[4,5,12,15] The greatest challenge of scaling sequencing down to the cellular level is capturing miniscule amounts of template and amplifying them to generate enough material for high-throughput sequencing.[10,15,20]

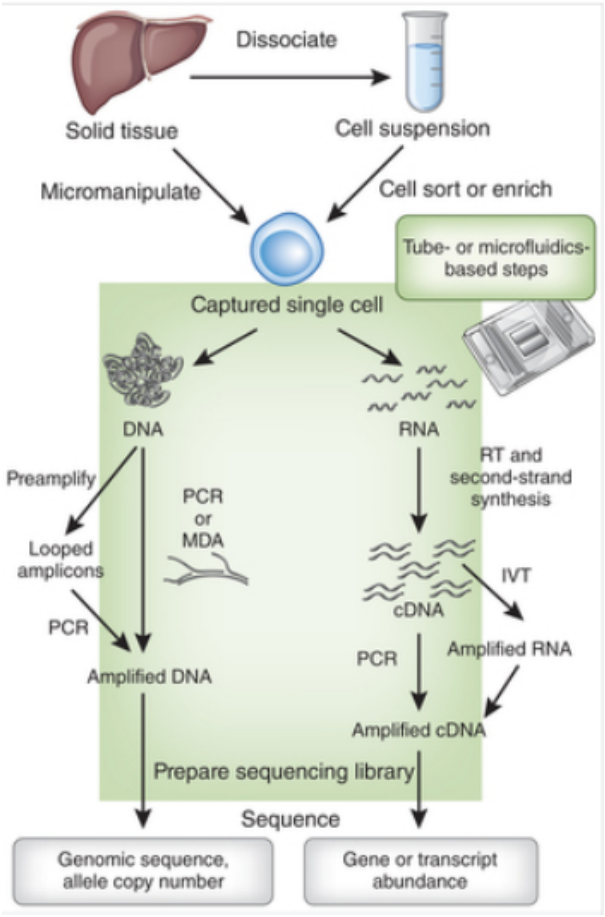

Figure 1: Workflow for amplifying and sequencing cellular DNA and RNA. RT: Reverse Transcription, IVT: in vitro transcription[15]

The amount of DNA or RNA in a single cell starts at a few picograms—not even close to the amount required by current sequencing machines.[4] Consequently, scientists must amplify these molecules, and in ways that minimize errors while surveying sequences as broadly and evenly as possible.[4] The cost and difficulty of maintaining fidelity and avoiding biases during heavy amplification is no simple task, but protocol improvements and increased commercial options are now facilitating the widespread adoption of single-cell sequencing approaches.[12] In 2013, Nature even selected single-cell sequencing as its Method of the Year, stating that "methods to sequence the DNA and RNA of single cells are poised to transform many areas of biology and medicine" through a "fresh wave of biological insights into development, cancer and neuroscience."[12]

Single-cell research in the past has relied principally upon microscopy to differentiate cell populations based upon phenotypic characteristics.[5] However, a growing body of scientific research has demonstrated that cells of the same size, shape and immunophenotype can still harbor significant differences in genomic variation and expression.[5,12,17] Now, novel methods of single-cell sequencing are enabling researchers to define new molecular profiles for subpopulations and to monitor these changes as cells are activated, develop, and respond to environmental stimuli, thereby revealing a complex ecology of heterogeneous cell states that coalesce to produce emergent system-level function.[5,16]

A typical human cell contains approximately 6 billion base pairs of DNA and 600 million bases of mRNA—an immense coding capacity.[5] Deep sequencing of DNA and RNA from single cells can read these blueprints for cellular function more comprehensively and at higher resolution than was previously possible.[5] DNA and RNA sequencing performed on tissue samples or cell populations can mask these significant biological differences between cells by averaging them or mistaking them for technical noise.[12] Single-cell sequencing methods offer a way to dissect this heterogeneity.[4,5] These approaches have already revealing a surprising level of mosaicism in somatic tissues such as the brain, and may lead to a better understanding of why some cells degenerate while adjacent cells are normal, or why some cells respond to certain drugs while others do not, ultimately providing critical insights into cell function and dysfunction within a micro-environmental context.[5]

Differences between cells are often even more pronounced at the RNA level, even within such seemingly uniform populations as immune cells purified on the basis of cell-surface markers.[12] Single-cell transcriptome profiling can identify biologically relevant differences in cells, even when cells may not be distinguishable by marker genes or cell morphology.[5,12] Moreover, single-cell sequencing makes rare cells and microbes more accessible to analysis, (provided that these cells can be isolated or enriched from their heterogeneous environments) and can be clinically applied to enhance preimplantation screening of embryos fertilized *in vitro*.[12]

Single cell sequencing promises to provide critical insights into diverse areas of biology and medicine, but its potential applications for cancer research and treatment are particularly evident.[13] Single-cell sequencing can reveal mutations and structural changes in the genomes of cancer cells, which can be used to describe the clonal structure of tumors and to trace the

evolution and spread of the disease, and can also be clinically applied to cancer diagnostics based on rare circulating tumor cells.[13]

Cancer Heterogeneity

Molecular methods are invaluable for cancer research due to the sheer breadth of information that can be obtained from a single tumor sample.[13] Microarrays are capable of analyzing thousands of transcripts or millions of single nucleotide polymorphisms (SNPs), and next-generation sequencing (NGS) can reveal genetic abnormalities at base pair resolution.[9,13] However, because most current procedures require bulk DNA or RNA from over 100,000 cells, they can only provide global information on the average state of the population of cells.[4,5,12,13]

This limitation is troubling, as solid tumors are complex mixtures of cells including non-cancerous fibroblasts, endothelial cells, lymphocytes, and macrophages that often contribute more than 50% of the total DNA or RNA extracted, and can complicate molecular classification by masking the signal from the cancer cells.[13] In addition, solid tumors are often composed of multiple clonal subpopulations, further confounding the analysis of clinical samples.[14,18] When multiple clones are present in a tumor, standard molecular assays reflect either an average signal of the population or only the signal from the dominant clone, which is not always the most malignant.[13] This becomes particularly important as molecular assays are employed for directing targeted therapy, as in the use of *ERBB2* (*Her2-neu*) gene amplification to identify patients likely to respond to Herceptin treatment in breast cancer, where 5-30% of all patients have been reported to exhibit such genetic heterogeneity.[19]

Sequential tumor analysis has also demonstrated that intratumor genetic heterogeneity dynamically evolves throughout the course of the disease.[17] For instance, in a 2011 *Nature* study by Navin *et al*. profiling the genomes of single cells for copy number variations revealed a punctuated model of tumor evolution involving bursts of genomic instability following a stable expansion of tumor mass rather than gradual accumulation of mutations, as was previously believed.[14] Such novel insights demonstrate how powerful single-cell methodologies can be for understanding the genomics of human cancers. This also has consequences for prognostic biomarker strategies, as the tumor subclone that may ultimately influence a patient's therapeutic outcome may evade detection because of its absence or low scarcity at diagnosis or because of its distance from the tumor biopsy site.[17]

The clinical implications do not end there. A study by Maley et al. correlated heterogeneity with the development of malignancy by demonstrating a strong correlation between the degree of clonal diversity and the probability of invasive progression of Barrett esophagus to esophageal lesions.[11] Other studies have shown that mutational status can predict resistance to treatment by preventing the binding of a drug to its target. For instance, mutant forms of the BCR-ABL fusion protein, which normally acts as a predictor a patient's response to imatinib mesylate, has been implicated in the relapse of chronic myeloid leukemia.[21] This pattern has also been observed in other tumor types such as melanoma and gastrointestinal stromal sarcomas[21], providing additional evidence that intratumor heterogeneity is closely tied to tumor progression and the development of resistance to therapy.

Isolating single cells

Studying a single cell requires isolating it from a tissue sample or cell culture in a way that preserves biological integrity. There are two key steps in the isolation process: 1) the tissue must be removed from the organism—typically by dissection or biopsy—and broken down into its constituent cells, often through enzymatic disaggregation, and 2) single cells must be placed into individual reaction chambers for lysis and further processing.[16]

Currently available methods for accomplishing this include micromanipulation, laser-capture microdissection (LCM) and flow cytometry (Fig. 2a-c).[13] Micromanipulation of individual cells using a transfer pipette or serial dilution is a readily available—albeit labor-intensive and error-prone—method and has been used for isolating single cells from culture or liquid samples such as sperm, saliva or blood.[13,16] Many tissues can be dissociated to produce cell suspensions, which are easier to handle and allow cells expressing specific markers to be enriched with a cell sorter.[13] Instruments that trap very rare cells on the basis of their surface markers are also being used to isolate tumor cells that circulate in blood.[13]

LCM allows individual cells to be isolated directly from tissue sections, making it clinically suitable.[13] This method requires tissues to be sectioned, mounted and stained so that they can be visualized during the isolation process.[13] LCM is useful for genetic profiling because it allows single cells to be isolated directly from morphological structures, such as ducts or lobules in the breast, but has drawbacks, in that some nuclei will inevitably be sliced in the course of tissue sectioning, causing loss of chromosome segments and generating artifacts in the

data.[13] Moreover, although LSM preserves the knowledge of the spatial location of a sampled cell within a tissue (unlike other methods involving tissue dissagregation), it is hard to capture a whole single cell without also collecting the materials from neighboring cells.[13,16]

Flow cytometry using fluorescence-activated cell sorting (FACS) is by far the most efficient method for isolating large numbers of single cells from liquid suspensions. Although it requires expensive appliances, FACS is readily available at most hospitals and research institutions, and is used routinely to sort cells from hematopoietic cancers.[13] Several instruments have even been optimized for sorting single cells into 96-well plates for subcloning cell cultures.[13] FACS has the added advantage that cells can be labeled with fluorescent antibodies or nuclear stains and sorted into different fractions for future analysis.[13] However, the downside is that FACS requires a large number of suspended cells as starting material, which could affect the yield in low-abundance cell subpopulations.[16] Rapid flow in the machine can also damage cells, which, producing difficulties if living cells are necessary for downstream protocols.[16] Most importantly, FACS and other microfluidic devices first require detaching cells from their microenvironments, which will perturb transcriptional states and affecting the results of RNA expression analysis.[13]

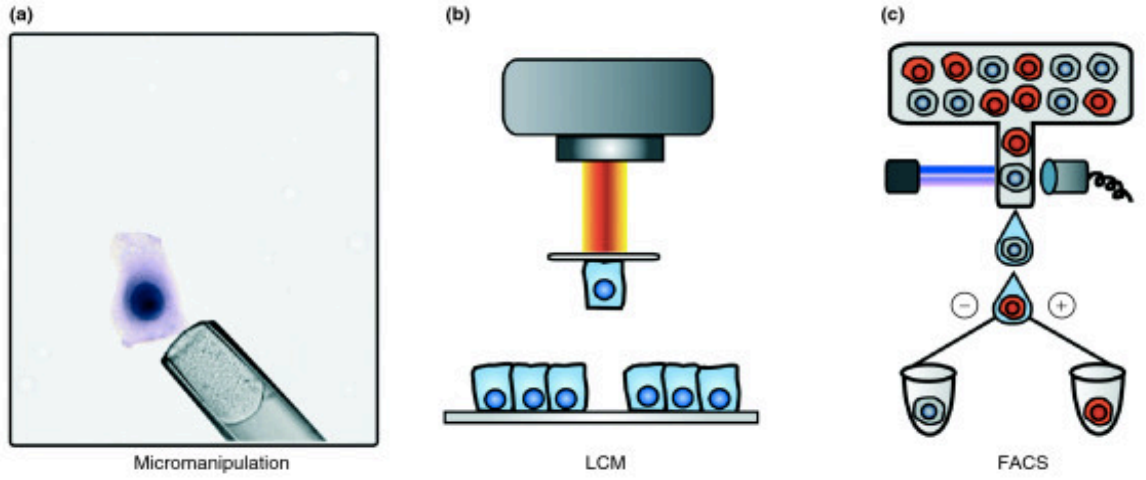

Figure 2: Techniques for isolating singe cells. (a) micromanipulation, (b) laser-capture micro-dissection (LCM), (c) fluorescence-activated cell sorting (FACS)[13]

The development of single cell protocols

Various approaches have been used to measure genome-wide information of individual cells, such as cytological approaches, array comparative genome hybridization (aCGH), and

single-cell sequencing (Fig. 4a-c). Cytological methods such as spectral karyotyping, fluorescence in situ hybridization (FISH) and Giemsa staining emerged in the 1970s in the fields of cytology and immunology and enabled the first qualitative analysis of genomic rearrangements in single tumor cells (Fig. 4a).[13] The advent of PCR in the 1980s enabled immunologists to investigate genomic rearrangements by directly amplifying and sequencing DNA from single cells.[13] Together, these tools provided the first insight into the genetic heterogeneity of solid tumors.[13]

However, PCR was limited in that it could not amplify the entire human genome in a single reaction, even with when augmented by advances such as primer extension pre-amplification.[13] A major milestone occurred with the discovery of two incredibly processive DNA polymerases (Phi29 and Bst polymerase), which could amplify the human genome over 1,000-fold through a mechanism called multiple displacement amplification (MDA).[9,13] In this reaction, random primers are extended by the phi29 DNA polymerase. As the polymerase extends the growing DNA strands, it displaces downstream strands resulting in a branching form of amplification.[9,10,13] This approach is now commercially available and has been actively used in whole genome amplification (WGA), as it reduces amplification bias by 3-4 orders of magnitude and yields more uniform coverage of the genome than previous PCR-based methods.[9,12,20]

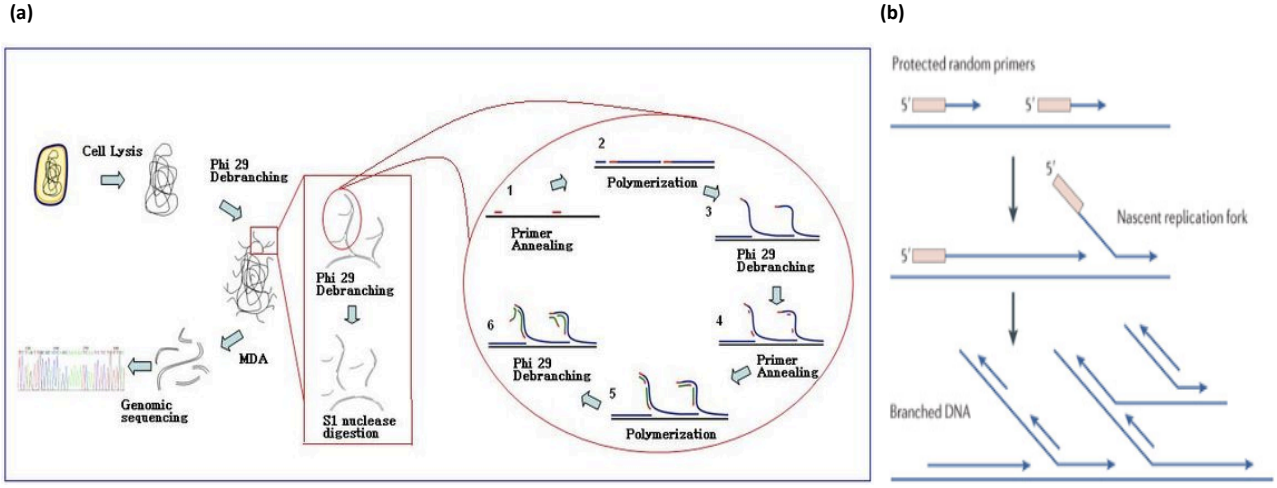

Figure 3: Multiple displacement amplification (MDA). (a) Schematic overview. Source: Qianli Ma (b) As the MDA reaction proceeds, repeated strand displacement and priming causes the hyper-branched network to expand dramatically, generating thousands or even millions of copies of the original DNA.[9]

Coupling WGA with array CGH enabled the measurement of genomic copy number in small populations of cells, and even single cells of various cancer types, including circulating tumor cells (CTCs), as well as both colon and renal cancer cell lines (Fig. 4b).[9,13] However, these studies had limited resolution and reproducibility, particularly since WGA does not amplify uniformly across the genome.[13,20] Rather, WGA fragments amplified from single cells are sparsely distributed across the genome, representing less than 10% of the unique human sequence.[9,13,20] As a result, there is no coverage for up to 90% of probes, resulting in decreased signal to noise ratios and high standard deviations in copy number signal.[9,13]

An alternative approach is to use next generation sequencing (NGS), which provides a non-targeted approach to sample the genome (Fig 4c).[13-16] Instead of differential hybridization to specific probes, sequence reads are integrated over contiguous and sequential lengths of the genome and all amplified sequences are used to calculate copy number.[13,16] A combination of NGS with FACS and WGA in a method called single-nucleus sequencing (SNS) has made it possible to calculate absolute copy number from single cells.[13,14] In 2011, Nicholas Navin and Michael Wigler's group at Cold Spring Harbor Laboratory, profiled large deletions or duplications of DNA called copy-number variants (CNVs) at 50-kilobase resolutions across the genomes of breast tumor cells from two individuals.[14]

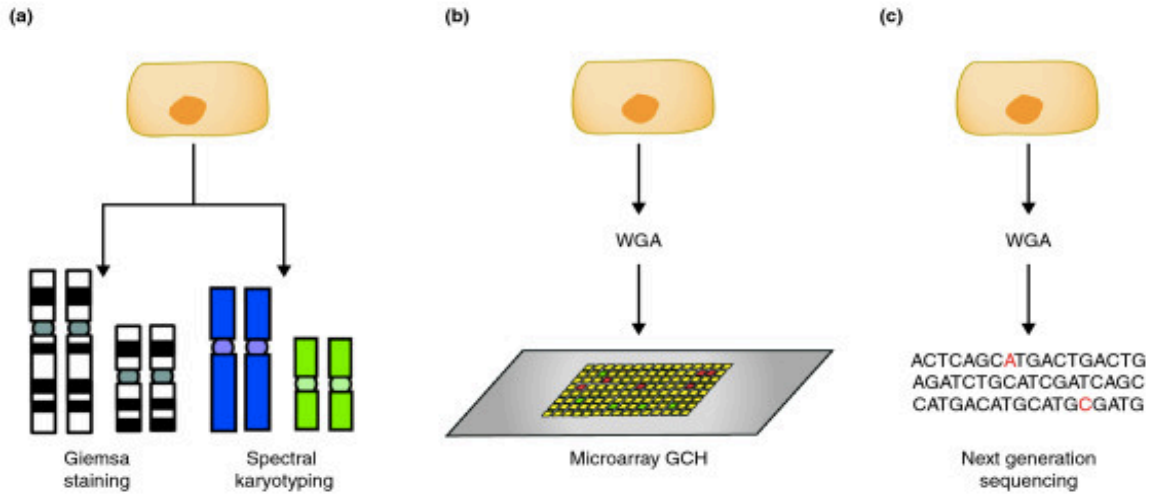

Figure 4: Techniques for genetic profiling. (a) cytological methods to visualize chromosomes in single cells, (b) whole genome based amplification (WGA) and microarray comparative genomic hybridization (CGH), (c) WGA and next generation sequencing (NGS)[13]

MALBAC

In a recent paper in *Science*, Zong *et al.* presented a new method called multiple annealing and looping-based amplification cycles (MALBAC), and demonstrated that it amplifies genomic DNA from single human cells with far less bias than previous methods.[10,20] This approach enhances the sequencing of both parental chromosomes with even coverage, increasing the accuracy with which heterozygous single-nucleotide variations can be genotyped.[10,20] MALBAC directly addresses the underlying causes of amplification bias by combining features of MDA and PCR (Fig. 5).[20] Primers consisting of eight nucleotides of random sequence and a common 27-nucleotide tag are annealed to template DNA and are extended by a polymerase in an isothermal strand-displacement reaction, similar to MDA.[20]

Unlike MDA, however, MALBAC employs a "quasilinear" amplification consisting of repeated short cycles of primer extension followed by denaturation at 94 °C so that initial priming events are more-evenly distributed over the course of multiple cycles.[20] In addition, the ends of the newly made amplicons base pair by means of the 27-nt sequence common to each primer, forming loops that inhibit the 3′ ends from serving as primers, thereby keeping the amplification process linear.[20] After five cycles of this, the resulting DNA is amplified by PCR to a level sufficient for sequencing.[10] Using MALBAC, Zong et al. obtained enough coverage to sequence 93% of the human genome and detect CNVs in a single cancer cell.[20]

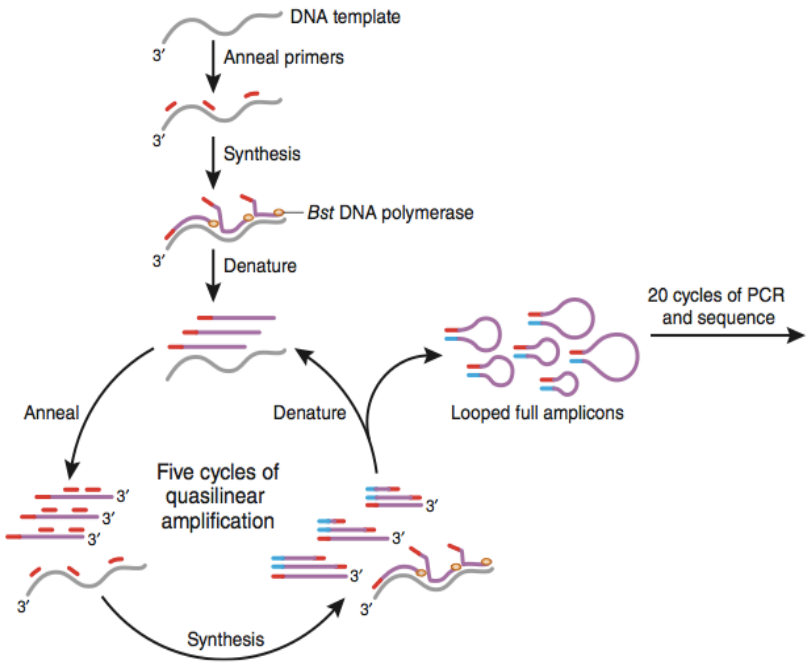

Figure 5: A simplified diagram of the MALBAC reaction[10]

Detailed studies concerning the reaction mechanisms of MALBAC or the role each of the individual steps plays in reducing amplification bias have yet to be conducted.[10] For instance, it is unclear how looping alters the kinetics of the reaction.[10] It will be important to evaluate how

much of the bias reduction is attributable to the cycling of initial priming or to the loop structure of the amplicons, particularly since these features could potentially be applied to enhance priming evenness in MDA and other existing methods.[10]

Additional research is necessary to properly investigate MALBAC's performance and applicability. Although the procedure has a low false-negative rate for detecting heterozygous loci, the false-positive rate was found to be about 40 fold higher than that of MDA, primarily due to the misincorporation of nucleotides by DNA polymerases.[20] This difference is likely because MDA uses the high- fidelity phi29 DNA polymerase while MALBAC uses two more error-prone DNA polymerases: the large fragment of *Bacillus stearothermophilus* (*Bst*) DNA polymerase for isothermal strand displacement and Taq DNA polymerase for PCR.[10,20]

To compensate for false-positives, Zong *et al.* sequenced several single cells from a highly homogeneous cell culture of clonal kindred cells.[20] DNA from three of these cells was amplified by MALBAC and analyzed by high-throughput sequencing.[20] Sequences of single cells were compared with the predominant sequence of a bulk DNA extraction from the culture, enabling the identification of rare cell-specific alleles.[20] Unless variants were found in all three cells, they were considered to be false positives.[20] Thus, it appears that statistical strategies using multiple single cells will still be required for many studies using MALBAC.[10] However, it is unclear how well this particular analysis can be carried out with more complex cell populations such as tumor cells.[10] There may also be ways to reduce the false-positive error rate by utilizing polymerases with higher-fidelity (e.g. phi29) or strong proofreading activity.[10] The 20 cycles of PCR that are required might also be optimized to reduce error rates, thereby maximizing coverage while minimizing uneven representation and error amplification that result from repeated cycles of PCR.[10]

A second limitation of MALBAC is that about 20–30% of known single-nucleotide variants could not be accurately genotyped even when multiple cells were pooled in a single MALBAC reaction, likely because sequences in certain contexts such as regions of secondary structure are difficult for DNA polymerases to transcribe.[10,20] A difficult template sequence for either of the DNA polymerases used would result in poor recovery in the final MALBAC product.[10,20] Research strategies implementing the MALBAC protocol would greatly benefit from a published analysis of the sequences in the human genome that are consistently underrepresented.[10]

Clinical Applications

Despite the recent advances and excitement surrounding it, single-cell sequencing is far from a routine experimental procedure and will require substantial improvements in reductions in the cost and time of sequencing before it can become useful in a clinical setting.[5] Fortunately, the cost of DNA sequencing is declining rapidly due to a combination of industrial competition and technological innovation.[12] Sequencing also has incredible potential for massive of samples using barcoding strategies.[6,13] Barcoding involves adding a specific 4-6 base oligonucleotide sequence to each library as it is amplified, so that samples can be pooled together in a single sequencing reaction, after which the reads are deconvoluted by their barcodes for downstream analysis.[6] With the current throughput of the Illumina HiSeq2000, it is possible to sequence up to 25 single cells on a single-flow cell lane, allowing 200 cells to be profiled in a single run.[13] Furthermore, by decreasing the genomic resolution of each single-cell copy number profile, it becomes feasible to profile hundreds of cells in parallel on a single lane, or thousands on a run, making single-cell profiling economically viable for clinical applications.[6,13]

Cell Lineage Reconstruction

Key questions regarding disease progression and responsiveness to treatment in cancer patients remain unanswered despite decades of intense research and scrutiny, such as the origin of metastases or cause of cancer relapse after therapy.[13] Finding these answers will likely depend on a greater understanding of the understanding the emergence and distribution of driver mutations in the context of a patient's cancer cell lineage tree[17]—an approach that can be greatly facilitated by single cell sequencing.

In one study by Anderson *et al*., heterogeneity and tumor origin in acute lymphoblastic leukaemia were studied by assaying the occurrence of chromosomal abnormalities in individual cells using fluorescence *in situ* hybridization (FISH), enabling an analysis of subclonal architecture during cancer progression.[2] More recently, sequencing of hundreds of single nuclei was used to approximate copy-number profiles for single breast cancer cells, permitting the reconstruction of tumor evolutionary history and population structure.[14]

Detection of rare tumor cells

A major application of single-cell sequencing will likely be in the detection of rare tumor cells in clinical samples with fewer than a hundred available cells, such body fluids (e.g. lymph, blood, sputum, urine, or vaginal or prostate fluid), or clinical biopsy samples.[13] Patients with breast cancer often undergo fine-needle aspirates, nipple aspiration, ductal lavages or core biopsies, but genomic analysis is rarely applied to these samples due to limited DNA or RNA.[13] In early stage breast cancers, which are detected by these methods, typically only 5% to 10% of patients progress to invasive carcinomas, making it difficult for oncologists to determine how aggressively to treat individual patient.[8] Measuring tumor heterogeneity in these limited clinical samples using single cell sequencing may provide critical prognostic information on whether these tumors will evolve into invasive carcinomas, and could lead to more informed oncological treatment decisions.

Screening for rare tumor cells in sputum, sperm ejaculates, or vaginal fluid may also improve the early detection of lung, prostate, and ovarian cancers, respectively.[13] The genomic profile of these cells may provide useful information on the lineage of the cell and the organ of origin. Moreover, if the genomic copy number profiles of rare tumor cells accurately represent the genetic lesions in the primary tumor, they may also provide an opportunity for targeted therapy.[13] Prior studies have reported that classes of genomic copy number profiles correlate with survival[7], and that the profiles of rare tumor cells may consequently have predictive value in assessing the severity of the primary cancer from which they originated.[13]

Diagnostic value of CDCs and DTCs

Another significant clinical application of single-cell sequencing may lie in the genomic profiling of copy number or sequence mutations in circulating tumor cells (CTCs) and disseminated tumor cells (DTCs). Although whole genome sequencing of single CTCs or DTCs has yet to be achieved, data obtained from future methods may one day provide important information for monitoring and diagnosing cancer patients.[13] Unlike other circulating cells, CTCs often contain unique epithelial surface markers which could potentially enable non-invasive “fluid biopsies” that would provide both an indication of cancer activity in a patient as well as genetic information that could direct the course of treatment.[13] Investigating CTCs with single-cell sequencing has the advantages of allowing the resolution of impure mixtures and the analysis of scant amounts of DNA.[13] Even a single CTC in a typical 7.5 ml blood draw can be

examined to construct a genomic profile of copy number variations, allowing physicians to identify genomic amplifications of oncogenes and deletions of tumor suppressors.[13] By profiling multiple samples from the primary tumor, secondary metastases and CTCs of patients, it also becomes possible to trace the evolutionary lineage of cancer cells and infer their site of origin.[17] Recent studies have shown that many patients with non-metastatic primary tumors also show evidence of CTCs.[1,3] If these cells share the mutational profile of the primary tumors from which they originate, they could provide a powerful, non-invasive approach for identifying early signs of cancer.[13] In the future, detecting and profiling CTCs or DTCs in normal patients—perhaps even as part of a routine checkup—may facilitate detection of primary tumors, which could then be located for biopsy and treatment using magnetic resonance imaging or computed tomography.[13] CTC monitoring in metastatic patients would also have important applications in monitoring residual disease after adjuvant therapy to ensure that the cancer does not return.[13]

Response to chemotherapy

Tumor heterogeneity may be a decisive factor in a patient's response to chemotherapy. Due to evolutionary pressures, tumors with the most diverse allele frequencies have the highest probability of surviving a catastrophic selection pressure such as a cytotoxic agent or targeted therapy.[17] The question of whether resistant clones exist prior to treatment in the primary tumor or emerge in response to adjuvant therapy by acquiring *de novo* mutations has important implications for cancer treatment and has previously been investigated in genetic profiling studies of cervical and ovarian cancers.[13] Although both studies detected some tumors with pre-existing resistant subpopulations that increased following treatment, the fact that the data was derived from population-based signals from millions of cells suggests that these results may underestimate the total frequency of resistant clones in primary tumors.[13] As such, these questions are better addressed using single-cell sequencing methods, which can provide a fuller picture of the extent of intratumor genomic heterogeneity.

Knowing the degree of genomic heterogeneity may also provide useful prognostic information for patients given the option of chemotherapy.[13,16] Theoretically, patients with monogenomic tumors will respond better to treatment compared to patients with polygenomic tumors, as the latter has a higher probability of developing or having resistant clones.[13,16,17] Awareness of the degree of intratumor heterogeneity (which may serve as a proxy for the

efficacy of treatment) could help patients decide whether or not it is worth receiving chemotherapy, given the brutal side effects associated with the procedure.[13] Moreover, single-cell sequencing can in principle also provide a higher sensitivity for detecting rare chemoresistant clones in primary tumors, further increasing its prognostic value in that respect.[13,16]

Conclusions

While genomic analysis of bulk tissue samples can provide a broad overview of a patient's cellular aberrations, they are unable to determine whether the cells in a tumor contain the same full set of mutations or different subsets of mutations that in combination drive cancer progression.[4,5,13,16] The more sensitive measurements provided by single-cell sequencing can overcome these research barriers to offer an unparalleled view of tumor evolution and heterogeneity and provide a means of detecting and analyzing the genomes of rare cancer cells. The medical applications are diverse, ranging from tumor cell lineage reconstruction to early cancer detection to CTC monitoring during treatment of metastatic patients.[13,16] The development of single-cell protocols also has the potential to catalyze decisive jumps in genomic copy number profiling by eliminating the locus restrictions of FISH probes, thereby enabling the rapid identification of thousands of cancer genes to better inform treatment decisions.[13]

Some major hurdles in applying single-cell methods to clinical practice will be to innovate multiplexing strategies that allow rapid and inexpensive profiling of large numbers of individual cells, and to develop these methods for paraffin-embedded (as opposed to frozen) tissues, since many samples are clinically processed in this manner.[13] Despite these challenges, single cell technologies are maturing at a rapid pace, and their application in clinical settings is likely to improve all three major themes of oncology: detection, progression, and prediction of therapeutic efficacy.[12,13] As the cost and ease of examining individual cells continues to improve, single-cell sequencing will become increasingly available to researchers and clinicians as a standard tool for understanding cancer biology at unprecedented resolution.

References

1. Allard, W. J. "Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but Not in Healthy Subjects or Patients With Nonmalignant Diseases." *Clinical Cancer Research* 10.20 (2004): 6897-904. Print.

2. Anderson, Kristina, Christoph Lutz, Frederik W. Van Delft, Caroline M. Bateman, Yanping Guo, Susan M. Colman, Helena Kempski, Anthony V. Moorman, Ian Titley, John Swansbury, Lyndal Kearney, Tariq Enver, and Mel Greaves. "Genetic Variegation of Clonal Architecture and Propagating Cells in Leukaemia." *Nature* 469.7330 (2010): 356-61. Print.

3. Bidard, F.C., C. Mathiot, S. Delaloge, E. Brain, S. Giachetti, P. De Cremoux, M. Marty, and J.-Y. Pierga. "Single Circulating Tumor Cell Detection and Overall Survival in Nonmetastatic Breast Cancer." Annals of Oncology 21.4 (2010): 729-33. Print.

4. Chi, Kelly Rae. "Singled out for Sequencing." *Nature Methods* 11.1 (2013): 13-17. *Nature.com*. Nature, 30 Dec. 2013. Web. 12 Mar. 2014. <http://www.nature.com/nmeth/journal/v11/n1/full/nmeth.2768.html>.

5. Eberwine, James, Jai-Yoon Sul, Tamas Bartfai, and Junhyong Kim. "The Promise of Single-cell Sequencing." *Nature Methods* 11.1 (2013): 25-27. *Nature.com*. Nature, 30 Dec. 2013. Web. 12 Feb. 2014. <http://www.nature.com/nmeth/journal/v11/n1/full/nmeth.2769.html>.

6. Erlich, Y., K. Chang, A. Gordon, R. Ronen, O. Navon, M. Rooks, and G. J. Hannon. "DNA Sudoku--harnessing High-throughput Sequencing for Multiplexed Specimen Analysis." *Genome Research* 19.7 (2009): 1243-253. *Genome Research*. Cold Spring Harbor Laboratory Press, 2009. Web. 13 Feb. 2014. <http://genome.cshlp.org/content/19/7/1243.long>.

7. Hicks, J., A. Krasnitz, B. Lakshmi, N. E. Navin, M. Riggs, E. Leibu, D. Esposito, J. Alexander, J. Troge, V. Grubor, S. Yoon, M. Wigler, K. Ye, A.-L. Borresen-Dale, B. Naume, E. Schlicting, L. Norton, T. Hagerstrom, L. Skoog, G. Auer, S. Maner, P. Lundin, and A. Zetterberg. "Novel Patterns of Genome Rearrangement and Their Association with Survival in Breast Cancer." *Genome Research* 16.12 (2006): 1465-479. *Genome Research*. Cold Spring Harbor Laboratory Press, 2006. Web. 13 Mar. 2014. <http://genome.cshlp.org/content/16/12/1465.long>.

8. Kerlikowske, K. "Characteristics Associated With Recurrence Among Women With Ductal Carcinoma In Situ Treated by Lumpectomy." *CancerSpectrum Knowledge Environment* 95.22 (2003): 1692-702. *Journal of the National Cancer Institute*. Oxford Journals, 26 Sept. 2003. Web. 13 Mar. 2014. <http://jnci.oxfordjournals.org/content/95/22/1692.long>.

9. Lage, Jose M. "Whole Genome Analysis of Genetic Alterations in Small DNA Samples Using Hyperbranched Strand Displacement Amplification and Array-CGH." *Genome Research* 13.2 (2003): 294-307. *Genome Research*. Cold Spring Harbor Laboratory Press, 2003. Web. 13 Mar. 2014. <http://genome.cshlp.org/content/13/2/294.long>.

10. Lasken, Roger S. "Single-cell Sequencing in Its Prime." *Nature Biotechnology* 31.3 (2013): 211-12. *Http://bernstein.harvard.edu*. Harvard University, Mar. 2013. Web. 12 Mar. 2014. <http://bernstein.harvard.edu/pages/news/nbt.2523.pdf>.

11. Maley, Carlo C., Patricia C. Galipeau, Jennifer C. Finley, V. Jon Wongsurawat, Xiaohong Li, Carissa A. Sanchez, Thomas G. Paulson, Patricia L. Blount, Rosa-Ana Risques, Peter S. Rabinovitch, and Brian J. Reid. "Genetic Clonal Diversity Predicts Progression to Esophageal Adenocarcinoma." *Nature Genetics* 38.4 (2006): 468-73. *Http://evolution.binghamton.edu*. 2006. Web. 13 Mar. 2014. <http://evolution.binghamton.edu/evos/wp-content/uploads/2009/08/Maley2.pdf>.

12. "Method of the Year 2013." *Nature Methods* 11.1 (2013): 1. *Nature.com*. Nature, 30 Dec. 2013. Web. 12 Mar. 2014. <http://www.nature.com/nmeth/journal/v11/n1/full/nmeth.2801.html>.

13. Navin, Nicholas, and James Hicks. "Future Medical Applications of Single-cell Sequencing in Cancer." *Genome Medicine* 3.5 (2011): 31. *Genomemedicine.com*. Genome Medicine, 31 May 2011. Web. 11 Mar. 2014. <http://genomemedicine.com/content/3/5/31#sec5>.

14. Navin, Nicholas, Jude Kendall, Jennifer Troge, Peter Andrews, Linda Rodgers, Jeanne Mcindoo, Kerry Cook, Asya Stepansky, Dan Levy, Diane Esposito, Lakshmi Muthuswamy, Alex Krasnitz, W. Richard Mccombie, James Hicks, and Michael Wigler. "Tumour Evolution Inferred by Single-cell Sequencing." Nature 472.7341 (2011): 90-94. Print.

15. Nawy, Tal. "Single-cell Sequencing." *Nature Methods* 11.1 (2013): 18. *Www.nature.com*. Nature, 30 Dec. 2013. Web. 12 Mar. 2014. <http://www.nature.com/nmeth/journal/v11/n1/full/nmeth.2771.html>.

16. Shapiro, Ehud, Tamir Biezuner, and Sten Linnarsson. "Single-cell Sequencing-based Technologies Will Revolutionize Whole-organism Science." Nature Reviews Genetics 14.9 (2013): 618-30. Print.

17. Swanton, Charles. "Intratumor Heterogeneity: Evolution through Space and Time." *Journal of Cancer Research* 74.5 (2012): n. pag. *Cancer Research*. American Association for Cancer Research, 1 Oct. 2012. Web. 12 Mar. 2014. <http://cancerres.aacrjournals.org/content/72/19/4875.full>.

18. Torres, Lurdes, Franclim R. Ribeiro, Nikos Pandis, Johan A. Andersen, Sverre Heim, and Manuel R. Teixeira. "Intratumor Genomic Heterogeneity in Breast Cancer with Clonal

Divergence between Primary Carcinomas and Lymph Node Metastases." *Breast Cancer Research and Treatment* 102.2 (2007): 143-55. Print.

19. Vance, Gail H., Todd S. Barry, Kenneth J. Bloom, Patrick L. Fitzgibbons, David G. Hicks, Robert B. Jenkins, Diane L. Pearsons, Raymond R. Tubbs, and Elizabeth H. Hammond. "Genetic Heterogeneity in HER2 Testing in Breast Cancer: Panel Summary and Guidelines." *Archives of Pathology and Laboratory Medicine* 133.4 (2009): 611-12. *Archivesofpathology.org*. College of American Pathologists, 2009. Web. 13 Mar. 2014. <http://www.archivesofpathology.org/doi/full/10.1043/1543-2165-133.4.611>.

20. Zong, C., S. Lu, A. R. Chapman, and X. S. Xie. "Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell." *Science* 338.6114 (2012): 1622-626. *Http://www.ncbi.nlm.nih.gov/*. NCBI, 21 Dec. 2012. Web. 12 Feb. 2014. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3600412/>.

21. Farhangfar, Carol J., F. Meric-Bernstam, J. Mendelson, G.B. Mills, and A.K. Lucio-Eterovic. "The Impact of Tumor Heterogeneity on Patient Treatment Decisions." *Clinical Chemistry* 59.1 (2013): 38-40. www.*clinchem*.org/. 9, Nov, 2012. Web. 12 Mar. 2014. <http://www.clinchem.org/content/59/1/38.full.pdf+html>.