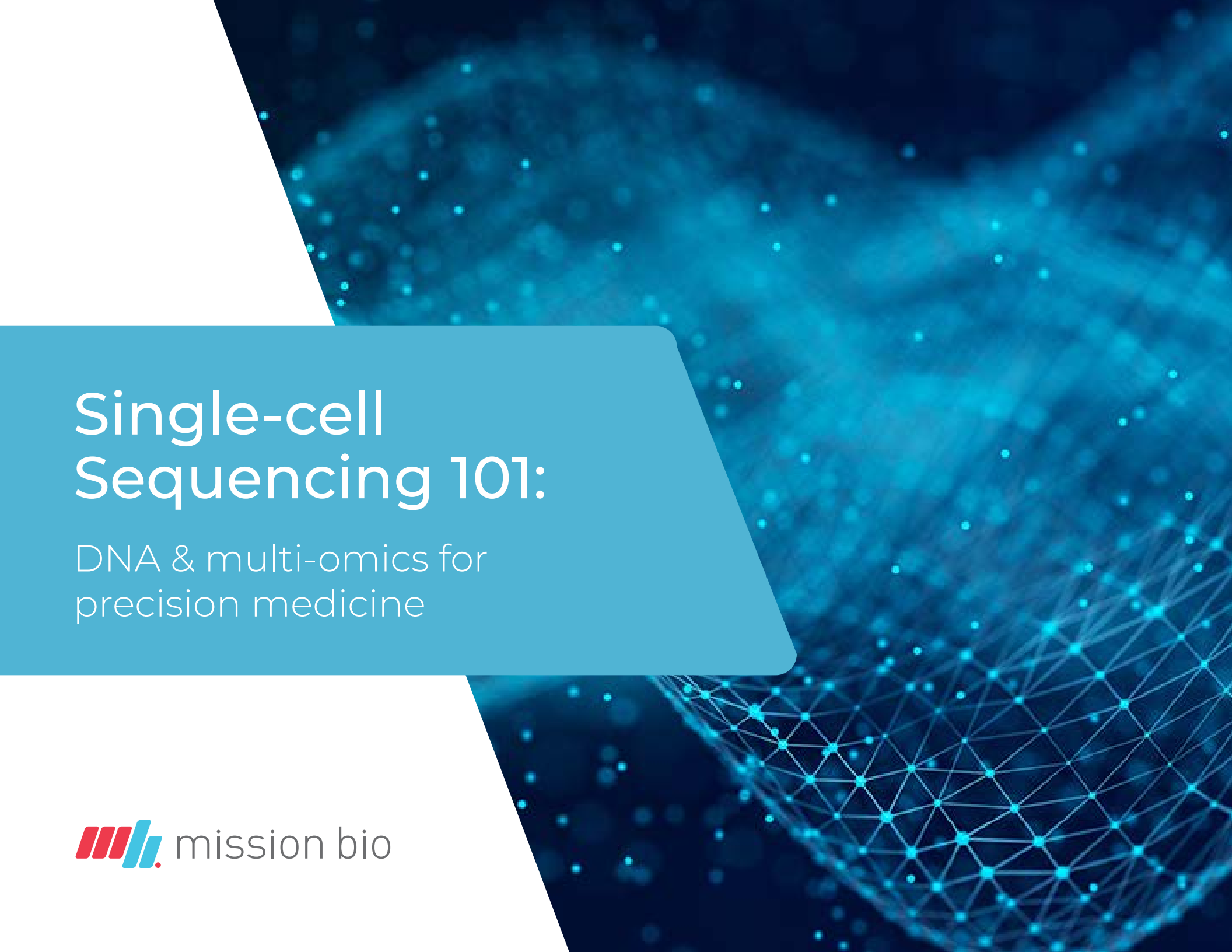

Single-cell Sequencing 101:

DNA & multi-omics for precision medicine

mission bio

Table of Contents

Introduction

Complex and elusive, cancer has been the focus of intense research for decades. Not only is this disease the second leading cause of mortality in the USA, but it pervades virtually every population worldwide.[1] The challenge of cancer has been met with an enormous effort to identify the factors that underlie the onset, progression, and relapse of the disease. This global endeavor and concurrent advances in biotechnologies have led to many breakthroughs in developing new therapies and enhancing treatment strategies.

However, significant hurdles remain. A major challenge to defeating cancer is its continual evolution and inherent heterogeneity. Each tumor is a unique melting pot of clones that harbor different sets of mutations, and this variation has significant implications for how tumors grow and respond to therapy. For instance, the presence of a few cells that are resistant to treatment can bring about relapse after a tumor has been seemingly eradicated. Thus, characterizing the underlying clonal architecture of tumors is key to developing cures.

Delving deeper into cancer requires a progressive approach that preserves the resolution of each cell. Currently, most clinics rely on bulk NGS metrics that average mutation frequencies across a tumor sample. Using these measurements to determine clonal structure is severely limited and based on inference. To truly uncover the complexity of cancer, we need a "more powerful microscope" (Fig 1).

Fortunately, innovations in single-cell technologies have enabled us to peer deeper into cancer with unprecedented resolution and sensitivity. With single-cell DNA sequencing (scDNA-seq), the genetic structure of tumors can be fully resolved. Furthermore, powerful single-cell multi-omics can now correlate genotype and immunophenotype — enabling the comprehensive characterizations of tumor cells. These innovations have particular promise in the translational space, where fundamental research on cancer complexity can directly inform actions in the clinic. Additionally, single-cell analyses can accelerate current pharmacological workflows, such as those for cell and gene therapies. With the ability to obtain genotype and phenotype information in a cell-specific manner, the era of single-cell 'omics has opened the door to a "resolution revolution."

In this eBook, we explore how single-cell DNA sequencing and multi-omics are rapidly elevating how cancer is both studied and treated. These technologies are being quickly adopted in the hematological oncology field, as liquid tumors are easier to sample and study over time. While most of our examples involve heme malignancies, we also discuss the development of this technology for solid tumors.

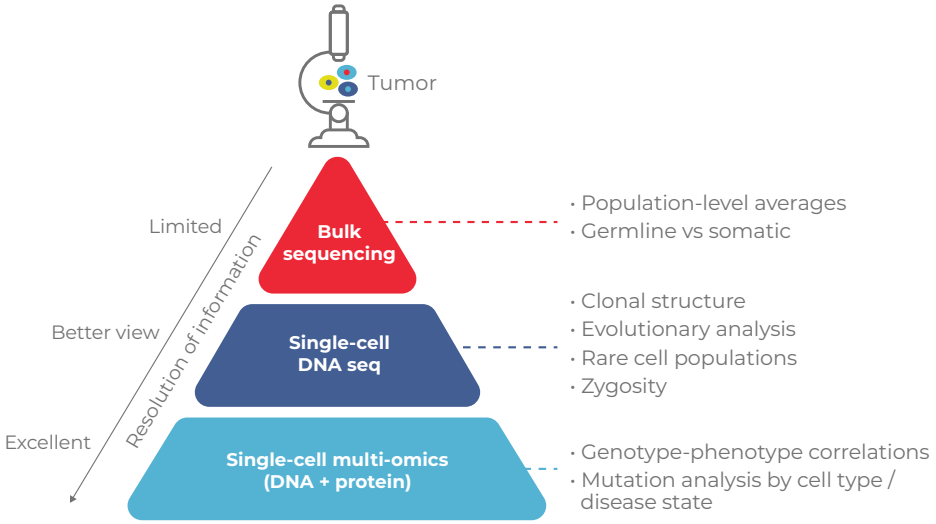

Figure 1. Peering deeper into cancer.

Different technologies facilitate varying levels of resolution for uncovering cancer complexity. Whereas bulk sequencing portrays only population-level metrics, single-cell analyses reveal fine-grained aspects of clonal architecture.

Cancer is an Evolving Disease

Similar to Darwinian evolution at the species level, cancer is subject to natural selection and other evolutionary phenomena. Because genetic instability is a common feature of cancer cells, mutations are commonplace. Selection acts on this genetic variation, enabling cells with favorable phenotypes to propagate. Through an iterative process of mutation and cell expansion, tumors generate genetic and phenotypic heterogeneity. This cell-to-cell variation contributes to the complexity of the disease. In this section, we explore how tumors emerge and change over time.

Clonal Evolution Affects Tumor Heterogeneity

The path to cancer is often a protracted process with mutations accruing over years before a cell becomes malignant. Once the "right" combination of mutations initiates neoplastic change, however, a tumor begins to grow. There are two main models of tumor evolution: linear and branched.[1] In **linear evolution**, each new driver mutation enables a selective sweep, in which the mutant cell — called a **clone** —outcompetes cells that don't have the mutation. Mutations are accumulated in a cumulative and stepwise fashion as each new clone replaces the one before it (Fig 2). Note that in this model of evolution, the cell population is largely homogeneous, with most cells containing the same **clonal mutation(s)** (i.e., passed down from a common parental clone).

In contrast, tumors may undergo **branched evolution**, a process in which different cells independently acquire mutations (called **subclonal mutations**) that confer a fitness advantage. In this case, the population does not turn over with each new clone. Rather, multiple **subclones** co-exist and evolve in parallel. This process

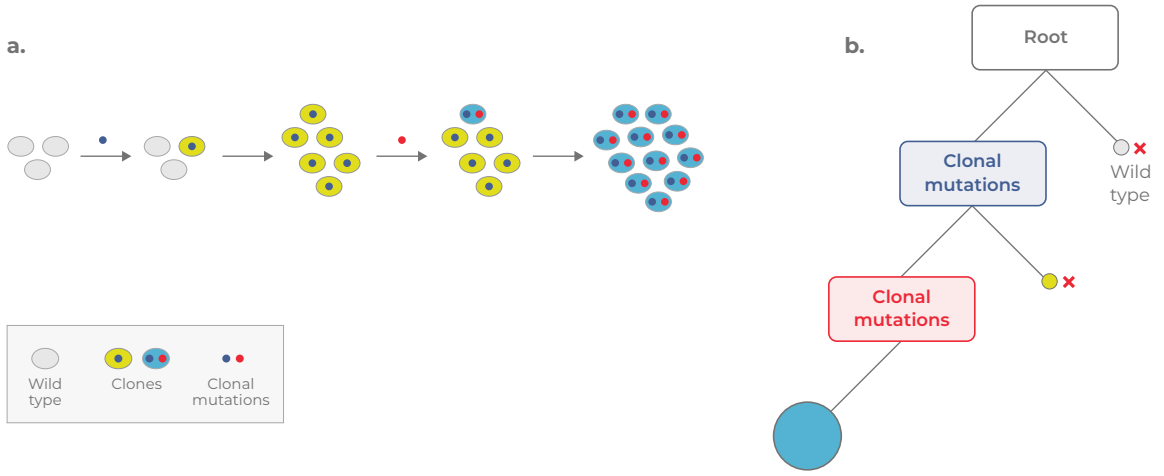

Figure 2. Linear evolution.

a. In linear evolution, each mutation confers a strong selective advantage, enabling each clone to replace the clone preceding it. **b.** The corresponding phylogenetic tree appears linear. Mutations are acquired in a cumulative fashion, with the most recent clone (blue circle) containing all of the mutations acquired previously.

generates intratumoral diversity over time. To get a better idea of how this occurs, let's walk through an example.

Consider a single healthy cell that acquires a driver mutation (e.g., *IDH1*) that enables its expansion. This cell is called a **founding clone** because it initiates tumorigenesis. As the clone propagates it gives rise to identical cells that contain the same clonal mutation. On a tumor phylogenetic tree, the founding clone resides on the "trunk" (Fig 3).

At a subsequent time point, a cell in the founding population may acquire a subclonal mutation, giving rise to a new subclone that propagates in parallel. Over time, more subclones may branch off, increasing the heterogeneity of the tumor. These genetically distinct subpopulations are represented by the "branches" on the tumor phylogenetic tree. In the example portrayed in Figure 3, three subclones co-exist together, with one **dominant clone** comprising the largest fraction of the tumor.

The evolutionary journey of each tumor is unique, with multiple factors determining the pattern of evolution (linear vs branched) and level of heterogeneity at any given point in time. Natural selection continues to act on the cells in the context of their microenvironment. Clonal populations with favorable phenotypes will increase in number, while those with neutral or deleterious characteristics remain stable or die out altogether.

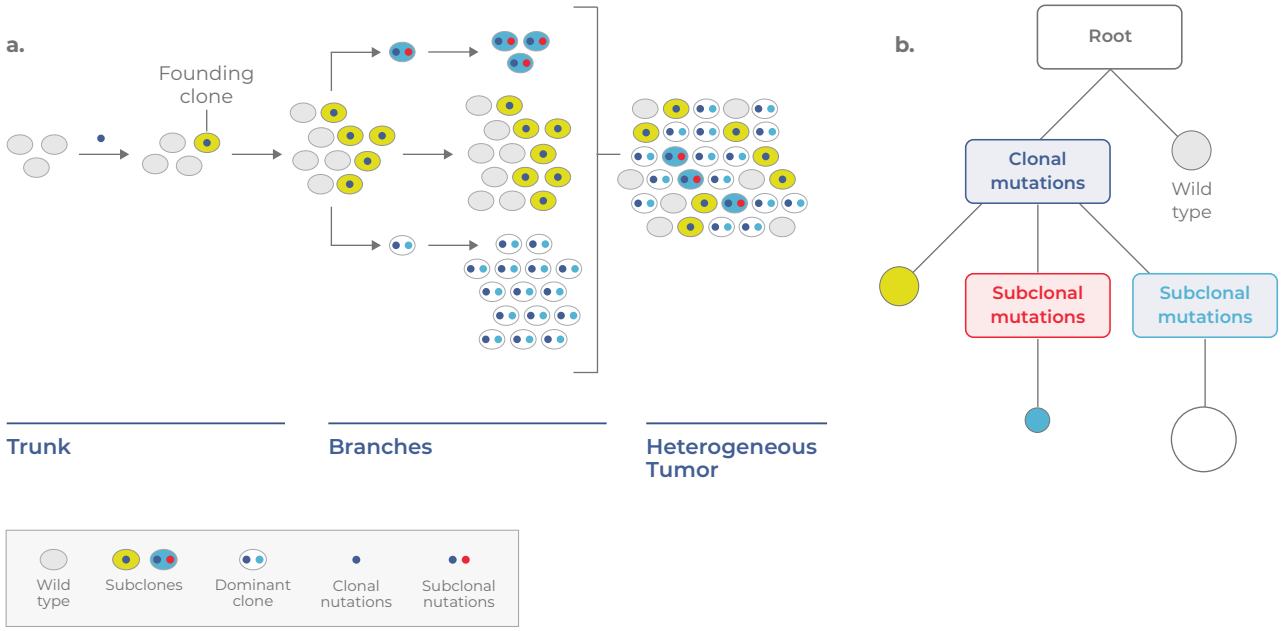

Figure 3. Branched evolution.

a. In branched evolution, subclones acquire advantageous mutations and evolve in parallel. Because different subclones harbor different sets of mutations, their presence increases the genetic heterogeneity of the tumor. **b.** The corresponding phylogenetic tree appears branched, with coexisting subclones containing different combinations of mutations.

Genetic drift and environmental changes also impact tumor composition. The introduction of an anti-cancer treatment into the tumor environment, for example, can dramatically remodel tumors by killing some cells and enabling the growth of others (we will explore this more in chapter 3). Overall, the dynamic evolution of tumors leads to their ever-changing **clonal architecture**.

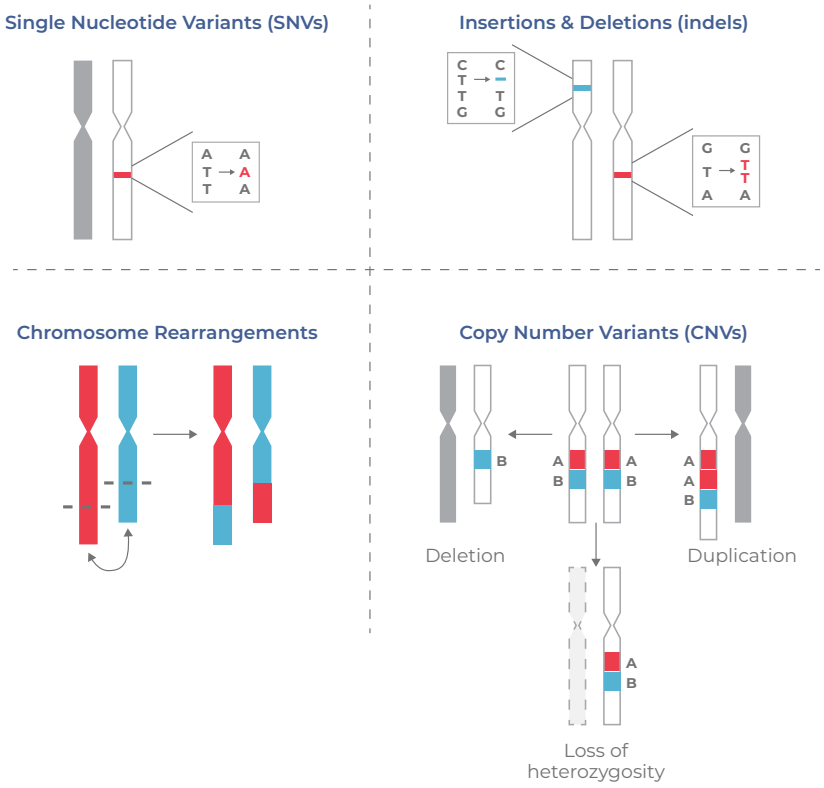

Figure 4. Types of DNA aberrations detectable by single-cell DNA sequencing.

Cancer is associated with a variety of DNA aberrations, from small mutations (SNVs, indels) to chromosomal rearrangements like translocations and gene fusions. Copy number variations (CNVs) of genes or entire chromosomes also contribute to tumor heterogeneity.

Mutation Type Contributes to Tumor Diversity

Variability in the *types* of mutation across cells also contributes to tumor heterogeneity. Tumors cells may contain a variety of lesions, ranging from small-scale mutations, like **single-nucleotide variants (SNVs)** and **insertions/ deletions (indels)** to large-scale chromosomal aberrations like **translocations** and **gene fusions** (Fig 4).

Although common in healthy cells, **copy number variants (CNVs)** occur more frequently in cancer. Types of CNVs include deletions, duplications, as well as **loss of heterozygosity (LOH)** (i.e. only one copy of an allele is present). Altering the number of gene copies — whether duplicating an oncogene or deleting a tumor suppressor — can result in uncontrolled cell proliferation.

As one might expect, cell-to-cell variation in mutation type and CNV increases the diversity of clones. This information may be clinically relevant, as the copy number of a particular gene may inform how virulent or drug-resistant a particular subclone is. For instance, the amplification of the oncogene *CKS1B* in multiple myeloma (MM) patients is associated with a poorer response to the chemotherapy, cisplatin.[2] Knowing how genetic lesions vary across tumor cells thus gives a rare view into the strengths and weaknesses of tumors.

Clonal Architecture and Evolution Has Clinical Implications

Many studies to date indicate that the evolutionary trajectories of cancer are complex leading to diverse clonal architecture. For instance, various whole-genome studies suggest that myeloid malignancies including acute myeloid leukemia (AML), MDS, MM, myelodysplastic syndrome (MDS), and chronic lymphocytic leukemia (CLL) have a branched pattern of evolution.[3]

This complexity makes it a formidable foe. Studies have shown that higher clonal diversity is associated with greater disease severity and poorer survivorship.[4] Moreover, this variation often exists not only within a tumor, but also across tumors within a patient, and across patients in a population.

But untangling the complexity of cancer can even the playing field. Advances in single-cell technology are enabling researchers to precisely reconstruct tumor phylogenies, enabling the identification of the main mutational drivers of disease (covered in Chapter 3).[4,5] Tumor heterogeneity also has profound implications for diagnosis, prognosis, and treatment of cancer. But before we jump into these applications, let's first review how single-cell DNA sequencing and multi-omics are conducted.

VOCABULARY

Driver mutation: a mutation that is causally implicated in oncogenesis (gives the cell a growth advantage).

Clone: cells that are genetically identical.

Founding clone: clone that acquires the tumor-initiating mutation.

Subclone: a clone that is descended from another clone but has acquired additional mutation(s).

Dominant clone: the clonal population that occurs at the highest frequency in the tumor.

Clonal architecture: the subclones that comprise a tumor and their frequencies.

Clonal evolution: the process by which different subclones within a tumor change over time. Clonal evolution is shaped by stochastic factors and natural selection.

Single nucleotide variants (SNVs): variants characterized by a substitution of one nucleotide at a specific genomic position.

Insertions/ deletions (indels): insertion or deletion of nucleotides into the genome.

Copy number variants (CNVs): variants characterized by the number of repeated sections of the genome (deletions/ duplications, loss of heterozygosity)

Loss of Heterozygosity (LOH): the loss of an allele from a heterozygous state, leaving only one allelic copy.

Translocation: a piece of one chromosome breaks off and attaches to another.

Gene fusion: a translocation event that places two formerly separate genes adjacent to one another. Some fusion genes encode fusion proteins that are implicated in cancer.

Uncovering the Architecture of Cancer with Single-cell Multi-omics

The heterogeneity and dynamism of cancer present formidable challenges to understanding and treating the disease. As discussed in the last section, evolution often leads to considerable genetic variation across clones. But the complexity of tumors does not stop at the level of DNA. Intratumoral phenotypic variation adds another layer of information that can be useful for research and clinical purposes. In this section, we discuss how single-cell DNA sequencing resolves variation at the genetic level, and how a powerful multi-omics approach correlates genotype and immunophenotype.

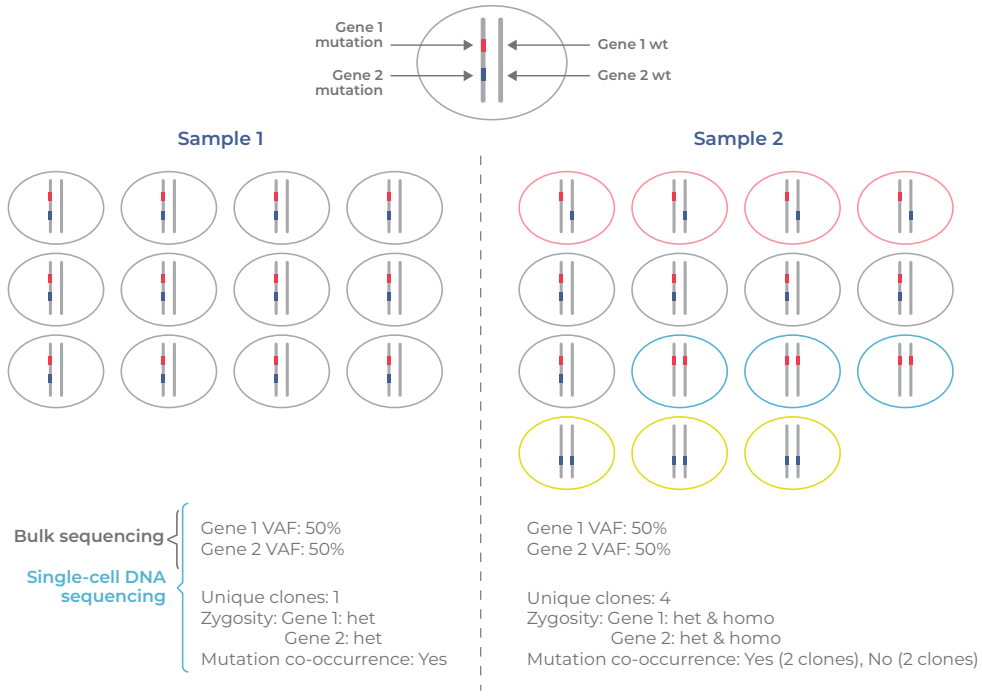

Figure 5. Bulk NGS vs scDNA-seq information on clonal structure.

Consider two tumor samples (1 and 2) with clones that contain mutations in gene 1 (dark blue) and/or gene 2 (red). Sample 1 contains a single clone whereas sample 2 contains four unique clones (red, gray, blue, yellow) that differ in gene mutations and zygosity. According to bulk-calculated VAFs, both samples appear identical. Only single-cell sequencing identifies the differences in structure between the samples with regard to clone number, mutational co-occurrence, and zygosity.

Single-cell DNA Sequencing Unveils the Genetic Variation

Deciphering the patterns of somatic mutations across clonal populations provides foundational insights into cancer. Traditionally, oncologists have relied on **bulk next-generation sequencing (bulk NGS)** to identify gene mutations present in tumor cells. This technique involves pooling the DNA across cells and measuring the **variant allele frequencies (VAFs)**, i.e., the percentages of variant DNA molecules. Bulk sequencing is useful to confirm whether candidate genes are mutated and whether these mutations are germline (inherited) or somatic in origin. But because the cell identities are not preserved, bulk NGS is unable to measure patterns of co-mutation or resolve cell-to-cell variation in zygosity and copy number (Fig 5). This leaves researchers and clinicians with a black box when it comes to understanding the clonal diversity in their sample.

Fortunately, innovative **single-cell DNA sequencing (scDNA-seq)** technology has now made it possible to elucidate the genetic heterogeneity of tumors. Sequencing may cover the entire genome or a subset of clinically relevant genes (gene panels). Although the former approach is appropriate for some objectives, targeted sequencing offers particular advantages for profiling tumor diversity. Because cancer continuously evolves, tumors often contain mutations that are relatively rare when they first emerge. Yet, detecting these rare mutations (and the clones that carry them) may be of significant clinical importance (e.g., minimal residual disease (MRD) and therapeutic resistance; Chapter 3). In order to detect rare clones with confidence, high sequencing depth is necessary. Targeted sequencing thus offers a cost-effective solution.

Mission Bio's **Tapestri Platform** enables highly sensitive targeted DNA sequencing at the single-cell level. By incorporating a cell-specific barcode to the amplified target sequences, this droplet-based technology measures genetic lesions (SNVs, indels, chromosomal rearrangements) and CNVs in each cell. Other structural characteristics like zygosity, mutational co-occurrence, and the presence of rare cell populations are also identified (Figs 5-6). This powerful technology has unlocked a host of new opportunities to understand how the mutational landscape of tumors fluctuates over time. But Tapestri's capabilities do not stop there. Let's next look at this platform's powerful multi-omics capability.

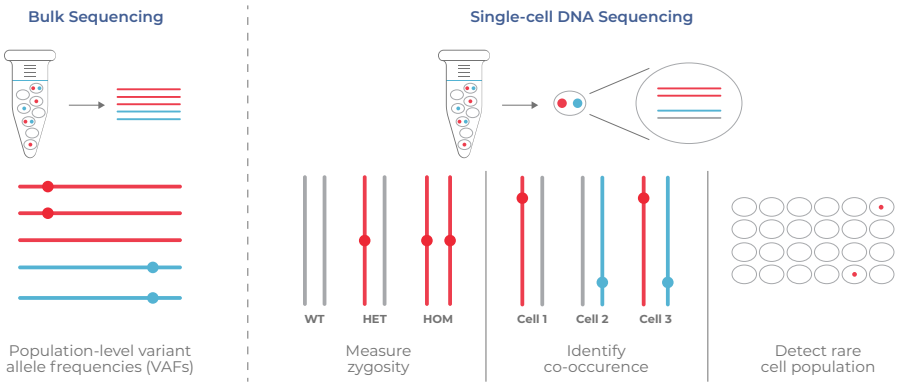

Figure 6. Information from bulk sequencing vs. single-cell DNA sequencing.

Bulk NGS only provides an average readout of mutations (reported as variant allele frequencies, VAFs) and does not resolve the clonal structure of tumors. Alternatively, scDNA-seq assesses each cell and reveals the co-mutations and zygosity across clones. Rare cell populations are also detected.

The Power of a Multi-omics Approach

Whereas measuring one parameter — like DNA — in a cell-specific manner provides an expanded view of cancer, a single analyte may not be sufficient to resolve all the clonal populations of a tumor. Single-cell multi-omics workflows, which simultaneously measure multiple analytes, often paint a more vivid picture. Combinations of

different molecules, whether DNA, RNA, or proteins, facilitate the dissection of tumors at a more granular level.

Cell-surface markers are a particularly informative analyte. Adorning the plasma membrane of each cell, these antigens play important roles in cell signaling and immune surveillance. Their expression also correlates with certain cell characteristics like cell type, stage of development, and disease state. The marker profile for a given cell is called its **immunophenotype**.

When it comes to cancer, measuring immunophenotypes has considerable value. Tumors are not only genotypically heterogeneous but also phenotypically diverse. A given sample may contain healthy and pathogenic cells of different lineages and states. The simultaneous measurement of genotype and immunophenotype (i.e., **proteogenomic profiling**) classifies these cells into groupings that can be filtered and more closely investigated (Figs 7-8). Moreover, obtaining this information across thousands of cells portrays a thorough depiction of the tumor's clonal diversity.

Conventional methods of proteogenomic profiling are laborious and require that samples be used in several assays across different instruments. In contrast, the Tapestri Platform enables the high-throughput measurement of DNA + protein of individual cells in a single assay (Fig 9).

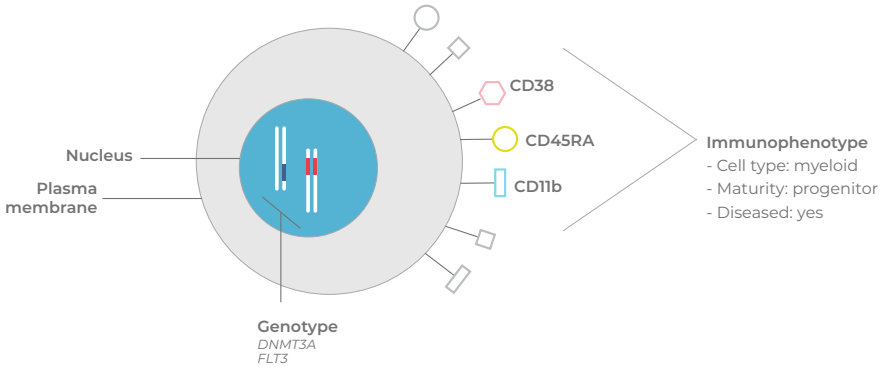

Figure 7. Genotype & immunophenotype.

In this simplified example, information about mutations in cancer-associated genes (*DNMT3A* and *FLT3*) and immunophenotype suggests that the cell is a myeloid progenitor in a malignant state.

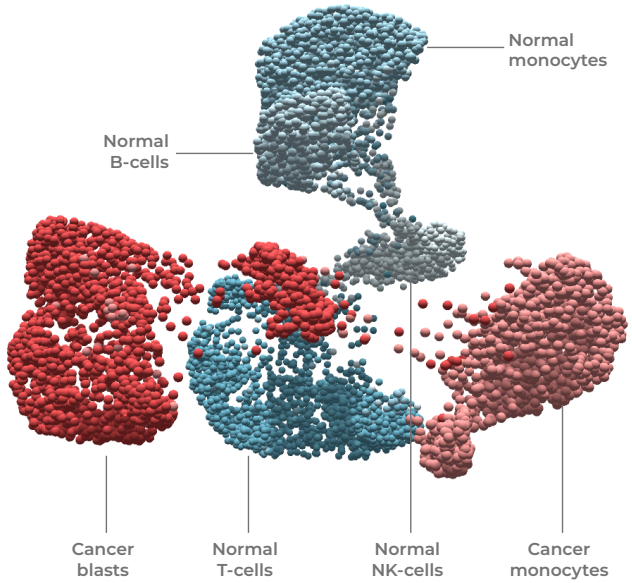

Figure 8. Tumors are genotypically & phenotypically complex.

The use of DNA + protein multi-omics was used to resolve cell type (T-cell, B-cell, blast, monocyte, NK) and state (normal vs cancer) in this leukemia sample.

The workflow begins with the selection of gene and protein panels. The former includes primers for targeted loci and the latter comprises **antibody-oligo conjugates (AOCs)** that bind specifically to target cell-surface proteins (Fig 10a). A short oligo attached to each antibody designates the identity of the corresponding antigen. The cell sample is first stained with a protein panel (preselected AOCs) and then deposited into a Tapestri cartridge where individual cells are encapsulated in oil droplets. Within each droplet, target genes and antibody-specific oligos are amplified with a cell-specific barcode. Libraries are then sequenced by NGS, and cell-specific mutational and protein profiles are reconstructed using Tapestri software (Figs. 9, 10b).

Multi-omics assessment with Tapestri yields rich datasets that enable robust cellular classifications based on genotypic and immunophenotypic profiles. The dataset in Figure 11 depicts distinct groupings of four cell lines based on SNV, CNV, and protein information. The plot that utilizes all three parameters portrays the most distinct clusters. This example illustrates how multi-modal analyses can reveal classifications that are not apparent with a single analyte.

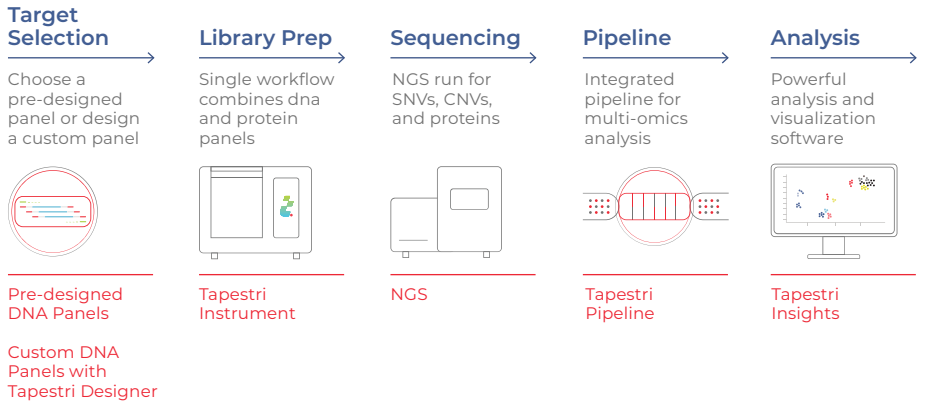

Figure 9. Single-cell DNA + protein multi-omics workflow.

The workflow begins with the selection of DNA and protein panels, followed by target amplification and library preparation on the Tapestri Platform. NGS yields DNA and protein data that are then reconstructed and visualized at single-cell resolution.

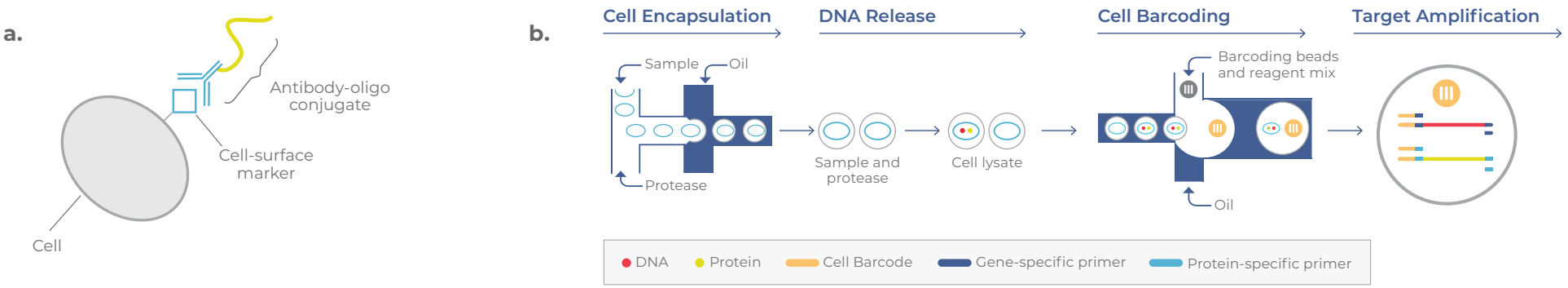

Figure 10. Single-cell DNA + protein targeting.

a. Antibody-oligo conjugates are used to tag each cell-surface marker with an oligonucleotide **b.** After cells are stained with a protein panel, the sample is deposited in a Tapestri cartridge, where individual cells are encapsulated in oil droplets with a protease (releases the DNA). Targeted DNA and protein-specific oligos are tagged with a cell-specific barcode during target amplification.

Single-cell DNA sequencing and multi-omics are enabling cancer to be scrutinized with an unprecedented level of resolution. Oncologists are quickly capitalizing on the use of these exciting technologies to interrogate various heme malignancies and solid tumors.[4-8] In the next chapter, we explore uses of single-cell DNA and multi-omics in cancer research, with a particular focus on translational applications for the clinic.

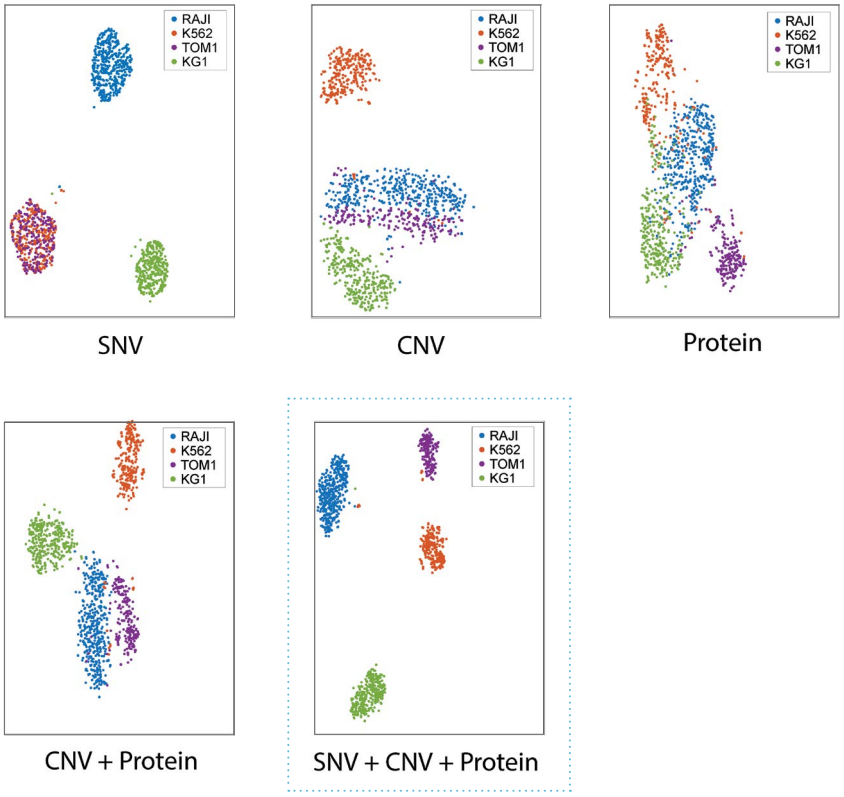

Figure 11. Single-cell multi-omics resolves distinct cell populations.

t-SNE projections of single-cell DNA + protein data showing clustering of RAJI, K562, TOM1, and KG1 cell lines. Note that all combining genotype and protein information yielded the most distinct clusters (blue box).

VOCABULARY

Bulk next-generation sequencing (bulk NGS): a technique where pooled DNA is sequenced and averaged across cells.

Variant allele frequency (VAF): the percentage of mutant alleles in the sample (the proportion of reads that match a particular gene variant divided by the coverage at the locus).

Single-cell DNA sequencing: Cell-specific DNA targeting and NGS.

Tapestri Platform: a single-cell DNA and multi-omics platform by Mission Bio.

Immunophenotype: a cell's profile of cell-surface markers.

Proteogenomic profiling: the simultaneous measurement of DNA and protein (in this case cell-surface markers).

Antibody-oligo conjugate (AOC): an antibody linked to a short oligonucleotide.

Interested in reading more about DNA + protein multi-omics capabilities? See our app note.

Clinical Applications of Single-cell Analysis

Translational research is the bridge that links fundamental discovery to clinical application. Profiling cancer leads to many advancements, including biomarker discovery and characterizing disease processes. This information lends insight into how we can better detect and surveil disease, segment patients into clinical trials, and combat resistance to therapies. Here, we outline three salient research areas that are utilizing single-cell analyses in ways that can be translated to the clinic.

Uncovering Evolutionary Trajectories

A growing area of interest in cancer research is the reconstruction of tumor evolutionary histories. Because single-cell analysis assigns mutations to different clones, phylogenetic trees can be reconstructed based on patterns of co-mutation.

Reassembling clonal histories not only reveals the overall pattern of evolution (linear vs. branched) but also enables researchers to understand other important aspects of disease progression. For instance, disease-initiating mutations, as well as patterns of mutual exclusivity and co-occurrence, are identified. Assessing cooperative mutations that give cells a fitness advantage is particularly illuminating. For instance, in the AML phylogeny depicted in Figure 12, the co-mutations, *DNMT3A* and *NPM1*, enabled a clone (dark blue circle) to expand and compose about 30% of the tumor.

Phylogenetic studies that leverage single-cell analysis are contributing valuable insights into cancer. For instance, Miles et al. (2020) revealed the chronological order of mutational acquisition as well as cooperative mutations that drove clonal expansion in AML. This study also found that evolutionary patterns differed across stages of disease progression.[5] In another AML study, Morita et al. (2020) discovered linear and branched patterns of evolution, with some subclones independently acquiring mutations in the same genes (convergent evolution). Proteogenomic profiling in both of these studies enabled the simultaneous characterization of mutational profile and cell type/state in an evolutionary context.[4]

These studies are shedding new light on the behavior of cancer. Not only do these endeavors contribute foundational knowledge to malignant transformation, but there is mounting evidence they have prognostic value. For instance, higher intratumoral heterogeneity is associated with the development of drug resistance and poorer prognosis in AML.[9,10] In the heme field alone a

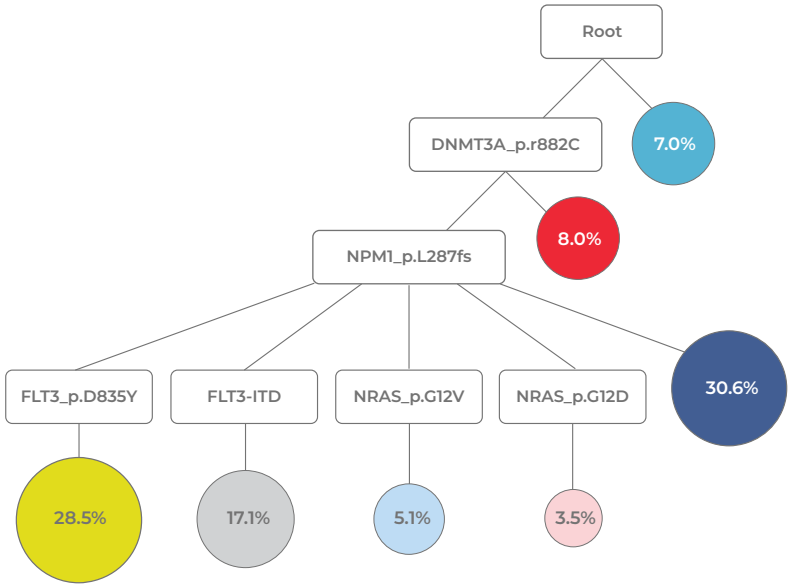

Figure 12. Evolutionary analysis reveals clonal dynamics.

In this AML phylogenetic tree, it is evident that early mutations (*DNMT3A* and *NPM*) are clonal and subsequent mutations (*FLT3* and *NRAS*) are subclonal. Clonal expansion events are also evident (yellow circle: *DNMT3A-NPM1-FLT3* mutant; dark blue circle: *DNMT3A-NPM1* mutant).

number of indications are actively being investigated, including CH, AML, MPN, MDS, multiple myeloma, CLL, and ALL.[4,5, 11-13], As tumors continue to be profiled across patients, general patterns of evolution will likely become apparent.

Therapeutic Resistance

Although anti-cancer therapies often succeed in bringing patients into **complete remission (CR)**, tumors often relapse at a later time point. The failure to eliminate cancer entirely is often due to the heterogeneity and dynamism of the disease. Tumor cells can evade treatment through **primary resistance**, in which pre-existing mutations offer protective properties. Alternatively, **secondary resistance** may emerge through mutations acquired during the course of treatment. Because most tumor cells succumb to therapy, cells that survive have little competition for resources. Even small initial quantities of resistant cells can rapidly expand and drive relapse. The ability for tumors to adapt in this fashion lies at the root of therapeutic failure.

Though cancer's uncanny ability to evade treatments remains a significant challenge, understanding this process may enable us to beat cancer at its own game. Single-cell DNA sequencing offers considerable value over traditional bulk analysis. Because bulk is unable to unambiguously delineate mutational co-occurrence across different clones, it is not possible to determine whether *all* clones will be targeted by a given therapy. Single-cell DNA sequencing can detect high-risk clones early on so that treatment plans can be adjusted to reduce the likelihood of relapse.

To illustrate the value of single-cell analysis, consider the example of a longitudinal study of a relapsed AML patient who received the targeted *FLT3* inhibitor therapy, gilteritinib (Fig 13). At the time the patient was treated, peripheral blood was tested at three timepoints (diagnosis, during therapy, and at relapse) using bulk NGS. Because bulk analysis cannot resolve clonal structure, there was no way to determine the clones present at each time point. When the blood samples were analyzed via scDNA-seq in a retrospective study, the tumor architecture was revealed.

This lack of clonal resolution in the initial study meant that clones that led to relapse were missed. For instance, at the diagnosis timepoint, the light blue clone was effectively targeted by gilteritinib while the dark blue clone was not (due to lacking the *FLT3* target). This latter clone expanded post-therapy and contributed to relapse. Similarly, a novel clone (red) that emerged during treatment had acquired resistance to gilteritinib via an *NRAS* mutation. This

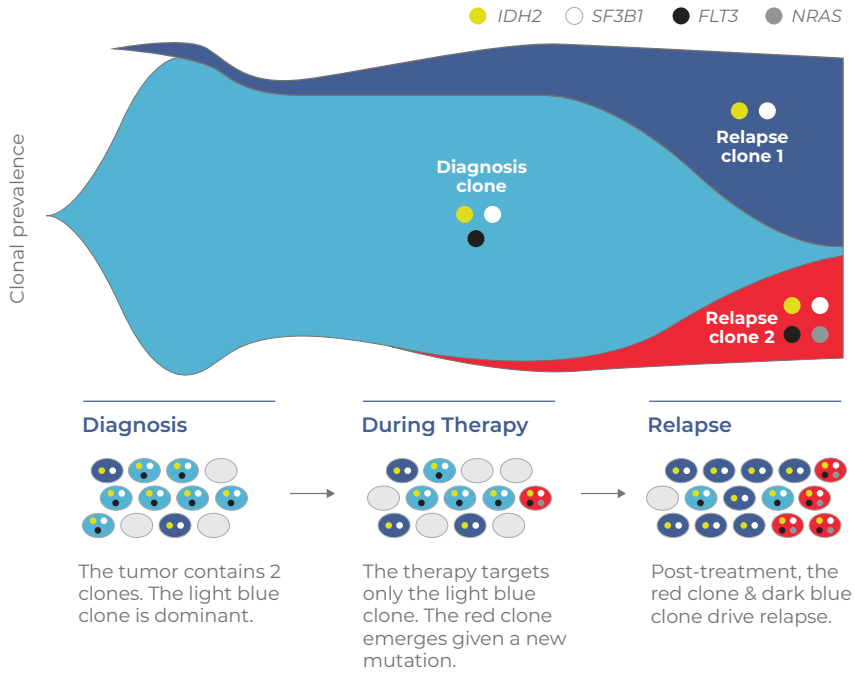

Figure 13. Clonal structure underlies resistance mechanisms.
A retrospective study that used scDNA-seq to analyze blood at three timepoints (diagnosis, during treatment, relapse). Single-cell analysis detected clones (dark blue and red) that ultimately drove relapse in the patient.

clone was too rare to be detected by bulk NGS during treatment (comprising only 0.1% of the tumor). Hence, it was free to expand unabated and contributed to relapse.

Using single-cell data to inform personalized treatment strategies holds promise for reducing rates of relapse. By resolving the clonal structure of a tumor prior to treatment, combinational therapies can be developed to ensure all clones are targeted. Surveillance during treatment can also ensure that any new therapy-resistant clones are identified. An increasing number of studies are using scDNA-seq to uncover mechanisms of resistance in clinical trials, further underscoring the potential of this technology in a therapeutic setting. [4,5,8,14]

Minimal Residual Disease (MRD)

As we saw in the example above, sometimes resistant cells escape therapy and drive relapse. Even for patients that have achieved CR, small quantities of remaining pathogenic cells — called **minimal residual disease (MRD)** — often remain. For myeloid malignancies, MRD is a powerful predictor of treatment outcomes. Data from clinical trials have revealed that the presence of MRD is indicative of both faster relapse and lower overall survival [15] (Fig 14). MRD is becoming established as a surrogate marker for relapse-free survival and could one day be used as a regulatory endpoint for drug approval.[16]

Given its clinical implications, MRD detection assays are commonly performed following induction chemotherapy. If malignant clones are detected and characterized, early action can be taken to intervene. Conventionally, MRD is identified by either genotype or immunophenotype. Traditional genotypic methods like PCR, qPCR, or bulk NGS detect mutation frequencies but do not assign aberrations to distinct clones. Without this clonal structure, the identification of malignant cells is limited.

MRD may also be detected through immunophenotyping (typically through multi-parameter flow cytometry). In AML, for instance, MRD is identified by flow-sorting cells based on the abnormal pattern of markers expressed by blast cells. However, the sensitivity of this assay is limited, as common markers are expressed by both healthy and leukemic cells.

Single-cell multi-omics (DNA + protein) analysis combines the strengths of both assays into a single workflow. The use of multiple analytes to detect aberrant cells enhances the detection of residual disease.[17,18] For example, Dillon et al. (2020) reported an AML case

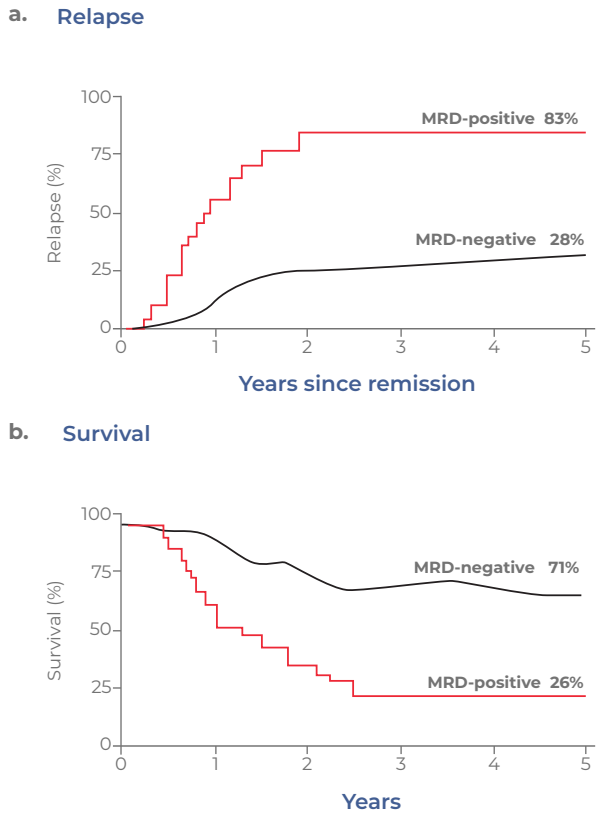

Figure 14. MRD status correlates with clinical outcomes.

In these hypothetical Kaplan-Meier curves, patients with MRD tend to have a higher incidence of relapse and lower overall survival.

in which leukemic cell-surface markers were present on only half of cells with malignant genotypes.[18] Here the use of DNA + protein multi-omics enabled MRD to be characterized to a fuller extent than either analyte could have alone. In the clinic, this highly-specific information can be used to guide personalized treatment regimens.

Clonal Hematopoiesis of Indeterminate Potential (CHIP)

Because cancer development is characterized by the gradual accumulation of somatic mutations, premalignant cells contain only subsets of pathogenic genes. In the blood, hematopoietic stem cells and progenitor cells (HSPCs) may acquire leukemia driver mutations in a process called **clonal hematopoiesis of indeterminate potential (CHIP)**. Although these mutations cause clonal expansion of HSPCs, the cells are not yet malignant. However, they are predisposed to acquiring additional mutations that lead to full transformation.

CHIP is an age-driven process occurring in about ~ 20% of the US population. Somatic mutations in epigenetic genes (occurring at ≥ 0.02 VAF) appear in middle age and increase over time (Fig 15).[19] Though CHIP itself is not a disease, people with the condition are at higher risk for developing AML or other hematological malignancies. Having CHIP has also been correlated with a higher risk of cardiovascular disease and reduced lifespan.

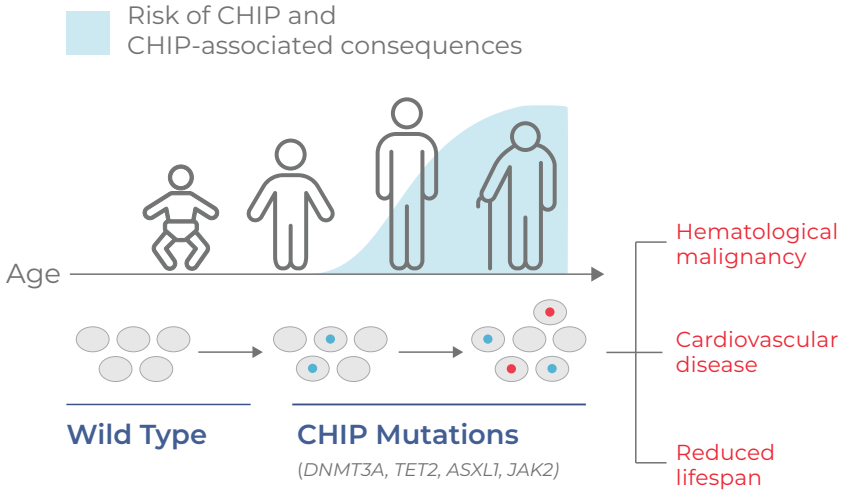

Figure 15. CHIP progresses with age.

CHIP is a non-malignant condition characterized by mutation and clonal expansion of blood cells, beginning at middle age. CHIP is associated with a higher risk of developing hematological malignancies and cardiovascular disease, as well as a reduced lifespan.

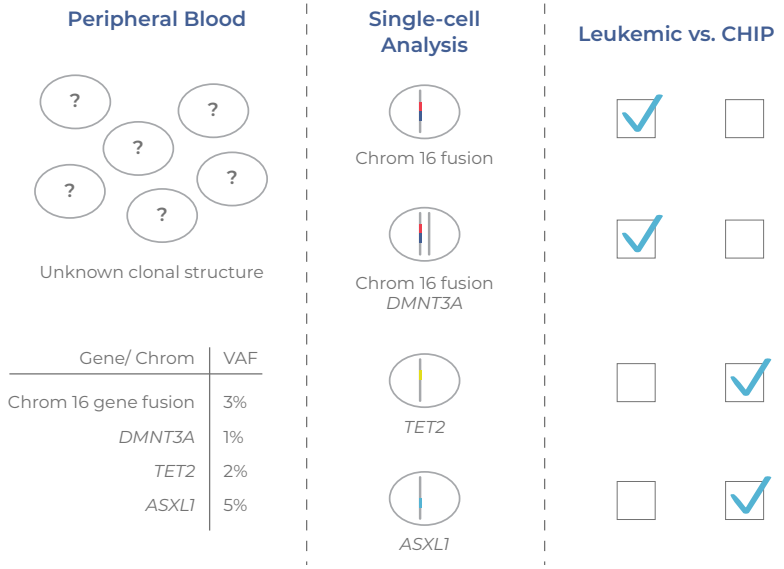

Figure 16. Single-cell analysis distinguishes CHIP clones.

Pre-malignant (CHIP) and malignant clones cannot be distinguished from VAFs alone. By resolving the mutational structure across clones, scDNA-seq enables the identification of CHIP and leukemic clones.

The etiology of CHIP is not well understood. For clinicians, one area of particular interest is the ability to differentiate CHIP (premalignant) and oncogenic clones. Because CHIP cells have leukemia-associated mutations (e.g. *DNMT3A*, *TET2*, *AXL1*, and *JAK2*) they may confound MRD detection. However, the *clonal structure* of these mutations can be used to verify whether cells are pathogenic or not.

These structural analyses can be achieved using scDNA-seq.[17] As an illustrative example, Figure 16 shows a cell population with several genetic aberrations at low frequencies. While the substructure of the tumor cannot be determined from the VAFs alone, it can be resolved with scDNA-seq. Here, the two leukemic clones have a gene fusion in chromosome 16 (a defining feature of AML). The CHIP clones have single mutations in either *TET2* or *ASKL1* that do not co-occur with the chromosome 16 fusion (indicating they are not leukemic).

Single cell DNA+ protein multi-omics can also be leveraged to distinguish MRD from CHIP clones. By layering of genotypic and immunophenotypic information, malignant and non-malignant clones can be identified with greater confidence. Indeed, Dillon et al. (2020) used proteogenomic profiling to accurately distinguish and CHIP and AML MRD.[18]

VOCABULARY

Complete remission (CR): no cancer is detected by standard lab tests and imaging.

Primary resistance: pre-existing mutations confer resistance to a therapeutic.

Secondary resistance: mutations that emerge during treatment confer resistance to a therapeutic.

Minimal residual disease (MRD): small quantities of cancer cells that remain in the body following cancer treatment.

Clonal hematopoiesis of indeterminate potential (CHIP): the process of mutational acquisition and subsequent clonal expansion of hematopoietic stem cells (HSCs), resulting in non-malignant clones.

Single-cell DNA and multi-omics analyses are advancing our knowledge of myeloid malignancies and solid tumors. Learn more in these app notes about profiling MDS and AML, and metastatic melanomas.

Single-cell Analysis of Genome Engineering

Many of today's cutting-edge research methods and therapies rely on the engineering of cells. By precisely altering DNA, genome engineering makes it possible to probe the relationship between a gene and its function, repair dysfunctional genes, or add functionality to cells. Changing the genome often involves some degree of variability. In some cases, different alterations are intentionally made across cells. Unwanted variation, on the other hand, may impact experimental results or therapeutic development. In this section, we discuss how scDNA-seq is used to measure precise genomic changes in individual cells.

Methods of Engineering Cells

Several genome engineering techniques can add exogenous genes or edit existing ones. In the former case, the addition of a transgene may offer new functionality or provide a working copy of a dysfunctional gene. Viral vectors, including adeno-associated viruses (AAV) and retroviruses (e.g., lentiviruses), are often used to ferry transgenes into cells and insert them into the genome (Fig 17a).

DNA can also be edited directly using viruses or nuclease-based tools, such as **CRISPR/Cas9** (Fig 17b). In this system, a single guide RNA (sgRNA) specifies a site in the genome for Cas9 to cut the DNA. Depending on the experimental goals, either **homology-directed repair (HDR)** or **non-homologous end joining (NHEJ)** is co-opted to make the desired edit while repairing the break. HDR is used to either change nucleotides or "knock in" new sequences, while NHEJ is used to disrupt genes (knockouts) by introducing indels into target sites. This latter method is useful for functional studies and loss-of-function (LOF) screens.

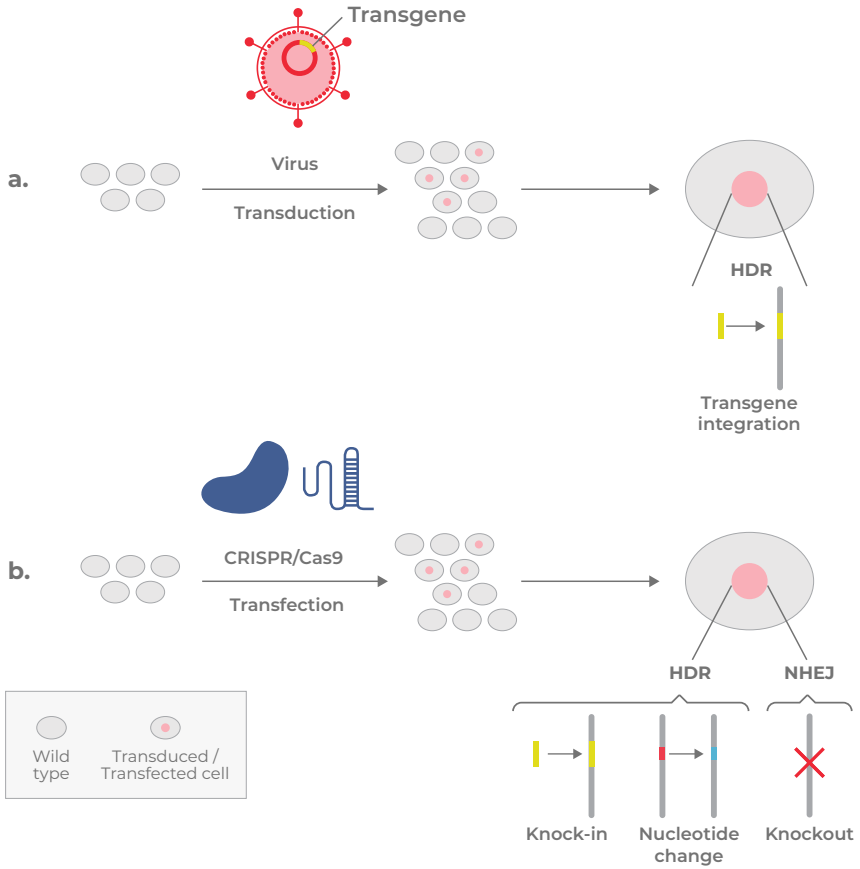

Figure 17. Genomic manipulation by viruses and gene editors.

a. Viruses are often used to insert transgenes into the genomes of cells through HDR **b.** CRISPR/Cas9 is used to either insert sequences or change nucleotides through HDR disrupt gene function (knockout) through NHEJ.

Loss-of-function Screening

CRISPR **loss-of-function (LOF) screens** are an invaluable tool for functional genomics studies. These screens pair the systematic ablation of a large number of genes with phenotypic analysis in order to characterize the genetic underpinnings of cellular pathways. LOF studies are crucial for understanding disease processes, discovering new drug targets, and assessing the genes that underlie the resistance or sensitivity of cancer to therapies.

Although the applications for CRISPR screening are quickly growing, typical screening protocols are limited in their ability to capture the heterogeneity of editing across cells. Typically, the average on-target editing efficiency is only measured at the population level before a screen. Thus, the effect of cell-to-cell variation for on- and off-target editing on phenotypic outputs is largely unknown.

Recently, there have been some advances in single-cell pooled screening with a transcriptome readout.[20-23] While these screens have gained valuable insights, they also have some drawbacks. Significantly, the CRISPR-induced edits to the genome are not directly measured. Rather, the identity of the target is indicated by sequencing the integrated sgRNA sequence in cells. Thus, the type of edit is never identified, nor are any off-target changes to the genome.

The integration of scDNA-seq into screening workflows could solve this issue. Because scDNA-seq measures the actual changes to DNA, the on- and off-target edits can be directly identified and correlated with other analytes. With Tapestri technology, multi-omic screens that simultaneously measure genomic edits and cell-surface markers in individual cells are now possible. This enables the high-throughput annotation of mutations in different cell states associated with cancer.

Single-cell multi-omic screens that measure the genetic edits, transcriptome, and proteome within a single cell, are being developed. These advances will undoubtedly enhance our understanding of disease states and provide invaluable information for pharmaceutical development.

Disease Modeling

Disease modeling often involves recapitulating the underlying mutations of a disease so that pathogenesis can be studied and therapies can be tested. Different types of cancer are associated with different mutations in different genes, with clonal architecture adding an additional layer of complexity. Modeling cancer using cell-based or animal models means that these various levels of variation must be re-created. Fortunately, genome engineering tools can be used to generate mutations recurrent in different types of cancer. Furthermore, multiple mutations can be made to single cells in order to recreate co-mutated clones found in tumors.

Mutation combinations are often engineered in a random fashion across cells. Thus, the genetic alterations must be analyzed in each cell in order to fully characterize the clonal composition of the model. Single-cell DNA sequencing can be used to directly analyze the genetic alterations produced by viruses and gene editors with single-cell resolution.[24] The use of scDNA-seq has already been successfully used to evaluate several cancer models. In one example, Dr. John Lee (Fred Hutchinson Cancer Center) used lentivirus to generate random combinations of loss- and gain-of-function mutations across urothelial cells, building an organoid model of bladder cancer[25] After transplantation into mice and a period of in vivo selection, cDNA-seq was used to identify specific combinations of mutations that were highly correlated with tumorigenesis.

In another experiment, multiple mutations typical of acute erythroid leukemia (AEL) were generated in mouse hematopoietic stem and progenitor cells using CRISPR.[26,27] After cells from a primary tumor were serially transplanted into mice, scDNA-seq was used to identify primary (CRISPR-generated) and secondary mutations that drove AEL tumorigenesis.

Dr. Ten Hacken et al. (2020) also used CRISPR to build a cancer model. Using a murine cell line, she generated LOF mutations recurrent in chronic lymphocytic leukemia (CLL) in random combinations [13] With scDNA-seq, her team was able

to simultaneously measure zygosity, on-target, and off-target editing in single-target cells, as well as various patterns of co-mutation in multiplexed cells. Lastly, she compared changes in clonal architecture before and after edited cells were transplanted into mice. This study demonstrated the capability of scDNA-seq to analyze up to six concurrent targets in single cells, as well as predicted off-target activity.

Cell and Gene Therapies

Cell and gene therapies (CGT) are rapidly expanding the treatment landscape for genetic diseases, including cancers and inherited diseases like sickle-cell anemia and Duchenne's muscular dystrophy. Cell therapies involve treating disease by transferring cells with a desired function into the body. The cells may originate from the patient (**autologous**), a donor (**allogeneic**), or an established cell line. This approach requires that cells be collected outside the body (ex vivo) prior to introduction into the patient. Gene therapies involve the transfer of DNA or the direct alteration of the patient's genome in order to treat a disease.

Some therapies are classified as either a cell or a gene therapy, while others are a combination of both. For example, chimeric antigen receptor T-cell (CAR T-cell) therapy involves harvesting T-cells from a patient and genetically modifying them to express synthetic antigen receptors on their surface. The edited T-cells are returned to the patient, where they mark tumor cells for destruction.

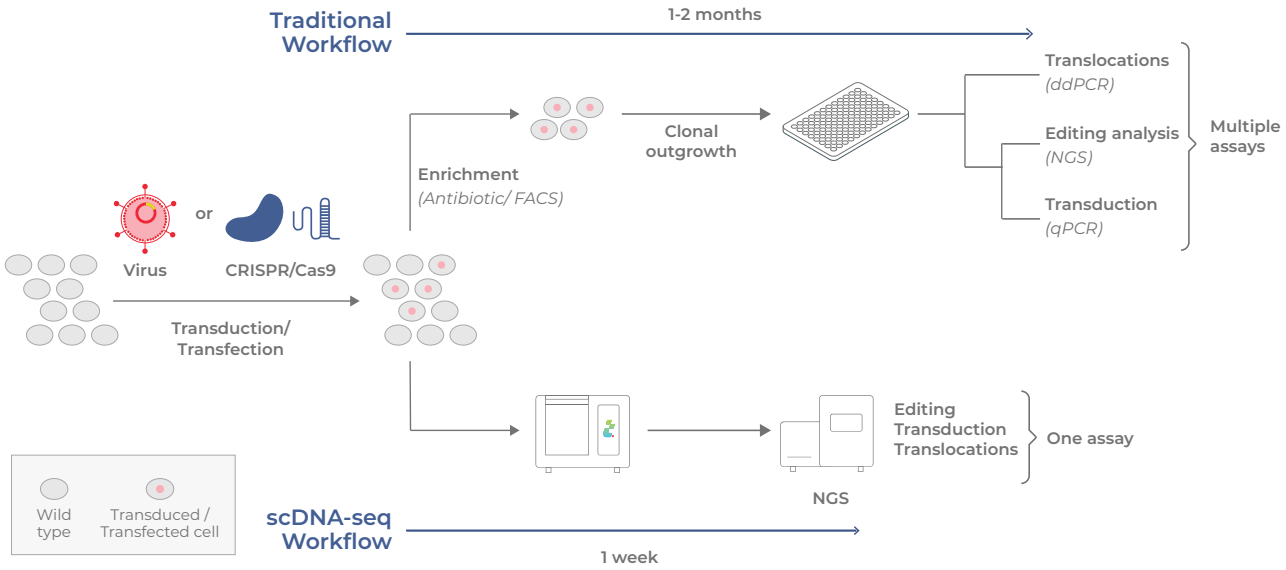

Figure 18. Conventional single-cell analysis vs scDNA-seq used for cell and gene therapies.

scDNA-seq of CGTs is needed in order to assess variation in transduction efficiency, on- and off-target editing, and translocation events. **a.** In the conventional workflow, transduced /transfected cells are enriched and clonally outgrown for several weeks. Assays that assess translocations and genome alterations are conducted separately. **b.** With scDNA-seq, no clonal outgrowth is necessary and all analyses are combined in a single assay.

Viruses and gene editors are commonly used to modify the genomes of cells used in cell and gene therapies. Yet, the genetic heterogeneity produced by these alterations may have consequences to the efficacy and safety of the therapy. Analyzing single cells using conventional techniques is a long and laborious process (Fig 18a). Manipulations that involve viral or plasmid vectors often involve a time-consuming step to enrich successfully transduced/transfected cells. Furthermore, to identify cell-specific edits, individual cells must be separated and clonally outgrown (often taking weeks to complete). Once clones have been expanded, the analysis of genome alterations/ transduction efficiency is typically accomplished through sequencing or qPCR. Unwanted translocation events, which can potentially cause safety issues, require a separate assay (ddPCR or qPCR).

Single-cell DNA sequencing offers several advantages over traditional workflows (Fig 18b). Because each cell is measured individually, the enrichment and clonal outgrowth steps can be bypassed, thus shortening the time between the genomic manipulation and analysis from weeks to days. Additionally, the analysis of transduction efficiency/ editing and unintended translocation events can be combined into a single assay.[28, 29]

Single-cell analysis can be leveraged in multiple stages of the CGT pipeline, including protocol optimization, release testing, and manufacturing. By expediting these key steps, single-cell technology is helping advance the next generation of medicine.

Want to learn more about how the Tapestri Platform enables cell and gene therapies? Check out our app notes on single-cell analysis of transduction and gene editing.

VOCABULARY

CRISPR/Cas9: a genome engineering tool composed of single guide RNA (sgRNA) and a Cas9 nuclease.

Homology-directed repair (HDR): a type of DNA repair that utilizes a DNA template to precisely repair double-strand breaks. It is leveraged in genome engineering to knock in nucleotides or make changes to the genome.

Non-homologous end joining (NHEJ): an error-prone type DNA repair that often inserts/deletes nucleotides at the break site. It is used in genome engineering to knock out genes.

CRISPR loss-of-function (LOF) screen: a screen that utilizes CRISPR to systematically ablate a large set of genes so that the phenotypic consequences can be assessed.

Disease modeling: the recapitulation of diseases using cell-based or organism-based models. This is often accomplished through genetic disruption or modification.

Cell and gene therapies (CGTs): a class of therapies that encompasses cell therapies (cells are used to introduce a therapy into the body), gene therapies (genetic alterations that have a therapeutic effect are made directly in the patient), or a combination of both.

Autologous: cells are sourced from the patient.

Allogeneic: cells are sourced from a matched donor.

Conclusion

Cancer is a complex disease. Tumors comprising genetically diverse clones fluctuate over time in response to environmental challenges. The genotypic heterogeneity of tumors forms the foundation of therapeutic resistance and is a significant roadblock to curative solutions.

In order to move the needle towards better treatment strategies, it is necessary to “see” tumors with greater resolution. Whereas conventional bulk techniques only report population-level metrics, single-cell multi-omics uncover nuanced clonal architecture.

New insights about cancer complexity not only enhance our undestanding the disease, but also inform clinical action. Additionally, single-cell analysis is driving innovations in disease modeling, screening, and cell and gene therapies.

With the ability to profile cancer at a more granular level, single-cell technology is enabling our progress towards precision medicine. New breakthroughs will continue to open the door to a hopeful future where curative therapies are within reach.

References

1. Nowell P. The clonal evolution of tumor cell populations. *Science*. 1976;194:23-28.
2. Shaughnessy J. Amplification and overexpression of CKS1B at chromosome band 1q21 is associated with reduced levels of p27Kip1 and an aggressive clinical course in multiple myeloma. *Hematology*. 2005;10 Suppl 1:117-26.
3. Landau DA et al. Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia*. 2014 Jan;28(1):34-43.
4. Morita K et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat Commun*. 2020 Oct 21;11(1):5327.
5. Miles LA et al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature*. 2020 Nov;587(7834):477-482.
6. Demaree B et al. Joint profiling of proteins and DNA in single cells reveals extensive proteogenomic decoupling in leukemia. bioRxiv 2020.02.26.967133.
7. App Note: Single-cell DNA analysis with nuclei isolated from solid tumor tissue. Mission Bio.
8. Gao Y et al. V211D Mutation in MEK1 Causes Resistance to MEK Inhibitors in Colon Cancer. *Cancer Discov*. 2019 Sep;9(9):1182-1191. Epub 2019 Jun 21.
9. de Bruin EC et al. Intra-tumor heterogeneity: lessons from microbial evolution and clinical implications. *Genome Med*. 2013 Nov 22;5(11):101.
10. Bochtler T et al. Clonal heterogeneity as detected by metaphase karyotyping is an indicator of poor prognosis in acute myeloid leukemia. *J Clin Oncol*. 2013 Nov 1;31(31):3898-905.
11. Maia C et al. Biological and clinical significance of dysplastic hematopoiesis in patients with newly diagnosed multiple myeloma. *Blood*. 2020 Jun 25;135(26):2375-2387.
12. Albertí-Servera L et al. Single-cell DNA amplicon sequencing reveals clonal heterogeneity and evolution in T-cell acute lymphoblastic leukemia. *Blood*. 2021 Feb 11;137(6):801-811.
13. ten Hacken E et al. High throughput single-cell detection of multiplex CRISPR-edited gene modifications. *Genome Biol*. 2020;21(1):266. Published 2020 Oct 20.
14. McMahon CM et al. Clonal selection with RAS pathway activation mediates secondary clinical resistance to selective FLT3 inhibition in acute myeloid leukemia. *Cancer Discov*. 2019 Aug;9(8):1050-1063.
15. Jongen-Lavrencic M et al. Molecular minimal residual disease in acute myeloid leukemia. *N Engl J Med*. 2018 Mar 29;378(13):1189-1199.
16. Landgren O. MRD testing in multiple myeloma: from a surrogate marker of clinical outcomes to an every-day clinical tool. *Semin Hematol*. 2018 Jan;55(1):1-3.
17. Ediriwickrema A et al. Single-cell mutational profiling enhances the clinical evaluation of AML MRD. *Blood Adv*. 2020 Mar 10;4(5):943-952.
18. Dillon LW et al. Personalized single-cell proteogenomics to distinguish acute myeloid leukemia from non-malignant clonal hematopoiesis. medRxiv 2020.10.22.20216028,
19. Gibson CJ, Steensma DP. New insights from studies of clonal hematopoiesis. *Clin Cancer Res*. 2018 Oct 1;24(19):4633-4642.
20. Datlinger P et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods*.14, 297–301 (2017).
21. Dixit A et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screen. *Cell*. 15;167(7):1853-1866.e17 (2016)
22. Hill AJ et al. On the design of CRISPR-based single-cell molecular screens. *Nat. Methods*.15, 271 (2018).
23. Replogle JM et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol*. 38, 954-961 (2020).
24. App Note: Measuring the efficiency of CRISPR genome-editing systems using the Tapestri Platform and Tapestri single-cell DNA custom panels. Mission Bio.
25. App Note: Tapestri Platform resolves clonality of heterogeneous mouse organoid cancer model through single-cell DNA sequencing of lentiviral barcodes. Mission Bio.
26. App Note: Single-cell DNA analysis of a novel acute erythroid leukemia (AEL) CRISPR induced mouse model. Mission Bio.
27. Iacobucci I et al. Modeling and targeting of erythroleukemia by hematopoietic genome editing. *Blood*. 2021 Jan 19:blood.2020009103.
28. App Note: In-depth quantification of cell and gene therapy DNA integrations enabled with single-cell sequencing. Mission Bio.
29. App Note: Deep understanding of cell and gene therapy genome editing protocols enabled with single-cell sequencing. Mission Bio.

mission bio

eBook_Single-Cell DNA & Multi-omics_RevA

CONTACT US TO LEARN MORE
www.missionbio.com

Your local contact in Switzerland:

bucher biotec

Bucher Biotec AG
Viaduktstrasse 42
CH-4051 Basel
Tel. 061 269 1111
info@bucher.ch
www.bucher.ch