Trends in Cancer

CellPress

Review

Implications of Tumor Clonal Heterogeneity in the Era of Next-Generation Sequencing

Meagan A. Jacoby,[1] Eric J. Duncavage,[2,3] and Matthew J. Walter[1,3,*]

Recent whole-genome sequencing (WGS) studies have demonstrated that tumors typically comprise a founding clone and multiple subclones (i.e., clonal heterogeneity is common). The possible combination of mutations in each tumor clone is enormous, making each tumor genetically unique. Clonal heterogeneity likely has a role in cancer progression, relapse, metastasis, and chemoresistance due to functional differences in genetically unique subclones. In current clinical practice, gene mutations are only classified as being present or absent, ignoring the clonal complexity of cancers. In this review, we address how tumor clonality is measured using next-generation sequencing (NGS) data, highlight that clonal heterogeneity is common across multiple tumor types, and discuss the potential clinical implications of tumor clonal heterogeneity.

Clonal Evolution in Cancer

Advances in massively parallel digital sequencing, referred to commonly as NGS, have allowed researchers to rapidly define the genomic landscape of multiple human cancers. These large-scale genome-wide sequencing studies have demonstrated that genetic diversity exists across tumor types and among individuals with the same type of cancer (intertumor genetic heterogeneity) as well as within a tumor comprising an admixture of a founding clone and subclones each carrying a unique combination of mutations within a single patient (intratumor genetic heterogeneity). Recent studies have shown that the intratumor clonal heterogeneity within an individual patient with cancer has clinical implications for treatment response and outcomes.

The idea that cancer results from the sequential selection of mutant subpopulations derived from a common ancestor was postulated by Nowell decades ago using cytogenetic data [1]. The initiating event, presumably a mutation in a driver gene, leads to the expansion of a cell that ultimately becomes the founding clone detected at diagnosis. At diagnosis, there is a complement of mutations shared by all tumor cells (clonal mutations) through which daughter subclones can trace their lineage, which we refer to here as the founding clone. It is possible that, during tumorigenesis, a more distant ancestral clone existed that subsequently became extinct; thus, the founding clone identified at diagnosis may represent the most recent common ancestor that can be detected, as previously reviewed [2]. Ongoing accumulation of acquired mutations in cells from the founding clone gives rise to subclones that contribute to the genetic complexity of a tumor. In addition to driver mutations, many random mutations that are presumed to be neutral for cancer pathogenesis (i.e., passenger mutations) are present in the founding clone and subclones. These random mutations are acquired during the normal aging process and are carried forward when a tumor cell expands [3]. The genetic diversity that results from the staggering number of possible combinations of driver and

Trends

High-resolution NGS studies have provided evidence of the genetic complexity for multiple tumor types. Most tumors comprise a founding clone and one or more subclones that result from the sequential selection of mutant subpopulations.

Each patient presents with a genetically unique tumor due to the enormous number of possible gene mutation combinations that exist.

The presence of genetically diverse subclones has implications for diagnosis, prognosis, and therapy. Future trials are needed to determine whether the clinical response to a mutation-specific targeted therapy is impacted by whether the mutation is in the founding clone or a subclone.

Measuring and monitoring a single gene mutation using NGS may underestimate tumor burden if the mutation is in a subclone.

The optimal sequencing platform used to identify and monitor tumor clonal architecture depends on the breadth and depth of coverage that is needed, cost, and the expected number of mutations present in a tumor.

[1]Department of Medicine, Division of Oncology, Washington University School of Medicine, St Louis, MO, USA
[2]Department of Pathology and Immunology, Washington University School of Medicine, St Louis, MO, USA

passenger mutations makes every tumor and clone genetically unique, providing a molecular fingerprint of each individual's cancer.

NGS efforts using high-resolution (single-base) techniques support Nowell's model of cancer clonal evolution where a founding clone gives rise to a subclone(s). Tumor clones can follow both a linear and complex branching pattern of evolution (addressed in recent reviews [2,4,5]), likely shaped by inherent properties conferred by somatic mutations and selective pressure from therapies. It has been suggested that the subclonal genetic composition, as well as the size, number, and stability, of the subclones is important for disease progression and success or failure of therapy. Due to space constraints, in this review we focus on recent genome-sequencing efforts that highlight the genetic diversity of adult cancers using DNA-based sequencing assays and provide insight into intratumor clonal heterogeneity and evolution, with an emphasis on studies using WGS and selected examples using whole-exome sequencing (WES). Detecting and monitoring subclonal evolution may ultimately have diagnostic, prognostic, and therapeutic implications for patients with cancer.

Imputing Tumor Clonality from NGS Data

Results from NGS methods can be used to identify mutations present in a patient's tumor and determine in which clone a mutation occurs. Compared with Sanger sequencing, which generates data from a mixture of multiple DNA molecules present in a sample, NGS reads originate from single DNA molecules. NGS data provides digital read counts proportional to the number of DNA molecules that contain a reference sequence base relative to those that have a variant sequence. By dividing the number of observed DNA molecules with a sequence variant (i.e., variant reads) by the total number of DNA molecules identified (i.e., total reads), the variant allele fraction/frequency (VAF) can be calculated (i.e., VAF = variant reads/total reads) (Figure 1). Somatic mutations, defined as sequence variants present only in the tumor and not in the normal tissue (i.e., constitutional or inherited DNA), can be identified by comparing sequence data from the tumor tissue and nontumor tissue from the same individual. It is necessary to include matched normal DNA from the same individual to confidently identify somatic mutations when

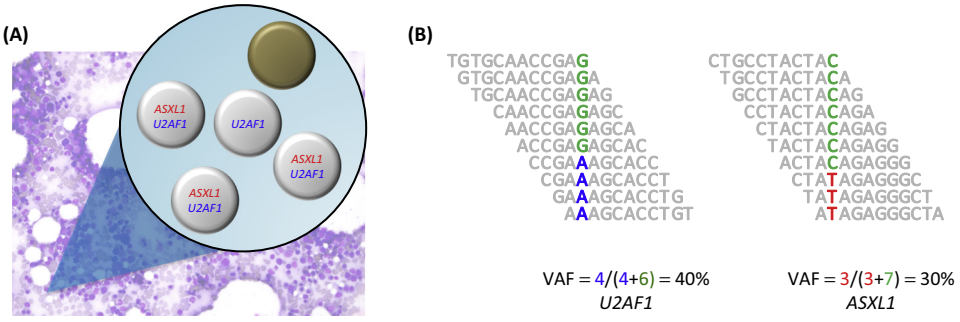

Figure 1. Calculating Mutation Variant Allele Fractions (VAFs) from Next Generation Sequencing (NGS) Reads. (A) Representative bone marrow aspirate from a patient with myelodysplastic syndrome is shown. An expanded view of bone marrow cells is provided to represent the proportion of cells with various mutations. Four tumor cells (gray) and one normal cell (brown) are depicted. A heterozygous U2 small nuclear RNA auxiliary factor 1 (*U2AF1*) mutation occurs in four tumor cells and a heterozygous additional sex combs like transcriptional regulator 1 (*ASXL1*) mutation occurs in three tumor cells that also harbor a *U2AF1* mutation. The *U2AF1* mutation occurs in the founding clone and the *ASXL1* mutation in a subclone. (B) Ten sequencing reads are shown for each gene mutation with the reference nucleotide in green and the variant nucleotide in blue (*U2AF1*) or red (*ASXL1*). NGS of bulk genomic bone marrow DNA demonstrates a *U2AF1* mutation present in four out of ten total reads (VAF = 40%) and an *ASXL1* mutation present in three out of ten reads (VAF = 30%). The fraction of cells in the bone marrow harboring each mutation is directly proportional to the VAF of heterozygous mutations occurring in the diploid part of the genome [e.g., 80% of cells harbor a *U2AF1* mutation (VAF 40% × 2 = 80% of cells) and 60% of cells harbor an *ASXL1* mutation (VAF 30% × 2 = 60% of cells)].

[3]Alvin J. Siteman Cancer Center, Washington University School of Medicine, St Louis, MO, USA

*Correspondence: mjwalter@dom.wustl.edu (M.J. Walter).

using WGS and WES platforms. When bulk tumor samples are sequenced for a large number of genes, the number of unique clones can be imputed by clustering mutation VAFs (Figure 2). Mutations in the diploid part of the genome that occur with the same VAF usually co-occur in the same cell (e.g., two genes with heterozygous mutations and VAFs of 40% are likely present in the same cell). While imputing the number of tumor clones present in a bulk sample using this approach has been validated using single-cell sequencing studies [6,7], additional studies indicate that sequencing bulk tumor samples may underestimate subclonal complexity in certain instances [8,9]. The size of each clone (based on their VAFs) can also provide information regarding the temporal acquisition of mutations by comparing the VAFs among clones. For example, a cell within a clone containing a U2 small nuclear RNA auxiliary factor 1 (*U2AF1)* gene mutation with an average VAF of 40% (i.e., a founding clone with a heterozygous mutation present in 80% of cells) gives rise to a clone with an additional sex combs like transcriptional regulator 1 (*ASXL1)* mutation with an average VAF of 30% (i.e., a subclone with a heterozygous mutation present in 60% of cells) (Figure 1). In this example, the subclone contains all the mutations present in the founding clone (e.g., *U2AF1* and others), in addition to more recently acquired subclone-specific mutations (e.g., *ASXL1* and others) (Figures 1 and 2).

Cancers are Clonally Heterogeneous and Genetically Unique

WGS of the first cancer genome was reported for a patient with *de novo* acute myeloid leukemia (AML) (i.e., AML occurring without an antecedent blood cancer) [10]. Subsequently, WGS of 50 *de novo* AML genomes was performed by The Cancer Genome Atlas (TCGA) project and revealed that at least one subclone was detected with the founding clone in more than half of the

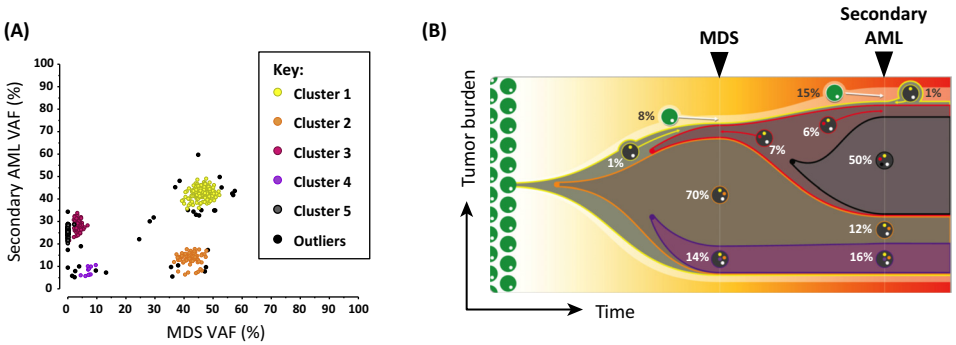

Figure 2. Imputing Tumor Clonality Using Mutation Variant Allele Fractions (VAFs). (A) The VAFs of 477 validated mutations in Subject UPN298273 are shown at the myelodysplastic syndromes (MDS) stage and after progression to secondary acute myeloid leukemia (AML). Each dot on the plot represents a mutation. VAFs were adjusted for chromosomal copy number. Unsupervised clustering of individual mutations identified five distinct mutation clusters, representing unique tumor clones. A change in the size of a clone between MDS and secondary AML is represented by different mutation VAFs. (B) A model summarizing clonal evolution from the MDS stage to the secondary AML stage in Subject UPN298273 using the VAF data from (A.) Green cells at the left side represent normal hematopoietic stem/progenitor cells (HSPCs). Each HSPC has its own unique set of mutations randomly acquired during normal aging (white dot). A transforming event occurs and a single cell gains a growth advantage and expands, carrying forward all the pre-existing mutations (white) in addition to 192 cluster 1 mutations (yellow). Cells in clone 1 contain cluster 1 mutations. Cells in clone 2 (orange) originated from a single cell in clone 1 (since all cluster 1 mutations are heterozygous and present in nearly all secondary AML cells) and, therefore, contain all cluster 1 and 2 mutations. Cells in clone 3 (red) originated from a single cell in clone 1 and contain all cluster 1 and 3 mutations. Cells in clone 4 (purple) originated from a single cell in clone 2 and contain all cluster 1, 2, and 4 mutations. Cells in clone 5 (black), the last clone to emerge (comprising 50% of the bone marrow cellularity in secondary AML), contain cluster 1, 3, and 5 mutations. During transformation to secondary AML, clone 2 (orange), carrying a nucleophosmin (*NPM1)* mutation, collapses from 70% of cells in MDS to 12% of cells in secondary AML, while clone 5 emerges (black). Tracking only the *NPM1* mutation VAF in this patient would underestimate the tumor burden and therapeutic response over time, highlighting a limitation of monitoring clonality and tumor burden using a single driver gene. Modified, with permission, from [14].

patients tested, using coverage models that were not capable of detecting subclones or common ancestral clones that comprised less than 10–15% of the cells in the sample [11]. Additional studies have consistently verified that tumor clonal heterogeneity is common in *de novo* AML [3,7,12] (Table 1). By performing deep WGS (approximately 312× coverage), it was

Table 1. Selected WGS Studies of Clonal Heterogeneity in Adult Cancer[a]

Disease	Methods	Control Tissue	Comments	Refs
AML	WGS of eight paired diagnostic and relapse cases	Skin	Founding clone was always present at relapse and exhibited clonal evolution	[40]
	WGS of 50 cases	Skin	>50% of cases contained a subclone(s) in addition to founding clone	[11]
	WGS of 24 cases[b]	Skin	AML typically contains a subclone(s) and always a founding clone; hematopoietic stem cells acquire random mutations with age that are captured by expanding clone	[3]
	WGS of 19 cases[c]	Skin	AML subclones exhibit functional heterogeneity	[7]
	WGS of 58 cases	Skin	AML typically contains a subclone(s) and always a founding clone; leukemia-associated mutations can persist at morphologic remission	[12]
Secondary AML	WGS of 15 cases	Skin	Progression to secondary AML characterized by persistence of MDS founding clone and expansion of a subclone(s)	[14,15]
Therapy-related AML (t-AML)	WGS of 22 cases[d]	Skin	Number of subclones in t-AML similar to *de novo* AML; subclones harboring *TP53* mutations can expand preferentially after therapy	[43]
Lung (non-small cell lung cancer)	WGS of 17 cases	Adjacent normal lung	Multiclonal tumor in ten of 17 cases; *KRAS* and *EGFR* mutations present in founding clone	[38]
Melanoma (metastatic)	WGS of 13 cases	Peripheral blood mononuclear cells	Most tumors were multiclonal; *NRAS* and *BRAF* mutations generally clonal; example of tumor with distinct mutational signatures in founding clone and subclone	[34]
Prostate (metastatic)	WGS of ten cases	Normal liver, spleen, or kidney	Considerable clonal heterogeneity exists in prostate cancer; metastasis to metastasis spread by both monoclonal and polyclonal seeding occurs	[39]
Gastric	WGS of 49 cases	Peripheral blood mononuclear cells	Clonal heterogeneity is present in gastric cancer	[37]
Esophageal adenocarcinoma (EAC)	WGS of 23 paired Barrett's esophagus and EAC cases	Blood or normal esophagus	Barrett's esophagus is polyclonal; dysplasia can develop from multiple clones;<20% overlap of single nucleotide variants in paired samples in most cases	[36]
Breast (primary)	WGS of ten cases	Blood	Subclonal heterogeneity varied among tumors	[26]

[a]Minimum of five cases sequenced with standard approximately 30× coverage of primary tumor samples.
[b]Two cases previously reported.
[c]Four cases previously reported.
[d]One case previously reported.

shown that deep sequencing substantially improves the discovery of variants across a range of VAFs, and allows for a more definitive model of tumor clonal architecture [13]. Collectively, the data indicate that all AML samples contain multiple clones at presentation. Similarly, standard WGS studies of 15 secondary AML genomes [AML arising from an antecedent myelodysplastic syndrome (MDS)] demonstrated that secondary AML always contained a founding clone and one or more subclones [14,15] (Figure 2). This intratumoral genetic complexity has also been reported for other hematological malignancies using WGS or WES, including MDS [15–17], chronic lymphocytic leukemia (CLL) [18–20], multiple myeloma [21,22], and non-Hodgkin lymphoma [23].

Solid tumors, including but not limited to, renal cell, colorectal, melanoma, head and neck, lung, pancreatic, gastric, esophageal, and breast, also exhibited clonal heterogeneity when multiple regions of the bulk tumors were interrogated using NGS platforms [24–39] (Table 1). For example, clonal heterogeneity was evident in renal cell carcinoma using WES to sequence multiple regions within the primary tumors from ten patients where approximately 70% of somatic mutations were not detected uniformly throughout an individual tumor [24,25]. The authors showed that approximately 10% of the regions sequenced to approximately 100× demonstrated subclones in one region that were the most prevalent clone in another tumor region [25]. The authors found that clonal heterogeneity appeared to increase with the number of biopsies, indicating that a single biopsy probably underrepresents the complexity of a tumor [25]. Independent studies of localized lung cancer using WGS or WES also demonstrated clonal heterogeneity [29,30,38]. Using multiregion sequencing, there were several examples of driver mutations that were in the founding clone in one region and a subclone or absent in a different region [30], similar to renal cell carcinoma.

The clonal evolution of breast cancer was also demonstrated using multiregion sequencing. Targeted candidate gene sequencing (360 genes, 166× coverage) of multiple geographically distinct areas from an individual tumor revealed that eight out of 12 tumors had spatial clonal heterogeneity [26]. In four tumors, subclonal driver mutations were identified that were absent in the majority of other sampled regions despite >1000× coverage. When this approach was extended to 50 primary tumors using targeted gene capture, subclonal heterogeneity was again observed in a subset of tumors [26].

Collectively, clonal evolution is common in hematologic and solid tumors and presents a diagnostic challenge, especially in solid cancers, which may harbor regional heterogeneity in subclone distribution and the presence of actionable target genes. All cancers likely contain multiple clones at presentation. However, the ability to detect these clones remains limited by the sequencing coverage depths obtained and spatial heterogeneity. Sampling bias in solid tumors may lead to difficulty detecting founding clones. Clonal evolution presents a therapeutic challenge, because whether a mutation is in the founding clone or a subclone is likely to have an impact on the efficacy of targeted therapies (discussed below).

Clonal Heterogeneity Has Implications for Tumor Progression, Relapse, and Metastasis

Progression of disease from MDS to secondary AML is a dynamic process shaped by subclone evolution and clonal selection. Using NGS of paired MDS and secondary AML genomes (with skin as control normal tissue), disease progression from MDS to secondary AML was always characterized by the persistence of the MDS founding clone and expansion of a subclone (Figure 2) [14,15]. In *de novo* AML, WGS of eight primary tumor and relapse pairs was used to define clonal architecture at presentation and at relapse [40]. Clonal evolution at relapse was present in one of two patterns: either the founding clone itself acquired additional mutations, or a subclone (derived from the founding clone) that was present at diagnosis underwent clonal

evolution by gaining new mutations and expanded at relapse. In addition, chemotherapy impacted the mutation spectrum at relapse, with a significantly higher frequency of transversion mutations in relapse samples compared with primary tumor samples, indicating that chemotherapy can shape clonal evolution.

WES studies using diagnostic samples of low-grade gliomas harboring isocitrate dehydrogenase 1 (*IDH1*) mutations and paired subsequent local recurrences revealed a spectrum of clonal evolution [32]. In some cases, the recurrent tumor acquired additional mutations that suggested it was derived directly from the initial glioma, while in others, the linear evolutionary relation between the primary tumor and recurrence was not clear. This suggested to the authors that an ancestral clone had independently given rise to the primary and recurrent gliomas. The authors also observed a temozolomide-induced mutation signature in six out of ten recurrent tumors from treated patients, again supporting that chemotherapy can shape clonal evolution of disease progression [32].

Clonal heterogeneity and subclone evolution in the primary tumor may also have a role in solid tumor metastases [26,31,39,41]. In breast cancer, WGS of samples from the same patient showed that metastatic disease arose from subclones detected in the primary tumor [26,41]. In pancreatic cancer, sequencing of multiple sections of the primary tumor from two patients was performed and showed that geographically distinct subclones were present that gave rise to anatomically separated metastases [31]. Similarly, subclones present in primary prostate cancers were involved in seeding metastases [39]. These studies consistently show that subclones can contribute to tumor progression, relapse, and metastasis in various cancers.

Subclonal Diversification Has Prognostic and Therapeutic Implications

Although data are still somewhat limited, studies across different tumor types are beginning to address whether the number of subclones and stability of subclonal architecture are important for prognosis. In solid tumors, WES performed on multiple areas of a limited number of localized lung adenocarcinomas showed that patients who relapsed had a larger fraction of subclonal mutations than patients without relapse (40% versus 17%, $P = 0.006$) [29]. WES of multiregion samples harvested from eight patients with esophageal cancer showed an association between high clonal heterogeneity and poor response to neoadjuvant chemotherapy [42]. The presence of a chemotherapy-induced mutation signature was also observed following platinum-based chemotherapy in these samples [42]. In CLL, patients with subclonal driver mutations had a shorter time to treatment or death than those without subclonal driver mutations, independent of other poor prognostic markers [18]. In a separate study, WGS of serial samples from three patients with CLL revealed that disease progression correlated with an expansion of a subclone in two patients [19]. The number of subclones has also been suggested to impact the risk of MDS progression to secondary AML. Patients with MDS at high-risk of progressing to secondary AML harbored more subclones than low-risk patients with MDS in one study [17], while another study observed a correlation between the number of driver gene mutations present in the MDS sample and leukemia-free survival [16]. By contrast, the degree of clonal heterogeneity in one breast cancer study did not correlate with response to neoadjuvant therapy, although only a limited number of cases were examined [26].

The driver gene present in a subclonal population may also have prognostic and therapeutic relevance. Studies in CLL examined the interaction between Tumor Protein P53 (*TP53*) mutations and cytotoxic chemotherapy. Rossi *et al.* identified that patients with minor subclones harboring a *TP53* mutation at diagnosis had an equally unfavorable prognosis as those with *TP53* mutations in their founding clone, and that the pre-treatment *TP53* subclones expanded to become the dominant clone at relapse [20]. Recently, Wong *et al.* also demonstrated treatment selection of hematopoietic cells with *TP53* alterations in patients with therapy-related MDS/AML

[43]. The *TP53* mutations discovered at diagnosis were found to be present at low frequencies (0.003–0.7%) in hematopoietic cells 3–6 years before the development of disease [43]. Furthermore, the *TP53* mutation was found before any chemotherapy in two of these cases where samples were collected before chemotherapy [43]. Therefore, chemotherapy likely contributes to the selection of rare pre-existing chemoresistant clones, which have a survival advantage after therapy.

Given that clonal heterogeneity is common in cancer, what mutations should be prioritized for therapeutic targeting? Pharmacologic targeting of a particular mutation may have different clinical impact depending on whether the mutation is in the founding clone or a subclone. Driver mutations in the founding clone are attractive drug targets [e.g., translocations in promyelocytic leukemia-retinoic acid receptor, *alpha (PML-RARA)* in acute promyelocytic leukemia, and breakpoint cluster region-ABL proto-oncogene 1, nonreceptor tyrosine kinase *(BCR-ABL)* in chronic myeloid leukemia], since they are present in all the tumor cells, including all subclones. However, targeting recurrently mutated cancer genes regardless of which clone they occur in will likely remain a challenge, since driver mutations are sometimes present in the founding clones in one cancer but in a subclone in another. Indeed, a recent survey of the clonal status of actionable driver events in nine major cancer types showed that mutations with clinical relevance could be in subclones of a tumor sample [44]. Given that mutation-specific targeted therapies are now being used in clinical trials, we can begin to address whether the clonal distribution of driver gene mutations matters. For example, if an AML sample harbors an *IDH1* mutation only in a subclone, treatment with an IDH1 inhibitor would not be predicted to affect the founding clone (Figure 3, Key Figure). This would likely result in only transient clinical benefit and potentially contribute to the expansion of the founding clone or subclones lacking *IDH1* mutations. Future clinical studies will have to address the best approach for incorporating the (sub)clonal status of actionable target genes into therapeutic decisions and monitoring patient outcomes.

Monitoring Tumor Burden Using Mutation VAFs

Identification of disease based on molecular markers (i.e., gene mutations) could lead to early initiation of treatment before clinical relapse or chemorefractoriness. The optimal approach to monitoring tumor burden, including what mutations to sequence, the sequencing platform, and the samples to sequence, is an active area of research. No single platform or approach will likely be ideal for all cancers. As discussed above, the regional clonal heterogeneity observed in solid tumors presents a challenge to studying clonal heterogeneity using material from a single biopsy site. One approach to overcome the difficulty of obtaining multiregion biopsies for solid tumors is to sequence plasma cell free DNA from patients. While this approach can identify the clonal complexity present in solid tumors, it remains technically challenging, and it is unclear whether DNA from all subclones is equally represented in the plasma [45]. By contrast, monitoring the subclonal architecture in hematopoietic cancers has an advantage over solid tumors because leukemia cells appear to freely mix in the bone marrow or blood [7]. Therefore, we focus on how mutation VAFs can be used to monitor tumor burden and subclonal architecture in hematopoietic cancers (Figures 2 and 3).

While sequencing bulk bone marrow samples from patients with leukemia has been used to impute tumor clonality, this approach may underestimate the clonal heterogeneity of some samples. For example, subclones containing mutations with similar VAFs may not be distinguishable from each other based on sequencing of a single time point. Sequencing single cells or serial samples from the same individual could be used to overcome this limitation. Paguirigan *et al.* showed that multiple subclones may exist with similar VAFs by performing single cell sequencing for fms-related tyrosine kinase 3 (*FLT3)* and nucleophosmin 1 (*NPM1)* mutations in patients with AML [9]. Similarly, single cell sequencing of samples from patients with ALL found

Key Figure

Tumor Burden Monitoring Using Next-Generation Sequencing (NGS)

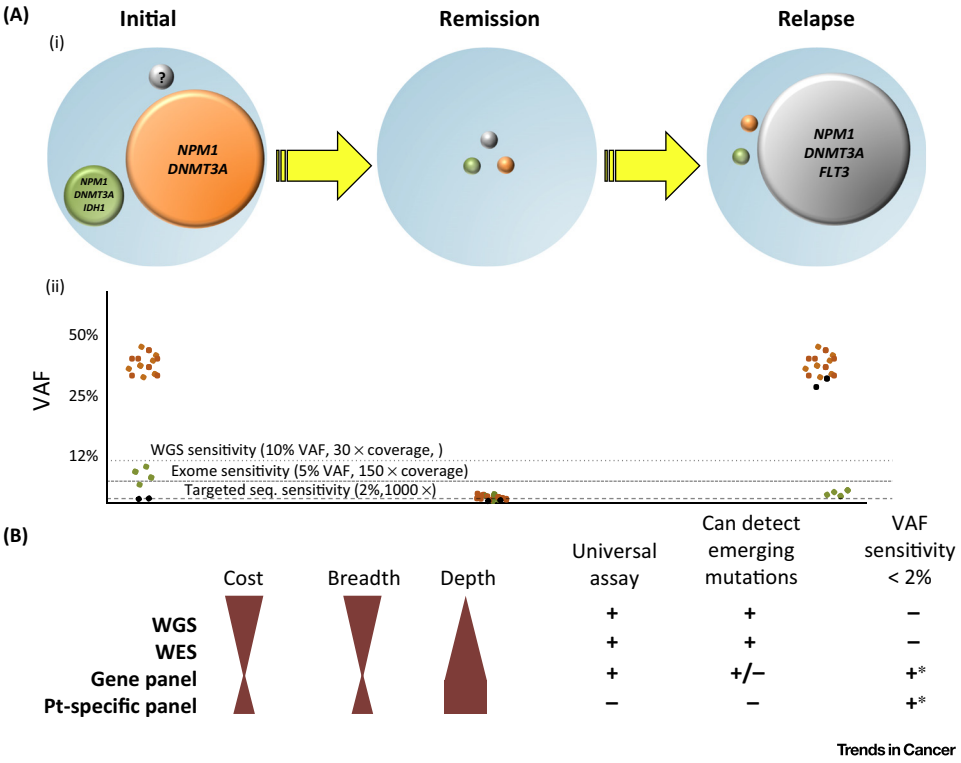

Figure 3. (A) We have created these data for illustrative purposes. (i) In this acute myeloid leukemia (AML) example, sequencing identifies a founding clone containing multiple mutations (orange), including a nucleophosmin (*NPM1)* and DNA (cytosine-5-)-methyltransferase 3 alpha *(DNMT3A)* mutation, and a subclone (green) containing the founding clone mutations and subclone-specific mutations, including an isocitrate dehydrogenase 1 (*IDH1)* mutation. An additional small subclone (black) is present below the limit of detection using standard next-generation sequencing (NGS) methods [whole-genome sequencing (WGS), whole-exome sequencing (WES), or targeted gene panels, indicated by the broken lines in (ii)]. During remission, the mutation variant allele fraction (VAFs) of the founding clone (orange dots) and subclone mutations (green dots) decrease and are undetectable using standard WGS and WES. Many of these mutations, excluding the black subclone mutations, are detectable by targeted sequencing (1000× coverage, 1–2% VAF sensitivity). At relapse, a subclone (black) emerges containing founding clone mutations and a fms-related tyrosine kinase 3 (*FLT3)* mutation. Using ultra-sensitive sequencing techniques, the clone that dominates at relapse (black) in fact pre-existed at initial diagnosis at extremely low levels below the limit of detection using standard methods. (B) A comparison of sequencing methods for tumor burden and clonality monitoring. While WGS offers the greatest sequencing breadth, standard coverage is generally only 30×, limiting detection of mutations to those with a VAF of at least 10%. The breadth of typical WES platforms is approximately 1% of the genome (proportional to the annotated coding region) and typically provides approximately 150× coverage. This allows for detection of mutations with VAFs of at least 5%. 'Targeted' or 'gene panel'-based sequencing is generally limited to 50–500 genes providing minimum sequencing breadth, but can achieve high coverage (1000×) for a limited cost compared with WGS and WES. Once a WGS and WES platform is credentialed, it can be used universally for all patients (i.e., universal assays). Panel-based sequencing can include 'general' panels that include the most commonly mutated genes in a particular cancer (i.e., universal assay), or 'patient-specific' panels (i.e., nonuniversal) that are custom designed to detect all somatic mutations originally identified using broader WGS or WES approaches. For tumor-monitoring applications in AML, the universal 'one-size fits most' gene panel is limited to detecting and monitoring two or three mutations on average in a patient. By comparison, patient-specific panels allow for monitoring all mutations originally identified using WGS or WES, potentially providing greater accuracy to monitor clones, but with added costs of assay design and implementation challenges. As opposed to unbiased WGS and WES, patient-specific sequencing panels cannot detect mutations that were not identified in the original discovery sample (i.e., new emergent mutations are missed in serial samples). *An advantage of targeted panels is that they can be used in conjunction with error-corrected sequencing approaches to potentially identify mutations with VAFs well below 2%, given sufficient coverage depths are obtained (>10 000×) [47,48].

that clonal heterogeneity was higher than predicted based on bulk sample sequencing [8]. While clones with similar mutation VAFs can be distinguished using single cell sequencing, this technology remains technically challenging and is too expensive to implement clinically. A more feasible approach to resolve independent clones with similar VAFs is to monitor changes in VAFs in serial samples. Two clones that display differential growth properties or response to therapy will have their VAFs diverge over time. Therefore, sequencing serial samples may provide critical information regarding subclonal response to therapy, and the utility of this approach could be tested in clinical trials.

Monitoring tumor burden through measurements of VAF could contribute to clinical decision making. For example, a decrease in mutant VAF would be expected to correlate with a response to therapy. However, knowledge of the subclonal architecture of the tumor may be critical in estimating the response, because monitoring a single gene mutation can underestimate overall tumor burden if the mutation is in a subclone. As an example, if *NPM1* (Figure 2B, present in the orange subclone) were the only mutation being monitored in a patient, the tumor burden would be predicted to be declining over time. However, the overall tumor burden is in fact unchanged because additional subclones are expanding as the mutant NPM1 clone is contracting (Figure 2B). Thus, tracking only the *NPM1* mutation VAF in this patient would underestimate the tumor burden and therapeutic response over time, highlighting a limitation of monitoring clonality and tumor burden using a single driver gene. Using mutation VAFs to impute tumor clonality and to track the contraction and emergence of subclones in response to therapy will likely prove useful in answering questions such as: must the founding clone be eradicated to achieve a durable remission? Does control of subclonal populations impart clinical benefit in the absence of eradication of the founding clone? And, does the therapy chosen affect the subclonal populations that expand at relapse and/or progression? Thus, clinical trials are necessary to define the role that tumor clonality assessment and disease burden monitoring has on therapeutic decisions.

Several variables should be considered when choosing a sequencing method to discover and monitor changes in tumor subclonal composition (i.e., tumor clonality). While most studies have relied on NGS-based methods, including WGS, exome sequencing, or high-coverage targeted sequencing, alternative technologies exist, including droplet digital PCR (ddPCR) and COLD PCR [43,46]. The major advantage of PCR-based methods to monitor changes in tumor subclonal composition is that they can provide sensitivities higher than standard NGS sequencing (on the order of one mutated cell in 1000 normal cells or better). Such methods are ideal in the clinical setting for the detection of highly recurrent 'hotspot' mutations, such as *BRAF* V600E mutations in melanoma, *KRAS* codon 12 mutations in lung cancer, or *IDH1/2* mutations in central nervous system (CNS) malignancies, but are less well suited for the detection of nonrecurrent mutations, such as those in tumor suppressor genes. For NGS-based methods, there are multiple factors to consider when choosing an assay, including the number of mutations that can be discovered (dependent on the breadth of coverage) and the limit of detection at which a mutation or clone can be identified (dependent on the depth of coverage) (Figure 3). WGS remains the most comprehensive and unbiased platform to detect and monitor clonal heterogeneity, but the depth of coverage for standard WGS is generally approximately 30× due to cost, which limits the discovery of mutations to those with VAFs of >10%. By contrast, targeted sequencing panels can be used to achieve very high coverage depths (often >1000×) of 50–400 genes at a fraction of the cost of WGS. With adequate coverage, these targeted panels are capable of detecting mutations with VAFs of approximately 1%; however, they often have insufficient sequencing breadth to discover enough mutations to confidently reconstruct the subclonal architecture of a tumor. Exome sequencing reagents, which target between 30 and 70 megabases of coding sequencing space, provide an alternative approach. Standard WES platforms produce higher sequencing coverage than WGS (generally on the order of 100–200× coverage) at a reasonable cost, allowing detection of mutations with VAFs as

low as 5% (Figure 3). However, as opposed to WGS, genetic changes located in noncoding DNA regions, which could be important, will be missed.

For cancer types with a high mutation rate, including some lung cancers and melanomas, exome or even panel-based targeted sequencing can identify a sufficient number of mutations to reconstruct the subclonal architecture of a tumor. However, this is not the case for cancers with lower mutation rates, including MDS and AML, where panel-based sequencing may only identify three or four mutations per sample [15–17]. A hybrid approach could be considered for many cancers. Initial diagnostic tumors could be sequenced using platforms providing adequate sequencing breadth (i.e., WGS or WES), followed by monitoring of serial samples using gene panels targeting the patient-specific mutations identified at diagnosis. While this approach has its advantages, it will not detect new mutations or clones that expand during progression or relapse that were not detected in the original diagnostic sample (Figure 3). In addition, validation of patient-specific PCR or hybridization-capture assays is cumbersome and may not be clinically feasible. While many questions remain regarding the best approach to monitor mutations and clones in tumor samples, progress has been made in several cancers, including AML.

Recently, sequencing results were used to identify patients with AML in remission who were at an increased risk of future relapse. Somatic mutations were first identified in a patient's diagnostic AML sample using WGS or WES. Next, these mutations were genotyped in their paired remission bone marrow sample following induction chemotherapy using WES or a targeted patient-specific NGS gene panel [12]. This approach achieves high coverage depths with less overall cost and allows for the identification of clones with VAFs of <2% in remission samples. Detection of persistent leukemia-associated mutations in at least 5% of bone marrow cells (VAFs = 2.5%) in remission samples obtained approximately 30 days after induction chemotherapy was associated with worse outcomes. The data suggest that molecular monitoring of tumor burden using mutation VAFs obtained from NGS results improves risk stratification for patients with AML. Future clinical trials will be required to determine how best to broadly implement and monitor tumor clonal heterogeneity and implement these findings into the clinical care of cancer patients in real time.

Concluding Remarks

Large-scale sequencing efforts have defined the genetic landscape of cancers, increased our understanding of the genetic diversity across tumor types, and identified recurrently mutated driver genes and genetic pathways in cancer. Cancer is a mixture of subclonal populations derived from a founding clone, each with different inherent properties and responsiveness to therapy. Studies support the idea that relapsed disease is often derived from an emergent subclone that can be molded by selection pressure from chemotherapy. Thus, it is increasingly apparent, as Nowell predicted, that each patient is an 'individual therapeutic problem'. Measuring changes in clonal heterogeneity in response to therapy should be included in clinical trials going forward. A better understanding of how subclonal diversity at presentation, or after therapy, impacts outcomes could ultimately provide improved prognostic information that may alter treatment choices for patients with cancer (see Outstanding Questions).

Outstanding Questions

Is the number of tumor clones (i.e., level of clonal heterogeneity) present in a diagnostic sample predictive of a patient's response to specific chemotherapies? Future clinical trials could incorporate this into the study design.

Is the clinical response to mutation-specific targeted therapies influenced by whether the mutation is in the founding clone or a subclone? This must be tested in a uniform cohort of patients all receiving the same treatment and follow-up monitoring.

Is it necessary to therapeutically target a mutation in the founding clone to cure cancer? It is possible that altering the clonal architecture of a tumor by targeting multiple subclones could provide clinical benefit without eradicating the founding clone.

What is the best NGS approach to maximize the number of clones that can be detected in a cancer? Accurate determination of the number of clones in a tumor requires detection of a large number of mutations. A balance between sequencing breadth and depth needs to be considered when choosing an approach to detect tumor clones.

What is the best NGS approach to serially monitor tumor burden in patients? The utility of serially monitoring tumor burden using sequencing will have to be tested in prospective clinical trials to determine whether sequencing provides information independent of standard clinical parameters.

Does the detection of subclones that comprise <5% of the cells in a tumor have important clinical significance (i.e., how sensitive does the NGS platform need to be)? If detecting and tracking rare subclones is important, the development and implementation of clinically applicable sequencing methods will have to be tested in clinical trials.

Acknowledgments

M.J.W. is a Scholar of the Leukemia and Lymphoma Society. We thank Tim Graubert and Tim Ley for helpful scientific discussions. We apologize to authors whose works are not cited due to space constraints.

References

1. Nowell, P.C. (1976) The clonal evolution of tumor cell populations. *Science* 194, 23–28
2. Yates, L.R. and Campbell, P.J. (2012) Evolution of the cancer genome. *Nat. Rev. Genet.* 13, 795–806
3. Welch, J.S. *et al.* (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150, 264–278
4. Landau, D.A. *et al.* (2014) Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia* 28, 34–43

5. McGranahan, N. and Swanton, C. (2015) Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* 27, 15–26
6. Hughes, A.E. *et al.* (2014) Clonal architecture of secondary acute myeloid leukemia defined by single–cell sequencing. *PLoS Genet.* 10, e1004462
7. Klco, J.M. *et al.* (2014) Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. *Cancer Cell* 25, 379–392
8. Gawad, C. *et al.* (2014) Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17947–17952
9. Paguirigan, A.L. *et al.* (2015) Single-cell genotyping demonstrates complex clonal diversity in acute myeloid leukemia. *Sci. Transl. Med.* 7, 281re282
10. Ley, T.J. *et al.* (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72
11. The Cancer Genome Atlas Research Network (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 368, 2059–2074
12. Klco, J.M. *et al.* (2015) Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. *JAMA* 314, 811–822
13. Miller, C. *et al.* (2015) Optimizing cancer genome sequencing and analysis. *Cell Syst.* 1, 210–223
14. Walter, M.J. *et al.* (2012) Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366, 1090–1098
15. Walter, M.J. *et al.* (2013) Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* 27, 1275–1282
16. Papaemmanuil, E. *et al.* (2013) Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 122, 3616–3627 quiz 3699
17. Haferlach, T. *et al.* (2014) Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 28, 241–247
18. Landau, D.A. *et al.* (2013) Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 152, 714–726
19. Schuh, A. *et al.* (2012) Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* 120, 4191–4196
20. Rossi, D. *et al.* (2014) Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood* 123, 2139–2147
21. Egan, J.B. *et al.* (2012) Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood* 120, 1060–1066
22. Bolli, N. *et al.* (2014) Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* 5, 2997
23. Green, M.R. *et al.* (2013) Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood* 121, 1604–1611
24. Gerlinger, M. *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366, 883–892
25. Gerlinger, M. *et al.* (2014) Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* 46, 225–233
26. Yates, L.R. *et al.* (2015) Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* 21, 751–759
27. Cooper, C.S. *et al.* (2015) Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat. Genet.* 47, 367–372
28. Sottoriva, A. *et al.* (2015) A Big Bang model of human colorectal tumor growth. *Nat. Genet.* 47, 209–216
29. Zhang, J. *et al.* (2014) Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346, 256–259
30. de Bruin, E.C. *et al.* (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 346, 251–256
31. Yachida, S. *et al.* (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 467, 1114–1117
32. Johnson, B.E. *et al.* (2014) Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* 343, 189–193
33. Mroz, E.A. *et al.* (2015) Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS Med.* 12, e1001786
34. Ding, L. *et al.* (2014) Clonal architectures and driver mutations in metastatic melanomas. *PLoS ONE* 9, e111153
35. Shah, S.P. *et al.* (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486, 395–399
36. Ross-Innes, C.S. *et al.* (2015) Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. *Nat. Genet.* 47, 1038–1046
37. Wong, S.S. *et al.* (2014) Genomic landscape and genetic heterogeneity in gastric adenocarcinoma revealed by whole-genome sequencing. *Nat. Commun.* 5, 5477
38. Govindan, R. *et al.* (2012) Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* 150, 1121–1134
39. Gundem, G. *et al.* (2015) The evolutionary history of lethal metastatic prostate cancer. *Nature* 520, 353–357
40. Ding, L. *et al.* (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481, 506–510
41. Ding, L. *et al.* (2010) Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* 464, 999–1005
42. Murugaesu, N. *et al.* (2015) Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov.* 5, 821–831
43. Wong, T.N. *et al.* (2015) Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* 518, 552–555
44. McGranahan, N. *et al.* (2015) Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci. Transl. Med.* 7, 283ra254
45. Diaz, L.A., Jr and Bardelli, A. (2014) Liquid biopsies: genotyping circulating tumor DNA. *J. Clin. Oncol.* 32, 579–586
46. Murphy, D.M. *et al.* (2013) NRAS mutations with low allele burden have independent prognostic significance for patients with lower risk myelodysplastic syndromes. *Leukemia* 27, 2077–2081
47. Schmitt, M.W. *et al.* (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14508–14513
48. Kinde, I. *et al.* (2011) Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9530–9535