Tracking Cancer Evolution through the Disease Course

Chris Bailey¹, James R.M. Black²,³, James L. Reading⁴, Kevin Litchfield⁵, Samra Turajlic⁶, Nicholas McGranahan²,³, Mariam Jamal-Hanjani³,⁷, and Charles Swanton¹,³,⁷

ABSTRACT

During cancer evolution, constituent tumor cells compete under dynamic selection pressures. Phenotypic variation can be observed as intratumor heterogeneity, which is propagated by genome instability leading to mutations, somatic copy-number alterations, and epigenomic changes. TRACERx was set up in 2014 to observe the relationship between intratumor heterogeneity and patient outcome. By integrating multiregion sequencing of primary tumors with longitudinal sampling of a prospectively recruited patient cohort, cancer evolution can be tracked from early- to late-stage disease and through therapy. Here we review some of the key features of the studies and look to the future of the field.

Significance: Cancers evolve and adapt to environmental challenges such as immune surveillance and treatment pressures. The TRACERx studies track cancer evolution in a clinical setting, through primary disease to recurrence. Through multiregion and longitudinal sampling, evolutionary processes have been detailed in the tumor and the immune microenvironment in non–small cell lung cancer and clear-cell renal cell carcinoma. TRACERx has revealed the potential therapeutic utility of targeting clonal neoantigens and ctDNA detection in the adjuvant setting as a minimal residual disease detection tool primed for translation into clinical trials.

INTRODUCTION

Over the last decade, the study of tumor evolution has expanded significantly, aided by the advent of technologies that enable rapid and relatively inexpensive sequencing of cancer exomes and genomes. This has shaped our understanding of tumor diversification and evolution, with genetic and nongenetic variation providing the basis for intratumor heterogeneity. By tracking both clonal (present within all tumor cells) and subclonal (present within a subset of tumor cells) mutations, researchers have found that tumors follow distinct evolutionary patterns. For instance, in some tumors there is evidence for a Darwinian pattern of evolution with selection of subclonal mutations, whereas in others there appears to be a neutral accumulation of mutations following a clonal driver event (1).

A key focus of research has been to link these observations to clinical outcomes, such as prognosis and treatment response, in order to understand the driving forces behind cancer progression and to develop personalized approaches to cancer medicine. The national observational longitudinal studies TRACERx (Tracking Cancer Evolution through Therapy) Lung (ClinicalTrials.gov Identifier: NCT01888601) and Renal (ClinicalTrials.gov Identifier: NCT03226886) have attempted to address this (2, 3). Observations from the first 100 patients recruited to the TRACERx Lung and Renal studies have been published in a series of reports, with data collection now complete for the first 421 patients in the TRACERx Lung study.

These studies involve sampling of primary and recurrent disease, as well as “liquid” biopsies to isolate circulating tumor DNA (ctDNA) and circulating tumor cells (CTC) from blood samples.

THE TRACERx STUDIES

The primary objectives of the TRACERx studies are to determine the extent of intratumor heterogeneity and establish its relationship with clinical outcome and the impact of
adjuvant and other therapies on the cancer genomic landscape at recurrence.

Multiregion sampling and sequencing of surgically resected specimens in TRACERx allow for the detailed assessment of intratumor heterogeneity. However, the majority of sampling at relapse or progression, where clinically safe and possible, is limited to single-lesion biopsies. Invasive biopsies are avoided if the procedure is deemed to be high risk, due to either the patient’s comorbidities or the site of disease. ctDNA analysis in this context can capture genetic heterogeneity in a manner not limited to a single site of disease. Furthermore, corecruitment into the Cancer Research UK (CRUK)-funded national research autopsy study, PEACE (Posthumous Evaluation of Advanced Cancer Environment; ClinicalTrials.gov Identifier: NCT03004755), allows for extensive sampling of all metastatic disease sites in the postmortem setting.

Although the benefit of retrospective analysis of data from significant resource projects such as The Cancer Genome Atlas (TCGA) and Pan Cancer Analysis of Whole Genomes cannot be overstated (4, 5), large-scale longitudinal prospective studies of patients and their cancer genomes with clinical histories (case report forms), imaging and pathology analysis matched with paired recurrence specimens, and ctDNA/CTC acquisitions are rare. Typically, genomic studies of cancer provide only a snapshot of the genomic landscape and seldom contextualize this temporally or spatially in a cohort of prospectively recruited patients in a real-world clinical setting with standard-of-care therapy to minimize confounding effects of treatment discordance across clinical sites. Furthermore, sampling, sequencing, and processing protocols may be non-uniform, and the plethora of computational tools for interrogating genomic data may introduce technical inconsistencies and limit the reliability of subsequent retrospective analyses. Crucially, independent pathologic review of specimens may not be performed, and clinical annotation is often absent or incomplete. Such studies frequently consider the cancer genome in isolation and in the absence of other features including its transcriptome or epigenome, as well as the immune microenvironment, metabolome, and microbiome. There is therefore an unmet scientific and clinical need for prospective, longitudinal evolutionary cohort studies of cancer “multiomics” with robust clinicopathologic correlates.

Here, we outline some of the lessons learned from TRACERx longitudinal analyses, the importance of future such studies in other tumor types, and the questions outstanding in the field, offering a road map for cancer evolutionary studies in the coming decade.

**SETTING UP A PROSPECTIVE LONGITUDINAL STUDY IN CANCER EVOLUTION**

TRACERx has galvanized a national research infrastructure linking surgical oncology, radiology, pathology, radiation, and medical oncology disciplines, with trial-specific procedures set up across multiple UK hospital sites to sample tumors with a view to performing detailed genomic and immunologic analyses alongside ctDNA and CTC analysis. Samples are tracked using a centralized database enabling sample oversight across all sites. Detailed case report forms are collected routinely from patients at each clinic visit, recording their general medical and oncologic history. In the adjuvant setting, patients are offered National Institute for Health and Care Excellence (NICE)-approved standardzed adjuvant protocols across all hospital sites according to disease stage, minimizing the confounding variables associated with treatments differing between centers. This has established a unique clinical, pathologic, and scientific data set. Collaborative efforts between the clinical trials center, hospital trusts, and research laboratories have been essential in overcoming some of the logistic challenges posed, from sample collection to patient follow-up.

Despite decades of research into non–small cell lung cancer (NSCLC), clinical outcomes for patients remain poor, with a five-year relative survival of 26.5% reported in the November 2019 Surveillance, Epidemiology, and End Results Program (SEER) data submission (https://seer.cancer.gov/csr/1975_2017/). Moreover, significant numbers of patients with early-stage disease will go on to relapse despite surgery (long-term survival is less than 50% following surgery; ref. 6) and adjuvant chemotherapy. Although molecular driver events such as TP53, KRAS, and EGF mutations, and genome instability, are well described in NSCLC (7), there remains an unmet need for insight into its biology and its relationship with the clinical course.

The structure of TRACERx Lung is illustrated in Fig. 1. Following a diagnosis of early-stage NSCLC, patients eligible for TRACERx Lung receive routine standard of care, undergoing surgery to resect their primary tumor, with or without adjuvant chemotherapy depending on the disease stage. Multiple regions from the primary tumor (and in some cases lymph nodes) are obtained for analysis. Patients are followed up for at least five years, and in patients with recurrent disease, tissue biopsies are obtained from the site of relapse where possible. TRACERx Lung has a target accrual of 842 patients from 14 hospital sites across the UK.

TRACERx Renal is a separate prospective multiregion study of primary tumors and paired metastatic lesions within 320 patients with clear-cell renal cell carcinoma (ccRCC). As with NSCLC, surgery can be curative in early-stage ccRCC, and survival rates are on average superior to those of NSCLC (exceeding 50%), but with significant heterogeneity in clinical outcome; approximately one third of patients with localized ccRCC will relapse after surgical resection. Of note, surgery plays a significant role in the management of metastatic ccRCC, offering the chance to examine the primary tumor across all stages of disease, as well as access to paired metastatic disease. Previous studies had highlighted the common driver events of this disease, including loss of chromosome 3p, which harbors four tumor suppressor genes: VHL, SETD2, BAP1, and PBRM1, all of which are frequently lost through mutation of the remaining allele. Significant intratumor heterogeneity of mutations and somatic copy-number alterations (SCNA) had also been described (8, 9); TRACERx Renal aimed to explore associations between this intratumor heterogeneity and the diversity of clinical outcomes.

The target accrual of 842 in the Lung cohort was required to detect at least a 23% relative risk reduction and a 10% improvement in five-year overall survival, comparing tumors with low and high intratumor heterogeneity and assuming a
Figure 1. Outline of TRACERx Lung and PEACE autopsy studies. A, Following a diagnosis of early-stage NSCLC, patients undergo surgery to resect their primary tumor (with adjuvant chemotherapy, depending on disease stage). Multiple regions from the primary tumor are sampled. In the event of recurrent disease, biopsies are obtained where possible. Patients who are enrolled in the PEACE study have multiple sites of disease sampled for analysis. Tumor samples are processed for WES, RNA-seq, and tissue microarrays. B, Recruitment sites for TRACERx and PEACE in the UK. C, The core analysis of tumor samples involves calling clonal and subclonal mutation and copy-number alterations, with construction of phylogenetic trees. Combining tissue microarray, RNA-seq data, and other techniques has facilitated the study of immune and spatial heterogeneity. PV-CTC, pulmonary vein circulating tumor cells.

For the TRACERx Renal study, a sample size of 320 was required to detect an association between intratumor heterogeneity and disease stage. Assuming a median overall disease survival of 15 months, this sample size was predicted to have enough power to detect a six-month difference in disease-free survival comparing low and high intratumor heterogeneity scores within stages I to III. Prospective studies of this size are essential to effectively integrate complex genomic and pathologic data and make meaningful, clinically applicable conclusions in the context of standard-of-care clinical practice while mitigating statistical problems such as multiple testing that can plague small or retrospective analyses.

The dynamic fitness landscapes of different tumors throughout the disease course can share certain properties and be classified accordingly. Mapping evolutionary trajectories of the cancer genome throughout the disease course requires consideration of the myriad ways in which tumors can adapt to selection pressures applied by a diverse tumor microenvironment. TRACERx uses multiregion whole-exome sequencing (WES) to study the mutation and copy-number landscape. This is complemented by RNA sequencing (RNA-seq), T-cell receptor sequencing, reduced-representation bisulfite sequencing, immunohistochemistry, fluorescent in situ hybridization, flow cytometry, and imaging mass cytometry to reconcile mutational, copy number, epigenomic, and transcriptomic heterogeneity in the context of the tumor and its immune microenvironment. Longitudinal blood sampling for CTCs and ctDNA allows for correlative genomic analysis between tissue and blood. Tissue microarrays are constructed for specialized immunohistochemistry assays, and fresh-frozen tumor samples are used for DNA and RNA extraction, improving the quality of downstream sequencing data relative to formalin-fixed, paraffin-embedded (FFPE) samples. Sequencing data are processed through a uniform bioinformatic pipeline. Published data are made available for clinical and scientific communities through a Data Access Committee overseen by the Cancer Research UK and University College London Cancer Trials Centre.

Patients who subsequently develop recurrent disease after initial curative surgery, where possible, undergo a biopsy to confirm relapse or to exclude a new cancer primary. Depending on the site of the disease, ease of access for biopsy, the performance status of the patient, and associated comorbidities, a decision can be made on the appropriateness of biopsy sampling. Despite the clinical indications for sampling in the relapse setting, all research biopsies are costed to support the interventional teams at the various hospital sites.

TRACERx patients who suffer disease relapse are, where sensitive and appropriate discussions allow, corecruited into the PEACE study, allowing for sampling from all sites of disease in the postmortem setting. The PEACE study has highlighted unique logistic challenges in relation to autopsies proceeding in a timely manner within the context of National Health Service mortuary services and resources.
The COVID-19 pandemic has caused widespread disruption to cancer care and cancer clinical trials (10), with an inevitable impact on patient recruitment and follow-up. Despite this, where possible, funding bodies have supported research staff to persevere in delivering the research promised by studies such as TRACERx and PEACE.

INSIGHTS FROM TRACERx

Study Background

Intratumor heterogeneity has long been described in cancer and proposed as evidence for evolutionary processes at play in cancer development and progression. Early observations from Nowell, who described phenotypic heterogeneity within cancers, and Goldie, who postulated a link between genomic instability, tumor heterogeneity, and drug resistance, supported a process of Darwinian clonal evolution in tumor biology (11, 12).

Subsequently, retrospective studies from our group and others, prior to TRACERx, demonstrated that intratumor heterogeneity results from branched tumor evolution (8, 13–17). Navin and colleagues used array comparative genomic hybridization to illustrate differences between ploidy profiles of different tumors (13). Tumors showing “polygenomic” profiles, indicating the presence of genomic heterogeneity, were common, and descent of different clones from a common ancestor was demonstrated. The advent of next-generation sequencing has also provided significant insight. For example, Greaves and colleagues demonstrated that acute lymphoblastic leukemia evolves through branched, nonlinear evolution, producing distinct subclones (14). Serial transplantation in mouse models demonstrated that these different clones were functional and able to potenti ate novel cancers. Our group reported intratumor heterogeneity within driver mutations for ccRCC and suggested selection of subclonal mutations late in tumor evolution (8). In addition, subclonal driver mutations were also a predictor of poor prognosis in a longitudinal cohort of chronic lymphocytic leukemia, and low-frequency mutations present prior to treatment were shown to contribute to resistance to targeted therapy (16, 18).

Before TRACERx, it was understood that SCNAs contributed to intratumor heterogeneity (19). Exploration of the SCNA landscape across cancer types had demonstrated an association between the presence of SCNAs, genomic instability, and poor clinical outcome, in terms of both survival and treatment resistance (20, 21). However, the impact of the rate of acquisition of these events, reflecting the degree of ongoing chromosomal instability (CIN), upon clinical outcome was not known. In addition, whole-genome doubling events had been characterized across cancer types (22), but the relevance of this event to tumor evolution, as well as the extent to which CIN provides the variation that acts as the substrate for selection, was unclear.

Studies of the prognostic or predictive relevance of intratumor heterogeneity had previously been limited to small, retrospective cohorts. TRACERx aimed to explore this question prospectively, attempting to address whether somatic heterogeneity might be associated with poor clinical outcome. The major findings are summarized in Table 1.

### Table 1. Summary of key findings from TRACERx

<table>
<thead>
<tr>
<th>Study</th>
<th>Summary</th>
<th>Tumor type</th>
<th>Samples</th>
<th>Key finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamal-Hanjani et al., 2017</td>
<td>Prospective analysis of intratumor heterogeneity</td>
<td>NSCLC</td>
<td>WES of 327 biopsies from 100 patients</td>
<td>Intratumor heterogeneity of copy-number events, but not mutations, is associated with adverse outcome.</td>
</tr>
<tr>
<td>Turajlic et al., 2018</td>
<td>Analysis of evolutionary trajectories</td>
<td>ccRCC</td>
<td>1,206 primary tumor regions from 101 patients</td>
<td>Different evolutionary subtypes, defined by features such as the clonality of driver mutations or chromosomal complexity, correlate with distinct clinical phenotypes.</td>
</tr>
<tr>
<td>Turajlic et al., 2018</td>
<td>Analysis of paired primary and metastasis</td>
<td>ccRCC</td>
<td>575 primary and 335 metastatic biopsies from 100 patients</td>
<td>Genomic features, such as chromosomal complexity and loss of 9p, promote metastatic competence.</td>
</tr>
<tr>
<td>Mitchell et al., 2018</td>
<td>Multiregion analysis timing landmark events</td>
<td>ccRCC</td>
<td>Whole-genome sequencing of 95 biopsies from 33 patients</td>
<td>Loss of chromosome 3p, an important early event in ccRCC development, can occur in childhood and precede tumor development by 30–50 years.</td>
</tr>
<tr>
<td>Lopez et al., 2020</td>
<td>Study of WGD events and deleterious events in tumor evolution</td>
<td>Pan-cancer</td>
<td>TRACERx 100 and TCGA cohorts</td>
<td>WGD is enriched in tumor types with a high rate of LOH and selected for when deleterious alterations are acquired at a high rate, suggesting WGD events may mitigate against the impact of deleterious alterations in tumor evolution.</td>
</tr>
</tbody>
</table>

(continued)
Table 1. Summary of key findings from TRACERx (Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Summary</th>
<th>Tumor type</th>
<th>Samples</th>
<th>Key finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watkins et al., 2020</td>
<td>Analysis of SCNAs in primary tumors and metastases that have undergone multi-region sampling</td>
<td>Pan-cancer</td>
<td>WES of 1,421 samples from 394 tumors, and 1,024 independent metastatic samples</td>
<td>SCNA heterogeneity provides the substrate for ongoing tumor evolution; more than one third of tumors show evidence of parallel evolution of SCNAs, and certain SCNAs recur at high frequency subclonally and in metastasis.</td>
</tr>
<tr>
<td>Rosenthal et al., 2019</td>
<td>Study of methods of immune evasion</td>
<td>NSCLC</td>
<td>WES, RNA-seq, and tumor-infiltrating lymphocyte estimates from 258 regions of 88 tumors</td>
<td>Tumors experience strong selection pressures to evade immune surveillance early in tumor evolution and develop heterogeneous methods of immune escape.</td>
</tr>
<tr>
<td>Joshi et al., 2019</td>
<td>Exploration of the T-cell receptor (TCR) repertoire</td>
<td>NSCLC</td>
<td>220 tumor regions, 64 non-tumor lung tissue samples, and 56 peripheral blood mononuclear cell samples from 72 patients</td>
<td>TCR heterogeneity varies between tumors and is influenced regionally by genomic heterogeneity; clonal nonsynonymous mutations are correlated with ubiquitous TCRs.</td>
</tr>
<tr>
<td>Ghorani et al., 2020</td>
<td>Study of the impact of tumor mutational burden upon T-cell differentiation</td>
<td>NSCLC</td>
<td>Multiregion flow cytometry from 31 patients from the TRACERx 100 cohort</td>
<td>TMB drives intratumoral T-cell differentiation and is associated with loss of progenitor-like and increased abundance of dysfunctional T-cell phenotypes; in NSCLC, this skewing is associated with poor clinical outcome.</td>
</tr>
<tr>
<td>AbdulJabbar et al., 2020</td>
<td>Spatial histologic analysis of lung cancer cohorts integrated with genomic and transcriptomic data</td>
<td>NSCLC</td>
<td>Multiregion histology from 100 TRACERx patients, and an independent cohort of 970 others</td>
<td>Tumors with multiple immune-cold regions, determined histologically, were associated with worse outcome and correlate with genomic features influencing the antitumor immune response.</td>
</tr>
<tr>
<td>Abbosh et al., 2017</td>
<td>Longitudinal ctDNA analysis throughout tumor evolution</td>
<td>NSCLC</td>
<td>TRACERx 100 cohort</td>
<td>Evolutionary dynamics of disease, including relapse and treatment resistance, could be profiled using ctDNA detection.</td>
</tr>
<tr>
<td>Chemi et al., 2019</td>
<td>Analysis of the relationship between CTCs derived from the pulmonary vein (PV-CTCs) and disease relapse</td>
<td>NSCLC</td>
<td>100 patients</td>
<td>The presence of PV-CTCs is associated with an increased risk of relapse. A case study suggested that clones found in PV-CTCs were more likely to be detected in metastasis than in the primary tumor.</td>
</tr>
<tr>
<td>Biswas et al., 2019</td>
<td>Investigation of impact of intratumor heterogeneity upon biomarker utility</td>
<td>NSCLC</td>
<td>RNA-seq from 156 regions of 48 tumors</td>
<td>Intratumor heterogeneity renders gene-expression biomarkers vulnerable to sampling bias, and a panel of clonally expressed genes can represent a robust biomarker in NSCLC.</td>
</tr>
</tbody>
</table>

**Heterogeneity in SCNAs and Clinical Outcome: The TRACERx Lung 100 Cohort**

A multiregion sequencing analysis from the first 100 prospectively recruited patients was published in 2017 and sought to explore the clinical implications of intratumor heterogeneity (7). Survival analysis showed that heterogeneity of SCNAs (as defined by the proportion of these events that were present only in a subset of cancer cells), but not mutations, predicted for poor outcome. This heterogeneity, detected through multiregion tumor sampling, may reflect the dynamic nature of CIN, suggesting the clinical importance of ongoing CIN in tumor evolution. Furthermore, an analysis of the intratumor heterogeneity, and timing, of
driver events revealed that most canonical driver mutations in NSCLC, such as in TP53, KRAS, and EGFR, were clonal and occurred prior to a whole-genome duplication event. However, other events such as mutations in PIK3CA and chromatin modifying genes were more likely to be subclonal and to occur later in tumor evolution. Importantly, with the help of phased copy-number information leveraged from multiregion sequencing, parallel evolution (where somatic events affecting the same gene occurred independently within distinct subclones) was seen among different haplotypes in copy-number events affecting genomic segments containing driver genes. This analysis added explicit evolutionary context to earlier observations about the clinical impact of intratumor heterogeneity of driver events and CIN (Fig. 2).

**Subsequent Insights into Chromosomal Instability and Cancer Evolution**

Within the TRACERx Lung 100 cohort, 76% of tumors showed evidence of a whole-genome duplication (WGD) event, which is consistent with observations in other cancer types (22). Although WGD has been linked to poor prognosis and increased subclonal diversity, a biological rationale for the frequency with which this event is seen in cancer had not been defined. One hypothesis is that WGD might mitigate the negative impact of consecutive deleterious mutations that accumulate within a cancer cell lineage. Therefore, within the TRACERx and TCGA NSCLC cohorts, this was explored in greater detail (23). In cancers with extensive loss of heterozygosity (LOH) across the genome, where the negative impact of deleterious mutations would in theory be more marked, WGD was enriched, and in simulations, WGD appeared to be selected preferentially when deleterious alterations were acquired at a high rate. Negative selection against deleterious mutations in essential genes acquired prior to WGD within genomic segments of LOH was also demonstrated using dN/dS, a tool to measure selection adapted from evolutionary biology (24, 25). These data suggest that WGD may be a selected event secondary to a “ratchet”-like phenomenon, mitigating the impact of the progressive accumulation of deleterious mutations in genes essential for lung cancer cell survival in regions of haploid LOH, brought about by excessive early CIN.

Although the finding in TRACERx that heterogeneity of SCNAs was associated with poor outcome suggests a role for CIN in fueling subclonal selection, this was not explicitly demonstrated. In a subsequent pan-cancer analysis of multiregion data, enabling phasing of SCNAs from 394 tumors, we explored this in greater detail, revealing extensive heterogeneity of SCNAs in primary and metastatic tumors, reflective of ongoing CIN during tumor evolution (26). Importantly, in the majority of tumors, and in particular in those with evidence of a prior WGD event, the evolution of SCNAs was consistent with positive and negative selection of these events. More than one third of tumors showed evidence of one or more parallel evolutionary events, in which the same genes were affected by independent SCNAs within different subclones. Frequent focal subclonal SCNAs that were often subject to parallel evolution were revealed, including gains at 5p15.33 and 8q24.1, containing TERT and MYC, respectively. In addition, in cases with both multiregion sampling of the primary and metastasis, subclonal copy-number events in the primary were found in all metastases, and in a combined analysis of paired and unpaired primary and metastatic samples, certain SCNAs that were frequently subclonal in primary tumors, such as those affecting CCND1 and MYC, were found to be significantly enriched within metastases in certain cancer types. This work builds upon observations from the TRACERx Lung 100 cohort and supports the importance of ongoing CIN in tumor evolution.

**Clinical Correlates of Cancer Evolutionary Subtypes**

In TRACERx Renal, the evolutionary dynamics of ccRCC were shown to precede the development of a tumor mass. A multiregion whole-genome sequencing analysis of 33 patients with ccRCC enabled inference of the timing of landmark events in the development of this tumor (27). 3p loss, a near-ubiquitous event in this disease, was shown in some cases to precede the growth of a clinically detectable tumor by 30 to 50 years, revealing this event as an attractive potential therapeutic target to reduce the incidence of this tumor.

The first interim analysis of the TRACERx Renal cohort defined seven evolutionary subtypes of ccRCC (3) that follow distinct trajectories following 3p loss and VHL inactivation, suggesting that in many cases cancer evolution is deterministic. Among these, there were tumors that were characterized by multiple clonal driver mutations and had evidence of early fixation of high levels of chromosomal complexity and limited ongoing evolution; BAP1-driven tumors that had similarly high levels of chromosomal complexity but with BAP1 as the sole additional driver; clonal expansion of PBRM1-mutant populations that led to frequent fixation of additional driver events, resulting in extensive intratumor heterogeneity, subclonal chromosomal complexity, and evidence of frequent parallel evolution events; VHL monodriver tumors that showed a paucity of additional drivers and minimal chromosomal complexity, and a subtype that in spite of 3p loss remained VHL–wild-type with cryptic early chromosomal complexity.

Defining the clinical behaviors of tumors with different evolutionary trajectories remains a key scientific question across tumor types. Although subclonal selection may be a feature of some cancers, others may acquire driver events at an early stage in tumor evolution and show little or no evidence of subclonal selection (28–30). Understanding the extent of subclonal selection from a biopsy at a single time point is challenging. Moreover, heterogeneity in the sampling and tissue processing (single vs. multiregion, FFPE vs. fresh-frozen samples), sequencing and analytic methods, and a lack of detailed clinical annotation can confound attempts to draw conclusions across studies about the relationships between evolutionary subtypes and clinical outcome. To address this problem, a uniform protocol accompanied by robust clinicopathologic information is required.

**Cancer Evolution and Metastasis**

TRACERx Renal explored the evolutionary dynamics of metastasis, in particular attempting to establish the extent to which subclonal selection during branched tumor evolution drives this process. In ccRCC, lymphatic and hematogenous
Figure 2. Tumor evolution in NSCLC. Evolutionary processes in NSCLC are outlined. Top, subclonal dynamics over time can be represented by a fish plot; however, a single sample in time provides only a snapshot. From this snapshot, tumor phylogeny can be inferred. Bottom, evolutionary processes generating immune and genomic heterogeneity are described as part of a “tree.” Events that occur in the “trunk” are clonal, i.e., they occur within every cell in the tumor. Through tumor evolution, subclones can emerge through selection; events that occur in these subclones are known as “branch” events.

Reconstruction of tumor phylogeny from a snapshot in time. Phylogenetic trees can be constructed from tumor samples through clustering of clonal and subclonal mutations that are subsequently ordered as a tree. Construction of phylogenetic trees is challenging and can be complicated by copy-number alterations and whole-genome doubling. Tumor evolution is a dynamic process: although much information can be gained from a single sample, multiregion and longitudinal sampling provides a more accurate picture of the evolutionary trajectory of the tumor.
dissemination can cause spread to the lungs, liver, adrenal glands, brain, and bones, as well as a local intravascular growth called a tumor thrombus; of the 98 patients included in this analysis, tumor thrombus was sampled in 25, lymph nodes sampled in 24, lesions of the adrenal glands in 20, bone in 17, lung in 13, and liver in 3 (31). Previous studies analyzed cohorts of matched and primary metastasis, whereas TRACERx Renal prospectively explored the evolutionary basis of metastasis (32). This has enabled the detailed comparison of primary tumor clones that go on to subsequently seed metastasis and those that seemingly fail to metastasize, revealing a selection for clones harboring extensive chromosomal complexity, as measured by Weighted Genome Integrity Index (wGII) scores (33, 34), in metastatic seeding.

In this analysis, driver mutations were not enriched or depleted within metastases (32). However, copy-number loss of 9p, containing the tumor suppressor CDKN2A, as well as 14q, was significantly overrepresented within metastatic lesions while existing mostly as a subclonal event within the corresponding primary tumors. 9p loss was associated with an aggressive disease phenotype and reduced progression-free and overall survival. Of descriptive note, the study also revealed a small number of cases with pancreatic metastases, with lower chromosomal complexity when compared with other metastatic sites, which were associated with latent metastases at periods of 15 years after primary diagnosis. This highlights not only the role that evolutionary subtypes play in determining clinical outcome, but also that subtypes may favor different metastatic patterns.

These observations were supported by sampling of multiple metastases in the PEACE study; in one patient who experienced rapid disease progression, a clone showing 9p and 14q loss seeded all 13 metastatic sites, while in another patient showing attenuated progression, indolent metastases to the pancreas lacked 9p and 14q loss, whereas subsequent multisite distant metastases had acquired these critical subclonal SCNAs.

**TUMOR EVOLUTION AND THE IMMUNE MICROENVIRONMENT**

**Tumor Evolution and Immune Escape**

The immune microenvironment plays a key role in shaping an evolving tumor, with positive selection favoring cancer cells that acquire immune evasion mechanisms (Fig. 2). These include mutations and LOH at HLA loci, disruption of antigen presentation machinery, and loss of neoantigens, which can be caused by DNA copy-number loss, transcript repression, and epigenetic silencing. Clinical and experimental data suggest that functional antitumor immunity exerts a strong selection pressure on evolving tumor subclones (35–37).

HLA class I genes (HLA-A, HLA-B, and HLA-C), located on the short arm of chromosome 6, encode cell-surface proteins that present peptides to CD8+ cytotoxic T cells as part of the class I major histocompatibility complex (MHC). These genes are highly polymorphic, thus creating diversity in immune responses at the population level. Loss of HLA class I genes reduces the ability of tumor cells to present antigens, providing a mechanism to escape immune surveillance (38). β2 microglobulin is also crucial for MHC class I binding and is located on the long arm of chromosome 15. The tool LOH in human leukocyte antigen (LOHHLA) was developed (39) in TRACERx to help quantify allele-specific HLA copy number from bulk sequencing data, which can be challenging given the polymorphic nature of the HLA locus. In TRACERx Lung, HLA LOH occurred in 40% of tumors, of which in 65% it was observed as a subclonal (branching) event. Due to the monoallelic loss of an HLA gene, there is a reduction (but not abrogation) of peptide presentation to tumor-infiltrating cytotoxic T cells; no tumor harbored loss of all six HLA class I alleles.

HLA LOH occurred as a focal loss (defined in this case as a non-arm event) with a higher than expected frequency, suggesting that this event is subject to positive selection in tumor evolution. In four cases, loss of the same HLA allele occurred as separate branch events in the same tumor, indicating parallel evolution that converges on HLA loss. HLA LOH was associated with the expansion of potentially antigenic mutations; in lung adenocarcinoma, subclonal nonsynonymous mutations were increased in tumors where HLA LOH had occurred. In tumor regions without HLA LOH, subclonal mutation burden was higher if HLA LOH had occurred in other regions of the same tumor, suggesting that increased mutational burden may drive selection pressure for HLA LOH. Furthermore, in lung adenocarcinomas, subclonal mutations were enriched for APOBEC mutational signatures (COSMIC single base substitution signatures 2 and 13) in tumors with HLA LOH. Furthermore, a trend toward enrichment of HLA LOH in brain metastasis was observed, suggesting that this event may be permissive for metastasis and/or associated with later-stage disease. In the study, tumors with clonal HLA LOH were associated with high immune infiltration and significantly elevated levels of PD-L1 staining by IHC, indicating that HLA LOH is a prevalent mechanism of immune evasion in tumors where immune activity and immune pressures are high (39).

The TRACERx Lung study has provided a platform to explore the role of neoantigen evolution in immune escape. In 2016, we found in NSCLC that tumors with a high clonal neoantigen burden were associated with increased overall survival (40). Building on previous work, clonal neoantigen burden was associated with high tumor immune infiltrate and increased PD-L1 expression. Moreover, clonal neoantigen burden was associated with clinical response to the checkpoint inhibitor pembrolizumab; however, this association was not significant if the neoantigen burden was subclonal. In some cases, subclonal neoantigens induced by cytotoxic chemotherapy were enriched in patients where response to checkpoint inhibition was poor. The concept that clonal neoantigens may be particularly immunogenic has since been supported by preclinical models (41) as well as data from patients with metastatic melanoma (42) and has led to the development of clonal neoantigen-targeting adoptive T-cell therapies (ClinicalTrials.gov Identifier: NCT04032847).

The intra- and intertumor heterogeneity of immune infiltration was characterized by integrating RNA-seq data with tumor-infiltrating lymphocyte (TIL) estimates in 258 regions from 88 TRACERx lung tumors (43). In this cohort, 28% of tumors had uniformly high levels of immune infiltration.
and 43% had uniformly low levels of immune infiltration, with disparate levels of infiltration in the remaining 28%. In lung adenocarcinoma, low immune infiltration was associated with increased subclonal diversity. This association between immune evasion and subclonal complexity is suggestive of an evolving tumor that has escaped immune surveillance.

To further explore the interplay between neoantigens and the immune response, we focused on neoantigen-directed mechanisms of immune escape, a term that encompasses both neoantigen depletion and disruption to antigen presentation. Potential mechanisms of neoantigen depletion that were analyzed included DNA copy-number loss, neoantigen transcript repression, epigenetic silencing of neoantigen sequences, and T cell-mediated elimination of neoantigen-expressing clones. In 43 of 88 TRACERx lung tumors, there was evidence of clonal neoantigen copy-number loss, and this was more likely to occur in regions of low immune infiltration. In highly immune-infiltrated tumors where HLA alleles were intact, there was a depletion in expressed predicted neoantigenic transcripts (relative to non-neoantigenic transcripts), thus proposing neoantigen transcript repression as an alternative mechanism of immune escape. Irrespective of immune infiltration, in consistently expressed genes essential for lung cancer survival, there was a reduction in neoantigens. Furthermore, there was a decrease in immunoediting (defined here as a reduction in the ratio of observed-to-expected neoantigens; ref. 44) from clonal to subclonal neoantigens in immune-cold tumors. This suggests that these tumors once contained an active immune microenvironment that became cold. Poorly expressed genes predicted to contain neoantigens exhibited an increase in promoter hypermethylation (compared with expressed genes containing neoantigens and the same genes without neoantigenic mutations), suggesting that promoter hypermethylation is an additional mechanism in neoantigen silencing.

Disruption to antigen presentation encompassing HLA LOH and mutations affecting the MHC complex, HLA enhanceosome, or peptide generation were found in 56% of lung adenocarcinomas and 78% of squamous cell carcinomas. By combining immune infiltration with these immune evasion mechanisms, the capacity for a tumor to evade the immune system was quantified. Tumors with low immune evasion capacity and uniformly high immune infiltrate with no evidence of DNA immunoediting and no disruption to antigen presentation were associated with improved disease-free survival. In a multivariate model, both low immune evasion capacity and clonal neoantigen burden were predictors of disease-free survival; however, subclonal neoantigen burden was not. The diversity of immune-escape mechanisms likely reflects the strength of selection pressures exerted by the immune system and highlights an important consideration for immunotherapeutic design.

**Intratumor Heterogeneity and the T-cell Landscape**

In the complex evolutionary arms race of tumor immunity, the immune system drives selection of evolving tumor subclones, and, reciprocally, the tumor mutational landscape may drive T-cell activation, clonal expansion, and differentiation. The expansion and spatial diversity of intratumoral T-cell receptors (TCR) were analyzed (45) by sequencing the α-chain and β-chain TCR repertoires from 220 tumor regions and 119 matched (nontumor lung or blood) samples from 72 patients in the TRACERx Lung 100 cohort. The most expanded intratumoral TCRs represented a higher proportion of the total TCR repertoire compared with those expanded in nontumor lung samples. Furthermore, the number of expanded TCRs correlated with nonsynonymous mutations in tumor regions, consistent with ongoing neoantigen-specific T-cell responses.

There were marked differences in TCR intratumor heterogeneity between tumors, whereby some tumors had a diverse regional TCR expression pattern and for others this pattern was more homogeneous. TCRs were therefore defined as ubiquitous or regional depending on their spatial distribution throughout multiregion tumor specimens and, interestingly, this was correlated with the number of clonal and subclonal nonsynonymous mutations, respectively, implying that ubiquitous TCRs may recognize clonal neoantigens. Interestingly, expanded TCR CDR3 sequences formed highly related clusters and a particular TCR clonal expansion was linked to amino acid composition, with regional TCR clustering far less prevalent than ubiquitous TCR clustering. There were also more unique DNA sequences encoding each ubiquitous expanded TCR when compared with regional TCR or random sampling, indicative of antigen-driven convergent recombination. The positive association between ubiquitous TCRs and a CD8+ Th1 transcriptional phenotype suggests that a cytotoxic T-cell response is linked with the presence of ubiquitous intratumoral TCRs. Clonal neoantigen-reactive T cells and cells bearing an exhausted phenotype also tended to harbor expanded TCR sequences.

Furthermore, expanded, ubiquitous TCR sequences in the tumor were preferentially found in peripheral T-cell clones, and TCRs in the blood displayed both contraction and expansion throughout the clinical course of disease, suggesting that longitudinal tracking of specific subsets of TCRs in the blood may in the future provide a method of monitoring dynamic intratumoral immune surveillance.

**Tumor Mutational Landscape and T-cell Differentiation**

T cells respond to antigen stimulation by activation, proliferation, and differentiation, manifesting in redistribution of progenitor–progeny subsets (46). The relationship between tumor mutational burden (TMB) and T-cell differentiation was explored in the TRACERx Lung 100 cohort (47). TMB is known to predict response to immune-checkpoint blockade, yet tumor-specific T cells often appear dysfunctional, suggesting neoantigens could both drive immunogenicity but also fuel chronic TCR stimulation that culminates in a shift in differentiation and accumulation of dysfunctional phenotypes. High dimensional flow cytometry analysis of multiregion TILs revealed highly diverse populations of CD4 and CD8 progenitor-like and dysfunctional TILs. Ghorani and Reading then integrated these data with paired WES data from the TRACERx Lung 100 cohort, unveiling that TMB correlated with T-cell subsets exhibiting evidence of antigen engagement, including dysfunctional T cells with high PD-1 expression, PD-1–expressing subsets with features of
terminally differentiated effector memory cells (CD57, granzyme B, and eomesodermin expression), and regulatory T cells with high coinhibitory receptor expression. Conversely, T-cell clusters lacking PD-1 with bystander or progenitor-like characteristics correlated negatively with TMB, implying a neoantigen-driven intratumoral differentiation process. Consistent with this, progenitor and dysfunctional T-cell subsets shared TCR sequences and dysfunctional T cells correlating with TMB were phenotypically and transcriptomically similar to neoantigen multimer reactive T cells. The progenitor and dysfunctional subsets were additionally shown to express canonical stem (TCF7) and exhaustion (TOX) transcription factors respectively by flow cytometry and single-cell RNA-seq. This putative program of neoantigen-driven differentiation skewing was also characterized in bulk tumor RNA-seq by identifying a gene signature that mapped loss of transcription factors TCF7 and LEF1 (TL-DS). The TL-DS signature correlated with TMB in tumors and poor clinical outcomes in TRACERx patients with lung adenocarcinoma, but not squamous cell carcinoma. When applied to TCGA data sets in a multivariable model, the signature was associated with adverse survival outcomes in six different tumor types, including NSCLC, raising the possibility that in the absence of immunotherapy, chronic neoantigen stimulation could fuel loss of progenitor T cells and gain of dysfunctional subsets, ultimately precipitating fatal immune failure.

Importantly, the burden of clonal, but not subclonal, mutations correlated with the loss of progenitor and gain of dysfunctional subsets, corroborating the observation that clonal neoantigens preferentially elicit immunoreactivity and suggesting that clonal neoepitopes may generate PD-1+ cells that could serve as a substrate for immunotherapies targeting the PD-1 axis (40, 48). These data highlight that neoantigens may represent a double-edged sword, triggering protective T cells and gain of dysfunctional subsets, ultimately precipitating fatal immune failure.

...
early detection, prognostication, and therapeutic intervention in NSCLC.

Clinical Utility of Circulating Biomarkers

The development of minimally invasive circulating biomarkers such as ctDNA and CTCs aims to improve approaches to early detection of relapse through minimal residual disease (MRD) monitoring and predictors of clinical outcome. Identification of patients with NSCLC at risk of relapse post-surgery may help identify those who would benefit from adjuvant therapies and potentially avoid unnecessary treatments and associated toxicities in those patients with a lower risk of disease recurrence. The prospective recruitment and longitudinal design inherent in TRACERx provided a platform to address this question. In creating bespoke multiplex PCR panels to detect clonal and subclonal single-nucleotide variants (SNV) in plasma, NSCLC relapse was profiled through ctDNA detection in the TRACERx Lung 100 cohort (52). Via the profiling of preoperative plasma samples, the determinants of ctDNA detection were identified. Necrotic tumors were more likely to release ctDNA, with non-adenocarcinoma subtype, lymphovascular invasion, and high Ki-67 proliferation index being independent predictors of ctDNA detection. Clonal SNVs were detected in all ctDNA-positive patients, and the plasma variant allele frequency correlated with tumor size. We utilized a panel of between 25 and 30 variants per patient to detect and characterize recurrent disease. Mutations in ctDNA were detected in 13 of 14 patients before or at confirmed radiologic relapse, with a median lead time of 70 days prior to imaging-confirmed recurrence. By mapping ctDNA-detected variants at recurrence back to the primary tumor WES data, we were able to identify the ancestral subclones leading to relapse. Finally, metastatic subclones were retrospectively tracked through ctDNA analysis in a patient who was co-recruited to the PEACE study, using a panel of 20 clonal SNVs with 4 to 15 subclonal SNVs per metastatic subclone. Using this 103-SNV panel, ctDNA was detected 151 days after surgery (315 days prior to clinical relapse), with 18 of 20 shared subclonal SNVs (present in three of nine metastatic clonal clusters) and two private subclonal SNVs detectable in ctDNA following clinical relapse on day 466. This bespoke approach is utilized as part of the Signatera assay (53).

A similar phylogenetic approach to MRD detection is now being implemented in the clinical trial setting. MERMAID-1 is a phase III trial (ClinicalTrials.gov Identifier: NCT04385368) in collaboration with ArcherDx and AstraZeneca. The trial aims to assess the efficacy of the PD-L1 inhibitor durvalumab in combination with chemotherapy in patients with resected stage II to III NSCLC. Following resection of the primary tumor, MRD detection through bespoke ctDNA panels will be used in the assessment of disease-free and overall survival. Furthermore, using the same approach, Powles and colleagues have demonstrated that patients with muscle-invasive urothelial cancer who are ctDNA-positive post cystectomy may benefit from atezolizumab (54). This analysis was done as part of the IMvigor010 trial, the Signatera assay utilizing a bespoke panel of 16 somatic mutations (55, 56).

Using the CTC capture platform CellSearch, CTCs taken from the pulmonary vein (PV-CTC) of patients from the TRACERx Lung 100 cohort were enumerated at surgery (57). CTCs were detected at a threshold of at least 1 PV-CTC in 7.5 mL of blood in 48% of patients. A high PV-CTC count (> 7 PV-CTCs per 7.5 mL blood) was an independent predictor for lung cancer relapse. Each doubling of PV-CTC count was associated with lung cancer–specific relapse, indicating a continuous relationship between PV-CTCs and clinical outcome. WES data from three primary tumor regions were compared with single-cell sequencing data from six single PV-CTCs [three of which were found to be EPCAM- and cytokeratin-expressing circulating epithelial cells (CEC)] and the metastatic relapse biopsy (10 months after surgery) in a single patient. They demonstrated a higher number of shared mutations between the PV-CTC and the metastatic sample compared with the PV-CTC and primary sample, with phylogenetic reconstruction demonstrating that the PV-CTC and metastatic samples came from the same branch. PV-CTCs taken at the time of surgery shared a common progenitor with the metastasis that was detected 10 months later, indicating that early dissemination of PV-CTCs may result in metastatic seeding.

Prognostic Signatures and Therapeutic Stratification

Prognostic gene-expression signatures have been utilized to predict clinical outcome in patients with a variety of cancer types; however, the interpretability of these biomarkers may be confounded by sampling bias driven by intratumor heterogeneity (58–60). By combining multiregion RNA-seq from 48 patients in the TRACERx Lung 100 cohort with TCGA and the Uppsala NSCLC data sets (61–63), the impact of RNA intratumor heterogeneity on gene-expression signatures was explored in lung adenocarcinoma (64). We demonstrated regional discordance of prognostic scores between biopsies from the same tumor, thus highlighting the vulnerability of such scoring approaches to sampling bias. By stratifying genes into quadrants based on intra- and intertumor heterogeneity and selecting a subset of genes that were expressed with low intratumor heterogeneity but high intertumor heterogeneity, we generated a 23-gene prognostic signature termed ORACLE (Outcome Risk Associated Clonal Lung Expression biomarker).

As a clonal expression biomarker, the ORACLE score was associated with outcome in the Uppsala cohort, in a multivariate model that adjusted for TNM stage, adjuvant treatment status, age, performance status, smoking history, sex, and Ki-67 staining. The genes that were stratified in the low intratumor heterogeneity but high intertumor heterogeneity quadrant (Q4) were enriched in proliferation pathways, such as nucleosome assembly and mitotic prometaphase. These genes retain transcriptional stability despite ongoing CIN and are subject to clonal copy-number gains early in tumor evolution.

ORACLE was able to classify patients with stage I lung adenocarcinoma into low-risk and high-risk groups. RNA intratumor heterogeneity also correlated strongly with subclonal copy-number diversity, indicating that ongoing chromosomal instability is a strong driver of transcriptomic heterogeneity. Finally, by comparing genes identified in each quadrant using this selection approach to genes with pan-cancer prognostic scores from the Prediction of Clinical Outcomes from Genomic Profiles (PRECOG) data set, we found
genes in Q4 were significantly enriched for prognostic genes in 19 of 39 malignancies. This approach to gene-expression profiling may improve refinement of prognostication including response to therapy.

From the Bench to Clinical Trials

Although TRACERx aims to establish the relationship between intratumor heterogeneity and clinical outcome, we are yet to understand the relevance of clonality in therapeutically targeting actionable driver events. DARWIN II (Deciphering Antitumor Response and Resistance With INtratumor Heterogeneity; ClinicalTrials.gov Identifier: NCT02314481) is a multiarm nonrandomized phase II trial that aims to investigate the role of predicted neoantigens and intratumor heterogeneity on response to anti–PD-L1 therapy, as well as outcomes in patients treated with targeted therapies against clonal or subclonal actionable alterations (Fig. 3).

Patients with relapsed disease in the TRACERx Lung study or those with available multiregion sequencing data from a primary NSCLC are eligible. Patients are stratified to specific therapies depending on their targetable alteration: BRAFV600 (vemurafenib), ALK/RET translocation (alectinib), and HER2 amplification (trastuzumab). All other patients are consented to the PD-L1 inhibitor atezolizumab. Sampling before and after treatment, including at autopsy, will facilitate the study of resistance to immune and targeted therapies. Longitudinal monitoring of predictive biomarkers of response and resistance may also guide future patient stratification and combination therapies. Recruitment to this trial is ongoing.

Despite recent improvements in NSCLC clinical outcomes with PD-1/PD-L1 checkpoint inhibitors, 45% to 50% of patients with metastatic disease do not achieve an optimal response with standard-of-care first-line chemotherapy plus
PD-1/PD-L1 combinations, and almost 70% patients experience disease progression or die within 12 months of treatment (65, 66). In this setting, second-line therapy is often associated with minimal clinical benefit. With evidence that clonal neoantigens drive T-cell reactivity, the first-in-human open-label, multicenter phase I/IIa trial to characterize the safety and clinical activity of autologous clonal neoantigen targeting T cells (cNeT) was announced in 2019. Developed by Achilles Therapeutics, the CHIRON and THETIS trials (ClinicalTrials.gov Identifier: NCT03997474 and NCT04032847) investigate the safety and clinical activity of a therapeutic product derived from autologous TILs primed to recognize patient-specific tumor clonal neoantigen epitopes in NSCLC and melanoma, respectively (Fig. 3).

**FUTURE PLANS TO ADDRESS UNANSWERED QUESTIONS**

Our understanding of cancer evolution has improved significantly over the last decade. Here, we outline outstanding questions in the field and areas of unmet clinical need that could be addressed in future studies.

**The Role of Chromosomal Instability in the Disease Course**

SCNAs provide a heterogeneous substrate for ongoing tumor evolution, and TRACERx has highlighted the relationship between heterogeneity driven by SCNAs and adverse clinical outcome. Future studies should aim to highlight specific copy-number events that may be prerequisites for certain evolutionary trajectories. Moreover, there is an ongoing effort to characterize the processes that generate SCNAs and drive CIN, understand cancer cell tolerance of them, and define their relationship with driver events. Developing the necessary data sets and tools required to measure the rate of CIN might help shed light on its exact role in the natural history of a tumor.

**Nongenetic Variation and Tumor Evolution**

Cancers may exploit transcriptomic and epigenomic variation during their evolution (67, 68). This can provide a tumor with alternate mechanisms of immune evasion as well as fueling adaptive evolution in response to selective pressures, such as targeted therapies (69). Understanding the evolutionary context of this nongenetic variation, and assessing epigenetic and transcriptomic heterogeneity, in both treated and untreated cancers must be addressed and will require a repertoire of studies using a variety of bulk and single-cell sequencing technologies.

**The Antitumor Immune Response as a Therapeutic Target**

Understanding which neoantigens stimulate an effective antitumor response is critical. Neoantigens that are clonal, highly expressed, and dissimilar to “self” with appropriate mitigation from HLA loss or copy-number loss tend to stimulate a more effective antitumor immune response (70). Moreover, neoantigens that are formed via frameshift mutations are more likely to be recognized by multiple HLA alleles. Understanding and leveraging high-quality neoantigens for treatment stratification promises to enhance immunotherapeutic approaches, including adoptive T-cell therapies and cancer vaccines (71). Evolving selection pressures exerted by the immune response that affect the transitions from preinvasive to invasive and early stage to disseminated disease may also have profound clinical relevance in the context of early detection and cancer progression.

**Understanding Clinical Correlates of Cancer Evolutionary Trajectories**

A key objective of TRACERx has been to understand the relationship between intratumor heterogeneity and clinical outcome. In TRACERx Renal, tumor subtypes with distinct evolutionary trajectories were described and associated with distinct clinical outcomes (3, 32). Expanding such analyses to large cohorts of cancers of other types and integrating data from preinvasive lesions, primary tumors, recurrences, and distant metastases may reveal novel insights into the evolutionary dynamics of cancer, support optimal disease classification, and facilitate new screening and treatment strategies. Longitudinal sampling, with liquid biopsies to study ctDNA as well as tissue biopsies at relapse or autopsy, may help us to understand the impact of treatment on cancers, as well as the development and evolution of therapy-resistant subclones. It is hoped that this might improve our ability to better classify early-stage disease and to predict the likely clinical course of the disease.

**Functional and Translational Validation of Observations**

Ascribing functional weight to observations made in bulk sequencing studies is important. Advanced technologies such as patient-derived organoid systems and genetically engineered mouse models will help validate these findings, including further elucidation of the importance of order in somatic events, quantification of nongenetic heterogeneity, and refined modeling of the immune interface in tumor evolution (72, 73). In addition, complex cellular imaging can provide additional spatial resolution of intratumor heterogeneity and the tumor microenvironment (74).

**Longitudinal, Prospective, and Comprehensive Clinical Studies to Augment Cancer Research**

TRACERx is the product of a national, comprehensive, multidisciplinary research infrastructure that links a diverse range of clinical specialties and scientific disciplines, made possible through extensive collaboration, uniform clinical care, and adherence to strict standard-of-care guidelines provided by a national health service. Studies of cancer evolution should endeavor to incorporate comprehensive clinical information with detailed genomic and immunologic analyses (Table 2). Establishing a sufficiently powered prospectively recruited cohort with robust sampling methods and detailed clinical annotation is important. Furthermore, considering the need for longitudinal sampling from early- to late-stage disease, with the incorporation of research autopsies, can help establish a data set encompassing the disease course in its entirety.

Maintaining a comprehensive and reproducible bioinformatic pipeline that is adaptable to different tissue types...
and sampling time points is crucial, with novel software and technological developments making portable analytic frameworks a reality. Finally, multiple samples within individual patients, whether they be from different regions of the same tumor, relapse, metastasis, from liquid biopsies or taken at autopsy, are likely to provide granularity to the inferences we can make regarding tumor evolution.

Outside of NSCLC and ccRCC, numerous studies have used multiregion or longitudinal sampling to address questions pertaining to intratumor heterogeneity and tumor evolution (75–81). Large, prospective, longitudinal clinical cohorts, however, are rare. The Glioma Longitudinal Analysis Consortium (GLASS) is a multi-institutional initiative that aims to profile longitudinal molecular trajectories of gliomas over several time points through treatment (82). Studies such as this are essential and are powered to address key, treatment-focused questions in the field. Developing large such as this are essential and are powered to address key, treatment-focused questions in the field. Developing large, prospective, longitudinal clinical frameworks a reality. Finally, multiple samples within individual patients, whether they be from different regions of the same tumor, relapse, metastasis, from liquid biopsies or taken at autopsy, are likely to provide granularity to the inferences we can make regarding tumor evolution.

TRACERx has helped to shed light on the evolutionary forces at work within lung and renal tumors. Longitudinal and multiregion sampling has helped to explore this at the molecular level, and some key observations are being tested in the clinic. However, many questions remain unanswered, and further studies will be critical to understanding the clinical relevance of evolutionary processes within cancer and the way in which we can leverage these to benefit patients.

Authors’ Disclosures

J.L. Reading reports personal fees from Achilles Therapeutics Ltd outside the submitted work; in addition, J.L. Reading has a patent for Modulation of T-cell cytotoxicity and related therapy pending. K. Litchfield reports personal fees from Roche Tissue Diagnostics and Monopteros and grants from CRUK TDL/Ono/LifeArc alliance outside the submitted work; in addition, K. Litchfield has a patent for indel burden and checkpoint inhibitor response pending. S. Tarajic reports a patent for Clear Cell Renal Cell Carcinoma Biomarkers P113326GB issued. N. McGranahan reports personal fees from Achilles Therapeutics outside the submitted work; in addition, N. McGranahan has a patent for PCT/GB2018/052004, PCT/EP2016/059401, and PCT/GB2020/052211 pending. M. Jamal-Hanjani reports non-financial support from Achilles Therapeutics and personal fees from Achilles Therapeutics outside the submitted work; in addition, M. Jamal-Hanjani has a patent for methods for lung cancer detection issued. C. Swanton is Royal Society Napier Research Professor (RP150154). His work is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001169), the Wellcome Trust (FC001169), and the Wellcome Trust (FC001169). C. Swanton is funded by Cancer Research UK (TRACERx, PEACE and CRUK Cancer Immunotherapy Catalyst Network), Cancer Research UK Lung Cancer Centre of Excellence, the Rosetrees Trust, Butterfield and Stoneygate Trusts, NovoNordisk Foundation (ID16584), Royal Society Research Professor’s Enhancement Award (RP/EA/180007), the NIHBI Medical Research Centre at University College London Hospitals, the CRUK-UCL Centre, Experimental Cancer Medicine Centre, and the Breast Cancer Research Foundation, USA (BCRF). His research is supported by a Stand Up To Cancer–LUNGevity–American Lung Association Lung Cancer Interception Dream Team Translational Research Grant (SU2C–AACR–DT23–17). Stand Up To Cancer (SU2C) is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the Scientific Partner of SU2C. C. Swanton also receives funding from the European Research Council (ERC) under the European Union’s Seventh Framework Programme (FP7/2007–2013) Consolidator Grant (FP7–THESEUS–617844), European Commission ITN (FP7–PloidyNet 607722), an ERC Advanced Grant (PROTEUS) from the European Research Council under the European Union’s Horizon 2020 research and innovation program (835297), and Chromavision from the European Union’s Horizon 2020 research and innovation program (665233). C. Swanton reports grants and personal fees from AstraZeneca (is a member of advisory board and Chief Investigator

### Table 2. A template for studies of tumor evolution

<table>
<thead>
<tr>
<th>Feature</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multicenter</td>
<td>Involvement of different centers helps with timely recruitment of a large cohort, enabling analysis of sufficient power for statistically meaningful observations.</td>
</tr>
<tr>
<td>Prospectively recruited</td>
<td>Prospective analyses are critical when making meaningful and robust inferences about clinical factors such as outcome.</td>
</tr>
<tr>
<td>Multiple samples</td>
<td>Analysis of different samples over time, such as through sampling at relapse or postmortem or ctDNA analysis, and space, using multiregion sequencing, can help to elucidate the evolutionary context of cellular events.</td>
</tr>
<tr>
<td>Uniform sampling method</td>
<td>It is critical that studies define a method for sampling, processing, and sequencing that is adhered to across sites to reduce technical inconsistencies.</td>
</tr>
<tr>
<td>Independent pathology review</td>
<td>Reduces impact of heterogeneous interpretation of histopathologic specimen.</td>
</tr>
<tr>
<td>Multiomics</td>
<td>Enables capturing of greater degree of tumor variation, including nongenetic events and microenvironmental alterations.</td>
</tr>
<tr>
<td>Detailed clinical annotation</td>
<td>Robust protocols for inputting clinical data reduce missing data. This is critical not only for prospectively defined variables of interest, but also for retrospective analysis, such as when examining the impact of clinicopathologic variables or environmental exposures.</td>
</tr>
<tr>
<td>Reproducible bioinformatic pipeline</td>
<td>Transparent, portable, and easily replicable analytic approaches improve efficiency and collaborative approaches to the study of tumor evolution.</td>
</tr>
</tbody>
</table>
for MeRmiD1 clinical trial), Bristol Myers Squibb, Pfizer, Roche-Ventana, Ono Pharmaceutical, grants from Archer Dx Inc., Boehringer Ingelheim, personal fees from Novartis, GSK, Celgene, MSD, Illumina, personal fees from Amgen, Genentech, Sarah Canon Research Institute, Medixci, and Bicycle Therapeutics, personal fees and stock options from GRAIL, stock options from Apogen Biotechnologies and Epic Biosciences, and has stock options and is co-founder of Achilles Therapeutics during the conduct of the study; grants and personal fees from AstraZeneca, Bristol Myers Squibb, Pfizer, Roche-Ventana, and Ono Pharmaceutical, personal fees from Novartis, personal fees from GSK, personal fees from Celgene, MSD, Illumina, Amgen, Genentech, Sarah Canon Research Institute, Medixci, Bicycle Therapeutics, grants from Archer Dx Inc, and Boehringer Ingelheim, other from GRAIL, Apogen Biotechnologies, Epic Biosciences, and personal fees and has stock options and is co-founder of Achilles Therapeutics outside the submitted work. No disclosures were reported by the other authors.

Acknowledgments

We would like to thank Tom Watkins, Alex Frankell, Rachel Rosenthal, Chris Abbo, Yinyin Yuan, and Kroopa Joshi for their thoughtful input and review of the manuscript. The fish plot was created using the fishplot R package from https://github.com/chrisamiller/fishplot/. The figures were created with the BioRender software at BioRender.com. The figures are unpublished original works, created by C. Bailey for the express purpose of publication in this Cancer Discovery review. C. Bailey holds an institutional BioRender license through the Francis Crick Institute.

Received October 29, 2020; revised December 21, 2020; accepted January 6, 2021; published first April 1, 2021.

REFERENCES

Tracking Cancer Evolution through the Disease Course


