ANNUAL REVIEWS

Annual Review of Cancer Biology

Tracing and Targeting the Origins of Childhood Cancer

Tim H.H. Coorens[1,2] and Sam Behjati[1,3,4]

[1]Wellcome Sanger Institute, Hinxton, United Kingdom; email: sb31@sanger.ac.uk

[2]Current affiliation: Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; email: tcoorens@broadinstitute.org

[3]Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom

[4]Department of Paediatrics, University of Cambridge, Cambridge, United Kingdom

Annu. Rev. Cancer Biol. 2022. 6:35–47

First published as a Review in Advance on December 21, 2021

The *Annual Review of Cancer Biology* is online at cancerbio.annualreviews.org

https://doi.org/10.1146/annurev-cancerbio-070620-091632

Keywords

pediatric cancer, genomics, transcriptomics, cell of origin, cancer evolution, phylogenetics

Abstract

Despite the success of treating childhood cancers with cytotoxic agents, novel therapeutic strategies are required to achieve the next leap in cure rates. A promising avenue may be to target the origin of childhood cancers. Here, we review recent advances in tracing the origins of pediatric tumors. Cancer-to-normal cell comparisons by single-cell mRNA sequencing reveal the fetal state of cancer cells, as well as their cell of origin. Recent phylogenetic analyses have uncovered large tissue-resident precursor clones to childhood cancers, which already possess key genomic alterations leading to tumor formation. Both the transcriptional fetalness and genomic status of the premalignant tissue bed provide further avenues for targeted therapy. Overall, these advances begin to describe the precise origins of pediatric tumors and pave the way for novel methods in detecting, treating, and perhaps even preventing childhood cancers.

1. INTRODUCTION

The evolution of the treatment of childhood cancer has been a roaring success. When the father of modern oncology, Sidney Farber, first trialed so-called antifolates in 1947, childhood leukemia was a universally fatal disease (Miller 2006). Now, seven decades later, the most common variant of childhood leukemia, standard-risk B cell acute lymphoblastic leukemia, is curable in more than 90% of children through modern treatment protocols that have evolved from arduous collaborative trials of multiagent treatments (Inaba & Mullighan 2020). This success is not confined to hematological malignancies. Across cancer diagnoses, 80% of children will be cured (Pizzo et al. 2016). However, this apparent triumph of modern pediatric oncology hides two important areas of stagnation and concern.

First, the prognosis of several cancer types remains stubbornly poor, despite intense basic biological and clinical research efforts. These include high-grade malignancies such as certain brain tumors, high-risk neuroblastomas, sarcomas, some renal tumors, and myeloid leukemias. Furthermore, children occasionally succumb to non-high-grade tumors that are resistant to cytotoxic treatment and are not amenable to surgical clearance. Examples of such low-grade tumors that may relentlessly relapse and progress include gliomas, fibromatoses, and immature teratomas.

Second, the limit of the tolerability of treatment intensification has probably been reached for most childhood cancer types. Survivors of childhood cancer often suffer from significant late effects. Data from the St. Jude Lifetime Cohort Study cohort suggest that late effects are universal among survivors of childhood cancer (Bhakta et al. 2017). Multimodal cytotoxic treatment does not necessarily translate into better outcomes, often because incremental survival benefits are undermined by immediate treatment-related toxicity, including death, or by mutilating long-term adverse effects (late effects), such as the arrest of neurocognitive development following craniospinal radiotherapy of the very young.

As successful as cytotoxic treatment may have been in getting us to the point of curing the majority of children diagnosed with cancer, the status quo is unacceptable. Furthermore, it would seem unlikely that further substantial progress will be made through trials of variations of therapeutic strategies that are more of the same. Novel strategies are required beyond cytotoxic treatment to achieve the next leap in cure rates. Such approaches may be derived from targeting the cancer genome or from mobilizing the antitumor immune response. A distinctive advantage of these approaches would be that they could be adapted from, and developed in parallel to, similar efforts in adults, as the underlying technical and biological principles are broadly the same. Another childhood cancer–specific approach could be to trace and target the origins of childhood cancer rooted in human development, which we discuss in this review.

1.1. Childhood Cancer is a Developmental Disease

The origin of childhood cancer lies in aberrant human development (Behjati et al. 2021, Filbin & Monje 2019). The evidence supporting this notion is extensive, albeit often imprecise. One major line of evidence is that childhood tumors resemble fetal human tissues morphologically. Max Wilms, for example, already noted in 1899 in his treatise on renal tumors the similarity of the childhood renal tumor, nephroblastoma (Wilms tumor), and embryonic lineages, which led him to conclude that these cancers were of fetal origin (Wilms 1899).

A second line of evidence is that the spectrum of childhood tumors is mostly unique and does not have adult correlates. Cancers of adults are generally of epithelial origin and increase in prevalence with age. Childhood tumors, by contrast, are overwhelmingly derived from nonepithelial lineages and show a predilection for specific, reasonably narrow postnatal age windows. When childhood and adult cancers are histologically similar, there are typically unique features that

distinguish these neoplasms. For example, childhood gliomas are often driven by hotspot mutations in type 3 histone genes, which rarely occur in adult gliomas (Mackay et al. 2017). Even for hematological malignancies, where cancer cells are morphologically and phenotypically very similar between adults and children, differences exist. In children, acute B cell lymphoblastic leukemia with childhood-specific genomic alterations predominates, whereas in adults myeloid malignancies and non-Hodgkin lymphoma are the most common hematological neoplasms (Downing & Shannon 2002).

A third line of evidence derives from explicit investigations into the origin of childhood cancer. However, until recently this has largely been based on work in model systems, as discussed below. There are, however, some notable direct observations in humans that indicate a developmental origin of childhood cancer. Most prominently, the work of Greaves (2018) led to the discovery of leukemogenic gene fusions in blood samples obtained at birth from children who developed overt leukemia later in life. The significance of this finding is twofold. First, it serves as a direct proof of cancer initiation before birth. Second, the discovery represents a measurable precursor of childhood cancer, which may in principle be amenable to detection by population-based screening.

1.2. Therapeutically Exploiting the Origin of Childhood Cancer

Given its origin in aberrant, stalled development, pediatric oncologists have long entertained and pursued the idea of differentiation treatment of childhood cancer. There are two examples of effective therapies that have entered clinical practice. One example is the retinoic acid derivative *cis*-retinoic acid, which is commonly used for the treatment of neuroblastoma in different treatment protocols. Neuroblastoma is a neural crest–derived cancer that typically arises from the developing adrenal medulla. It has been of particular interest for research into differentiation therapies, as certain subtypes of the disease, including a widely metastatic variant of infant neuroblastoma (4S), may spontaneously regress (van Noesel 2012). Moreover, resection specimens often exhibit gangliocytic differentiation following neoadjuvant treatment with cytotoxic chemotherapy, further testifying to the ability of neuroblastoma cells to mature. Differentiation of the neural crest into peripheral neural cells depends in part on retinoic acid signaling, providing the biological rationale behind retinoic acid treatment. The second example is all-*trans* retinoic acid, another retinoic acid derivative, which is a highly effective treatment of acute promyelocytic leukemia (APML), a rare leukemia of children and adults. APML is driven by rearrangements of the retinoic acid receptor, the dysregulation of which can be overcome through all-*trans* retinoic acid.

The developmental root of childhood cancer may be further exploited therapeutically by targeting fetus-specific antigens that are preserved in cancer cells and are (relatively) absent from postnatal tissues. The key exemplar in this regard is the disialoganglioside GD2, which is utilized by neuroblastoma cells and during sympathoadrenal development (Park & Cheung 2020). In postnatal tissues, the expression of GD2 is reasonably confined to the central and peripheral nervous systems (Kildisiute et al. 2021). Targeting GD2 through antibodies has become a routine treatment for high-risk neuroblastoma and is associated with few, generally manageable adverse effects (pain, bronchospasm). GD2 has furthermore been targeted through chimeric antigen receptor T cells in early-phase clinical trials (Straathof et al. 2020).

Finally, identifying the developmental root of childhood cancer may ultimately enable screening for, and prevention of, childhood cancer if neoplasms were formed via detectable precursors or were preventable in the first place through interventions during pregnancy. At first glance, the ideas of mass screening and prevention may appear naïve and unrealistic. Yet, screening programs for rare childhood diseases are already commonplace in the newborn screening that has been implemented in many countries. Furthermore, there is a precedent for preventing the formation of

a developmental disorder, namely spina bifida by folic acid supplementation during pregnancy. As preposterous as the concept of screening and prevention may therefore appear to be, it would not be outside the realm of possibility if the origins of childhood cancers were known, measurable, and targetable.

2. FINDING THE HUMAN ORIGIN OF CHILDHOOD CANCER

There is a vast body of research into modeling the developmental origins of childhood cancer in animals. For example, experiments in mice on medulloblastoma, the most common brain cancer of children, have revealed distinct developmental and anatomical origins of different subtypes of medulloblastoma (Gibson et al. 2010). Given that early embryogenesis and organogenesis are likely to be preserved evolutionarily, it would seem reasonable to suggest that these findings in mice may be relevant to human medulloblastoma, especially as murine and human tumors share morphological and transcriptional features. Nevertheless, beyond weaknesses of individual models and the generic criticism that mice are not humans, a principal shortcoming of modeling the developmental origin of childhood cancer is that it is largely attempting to recapitulate the unknown. For most childhood cancers, we do not know their precise origin in human development and subsequent pathogenesis. Animal models therefore may, or may not, recapitulate human developmental oncogenesis faithfully, irrespective of whether the fully formed mouse tumors resemble their human counterparts in the end. Therefore, there remains a fundamental need to directly determine the precise origin of childhood cancer in humans. Recent technical advances in single-cell mRNA sequencing and phylogenetic lineage tracing through somatic mutations have enabled such endeavors, which we discuss next.

2.1. Matching Cancer and Normal Cells by Single-Cell mRNA Sequencing

Cancer cells may retain transcriptional features of the cells from which they derive. Therefore, it is possible to gain insights into the origin of tumor cells by identifying the cell type that cancers most closely resemble (i.e., the normal cell correlate) (**Figure 1**). An important caveat of this line of reasoning is that the plasticity of the cancer cell transcriptome may obliterate mRNA traces of the cancer cell of origin or even change the cancer cell type. Nevertheless, normal cell correlates of childhood cancer have widely been considered to directly represent the cancer cell of origin. We prefer the more parsimonious view that the normal cell correlates primarily describe the differentiation state of cancer cells at the time of sampling.

There has been a recent upsurge in cancer cell–to–normal cell mRNA comparisons that aim to study the cell of origin of childhood cancers. These studies have been enabled by two key advances. The first is the rise of single-cell mRNA sequencing techniques, including plate-based, full-length transcript protocols and microfluidic systems that generate single-cell transcript counts, which are the predominant method utilized in single–cancer cell sequencing experiments (Aldridge & Teichmann 2020). Single-cell mRNA readouts are a prerequisite for cancer-to-normal comparisons, as the transcriptome of a bulk piece of tumor or normal tissue is a composite signal derived from millions of cells and multiple different cell types. Unless the transcriptomes of individual cells within a tissue can be captured, cancer-to-normal tissue mRNA comparisons based on bulk signals are invariably imprecise and may at best hint at the nearest normal cell correlates of cancers. The second key advance enabling systematic cancer cell–to–normal cell comparisons is the advent of the *Human Cell Atlas* project in 2016, a global endeavor to define all cell types of the human body at single-cell resolution, leveraging, in the first instance, mRNA sequencing (Rozenblatt-Rosen et al. 2017). Importantly, where legally permissible, this survey includes human fetal tissues

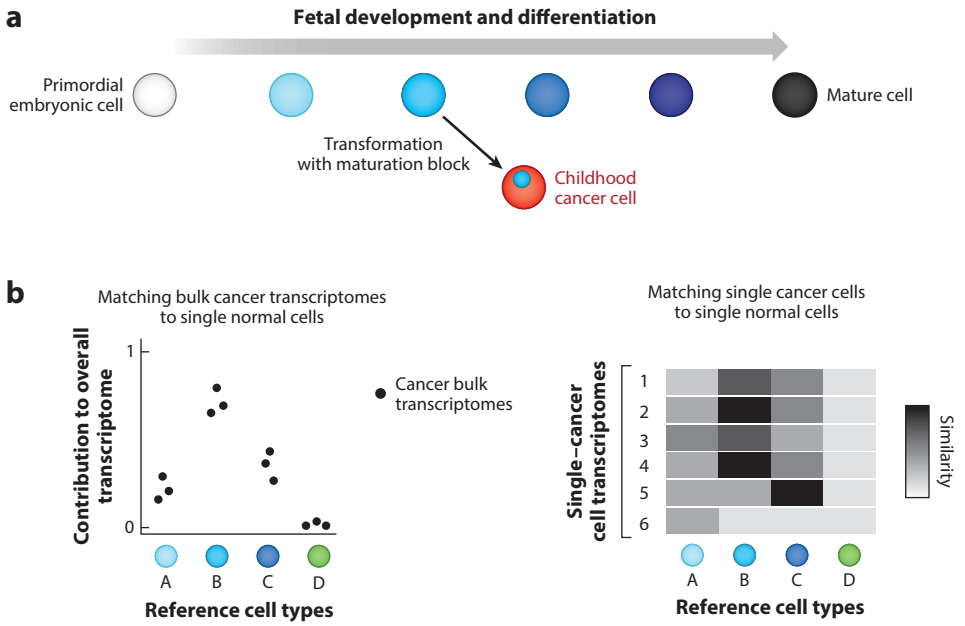

Figure 1

(*a*) Transcriptional remnant of the fetal cell of origin in childhood cancers. Childhood cancer is thought to arise as primordial embryonic cells develop and differentiate into mature cells. Cancer cells (*red cell*) may retain specific transcriptional features (*blue circle within red cell*) of their cell of origin. (*b*) A readout of the cancer cell of origin or normal cell correlate can be obtained through either comparing bulk cancer transcriptomes to single-cell-derived reference sets of normal cell types (*left*) or mapping the transcriptional readouts of individual cancer cells back to these reference sets (*right*). This latter method requires that the set of reference cell types is comprehensive and even contains several immature cell types, especially when tracing the origin of childhood cancers. Panel *b* adapted from Young et al. (2021).

(Behjati et al. 2018), which are the relevant reference for cancer cell–to–normal cell comparison of childhood tumors.

Experimental and analytical challenges of cancer cell–to–normal cell comparisons include the availability of fresh tumor tissues and the identification of cancer cells within single-cell data sets. Despite advances in single-nucleus RNA sequencing, the gold standard approach to generating single–cancer cell mRNA readouts remains fresh tumor tissue. In solid pediatric oncological practice, obtaining fresh viable tumor samples is challenging for several reasons, such as the rarity of cases, the competition of research with clinical needs over limited biopsy material, and the extensive necrosis that is present in resection specimens following neoadjuvant treatment with cytotoxic agents. In hematological practice this may be more straightforward, especially for leukemias that can even be sampled from peripheral blood. As a consequence, the scale of experiments remains mostly modest at present and confined to tens of specimens. A second challenge lies in identifying cancer cells within single-cell tumor data sets, as cancer cells may resemble non-neoplastic cells within tumor specimens. In some tumor types, markers are exquisitely sensitive and specific to cancer cells (e.g., CA9 in adult clear cell renal carcinoma), enabling cancer cell identification based on markers alone (Young et al. 2018). In most tumors, markers are more ambiguous. Furthermore, for some types of childhood cancer it has not been unequivocally resolved which cell types are cancerous. For example, in neuroblastoma there is a controversy as to whether Schwannian stroma or mesenchymal cells are malignant (Mora et al. 2001), thus necessitating a marker-independent

method for cancer cell identification. The definitive approach to identifying cancer cells in single-cell transcriptome data is to look in mRNA sequences for evidence of tumor-defining mutations, as determined from tumor DNA readouts (direct sequencing, copy number arrays). The predominant type of single-cancer cell mRNA data, non-full-length transcripts, only lends itself to copy number genotyping, as the proportion of the genome captured is far too limited for point mutation calling. Many methods for copy number calling in mRNA data build upon changes in gene expression. This approach may be problematic, however, as copy number variation may naturally coincide with regions of high gene expression in the cancer cell of origin (Kildisiute et al. 2021). As a consequence, expression-based copy number calling from single-cell data lacks specificity and may overcall cancer cells. The gold standard approach to calling copy number changes in single-cell sequencing data is to determine changes from imbalances of ratios of single-nucleotide polymorphisms (SNPs), captured by (high-quality) off-target cDNA reads (Fan et al. 2018, McCarthy et al. 2020). However, the latter approach is an arduous one requiring whole-genome sequencing of tumors (to derive heterozygous SNP ratios defining copy number changes) and is regrettably rarely utilized.

2.2. Insights from Matching Cancer and Normal Cells

These challenges notwithstanding, a myriad of single-cell mRNA sequencing studies have emerged in recent years that aim to trace the origin of childhood cancer, initially comparing cancer cells to fetal murine and, latterly, to human fetal cells. Childhood cancers that have been studied include brain tumors (Filbin et al. 2018, Gojo et al. 2020, Hovestadt et al. 2019, Jessa et al. 2019, Neftel et al. 2019, Tirosh et al. 2016), renal tumors (Young et al. 2018), and neuroblastomas (Dong et al. 2020, Jansky et al. 2021, Kameneva et al. 2021, Kildisiute et al. 2021), among others. For most tumor types, it has been possible to identify normal cell correlates that may represent the cancer cell of origin. Occasional examples of tumor types have emerged with no clear normal cell correlate. For instance, clear cell sarcoma of the kidney (CCSK) does not clearly resemble any particular mature or fetal kidney or more primordial embryonic cell, perhaps because the relevant normal cell has not yet been defined or because CCSK cells are transcriptionally distorted (Young et al. 2021). Another example without a clear match, in comparisons to developing brain cells, is a typical teratoid rhabdoid tumor, a childhood brain cancer. It has been suggested that this lack of a match may be due to an extracranial source of these tumors (Jessa et al. 2019). The overall conclusion that one may derive from these studies is that childhood cancer cells very clearly resemble fetal cells, substantiating in quantitative molecular terms the long-held view that pediatric neoplasms are stuck in development, which then defines a specific maturation block for therapeutic exploitation.

Mining single-cell mRNA analyses of childhood cancer cells for possible therapeutic targets has enabled the discovery of fetal genes that are utilized by childhood cancer cells but not in postnatal tissues, akin to the aforementioned example of GD2 expression in neuroblastoma. For instance, a recent single-cell mRNA study of neuroblastoma and fetal medullae identified such fetal transcripts utilized by neuroblastoma cells, including some that exhibit an even more restricted expression pattern in normal postnatal tissues than GD2 (Kildisiute et al. 2021). On some level, these observations were surprising. One might have reasonably expected that previous bulk analyses of neuroblastoma would have exhausted such findings, highlighting the power of quantitative molecular comparisons of individual cells.

Overall, these cancer cell–to–normal cell comparisons begin to describe the fetal state, and perhaps the origin, of childhood cancers and outline some of the possibilities for targeting this

fetalness or origin therapeutically. In view of the ambiguity and plasticity of cellular transcription, tracing the origin of childhood cancer definitively, however, requires phylogenetic approaches.

2.3. Phylogenetic Lineage Tracing from Somatic Mutations

To trace the origins of childhood cancers, one can leverage naturally occurring mutations of the DNA. All cells in the human body sustain damage to their genomes continuously throughout life, starting from the zygote. While the vast majority of this DNA damage is correctly repaired, postzygotic (somatic) mutations accumulate due to infidelities of repair and replication with each cell division. Therefore, every cell of the human body contains a set of somatic mutations that serves as a record of its past. Accordingly, somatic mutations shared between different cells suggest a common developmental lineage, since both are descendants of the cell that originally acquired those mutations. In principle, this approach can be scaled up to tissues, organs, and entire organisms to determine the developmental origin of each cell, including neoplastic expansions. These developmental relationships can be visualized as an extensive phylogeny, with the zygote at the tree's root, all sampled cells at its tips, and its nodes representing ancestral cells.

The key challenge with such an experiment is obtaining an accurate readout of the set of somatic mutations present in normal tissues. Whole-genome sequencing of large chunks of tissues (i.e., a bulk approach) will typically only yield a handful of detectable somatic mutations from the first few cell divisions of life (Ju et al. 2017). Since such bulk samples are generally large aggregates of cells with a polyclonal origin, only the earliest mutations will have a high enough variant allele frequency (VAF) to be detected. The VAF of a mutation multiplied by ploidy reflects the proportion of the sample that harbors it, with a VAF of 0.5 indicating that the mutation pervades all cells and a VAF of 0.25 indicating the mutation pervades half of all cells. However, constructing a phylogeny from the VAF of early mutations is severely hampered because we cannot readily discern which mutations have co-occurred in the same cells.

To overcome this problem, it is necessary to obtain direct readouts of the set of somatic mutations present in single cells. However, a single cell does not possess enough DNA for sequencing, so this needs to be amplified to be suitable for mutation detection. In recent years, this has been achieved in three ways: (*a*) single-cell sequencing after biochemical whole-genome amplification of the DNA, (*b*) sequencing the progeny of a single cell expanded in vitro, or (*c*) sequencing a naturally occurring monoclonal population of cells (**Figure 2*a*–*c***). These three approaches are discussed in the paragraphs below.

Direct single-cell genome sequencing has been used to study somatic mutagenesis in different normal tissues, such as brain (Lodato et al. 2015, 2018). However, the resolution of single-cell genome sequencing at a base pair level remains low. While it allows for the detection of copy number variants (CNVs) (Cheng et al. 2011, Laks et al. 2019), high rates of allelic dropout and the introduction of artifactual mutations preclude the accurate discovery of single-nucleotide variants (SNVs) or short insertions and deletions (indels) for phylogeny reconstruction. CNVs are exceedingly rare in normal tissues (Moore et al. 2021), and thus SNVs and indels provide the main means to perform lineage tracing. However, this obstacle can be overcome by specifically interrogating single cells for known mutant sites such as those discovered through ultradeep sequencing of bulk samples, which will order these mutations into a phylogeny (Bizzotto et al. 2021, Breuss et al. 2020).

Single cells can be expanded into organoids, colonies, or cell lines, which will amplify the DNA of the founder cell into an amount that is suited to regular whole-genome sequencing. The process of in vitro culturing and expansion will introduce private mutations into these samples, but mutations shared between two or more independent cultures will be preserved and unadulterated

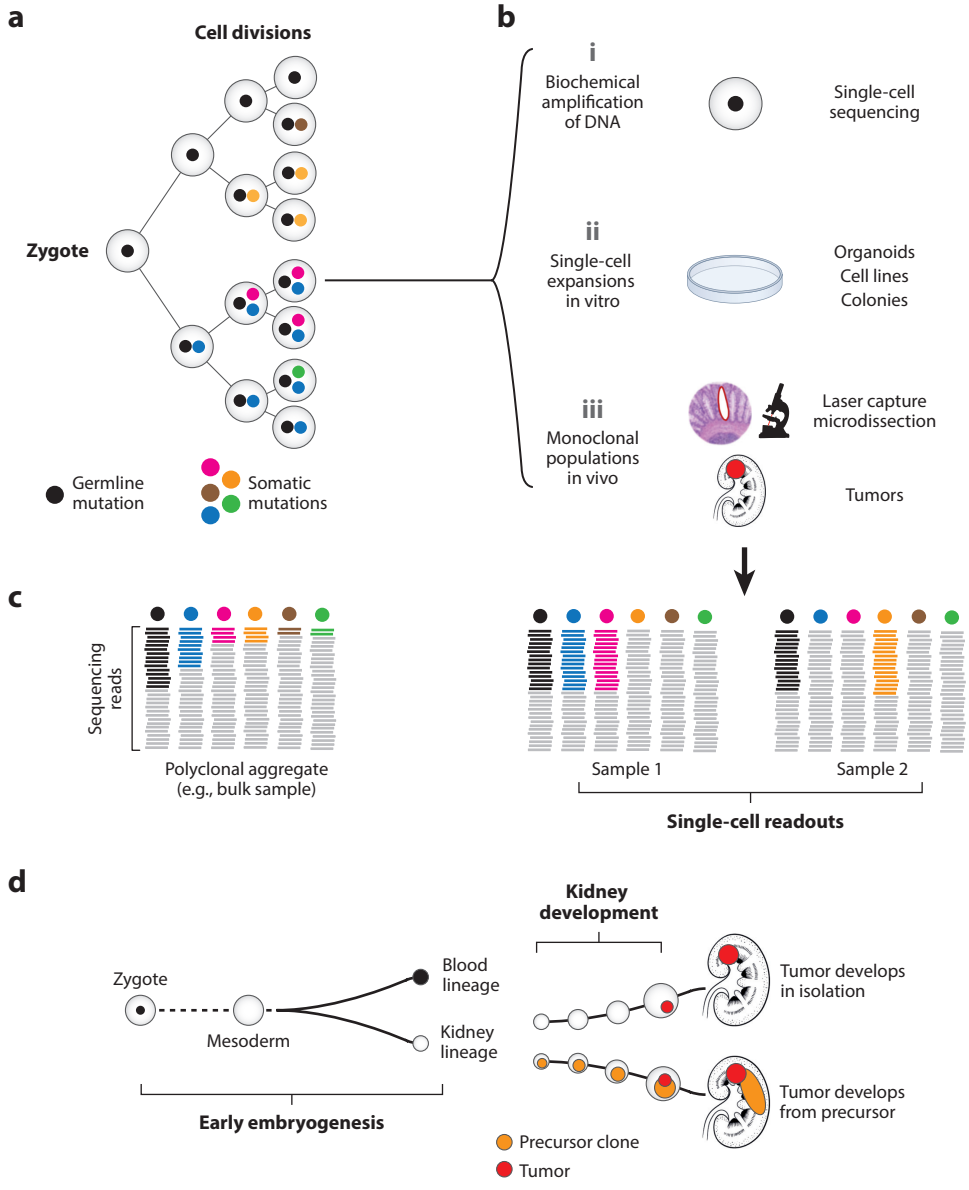

Figure 2

(*a*) In addition to inherited mutations, starting from the zygote, cells steadily acquire somatic mutations upon cell division, which become stable marks for the developmental history of these cells. Panel *a* adapted from Coorens et al. (2021b). (*b*) Readouts of somatic mutations in single cells are generally obtained in three ways: (*i*) biochemical amplification of the DNA in a single cell, (*ii*) expanding single cells into large clones in vitro, or (*iii*) sequencing naturally occurring monoclonal populations of cells. Histology image adapted from Coorens et al. (2021a). (*c*) Postzygotic mutations quickly become undetectable in polyclonal bulk samples, but single-cell readouts of somatic mutations can allow for the detection of early embryonic mutations and enable these to be ordered into a phylogeny. Panel *c* adapted from Coorens et al. (2021b). (*d*) Lineage tracing of childhood cancers with respect to different normal tissues can reveal whether the tumor has arisen in isolation or is derived from a (large) precursor clone, as is the case for Wilms tumors derived from expansion with hypermethylation of *H19*. Panel *d* and kidney image in panel *b* adapted from Coorens et al. (2019).

by the experimental procedure. Therefore, this approach is well suited for lineage tracing and has been used extensively for this purpose. The initial proof-of-principle study of mutation-based lineage tracing used this approach to study early murine development (Behjati et al. 2014). It has since been applied to construct phylogenies of early embryogenesis from epidermis-derived cell lines (Fasching et al. 2021, Park et al. 2021), as well as the development and dynamics of adult (Lee-Six et al. 2018) and fetal (Chapman et al. 2021) blood.

Recently, it has become possible to excise small populations of cells (approximately 200–1,000) via laser capture microdissection (LCM) and to subject these to high-fidelity whole-genome sequencing (Ellis et al. 2021). This approach can be used on naturally occurring monoclonal populations such as colonic crypts (Lee-Six et al. 2019, Olafsson et al. 2020, Robinson et al. 2021), endometrial glands (Moore et al. 2020), and patches of placental trophoblasts (Coorens et al. 2021b). This allows for a direct readout of the somatic mutations present in the founding cell of this unit. While not every organ is arranged in monoclonal histological units, as demonstrated in recent body-wide surveys (Moore et al. 2021), it has been possible to build extensive phylogenies of embryonic development across human tissues (Coorens et al. 2021a). An enormous advantage of this approach is that spatial information of the sequenced LCM cuts is retained, allowing for a precise histological interrogation of the genomic landscape.

Beside monoclonal populations of normal cells, a cancer is also a clonal outgrowth. The somatic mutations in the founder cell of a tumor represent a complete record of its history, from the zygote through its life as a precancerous cell to the establishment of the transformed tumor (Stratton et al. 2009) (**Figure 2*d***). Rooting the tumor as a tip in an extensive phylogeny of normal cells will yield the most complete picture of the developmental history of the cancer. However, cancers as a single clonal lineage enable interrogation of mutant sites in normal bulk tissues, such that the occurrence of these mutations can be ordered in time. For example, a mutation shared between a kidney tumor and blood likely represents a mutation that occurred before organogenesis, while one shared only with normal kidney tissue can signify a local, aberrant clonal expansion. Assessing the presence of tumor-specific mutations in bulk normal tissues is a more tractable experiment than the construction of extensive phylogenies to find the root of the cancer, but it is limited in its resolution. Nevertheless, the main insights from lineage tracing childhood cancers to date have been gained in this way (Coorens et al. 2019, 2020; Custers et al. 2021).

2.4. Insights from Lineage Tracing Childhood Tumors

In a first experiment, we traced the origin of Wilms tumor to large tissue-resident precursor clones (Coorens et al. 2019). We discovered somatic mutations that were shared between the Wilms tumor and corresponding normal kidney, but that were absent from blood. This feature was unique to Wilms tumors and not found in any other kidney cancer, either pediatric or adult, meaning that these somatic mutations delineate an early, aberrant clonal expansion. We termed this clonal nephrogenesis. Upon further analysis, we discovered that the clonal nephrogenesis was driven by early hypermethylation of the *H19* gene. *H19* resides within the 11p15.5 locus, an imprinted region of the genome. Methylation or loss of the maternal allele is a near-universal driver of Wilms tumor, as well as a primary cause of Beckwith-Wiedemann syndrome when present in all cells. Our study showed that an early cell acquires an epimutation of *H19* and subsequently generates a large clonal field from which the tumor eventually emerges. This early clone spans both kidneys in cases of bilateral Wilms tumor, which indicates that the hypermethylation happens very early in embryonic development, prior to the divergence of left and right renal lineages.

The initial discovery of an early clonal expansion predisposing to a childhood cancer prompts questions about the generality of this phenomenon. Recently, whole-genome sequencing of

hepatoblastoma and surrounding normal liver has revealed tissue-resident clones that already harbor loss of heterozygosity of the 11p15.5 locus in 10% of patients (Hirsch et al. 2021). This shows that, akin to Wilms tumors, hepatoblastoma can arise from precursor lesions with dysregulation of the *H19-IGF2* imprinting control region.

In addition, lineage tracing of malignant rhabdoid tumors (MRTs) revealed a pathogenesis very similar to the one observed for Wilms tumor (Custers et al. 2021). MRT was phylogenetically related to large precursor clones in normal nervous tissue that already lost both copies of *SMARCB1*, the sole driver event of MRT (Margol & Judkins 2014). These precursor lesions possess a cancer-like mutation burden and clonal composition even more pronounced than the clonal nephrogenesis observed in Wilms tumor. Intriguingly, no discernible genetic events with a phenotypic consequence distinguished the normal clone from the tumor. This might suggest that a nongenetic event, such as the transition to a so-called neoplasia-ready epigenetic state (Feinberg et al. 2006), might control the final transformation from normal to cancer cell.

It is of note that the vast majority of patients in whom we identify a precursor clone do not possess a germline mutation predisposing to cancer. Instead, the initiating step appears to be the acquisition of an early mosaic event, such as loss of *SMARCB1* or hypermethylation of *H19*. Unlike inherited predisposing mutations, mosaic drivers have the ability to create a differential fitness landscape across cells in the developing embryo or fetus and cause an aberrant clonal expansion of a single lineage. In other words, if all cells carry a driver mutation, none of them have an advantage over one another. However, if only some cells harbor a genomic alteration that increases their fitness, those cells have the opportunity to outcompete their unmutated counterparts. In such a scenario, this imbalance can potentially disrupt the physiological course of development and create large precursors lesions spreading through different organs.

This notion is reinforced by our study of bilateral neuroblastoma of the adrenal glands (Coorens et al. 2020). Here, left and right tumors only shared a small number of postzygotic mutations, all of which were also present in blood. Moreover, some tumors shared mutations with blood absent from the contralateral tumor. This indicates an early divergence of tumor lineages during the first few cell divisions of life and, hence, an independent emergence of both tumors. This is likely a consequence of the germline predisposition mutations present in these children.

The discovery of tissue-resident precursor clones that possess some of the key genomic features of their corresponding tumors has profound implications for the screening, treatment, and possible prevention of childhood cancers. Firstly, these precursors are delineated by precise genomic or epigenomic changes, which in principle allow them to be detected or even targeted for treatment. A preliminary experiment in MRT showed that reexpression of *SMARCB1*, as well as combined HDAC and mTOR inhibition, induces differentiation of MRT (Custers et al. 2021), opening up a possible avenue for differentiation therapies.

Secondly, the fact that these cells are morphologically and functionally normal, despite harboring driver mutations, hints at a process active in vivo to overcome their potential to form cancers. To illustrate, the risk of Wilms tumor drops dramatically after age 6 and is essentially zero beyond age 10 (Breslow et al. 1988). If clonal nephrogenesis represented a lifelong predisposition to Wilms tumor, such tumors would continue to appear throughout the entirety of childhood and long after. However, the absence of such prolonged increase in risk suggests that the predisposing effect of mosaic *H19* hypermethylation is transient and stoppable.

Overall, we have been able to discover a recurrent pattern of premalignant clonal expansions from which childhood cancers emerge, despite the limited resolution offered by comparison to bulk normal tissues. It is unlikely that these findings represent exceptions to the rule, as many childhood cancers might arise from small, local preneoplastic expansions. These would be impossible to detect using a bulk approach, but more sophisticated and detailed methods of spatial

genome sequencing provide a robust methodology for the next generation of phylogenetic lineage tracing experiments in childhood cancer.

3. FUTURE DIRECTIONS

Taken together, the results from comparing transcriptomes and genomes of childhood cancers to corresponding normal tissues have provided us with unprecedented insights into the origin and formation of pediatric malignancies. However, the current lines of inquiry only scratch the surface of what is possible and needed to fully realize the clinical potential of these findings. The future directions fall in roughly three categories.

Firstly, genomic and transcriptomic studies need to be rolled out systematically to investigate the origins of a much wider spectrum of childhood malignancies. The challenge posed here comes from the rarity of some childhood cancers.

Secondly, there is a need to leverage recent advances in genome sequencing (Ellis et al. 2021), which will enable exciting opportunities to root the development of childhood cancers in extensive phylogenies of normal tissues. The incorporation of genome sequencing of microscopic histological regions allows for a much higher resolution in the detection and spatial mapping of these precursor clones. This would also provide invaluable insights into the precise developmental point at which precursor clones emerge and their evolution thereafter.

Thirdly, to gain the most complete picture of the origins of childhood cancers, it is necessary to integrate genomic and transcriptomic data, which could be bolstered even further by the inclusion of epigenomic assays. Rather than the disconnected findings on the cancer cell of origin and the genetic lesions in precursor clones, such approaches would enable us to precisely quantify the differences between normal cells with and without malignant potential.

Together, these approaches provide us with a powerful means to identify and assess pediatric malignancy–specific targets for screening, therapy, and perhaps even prevention, which may herald a new chapter in the treatment of childhood cancers.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Aldridge S, Teichmann SA. 2020. Single cell transcriptomics comes of age. *Nat. Commun.* 11:4307

Behjati S, Gilbertson RJ, Pfister SM. 2021. Maturation block in childhood cancer. *Cancer Discov.* 11:542–44

Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, et al. 2014. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* 513:422–25

Behjati S, Lindsay S, Teichmann SA, Haniffa M. 2018. Mapping human development at single-cell resolution. *Development* 145(3):dev152561

Bhakta N, Liu Q, Ness KK, Baassiri M, Eissa H, et al. 2017. The cumulative burden of surviving childhood cancer: an initial report from the St. Jude Lifetime Cohort Study (SJLIFE). *Lancet* 390:2569–82

Bizzotto S, Dou Y, Ganz J, Doan RN, Kwon M, et al. 2021. Landmarks of human embryonic development inscribed in somatic mutations. *Science* 371:1249–53

Breslow N, Beckwith JB, Ciol M, Sharples K. 1988. Age distribution of Wilms' tumor: report from the National Wilms' Tumor Study. *Cancer Res.* 48:1653–57

Breuss MW, Yang X, Antaki D, Schlachetzki JCM, Lana AJ, et al. 2020. Somatic mosaicism in the mature brain reveals clonal cellular distributions during cortical development. bioRxiv 2020.08.10.244814. **https://doi.org/10.1101/2020.08.10.244814**

Chapman MS, Ranzoni AM, Myers B, Williams N, Coorens THH, et al. 2021. Lineage tracing of human embryonic development and foetal haematopoiesis through somatic mutations. *Nature* 595(7865):85–90

Cheng J, Vanneste E, Konings P, Voet T, Vermeesch JR, Moreau Y. 2011. Single-cell copy number variation detection. *Genome Biol.* 12:R80

Coorens THH, Farndon SJ, Mitchell TJ, Jain N, Lee S, et al. 2020. Lineage-independent tumors in bilateral neuroblastoma. *N. Engl. J. Med.* 383:1860–65

Coorens THH, Moore L, Robinson PS, Sanghvi R, Christopher J, et al. 2021a. Extensive phylogenies of human development inferred from somatic mutations. *Nature* 597:387–92

Coorens THH, Oliver TRW, Sanghvi R, Sovio U, Cook E, et al. 2021b. Inherent mosaicism and extensive mutation of human placentas. *Nature* 592:80–85

Coorens THH, Treger TD, Al-Saadi R, Moore L, Tran MGB, et al. 2019. Embryonal precursors of Wilms tumor. *Science* 366:1247–51

Custers L, Khabirova E, Coorens THH, Oliver TRW, Calandrini C, et al. 2021. Somatic mutations and single-cell transcriptomes reveal the root of malignant rhabdoid tumours. *Nat. Commun.* 12:1407

Dong R, Yang R, Zhan Y, Lai HD, Ye CJ, et al. 2020. Single-cell characterization of malignant phenotypes and developmental trajectories of adrenal neuroblastoma. *Cancer Cell* 38:716–33.e6

Downing JR, Shannon KM. 2002. Acute leukemia: a pediatric perspective. *Cancer Cell* 2:437–45

Ellis P, Moore L, Sanders MA, Butler TM, Brunner SF, et al. 2021. Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing. *Nat. Protoc.* 16:841–71

Fan J, Lee HO, Lee S, Ryu DE, Lee S, et al. 2018. Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome. Res.* 28:1217–27

Fasching L, Jang Y, Tomasi S, Schreiner J, Tomasini L, et al. 2021. Early developmental asymmetries in cell lineage trees in living individuals. *Science* 371:1245–48

Feinberg AP, Ohlsson R, Henikoff S. 2006. The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 7:21–33

Filbin MG, Monje M. 2019. Developmental origins and emerging therapeutic opportunities for childhood cancer. *Nat. Med.* 25:367–76

Filbin MG, Tirosh I, Hovestadt V, Shaw ML, Escalante LE, et al. 2018. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* 360:331–35

Gibson P, Tong Y, Robinson G, Thompson MC, Currle DS, et al. 2010. Subtypes of medulloblastoma have distinct developmental origins. *Nature* 468:1095–99

Gojo J, Englinger B, Jiang L, Hubner JM, Shaw ML, et al. 2020. Single-Cell RNA-seq reveals cellular hierarchies and impaired developmental trajectories in pediatric ependymoma. *Cancer Cell* 38:44–59.e9

Greaves M. 2018. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat. Rev. Cancer* 18:471–84

Hirsch TZ, Pilet J, Morcrette G, Roehrig A, Monteiro BJ, et al. 2021. Integrated genomic analysis identifies driver genes and cisplatin-resistant progenitor phenotype in pediatric liver cancer. *Cancer Discov.* 11:2524–43

Hovestadt V, Smith KS, Bihannic L, Filbin MG, Shaw ML, et al. 2019. Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* 572:74–79

Inaba H, Mullighan CG. 2020. Pediatric acute lymphoblastic leukemia. *Haematologica* 105:2524–39

Jansky S, Sharma AK, Korber V, Quintero A, Toprak UH, et al. 2021. Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma. *Nat. Genet.* 53:683–93

Jessa S, Blanchet-Cohen A, Krug B, Vladoiu M, Coutelier M, et al. 2019. Stalled developmental programs at the root of pediatric brain tumors. *Nat. Genet.* 51:1702–13

Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, et al. 2017. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* 543:714–18

Kameneva P, Artemov AV, Kastriti ME, Faure L, Olsen TK, et al. 2021. Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin. *Nat. Genet.* 53:694–706

Kildisiute G, Kholosy WM, Young MD, Roberts K, Elmentaite R, et al. 2021. Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. *Sci. Adv.* 7:eabd3311

Laks E, McPherson A, Zahn H, Lai D, Steif A, et al. 2019. Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. *Cell* 179:1207–21.e22
Lee-Six H, Obro NF, Shepherd MS, Grossmann S, Dawson K, et al. 2018. Population dynamics of normal human blood inferred from somatic mutations. *Nature* 561:473–78
Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, et al. 2019. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* 574:532–37
Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, et al. 2018. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* 359:555–59
Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, et al. 2015. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 350:94–98
Mackay A, Burford A, Carvalho D, Izquierdo E, Fazal-Salom J, et al. 2017. Integrated molecular meta-analysis of 1,000 pediatric high-grade and diffuse intrinsic pontine glioma. *Cancer Cell* 32:520–37.e5
Margol AS, Judkins AR. 2014. Pathology and diagnosis of SMARCB1-deficient tumors. *Cancer Genet*. 207:358–64
McCarthy DJ, Rostom R, Huang Y, Kunz DJ, Danecek P, et al. 2020. Cardelino: computational integration of somatic clonal substructure and single-cell transcriptomes. *Nat. Methods* 17:414–21
Miller DR. 2006. A tribute to Sidney Farber—the father of modern chemotherapy. *Br. J. Haematol.* 134:20–26
Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, et al. 2021. The mutational landscape of human somatic and germline cells. *Nature* 597:381–86
Moore L, Leongamornlert D, Coorens THH, Sanders MA, Ellis P, et al. 2020. The mutational landscape of normal human endometrial epithelium. *Nature* 580:640–46
Mora J, Cheung NK, Juan G, Illei P, Cheung I, et al. 2001. Neuroblastic and Schwannian stromal cells of neuroblastoma are derived from a tumoral progenitor cell. *Cancer Res.* 61:6892–98
Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, et al. 2019. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* 178:835–49.e21
Olafsson S, McIntyre RE, Coorens T, Butler T, Jung H, et al. 2020. Somatic evolution in non-neoplastic IBD-affected colon. *Cell* 182:672–84.e11
Park JA, Cheung NV. 2020. Targets and antibody formats for immunotherapy of neuroblastoma. *J. Clin. Oncol.* 38:1836–48
Park S, Mali NM, Kim R, Choi J-W, Lee J, et al. 2021. Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature* 597:393–97
Pizzo PA, Poplack DG, Adamson PC, Blaney SM, Helman L. 2016. *Principles and Practice of Pediatric Oncology*. Philadelphia: Wolters Kluwer
Robinson PS, Coorens THH, Palles C, Mitchell E, Abascal F, et al. 2021. Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. *Nat. Genet.* 53:1434–42
Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. 2017. The Human Cell Atlas: from vision to reality. *Nature* 550:451–53
Straathof K, Flutter B, Wallace R, Jain N, Loka T, et al. 2020. Antitumor activity without on-target off-tumor toxicity of GD2–chimeric antigen receptor T cells in patients with neuroblastoma. *Sci. Transl. Med.* 12(571):eabd6169
Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. *Nature* 458:719–24
Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, et al. 2016. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 539:309–13
van Noesel MM. 2012. Neuroblastoma stage 4S: a multifocal stem-cell disease of the developing neural crest. *Lancet Oncol.* 13:229–30
Wilms M. 1899. *Die Mischgeschwülste der Niere*. Leipzig, Ger.: Verlag Arthur Georgi
Young MD, Mitchell TJ, Custers L, Margaritis T, Morales-Rodriguez F, et al. 2021. Single cell derived mRNA signals across human kidney tumors. *Nat. Commun.* 12:3896
Young MD, Mitchell TJ, Vieira Braga FA, Tran MGB, Stewart BJ, et al. 2018. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* 361:594–99