Published in final edited form as:
Cell. 2011 April 1; 145(1): 25–29. doi:10.1016/j.cell.2011.03.019.

Mapping cancer origins

Richard J. Gilbertson
Departments of Developmental Neurobiology and Oncology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN, 38105, USA. Richard.Gilbertson@stjude.org

Summary

Cancer comprises a bewildering assortment of diseases that kill 7.5 million people each year. Poor understanding of cancer's diversity currently thwarts our goal of a cure for every patient, but recent integration of genomic and stem cell technologies promises a route through this impasse.

Why haven't we cured cancer?

Seventy years since the first reported use of cancer chemotherapy, malignancies are the second most common cause of death among children and adults. Why are the headlines rarely punctuated with cancer success stories? One explanation is that while classifying cancers is relatively straightforward, understanding the basis of cancer heterogeneity is complex. Over the years we have become quite proficient at cataloging cancer according to patterns of epidemiology and pathology. Each cancer is recognized to occur at a particular age, more frequently in one sex than another, and have a particular morphology—usually resembling the originating tissue. Advances in imaging and histology have enabled us to further segregate cancer diagnoses into distinct stages and grades that predict different treatment responses and prognoses.

Despite this exhaustive work, our attempts to understand the processes that generate the different forms of cancer have proved far less fruitful, hamstringing efforts to advance therapy in the clinic. Failure to determine the biological basis of histologically similar but clinically and molecularly distinct cancers (inter-tumoral heterogeneity) has proved especially limiting: preventing the development of preclinical models of the full spectrum of human cancers, and fostering a clinical trials culture that accepts 'all comers' with only the broadest categories of histological criteria to filter eligibility.

Our failure to define the origins of cancer subtypes is not for want of trying. However, our relatively crude understanding what drives cancers, coupled with uncertainty about initiating cell types has prevented investigators from making the jump from correlative observation to functional understanding. Recently, a string of publications suggest that the genomic revolution may provide a route through this impasse. Microarray technologies have transformed the depth with which we can interrogate cancers like leukemias (Ross et al., 2004), breast cancers (Sotiriou et al., 2003) and brain tumors (Gibson et al., 2010; Johnson et al., 2010; Northcott et al., 2010; TCGA, 2008), partitioning these diseases into robust subgroups according to genome wide patterns of gene expression, copy number alteration and mutation. These genomic profiles correlate with long recognized epidemiological,

pathological and clinical characteristics, provide fundamental biological insights, and detect molecular echoes of tumor origins.

Lessons from leukemia

Different types of chromosomal translocations—the principal oncogenic mutations in the blood—have long been associated with specific subtypes of leukemia. Genomic, stem cell and cancer assays have taught us important lessons about the basis of this 'matching'. First, the different forms of leukemia appear to arise from distinct points in the hematopoietic lineage that are susceptible to specific translocations. For example the *BCR-ABL* translocation seen in human chronic myeloid leukemia (CML) only initiates CML in uncommitted hematopoietic stem cells (HSCs) (Huntly et al., 2004), while translocations involving the *mixed-lineage leukemia (MLL)* gene can initiate acute leukemias in both HSCs and committed progenitor cells (Barabé et al., 2007; Chen et al., 2008; Krivtsov et al., 2006).

What is the biology behind this translocation-lineage stage matching? Comparative gene expression profiling suggests the answer might lie in the capacity of translocations to activate key leukemogenic programs. Extensive self-renewal is considered a requisite feature of leukemic stem cells. When committed, non-self-renewing granulocyte macrophage progenitors (GMP) are transduced with *MLL–AF9* they generate AML. The leukemic stem cells in this model retain a GMP-like gene expression profile, but they also acquire an aberrant self-renewal signature and self-renewal capacity, normally seen only in HSCs (Krivtsov et al., 2006). Since *BCR-ABL* does not appear to activate self-renewal, but rather enhanced cell proliferation and survival, its leukemogenic potential might be restricted to HSCs that already possess the capacity to self-renew (Huntly et al., 2004; Schemionek et al., 2010). Further probing of gene expression profile differences between normal and transformed hematopoietic cells has also highlighted new therapeutic opportunities. The transcriptome of *MLL-AF9* transformed GMPs encodes a reactivated Beta-catenin (Ctnnb1) signal that drives leukemogenic self-renewal, and that might be blocked for therapeutic gain (Wang et al., 2010).

Although it is intuitive that cancers arise from specific combinations of mutations and susceptible cell types, these landmark studies of leukemia demonstrate the power of genomic technologies to decipher this process. Importantly, these data demonstrate that mutations can activate oncogenic signals without globally reprogramming the initiating cell. As we shall see, the legacy of the initiating cell transcriptome within cancer cells can provide crucial clues to tumor origins as well as unmask novel therapeutic targets.

Charting cancer origins in solid tissues

The availability of assays for each stage in the hematopoietic lineage as well as the liquidity of blood has accelerated understanding of leukemogenesis beyond that of solid tumorigenesis. But studies of solid cancers are catching up. The rigid anatomical organization of solid tissues has allowed investigators to map cells that express transcriptomes similar to those seen in cancers, and improved techniques to isolate and culture cells in solid tissue hierarchies have advanced the study of these populations. These studies have identified cells in solid tissues that share the transcriptomes of solid tumors and might therefore initiate cancer (Figure 1).

Like most solid tissues, those of the central nervous system (CNS) give rise to a variety of cancers classified according to patterns of histology. Intracranial ependymomas are the third most common brain tumor of children, and carry a much worse prognosis than spinal forms of the disease that predominate in adults. Ependymomas contain transcriptomes and DNA

copy number alterations that correlate with tumor location (Johnson et al., 2010; Taylor et al., 2005). Similarly, gene expression profiles of medulloblastomas—the most common malignant pediatric brain tumor—carve these cancers into clinically and molecularly distinct subgroups, including the Sonic Hedgehog (SHH)-subtype driven by aberrant SHH signaling, and the highly curable WNT-subtype containing mutations in *CTNNB1* (Northcott et al., 2010).

To uncover the cellular origins of these diverse brain tumors, investigators tested the hypothesis that brain tumor subtypes inherit significant portions of their transcriptome from initiating CNS cells. Initial *in situ* hybridization analyses showed that the subventricular zone of the embryonic lateral ventricle and peri-canal region of the adult spine express the transcriptomes of human cerebral and spinal ependymoma, respectively (Taylor et al., 2005). Since these regions house neural stem cells (NSCs) the investigators looked for transcriptomic similarities between regionally and developmentally discrete mouse NSCs and human ependymomas (Johnson et al., 2010). Using a powerful new cross-species genomics algorithm, the investigators pinpointed embryonic cerebral and adult spinal NSCs as candidate cells-of-origin of cerebral and spinal ependymomas, respectively.

This approach has also provided clues to the origins of medulloblastomas, yielding the surprising insight that some of these tumors might arise outside of the cerebellum (Gibson et al., 2010). SHH-subtype medulloblastomas have been shown to arise from committed cerebellar granule neuron precursor cells (GNPCs) (Schuller et al., 2008; Yang et al., 2008). Not surprisingly therefore, *in situ* hybridization and cross-species genomics revealed a close match between SHH-subtype medulloblastomas and GNPC transcriptomes (Gibson et al., 2010). In stark contrast, the transcriptome of WNT-subtype medulloblastoma matched that of neural precursor cells of the lower rhombic lip and embryonic dorsal brainstem (Gibson et al., 2010). Remarkably, magnetic resonance imaging demonstrated that while human SHH-subtype medulloblastomas are confined to the cerebellum, WNT-subtype tumors invariably involve the dorsal brainstem, supporting further the notion that these different tumor types have distinct cellular origins (Gibson et al., 2010).

Clues to cancer origins are not just present in the transcriptomes of leukemias and brain tumors, but have also been uncovered through comparative profiling of breast cancers and the normal mammary gland. The ducts and lobules of the human breast are lined by two cell layers: an inner luminal cell population and a heterogeneous outer cell layer that includes basal progenitor cells. Basal-like breast cancers, so called for their basal-cell immunophenotype (cytokeratins 5/6, 14, and 17), are aggressive tumors, particularly prevalent among women with germline mutations in *BRCA1*. It seemed intuitive that basal-like tumors would arise from basal progenitor cells; however, recent data suggest these cancers actually arise from luminal progenitors. The pre-neoplastic breasts of *BRCA1* mutant woman contain an expanded population of aberrant luminal progenitor cells, and the transcriptomes of *BRCA1*-mutant breast tissue and basal-type breast cancers are more like that of luminal progenitors than other normal breast cell types (Lim et al., 2009).

Transcriptome mapping indentifies solid tumor initiating cells

Although the transcriptomes of certain normal and malignant solid tissue cells correlate, does this pinpoint cancer origins? Studies of mammary tissue have delineated a cellular hierarchy on which to frame this question in the breast. Cell transplant studies have identified a Lin^{-} $CD29^{hi}CD24^{+}$ mammary stem cell (MaSC) capable of reconstituting the entire breast via lineage-committed progenitors (e.g., $Lin^{-}CD29^{lo}CD24^{+}$ luminal progenitors) (Shackleton et al., 2006; Stingl et al., 2006). Initial studies indicate that different cells in the mammary hierarchy are susceptible to different mutations. For

example, transgenic expression of Wnt1 via the MMTV promoter generates heterogeneous breast cancers in mice that are preceded by the accumulation of aberrantly self-renewing MaSCs (Shackleton et al., 2006). In contrast, the mammary glands of MMTV-neu mice contain normal numbers of MaSCs and develop luminal breast cancers, while forced expression of Notch1 in MaSCs expands the luminal progenitor population, leading to basal-like breast cancers (Bouras et al., 2008).

Targeting tumor type-specific mutations to transcriptome-matched normal cells has provided direct evidence that comparative gene expression profiling can identify cancer origins (Figure 2). In the breast, conditional deletion of *Brca1* from mouse luminal progenitors, but not basal progenitors, produced tumors that mimic the histology and transcriptome of human *BRCA1*- mutant and sporadic basal-type breast cancers, thus confirming comparative gene expression profile predictions (Lim et al., 2009;Molyneux et al., 2010). Similarly embryonic cerebral NSCs that were predicted by transcriptome mapping to initiate cerebral ependymomas, generated these tumors when challenged with mutations found exclusively in this form of the human disease (Johnson et al., 2010). Further, mouse models have validated the surprising prediction that WNT-subtype medulloblastomas likely arise from progenitor cells in the dorsal brainstem (Gibson et al., 2010). Mutation of *Ctnnb1*—an invariable feature of WNT-subtype medulloblastoma—had little impact on progenitor cell populations in the cerebellum, but caused the abnormal accumulation of neuron precursor cells in the dorsal brainstem that progressed to form medulloblastomas that recapitulate the anatomy and gene expression profiles of human WNT-subtype medulloblastoma.

Further study will doubtless reveal significant exceptions and nuances in the relationship between the transcriptomes of normal cells and their malignant offspring. Nevertheless, integrated genomic and stem cell studies have provided a useful framework for investigating the origin of cancer subtypes.

In the right place at the wrong time

Tissues have been viewed largely as passive players in cancer pathogenesis; their risk of malignant transformation being dictated by heritable mutations and environmental exposures. But if cancers arise from preordained combinations of specific cell types and matched mutations, then the availability of either of these factors could dictate the epidemiology of cancer. In its broadest terms this concept is not new. The very different cancer types observed in children and adults have long been suspected to have their roots in development (Figure 1). However, evidence is emerging that more subtle shifts in the makeup of cell hierarchies might account for changes in cancer incidence. For example, it is tempting to speculate that the expanded population of aberrant luminal progenitors seen in the breasts of cancer-free patients carrying the *BRCA1* mutation provides a source of cells susceptible to additional transforming mutations, and thereby an increased risk of developing basal-type cancers (Lim et al., 2009;Molyneux et al., 2010).

Shifts in cell hierarchies might also explain the association between certain physiological states and cancer. For example, if the massive progesterone-induced increases in mouse MaSCs prove to occur in humans, then this may explain why breast cancer is associated with early menarche, late menopause and the inclusion of progestin in hormone replacement therapy (Joshi et al., 2010).

As genomic and stem cell technologies allow further dissection of cancer subgroups and their origins, it will be interesting to see if these studies provide answers to other key questions about cancer epidemiology. For example, do temporal changes in the GNPC lineage explain why SHH-subtype medulloblastoma incidence peaks in both early childhood and later life (Northcott et al., 2010)? And might the close matching of adult spinal NSC and

spinal ependymoma transcriptomes explain why these tumors occur almost exclusively in adults (Johnson et al., 2010)?

Toward cancer cures

Most biomedical discoveries rarely translate rapidly into improved patient care. However, recent discoveries that genomic tools can identify robust cancer subgroups, and point to the origins of these cancers, have immediate clinical relevance. This promise lies in the profound implications this information holds for the full spectrum of cancer research.

Non-specific cytotoxic treatments remain the mainstay of cancer therapy. Efforts to introduce more directed treatments that target mutant proteins in cancers have met with mixed results. The inefficiency of this process results in part from prior failures to adequately capture the heterogeneity of cancers in preclinical models. The integration of stem cell biology and genomics outlined in this review is producing multiple, highly accurate models of human cancer subgroups (Gibson et al., 2010; Johnson et al., 2010; Molyneux et al., 2010; Yang et al., 2008). Preclinical drug development using these models should allow investigators to better predict which forms of leukemias and solid cancers are most likely to respond to certain treatments (Figure 2). Such preclinical data could then be used to design genomic metrics that would direct treatments to the most appropriate patients in early clinical trials.

Cancer models built from specific cells-of-origin offer an additional advantage to drug developers. Therapies that cripple critical processes in cancer cells carry significant risks of damaging normal cell hierarchies. Understanding the origin of cancers might therefore allow the development of effective drugs with fewer side effects. For example, evidence that Ctnnb1 mediated self-renewal driven by *MLL-AF9* in AML is not required by adult HSCs, suggests CTNNB1 might be targeted therapeutically in this disease without incurring significant hematological toxicity (Wang et al., 2010).

Finally, the approaches outlined here promise to shed light on one of the greatest contemporary controversies in cancer research—the cancer stem cell hypothesis. Evidence that cancers are propagated and maintained by sub-populations of stem-like cancer cells has come largely from studies of human cancer xenografts in mice. But these systems do not allow lineage tracing of cancer development, and their interpretation is complicated by species differences. New mouse models of human tumors initiated from predefined and selected cells are enabling investigators to track tumorigenesis with much greater precision. Comparative genomic studies of initiating and daughter cancer cells should provide a more comprehensive view of the processes that cause and propagate cancer.

Acknowledgments

I apologize to those investigators whose work could not be referenced because of space limitations.

References

Barabé F, Kennedy JA, Hope KJ, Dick JE. Science. 2007; 316:600–604. [PubMed: 17463288]

Bouras T, Pal B, Vaillant F, Harburg G, Asselin-Labat ML, Oakes SR, Lindeman GJ, Visvader JE. Cell Stem Cell. 2008; 3:429–441. [PubMed: 18940734]

Chen W, Kumar AR, Hudson WA, Li Q, Wu B, Staggs RA, Lund EA, Sam TN, Kersey JH. Cancer Cell. 2008; 13:432–440. [PubMed: 18455126]

Gibson P, Tong Y, Robinson G, Thompson MC, Currle DS, Eden C, Kranenburg TA, Hogg T, Poppleton H, Martin J, et al. Nature. 2010; 468:1095–1099. [PubMed: 21150899]

Huntly BJ, Shigematsu H, Deguchi K, Lee BH, Mizuno S, Duclos N, Rowan R, Amaral S, Curley D, Williams IR, et al. Cancer Cell. 2004; 6:587–596. [PubMed: 15607963]

Johnson RA, Wright KD, Poppleton H, Mohankumar KM, Finkelstein D, Pounds SB, Rand V, Leary SE, White E, Eden C, et al. Nature. 2010; 466:632–636. [PubMed: 20639864]

Joshi PA, Jackson HW, Beristain AG, Di Grappa MA, Mote PA, Clarke CL, Stingl J, Waterhouse PD, Khokha R. Nature. 2010; 465:803–807. [PubMed: 20445538]

Krivtsov AV, Twomey D, Feng Z, Stubbs MC, Wang Y, Faber J, Levine JE, Wang J, Hahn WC, Gilliland DG, et al. Nature. 2006; 442:818–822. [PubMed: 16862118]

Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A, et al. Nat Med. 2009; 15:907–913. Epub 2009 Aug 2002. [PubMed: 19648928]

Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, Mackay A, Grigoriadis A, Tutt A, Ashworth A, et al. Cell Stem Cell. 2010; 7:403–417. [PubMed: 20804975]

Northcott PA, Korshunov A, Witt H, Hielscher T, Eberhart CG, Mack S, Bouffet E, Clifford SC, Hawkins CE, French P, et al. J Clin Oncol. 2010 Epub ahead of print.

Ross ME, Mahfouz R, Onciu M, Liu HC, Zhou X, Song G, Shurtleff SA, Pounds S, Cheng C, Ma J, et al. Blood. 2004; 104:3679–3687. Epub 2004 Jun 3629. [PubMed: 15226186]

Schemionek M, Elling C, Steidl U, Baumer N, Hamilton A, Spieker T, Gothert JR, Stehling M, Wagers A, Huettner CS, et al. Blood. 2010; 115:3185–3195. [PubMed: 20053753]

Schuller U, Heine VM, Mao J, Kho AT, Dillon AK, Han YG, Huillard E, Sun T, Ligon AH, Qian Y, et al. Cancer Cell. 2008; 14:123–134. [PubMed: 18691547]

Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, Wu L, Lindeman GJ, Visvader JE. Nature. 2006; 439:84–88. [PubMed: 16397499]

Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET. Proc Natl Acad Sci U S A. 2003; 100:10393–10398. Epub 12003 Aug 10313. [PubMed: 12917485]

Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li HI, Eaves CJ. Nature. 2006; 439:993–997. Epub 2006 Jan 2004. [PubMed: 16395311]

Taylor MD, Poppleton H, Fuller C, Su X, Liu Y, Jensen P, Magdaleno S, Dalton J, Calabrese C, Board J, et al. Cancer Cell. 2005; 8:323–335. [PubMed: 16226707]

TCGA. Nature. 2009; 455:1061–1068.

Wang Y, Krivtsov AV, Sinha AU, North TE, Goessling W, Feng Z, Zon LI, Armstrong SA. Science. 2010; 327:1650–1653. [PubMed: 20339075]

Yang ZJ, Ellis T, Markant SL, Read TA, Kessler JD, Bourboulas M, Schuller U, Machold R, Fishell G, Rowitch DH, et al. Cancer Cell. 2008; 14:135–145. [PubMed: 18691548]

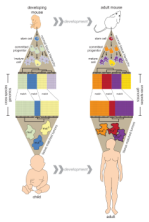

Figure 1. Cross-species genomics matches cells in developing and adult mouse tissues with human tumor subgroups

Matching of developing mouse tissues with childhood cancers (left). Different cell stages in an embryonic mouse tissue hierarchy are colored according to differentiation stage. Expression profiling segregates these cells according to transcriptome (top-left). Histologically similar but clinically distinct tumors from the corresponding childhood tissue express different transcriptomes driven by different mutations (mut^{1-3}) (bottom-left). Cross-species genomics matches human tumors with their candidate cells-of-origin in the corresponding developing mouse tissue. As development proceeds (right), the spectrum of normal cell types (top right) and cancers (bottom right) changes. The same approach shown for childhood cancers can match tumors with candidate cells-of-origin in adult tissues.

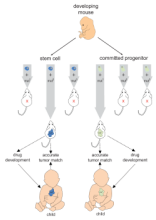

Figure 2. Modeling cancer heterogeneity

Subgroups of the childhood cancer shown in Figure 1 can be modeled accurately when transcriptome matched normal cell types are challenged with mutations found in the corresponding cancer. These models should prove extremely useful for developing new therapies for specific cancer subgroups.