

## Cancer Genome Sequencing—An Interim Analysis

Edward J. Fox, Jesse J. Salk, and Lawrence A. Loeb

Joseph Gottstein Memorial Laboratory, Department of Pathology, University of Washington, Seattle, Washington

### Abstract

With the publishing of the first complete whole genome of a human cancer and its paired normal, we have passed a key milestone in the cancer genome sequencing strategy. The generation of such data will, thanks to technical advances, soon become commonplace. As a significant number of proof-of-concept studies have been published, it is important to analyze now the likely implications of these data and how this information might frame cancer research in the near future. **The diversity of genes mutated within individual tumor types, the most striking feature of all studies reported to date, challenges gene-centric models of tumorigenesis.** Although cancer genome sequencing will revolutionize certain aspects of personalized care, the value of these studies in facilitating the development of new therapies, their primary goal, seems less promising. Most significantly, however, the cancer genome sequencing strategy, as currently applied, fails to characterize the most relevant genomic features of cancer—the mutational heterogeneity within individual tumors. [Cancer Res 2009;69(12):4948–50]

Cancer is a genomic disease associated with the accumulation of mutations (1). The prevailing models of tumorigenesis stress tumor progression as the result of sequential mutations in key cancer genes, each mutation driving a new round of clonal proliferation (2, 3). By cataloging all clonal mutations found in a variety of cancers, it is hoped that it will be possible to identify the key mutated genes of cancer and, in so doing, revolutionize basic and clinical cancer research (4). A series of studies have now been undertaken to annotate the cancer genomes of several different tumor types and it is timely to summarize the results and to evaluate the implications of these findings.

Early attempts at systematically identifying multiple somatic mutations within individual cancers yielded few prevalent clonal alterations (5–7). The relatively limited sequence coverage of these initial studies spurred on attempts at more comprehensive screens of the cancer genome. The first complete sequencing of all likely coding exons of a human cancer genome, conducted in breast and colon cancers (8, 9), concluded that these cancers contain a median of 84 and 76 clonal mutations, respectively, which alter protein function (Table 1). Although almost 10% of the 18,197 genes studied were detectably mutated in at least one specimen, each tumor displayed a unique and diverse profile of mutated genes, and other than those previously known (e.g., *TP53* and *APC*), no new prevalently mutated genes were identified. The authors of this study proposed that the cancer genome landscape is composed of a

handful of commonly mutated gene “mountains” but dominated by a vastly larger number of infrequently mutated gene “hills” (Fig. 1A), a view consistent with a large number of mutations, each providing a small fitness advantage, driving tumor progression (9). It is this clonal mutational heterogeneity, the authors concluded, which underlies the wide variation in tumor behavior and responsiveness to therapy.

Subsequent studies on diverse cancer types (Table 1) have generated the most extensive catalogs of somatic point mutations yet undertaken (10–14). The primary conclusion to be drawn from these studies is that most cancer types display substantial heterogeneity at the genetic level. Individual solid organ tumors harbor on average >50 nonsilent clonal mutations in the coding regions of different genes (colorectal  $\approx$  breast  $\gg$  pancreatic  $>$  glioblastoma multiforme), and although a few of these genes are mutated in a high proportion of tumors, the prevalence at which the majority are mutated among different tumors of the same cancer type is low. Additionally, these studies focused exclusively on identifying mutations in exons of known protein-coding genes. With the ever-increasing recognition that so-called “junk” DNA and intronic sequences contain functional elements, including regulatory regions and noncoding RNAs, the total number of clonal mutations with functional consequences cannot be easily estimated.

The characterization of the first hematopoietic cancer genome represents an important methodologic milestone in cancer genome sequencing—truly unbiased whole genome sequencing (15). Prior attempts at resequencing the tyrosine kinase gene family in acute myeloid leukemia (AML) yielded few mutations (16, 17). By exhaustively interrogating the cancer and paired normal genomes of a single AML patient using massively parallel sequencing, Ley and colleagues (15) identified 500 to 1,000 nonsynonymous somatic mutations throughout the cancer genome, of which 10 mutations were in protein-coding genes. Although the overall number of protein coding mutations is considerably fewer than that described for solid organ tumors, similar to the mutational diversity of other tumor types, the genes identified are mutated in only a small fraction of AML cases. Surprisingly, none of the 8 newly identified genes were found to be mutated in a further 187 cases of AML.

**Although the absence of prevalently mutated genes is disappointing from a targeted therapeutic perspective, the major shortcoming of all the above studies is not in the complexity of the problem they are attempting to address but in the fact that, by design, they do not analyze the deeper heterogeneity within individual tumors. A primary tumor is itself genomically heterogeneous, with each cell having a unique mutational signature.** This genetic variation within a cancer cell population reflects the history and dynamics of clonal evolution and, importantly, serves as a reservoir of genetic diversity from which therapy-resistant clones may arise (18). Analysis of disseminated single cells in minimal residual disease has shown that there is a high level of genomic heterogeneity within individual lesions as well as between primary

**Requests for reprints:** Lawrence A. Loeb, Department of Pathology, University of Washington, Box 357705, 1959 Northeast Pacific Street, Seattle, WA 98195. Phone: 206-543-6015; Fax: 206-543-3967. E-mail: laloeb@u.washington.edu.

©2009 American Association for Cancer Research.  
doi:10.1158/0008-5472.CAN-09-1231

**Table 1.** Selected details of cancer genome sequencing studies

Study	Cancer type	Total no. genes analyzed	No. mutated genes	Total no. mutations (nonsilent)	Average no. mutations per tumor
Sjjoblom et al. (8) and Wood et al. (9)	Breast ( $n = 11$ )*	18,191	1,137	1,243	84
	Colorectal ( $n = 11$ )*		848	942	76
Greenman et al. (10)	Diverse ( $n = 210$ )	518	581	798	NA
Jones et al. (11)	Pancreatic ( $n = 24$ )	20,661	1,007	1,163	48
Parson et al. (12)	Glioblastoma ( $n = 21$ ) <sup>†</sup>	20,661	685	748 <sup>†</sup>	47
TCGA (13)	Glioblastoma ( $n = 91$ )	601	223	453	NA
Ding et al. (14)	Lung ( $n = 188$ )	623	348	1,013	NA
Ley et al. (15)	AML ( $n = 1$ )	Whole genome	10	500–1,000	NA

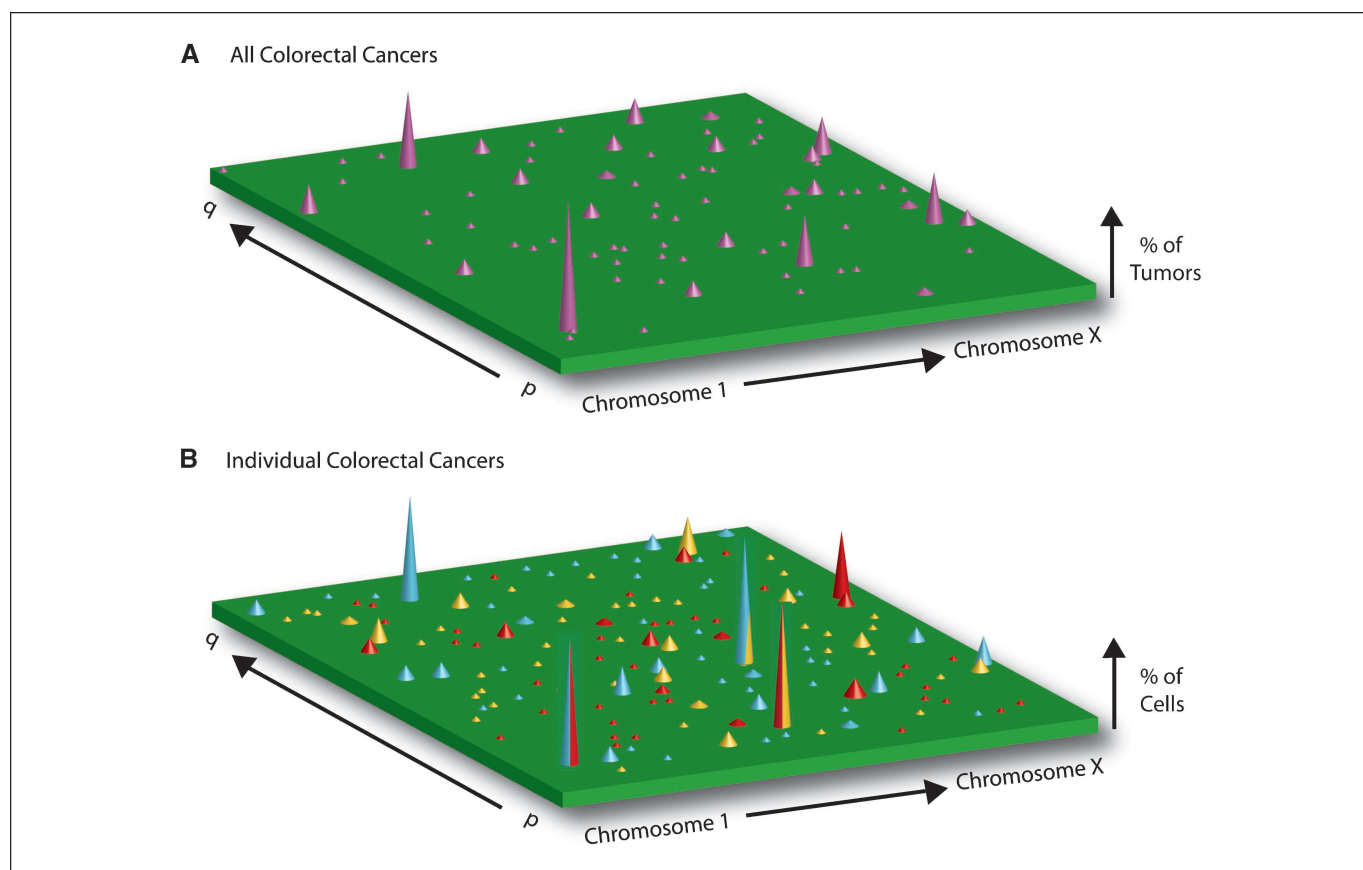
Abbreviation: NA, not applicable.

\*Discovery phase of study only.

†Excludes posttherapy samples and mismatch repair-deficient samples.

tumors and metastatic cells (19, 20). Although current “next-generation” sequencers can detect multiple rare subclones with frequencies as low as 1 in 5,000 (21), analyzing rarer subclones awaits an, as yet, undeveloped future generation of DNA

sequencing methodologies. It is this intraspecimen diversity, not detected by the cancer genome sequencing studies, that may be clinically most relevant and responsible for much of the diverse spectrum of clinical phenotypes. The question arises of whether



**Figure 1.** The mutational landscape of the cancer genome. *A*, the cancer genome landscape of colorectal cancer as proposed by Wood and colleagues (9) shows the mutational heterogeneity between individual tumors. The height of each violet peak represents the percentage of colorectal tumors found to carry a clonal mutation in a particular gene. The landscape comprises a few mountains, genes that are mutated in a large number of cancers (e.g., *TP53*; ref. 23), and a significantly larger number of hills, genes mutated in only a few individual tumors. Although there may be 50 or more genes clonally mutated in a given tumor, any given gene is rarely mutated in more than a few tumors. *B*, the mutational heterogeneity within individual tumors not detected by the cancer genome sequencing studies. The mutational landscape of three tumors is shown; each tumor is represented by a set of differently colored peaks (blue, red, and yellow). The height of each peak reflects the number of cells within a single tumor that have a particular mutated gene. Split colored peaks represent the small number of mutated genes shared in common between more than one tumor. Within an individual neoplasm, a few mutations are present throughout the population (*tall peaks*), a greater number are present in minority subclones (*short peaks*), and the majority are found in only one or a few cells (invisibly distributed throughout the *green background*).

there is, at a clinically meaningful level, a single consensus cancer genome per tumor.

We posit that a complete description of the cancer genome must capture the mutational heterogeneity within individual neoplasms as well as between different tumors. In addition to the clonal genomic landscape proposed, comprising mountains and hills of frequently and infrequently mutated genes within a given tumor type, additional mutational diversity exists within individual tumors themselves, with a large number of subclonal mutations being present in only one or a few cells (Fig. 1B). It is this mutational landscape, overlooked by the cancer genome sequencing studies to date, that is the basis for the wide variations in tumor behavior and responsiveness to therapy and represents the clinically most important features of the cancer genome.

New DNA sequencing technologies are undoubtedly important for guiding biological and medical research; however, many limitations exist. **To accurately quantify the subclonal diversity of a single tumor, it may be necessary to sequence millions of individual genome equivalents per tumor.** Present sequencing efforts only detect mutations that are present in the majority of the cells of a neoplasm. Until more advanced technologies become available or existing ones are retooled, a critical feature of the genomic landscape will continue to go unexplored.

The mutational heterogeneity of cancer, which has been recognized for many years (22) and now confirmed by DNA sequencing, underlies the complexity of the cancer genome landscape. **Currently, the heterogeneity within individual tumor types alone confounds many of the original expectations of the cancer genome strategy. A future of multiple targeted therapies and patient stratification, based on a mutational signature of defined key genes for each cancer type, seems less hopeful than initially anticipated. Although ongoing studies will continue to uncover the remarkable heterogeneity inherent in tumorigenesis, without addressing the deeper complexity of each individual cancer genome, we may not significantly affect patient care.**

## Disclosure of Potential Conflicts of Interest

L.A. Loeb: consultant/advisory board, Stratus, Inc. The other authors disclosed no potential conflicts of interest.

## Acknowledgments

Received 4/2/09; accepted 4/13/09; published OnlineFirst 6/9/09.

**Grant support:** NIH grants CA102029 and CA78885 (L.A. Loeb) and AG033485 (J.J. Salk) and National Cancer Institute/Health Research Board of Ireland Joint Research Fellowship (E.J. Fox).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

## References

- Weinberg RA. The biology of cancer. New York: Garland Science; 2006.
- Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell* 1990;61:759–67.
- Grady WM, Markowitz SD. Genetic and epigenetic alterations in colon cancer. *Annu Rev Genom Hum Genet* 2002;3:101–28.
- Pollack A. *New York Times*. 2005 March 28;Sect. A1.
- Wang TL, Rago C, Silliman N, et al. Prevalence of somatic alterations in the colorectal cancer cell genome. *Proc Natl Acad Sci U S A* 2002;99:3076–80.
- Bardelli A, Parsons DW, Silliman N, et al. Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* 2003;300:949.
- Wang Z, Shen D, Parsons DW, et al. Mutational analysis of the tyrosine phosphatome in colorectal cancers. *Science* 2004;304:1164–6.
- Sjblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;314:268–74.
- Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. *Science* 2007;318:1108–13.
- Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature* 2007;446:153–8.
- Jones S, Zhang X, Parsons DW, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008; 321:1801–6.
- Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008;321:1807–12.
- TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008;455:1061–8.
- Ding L, Getz G, Wheeler DA, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;455:1069–75.
- Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008;456:66–72.
- Tomasson MH, Xiang Z, Walgren R, et al. Somatic mutations and germline sequence variants in the expressed tyrosine kinase genes of patients with *de novo* acute myeloid leukemia. *Blood* 2008; 111:4797–808.
- Loriaux MM, Levine RL, Tyner JW, et al. High-throughput sequence analysis of the tyrosine kinome in acute myeloid leukemia. *Blood* 2008;111:4788–96.
- Loeb LA, Loeb KR, Anderson JP. Multiple mutations and cancer. *Proc Natl Acad Sci U S A* 2003;100:776–81.
- Stoecklein NH, Hosch SB, Bezler M, et al. Direct genetic analysis of single disseminated cancer cells for prediction of outcome and therapy selection in esophageal cancer. *Cancer Cell* 2008;13:441–53.
- Klein CA, Schmidt-Kittler O, Schardt JA. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci U S A* 1999;96:4494–9.
- Campbell PJ, Pleasance ED, Stephens PJ, et al. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A* 2008; 105:13081–6.
- Loeb LA, Springgate CF, Battula N. Errors in DNA replication as a basis of malignant changes. *Cancer Res* 1974;34:2311–21.
- Harris CC. p53: at the crossroads of molecular carcinogenesis and cancer risk assessment. *Science* 1993;262:1980–1.