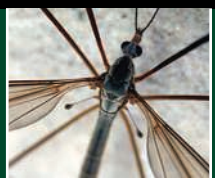
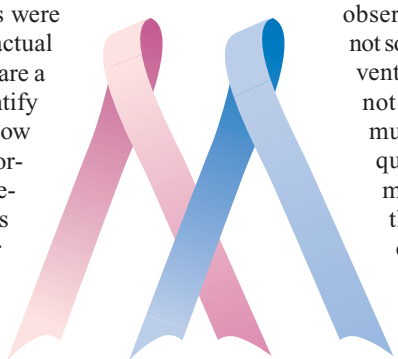


LETTERS

edited by Etta Kavanagh

Limits to the Human Cancer Genome Project?

ALTHOUGH THE EFFORTS OF VOGELSTEIN AND COLLEAGUES TO DEFINE THE MUTATIONAL landscape of two of the most common human cancers (breast and colon) are highly commendable ("The consensus coding sequences of human breast and colorectal cancers," T. Sjöblom *et al.*, Research Article, 13 Oct. 2006, p. 268), they also put into stark reality the challenges facing the Human Cancer Genome Project (HCGP). One wonders about the merits of such high-cost, low-efficiency, and ultimately descriptive-type "brute force" studies. Although previously unknown mutated genes were unearthed, the functional consequences of most of these and their actual role in tumorigenesis are unknown, and even with that knowledge we are a long way from identifying new therapeutic targets. Screening to identify potentially important genes in nonhematopoietic malignancies is now possible because of recent advances in transposon-based, unbiased, forward mutagenesis screens with potential for tissue-specific mutagenesis in mice (*1*, *2*). The advantage of such a system is that researchers can identify mutations that initiate, cooperate, and maintain a tumor by observing the development and progression of tumors in mice. Mouse models of human cancer are traditionally generated to contain known genetic changes identified in human cancers. Recent studies have demonstrated the power of the reverse, using abnormalities detected in cancer mouse models to study previously unknown, syntenic, genetic lesions and their significance in human cancers (*3*, *4*). These comparative oncogenomic approaches combined with unbiased mutagenesis screens should provide a list of high-priority targets that can then be studied comprehensively for mutations and epigenetic abnormalities in human cancer and validated as therapeutic targets. Such approaches may be more rational and cost-effective, allowing a better compromise between achieving the major goals of the HCGP and appropriation of funding to other worthwhile cancer efforts.



WEE J. CHNG

Division of Hematology-Oncology, Mayo Clinic, Scottsdale, AZ 85259, USA. E-mail: chng.wee@mayo.edu

References

1. L. S. Collier, C. M. Carlson, S. Ravimohan, A. J. Dupuy, D. A. Largaespada, *Nature* **436**, 272 (2005).
2. A. J. Dupuy, K. Akagi, D. A. Largaespada, N. G. Copeland, N. A. Jenkins, *Nature* **436**, 221 (2005).
3. M. Kim *et al.*, *Cell* **125**, 1269 (2006).
4. L. Zender *et al.*, *Cell* **125**, 1253 (2006).

IN THEIR RESEARCH ARTICLE "THE CONSENsus coding sequences of human breast and colorectal cancers" (13 Oct. 2006, p. 268), T. Sjöblom *et al.* present the most extensive data available on sequencing of DNA from human cancers. Sequencing of 465 Mb from 22 tumors yielded 189 cancer-associated genes and reveals the enormous diversity and complexity of cancer genomes. The findings that "[t]he vast majority of genes were not known to be mutated in tumors," the average tumor harbored 90 mutated genes, no gene was consistently mutated in either breast or colorectal tumors, and there was no uniform panel of mutated genes argue against massive DNA sequencing being an efficient method for target identification. The authors' statement that "the number of mutational events occurring during the evolution of human tumors … is much greater than previously thought" ignores a large literature on tumor heterogeneity and random mutations in human cancers (*1*).

There are important limitations to discovery of mutated genes by DNA sequencing. First, the vast majority of DNA sequence alterations are single-nucleotide polymorphisms, germline mutations, PCR errors, or DNA duplications. More than 99% of the nucleotide alterations observed by Sjöblom *et al.* were not somatic mutations. Second, conventional DNA sequencing does not detect nonclonal (random) mutations, and a consensus sequence is not informative of mutations in single cells. Although both random and clonal mutations could drive tumorigenesis, random mutations may also account for tumor cell heterogeneity, metastasis, and drug resistance (*2*). Third, expansion of the Human Cancer Genome Project must be rigorously justified in the context of diminished funding for the investigator-initiated grants that are critical for generating new approaches to improve cancer treatment and prognosis.

LAWRENCE A. LOEB AND JASON H. BIELAS

Department of Pathology, University of Washington, Seattle, WA 98195, USA.

References

1. L. A. Loeb, C. F. Springgate, N. Battula, *Cancer Res.* **34**, 2311 (1974).
2. J. H. Bielas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18238 (2006).

IN THEIR RESEARCH ARTICLE "THE CONSENsus coding sequences of human breast and colorectal cancers" (13 Oct. 2006, p. 268), S. Sjöblom *et al.* report an initial total of 816,986 putative nucleotide changes in tumors. Of these, 32% or 259,957 silent changes were discarded as insignificant. A second screen uncovered 133,693 changes, but we are not told how many were silent.

The winnowing process, which reduced the candidates to 191 changes, would probably have reduced the silent mutations proportionately. An independent screen of malignant gliomas after alkylation therapy found 26% silent mutations (*1*). Silent, or synonymous, mutations are assumed to be nonselective. A frequency of about 25% is expected if mutation is random. The frequency of silent mutations in the International Agency for Research on Cancer p53 database (*2*) is 4.41%, with a theoretical expectation of 23.5% for random mutation using the p53 selection of codons. Just as the low frequency of silent mutations in p53 indicates that this gene is selected during tumorigenesis, so their high frequency in this sequencing effort indicates that the vast majority of mutations occur randomly and are found in tumors as passengers. This conclusion in no way minimizes the importance of particular mutations for which the percentage of silent mutations is similar to that of p53. The Sanger Center Data Base, COSMIC, lists a total of 109 silent mutations out of 2335 mutations (4.67%) in eight cancer-related genes (e.g., CDKN2A, EGFR, KIT, and RB1) for which more than 100 total mutations had been recorded. New cancer-related genes may possibly be identified by methods similar to those used by Sjöblom *et al.* The question is whether massive sequencing is an efficient way to uncover them against the huge background of random mutational noise.

BERNARD S. STRAUSS

Departments of Molecular Genetics and Cell Biology and Radiation and Cellular Oncology, University of Chicago, Chicago, IL 60637, USA.

References

1. C. Hunter *et al.*, *Cancer Res.* **66**, 3987 (2006).
2. M. Olivier *et al.*, *Hum. Mutat.* **19**, 607 (2002).

Response

IMAGINE THAT NO MOLECULAR CANCER research had ever been performed before the Sjöblom *et al.* study on colorectal cancers. In a single stroke, the study would have identified all of the consensus coding sequence genes now known to play a role in this tumor type, including *TP53*, *APC*, *KRAS*, *SMAD4*, *FBXW7*, *EPHA3*, *SMAD2*, and *TGFBRII*. Many other genes not previously implicated in colorectal tumors were concomitantly discovered to be mutated at significant frequencies. These included genes that are central to the pathogenesis of other forms of cancer, such as *GNAS*, *NF1*, and *RET*. The study would also have provided unprecedented clues to the pathogenesis of inherited cancer syndromes such as familial adenomatous polyposis, neurofibromatosis, Li-Fraumeni syndrome, juvenile polyposis, and multiple endocrine neoplasia. Finally, it would have pointed to virtually all the major pathways currently known to be involved in neoplasia. Our study took less than a year to complete (once the technology was developed) and cost a tiny fraction of what was actually spent to discover a subset of the identified genes through conventional means. This effort demonstrated that unbiased genome sequencing is an extremely efficient way to discover cancer genes.

Chng, Strauss, and Loeb and Bielas raise a number of other points in their thought-provoking Letters, which we address here.

1) We have already pointed out that a large fraction of the mutations found in cancers were likely to be passenger mutations. Although the observed plethora of mutations may account for the clinical and biological heterogeneity of tumors, most of them are not likely to be integral to neoplasia. Precisely for this reason, we designed a measure (CaMP scores) to rank genes via their mutation frequency, taking into account gene size and nucleotide context. Those genes with the highest CaMP scores are the ones of most interest for future genetic and functional studies. All of the genes noted above ranked among the top 30 such genes.

2) Our data strongly support previous results indicating that most human cancers have a mutation rate that is similar to that observed in normal cells (*1*). A small fraction of cancers have higher mutation rates because of mismatch repair deficiencies. Such cancers were excluded from our study because mutations in these tumors are more difficult to interpret.

3) Our study was designed to identify those mutations that may drive tumorigenesis, i.e., clonal mutations. Such mutations have been shown to be directly responsible for tumor progression and are the only ones known to be useful as diagnostic and therapeutic targets. Once clonal mutations are identified in late-stage or therapy-resistant tumors, it could be useful to search for such alterations in earlier stage clinical samples wherein they may be present in a small fraction of tumor cells. However, genome-wide identification of alterations present in small tumor subpopulations is neither feasible nor desirable at this time.

4) Sequencing studies cannot replace functional studies in model organisms, and the latter studies are essential to reduce morbidity and mortality from cancer. But sequencing studies *can* guide functional studies by focusing them on genes that are likely to play a role in human cancers. We have demonstrated that such genes can be identified by relatively simple and inexpensive sequencing methods. As technology improves, such sequencing will become even more cost-effective. To turn the process on its head by trying to identify genes important in human cancer through functional studies, followed only later by sequencing, would be substantially more costly and less comprehensive.

TOBIAS SJÖBLOM,[1] SIÂN JONES,[1] LAURA D. WOOD,[1] D. WILLIAMS PARSONS,[1] JIMMY LIN,[1] THOMAS BARBER,[1*] DIANA MANDELKER,[1] REBECCA J. LEARY,[1] JANINE PTAK,[1] NATALIE SILLIMAN,[1] STEVE SZABO,[1] PHILLIP BUCKHAULTS,[2] CHRISTOPHER FARRELL,[2] PAUL MEEH,[2] SANFORD D. MARKOWITZ,[3] JOSEPH WILLIS,[4] DAWN DAWSON,[4] JAMES K. V. WILLSON,[5] ADI F. GAZDAR,[6] JAMES HARTIGAN,[7] LEO WU,[8] CHANGSHENG LIU,[8] GIOVANNI PARMIGIANI,[9] BEN HO PARK,[10] KURTIS E. BACHMAN,[11] NICKOLAS PAPADOPOULOS,[1] BERT VOGELSTEIN,[1] KENNETH W. KINZLER,[1] VICTOR E. VELCULESCU[1]

[1]Ludwig Center and Howard Hughes Medical Institute, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD 21231, USA. [2]Department of Pathology and Microbiology, Center for Colon Cancer Research, and South Carolina Cancer Center, Division of Basic Research, University of South Carolina School of Medicine, Columbia, SC 29229, USA. [3]Department of Medicine, Ireland Cancer Center, and Howard Hughes Medical Institute, Case Western Reserve University and University Hospitals of Cleveland, Cleveland, OH 44106, USA. [4]Department of Pathology and Ireland Cancer Center, Case Western Reserve University and University Hospitals of Cleveland, Cleveland, OH 44106, USA. [5]Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [6]Hamon Center for Therapeutic Oncology Research and Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [7]Agencourt Bioscience Corporation, Beverly, MA 01915, USA. [8]Soft Genetics LLC, State College, PA 16803, USA. [9]Departments of Oncology, Biostatistics, and Pathology, Johns Hopkins Medical Institutions, Baltimore, MD 21205, USA. [10]Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD 21231, USA. [11]Department of Radiation Oncology and Department of Biochemistry and Molecular

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Biology, Marlene and Stewart Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD 21201, USA.

*Present address: Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA.

Reference

1. C. Lengauer, K. W. Kinzler, B. Vogelstein, *Nature* **396**, 643 (1998).

Gene Expression and Ethnic Differences

IN HER ARTICLE "IN ASIANS AND WHITES, GENE expression varies by race" (News of the Week, 12 Jan., p. 173), J. Couzin explains that recent research has demonstrated that there is significant variation in gene expression in Asians and whites, although providing the disclaimer that "[g]enetic variation among races, long a political hot potato, has also been a scientific puzzle." Indeed, anthropologists and geneticists alike have been keenly interested in determining just what makes us different at the genetic level. The problem is how these differences are conceptualized. After decades of discussion, we are still left asking: what is a race? Race is truly a biological and taxonomic problem, not simply a sociological problem as many have argued. Now, it should be asked, what does it mean to say that gene expression varies between "Asians" and "whites." In the context of this research, "Asian" means either Japanese or Chinese, or both groups pooled together. The term "white" is used interchangeably with "Caucasian" and people of "European descent." This imprecise use of language further adds to the problem of not being able to accurately describe or classify the reference population. Who do these people represent? Certainly, both Asia and Europe are very large and diverse continents, both socially and biologically. The difference in gene expression between the Japanese and Chinese subjects cited in this report demonstrates that variation occurs not only on a continental level but also locally, which is indicative of patterns of ethnic variation rather than racial variation. We should not simply be asking how human groups vary in terms of genetic composition, but how those groups vary in terms of the social and biological processes that created them.

MICHAEL S. BILLINGER

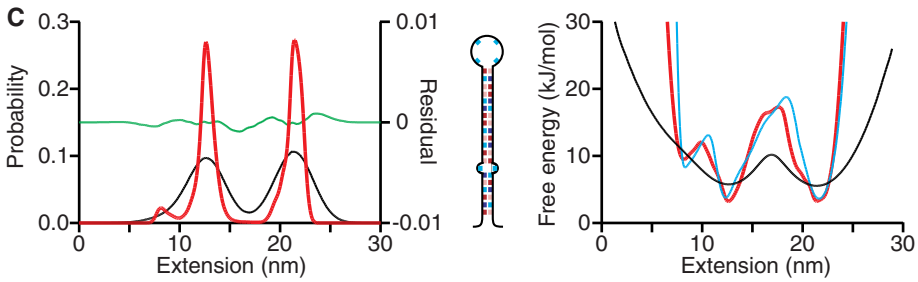
Edmonton, Canada.

CORRECTIONS AND CLARIFICATIONS

Brevia: "Ancient noncoding elements conserved in the human genome" by B. Venkatesh *et al.* (22 Dec. 2006, p. 1892). Table S1 should have been included in the Supporting Online Material (SOM), rather than posted on the author's Web site. Table S1 includes the coordinates of the conserved noncoding sequences within the elephant shark genome and the corresponding coordinates in the human genome. The elephant shark sequences (1.4x coverage) have been deposited at DDBJ/EMBL/GenBank under the project accession AAVX00000000. The version described in this paper is the first version, AAVX01000000. The authors have also submitted the traces of these sequences to the Trace Archive at the NCBI. Tables S8 and S9 should also have been included in the SOM, and the following statement should have been added to the acknowledgments in reference 8: "The sequences of the human and elephant shark noncoding elements are in Tables S8 and 9." Tables S1, S8, and S9 are now available with the SOM at www.sciencemag.org/cgi/content/full/314/5807/1892/DC1.

Research Articles: "The genome of the sea urchin *Strongylocentrotus purpuratus*" by Sea Urchin Genome Sequencing Consortium (10 Nov. 2006, p. 941). On pages 951 and 952, errors were made in renumbering authors' affiliations: Some changes were missed, and the affiliation for Nikki Adams was omitted. C. G. Elsik, T. Hibino, and V. D. Vacquier appear twice. C. G. Elsik is at Texas A&M University; T. Hibino is at the Sunnybrook Research Institute and Department of Medical Biophysics, University of Toronto; V. D. Vacquier is at the Scripps Institution of Oceanography. Corrected group affiliations, then individuals alphabetically: P. Kitts, M. J. Landrum, D. Maglott, K. Pruitt, A. Souvorov, National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894, USA. O. Fedrigo, A. Primus, R. Satija, Department of Biology and Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708, USA. Nikki Adams, Biology Department, California Polytechnic State University, San Luis Obispo, CA 93407, USA. C. Flytzanis, Department of Biology, University of Patras, Patras, Greece, and the Department of Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. B. E. Galindo, Biotechnology Institute, Universidad Nacional Autónoma de Mexico (UNAM), Cuernavaca, Morelos, Mexico 62250. J. V. Goldstone, Department of Molecular, Cellular, and Developmental Biology, University of California, Berkeley, Berkeley, CA 94720, USA. G. Manning, Razavi-Newman Center for Bioinformatics, Salk Institute for Biological Studies, La Jolla, CA 92186, USA. D. Mellott, Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Columbus Center, Baltimore, MD 21202, USA. J. Song, Department of Molecular and Cellular Biology and Biochemistry, Brown University, Providence, RI 02912, USA. D. P. Terwilliger, Department of Biological Sciences, George Washington University, Washington, DC 20052, USA. A. Wikramanayake, Department of Zoology, University of Hawaii at Manoa, Honolulu, HI 96822, USA.

Reports: "Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid" by M. T. Woodside *et al.* (10 Nov. 2006, p. 1001). On page 1002, the key in Fig. 2C is incorrect: In the legend for Fig. 2C, the descriptions of the color blocks should read, "Distance from F to U (black), F to I (blue), and I to U (red) plotted versus the mismatch location." The hairpin sequence shown in Fig. 3C is also incorrect. The corrected Fig. 3C is shown here.



News Focus: "Truth and consequences" by J. Couzin (1 Sept. 2006, p. 1222). Due to an editing error, the sentence on page 1223, "Kuersten and Padilla talked for about an hour and together examined the papers cited in the proposal" was incorrect. It should have read, "Kuersten and Padilla talked for about an hour and together examined the pages of the proposal."

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Obestatin, a Peptide Encoded by the Ghrelin Gene, Opposes Ghrelin's Effects on Food Intake"

N. Chartrel, R. Alvear-Perez, J. Leprince, X. Iturrioz, A. Reaux-Le Goazigo, V. Audinot, P. Chomarat, F. Coge, O. Nosjean, M. Rodriguez, J. P. Galizzi, J. A. Boutin, H. Vaudry, C. Llorens-Cortes

Zhang *et al*. (Research Articles, 11 November 2005, p. 996) reported that obestatin, a peptide derived from the ghrelin precursor, activated the orphan G protein–coupled receptor GPR39. However, we found that I^{125}-obestatin does not bind GPR39 and observed no effects of obestatin on GPR39-transfected cells in various functional assays (cyclic adenosine monophosphate production, calcium mobilization, and GPR39 internalization). Our results indicate that obestatin is not the cognate ligand for GPR39.

Full text at www.sciencemag.org/cgi/content/full/315/5813/766c

RESPONSE TO COMMENT ON "Obestatin, a Peptide Encoded by the Ghrelin Gene, Opposes Ghrelin's Effects on Food Intake"

Jian V. Zhang, Cynthia Klein, Pei-Gen Ren, Stefan Kass, Luc Ver Donck, Dieder Moechars, Aaron J. W. Hsueh

We cannot reproduce our original findings on obestatin binding and activation of GPR39 receptors in vitro. However, we can reproduce our original findings on the in vivo effects of obestatin in mice (decreases in food intake, gastric emptying responses, and body weight gain) under precise experimental conditions. Further studies are needed to reveal the exact relation between obestatin and the G protein–coupled receptor GPR39.

Full text at www.sciencemag.org/cgi/content/full/315/5813/766d