Decoding the evolution of a breast cancer genome

Philippe L. Bedard1,2 & Christos Sotiriou1,2,3*

Keywords: breast cancer; genomics

Shah et al have recently reported the first successful sequencing of the entire genome of a solid tumour (Shah et al, 2009). Philippe Bedard and Christos Sotiriou analyse their findings as well as the challenges of applying the study of cancer genomes to clinical cancer care.

Cancer is a clonal disease of acquired genetic alterations that impart a selective advantage for survival, proliferation, invasion, dissemination and immune escape (Fig 1). To date, mutations in 412 genes have been reported in human cancers (COSMIC: http://www.sanger.ac.uk/genetics/CGP/Census/), representing less than 2% of the 20,000 known protein-coding genes in the human genome (Stratton et al, 2009). These cancer genes were primarily identified through the screening of individual genes thought to be important for the pathogenesis of selected cancers, using the labour-intensive Sanger sequencing method that has been widely applied for more than 30 years. The formidable size of a diploid human genome—more than 6 billion bases in length—has prevented the sequencing of the entire genome of an individual cancer to determine its full spectrum of acquired mutations, but the recent development of massively parallel next-generation sequencing systems, able to generate volumes of deoxyribonucleic acid (DNA) sequence equivalent to up to 4 diploid human genomes on a weekly basis, has revolutionized the approach to cancer gene identification.

Shah et al reported the first successful sequencing of the entire genome of a solid tumour.«

In a recent issue of the journal Nature, Shah et al reported the first successful sequencing of the entire genome of a solid tumour (Shah et al, 2009). Shah and colleagues isolated DNA from a metastatic lobular breast cancer to generate aligned DNA sequence reads corresponding to a staggering 43.1-fold coverage of a haploid genome, or 141 billion nucleotides, using a high-throughput massively parallel DNA sequencing system (Illumina). They also sequenced the entire transcriptome (the set of all messenger ribonucleic acid (mRNA) transcripts expressed by the tumour) to yield an additional 6 billion nucleotides. After comparison with known libraries of DNA sequence variation and targeted re-sequencing of tumour DNA and normal lymphocyte DNA, the authors confirmed 32 somatic point mutations in the tumour. None of these point mutations was previously described and only 11 of the 32 genes were known to be mutated in other cancers. Surprisingly, none of the mutated genes was identified as a candidate cancer gene (CANC) genes through a mutational screen of known protein-encoding genes in 11 estrogen receptor (ER)-negative breast cancer cell lines (Sjoblom et al, 2006).

To determine the relevance of the individual mutations identified to the pathogenesis of breast cancer, Shah et al performed targeted sequencing of 9 mutation positions on an independent series of 192 breast cancers (including 112 lobular breast cancers). They did not find any identical mutations to the index tumour, however, three tumours (two lobular, one ductal) harboured non-synonymous variants/deletions affecting the ERBB2 kinase domain and two tumours (both lobular) contained different truncating variants affecting the HAUS3 gene. The ERBB2 gene encodes for the ErbB2/HER2/neu transmembrane growth factor receptor, known to be overexpressed and/or amplified in 20% of breast cancers. The detection of rare ERBB2 hotspot kinase domain mutations in non-ERBB2 amplified breast cancers is consistent with prior observations (Stephens et al, 2005), although its clinical relevance for predicting responsiveness to anti-HER2 directed therapy, such as trastuzumab, is unknown. HAUS3 is a subunit of the augmin complex of proteins that is required for centrosome formation and proper chromosome segregation during mitosis (Lawo et al, 2009). This is the first report of a mutation in this gene in a human cancer.
but the finding of HAUS3 truncating mutations in two additional lobular cancers suggests that its function may be important for preventing tumourigenesis. Through an analysis of DNA sequence read coverage, the authors were able to detect copy number alterations (CNAs) in the genome of the metastatic lobular cancer. Interestingly, 19 regions of high level DNA amplification were confirmed by fluorescence in situ hybridization (FISH), including a novel amplification not previously reported in human cancers or cell lines on chromosome 19 that involves the insulin receptor (INSR) gene.

By comparing sequencing of the transcriptome (RNA-seq) with the genome, the authors were able to identify alterations in two transcripts (COG3 and SRP9) that result in variant protein sequences not encoded by the genomic tumour DNA, suggesting that these variants were caused by RNA editing. RNA editing enzymes, such as ADAR, whose gene was highly expressed in this tumour, have been identified as important contributors to transcriptome diversity that regulates normal physiological processes and contributes to human disease (Li et al, 2009). This study adds to an emerging body of evidence that cancers may subvert several epigenetic processes, including RNA editing to adapt to environmental stress and promote tumour progression.

Targeted re-sequencing of the mutation amplicons was performed using genomic DNA from the primary tumour, obtained from the paraffin-embedded tumour blocks of the primary tumour removed 9 years earlier. By using massively parallel sequencing, the authors were able to estimate the frequency of mutations in the primary tumour. Only 5 of the 32 mutations (in ACB11, HAUS3, SLC24A4, SNX4 and PALB2) identified in the metastatic lesion were prevalent in the initial primary tumour. Interestingly, six additional mutations (in K1F1C, USP28, MYH8, MORC1, KIAA1468 and RNASEH2A) were detectable at lower frequencies in the primary tumour (1–13%). This provides important evidence of clonal heterogeneity at the time of initial diagnosis, with subsequent genomic evolution during the course of metastatic progression over a 9-year interval. The data also underscore the power of deep sequencing technologies to highlight the genomic heterogeneity present in cancers.

It is unclear whether these mutations were enriched during metastatic progression because of a greater ability to disseminate from the primary tumour site or were selected because of resistance to the adjuvant radiotherapy and endocrine therapy. This raises the tantalizing possibility that future adjuvant decision-making may take into account the presence of such rare treatment-resistant clones at the time of diagnosis, similar to the detection of low-abundance drug-resistant mutants in human immunodeficiency virus (HIV) to guide the selection of highly active anti-retroviral therapy (HAART) (Le et al, 2009). The detection of these rare mutations should be conducted in a targeted manner to assure the identification of these scarce variants.

Additional studies will be required to confirm the functional relevance of such low-abundance mutations that may be detected at time of diagnosis. It is interesting to note that the number of dominant mutations was considerably smaller in the primary tumour than the metastatic lesion. The relevance of these ‘early’ and ‘treatment-emergent’ mutations in the initiation and progression of the primary cancer is uncertain. The latter may not confer any selective advantage but be silent ‘passengers’ that do not contribute to cancer progression. These questions highlight the difficulty of making sense of the catalogue of mutations in an individual tumour made possible by whole-genome sequencing but on the other hand reveal a number of interesting issues and research lines that such sequencing approaches now open.

The falling cost of massively parallel sequencing systems will lead to an explosion of similar studies involving other tumour types in the coming years.
The challenge will be to translate the observations from these studies into validated biomarkers that can be used to guide clinical decision-making. Not only the presence but also the frequency at which a particular mutation occurs in a cancer can be estimated and will be relevant to the outcome of a disease. This may require the sequencing of hundreds of cancers with shared biological features, along with early integration of next-generation sequencing platforms into innovative clinical trials to clearly define the added value of personalized genomics to standard clinical cancer care.

References
Lawo S et al (2009) Curr Biol 19: 816-826.
Le T et al (2009) PloS One 4: e6079
Li JB et al (2009) Science 324: 1210-1213
Shah SP et al (2009) Nature 461: 809-813
Sjoblom T et al (2006) Science 314: 268-274
Stephens P et al (2005) Nat Genet 37: 590-592
Stratton MR et al (2009) Nature 458: 719-724