

Short communication

Next-generation sequencing

Jorge S Reis-Filho

Molecular Pathology Team, The Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

Corresponding author: Jorge S Reis-Filho, Jorge.Reis-Filho@icr.ac.uk

Published: 18 December 2009

This article is online at

<http://breast-cancer-research.com/supplements/11/S3/S12>

Breast Cancer Research 2009, **11(Suppl 3)**:S12

(doi:10.1186/bcr2431)

© 2009 BioMed Central Ltd

Abstract

Next-generation sequencing (also known as massively parallel sequencing) technologies are revolutionising our ability to characterise cancers at the genomic, transcriptomic and epigenetic levels. Cataloguing all mutations, copy number aberrations and somatic rearrangements in an entire cancer genome at base pair resolution can now be performed in a matter of weeks. Furthermore, massively parallel sequencing can be used as a means for unbiased transcriptomic analysis of mRNAs, small RNAs and noncoding RNAs, genome-wide methylation assays and high-throughput chromatin immunoprecipitation assays. Here, I discuss the potential impact of this technology on breast cancer research and the challenges that come with this technological breakthrough.

Introduction

Since the publication of the first draft of the human genome sequence [1,2], the field of genomics has changed dramatically. Most importantly, the availability of this information has led to a technological boom, with the development of high-throughput methods that could be used to interrogate the wealth of data available in the human genome and transcriptome. The fields of genomic and transcriptomic science have expanded at an unprecedented pace.

In the past decade we have witnessed the rise of microarrays, a technology that has been extensively applied to the study of cancer genomes and transcriptomes. Of all solid cancers, breast cancer has been the most comprehensively studied using these methods. Although some of the promises of microarrays have not materialised in the time frame some of the proponents of this technology have foreseen, the high-throughput data generated in microarray-based experiments have changed the way breast cancer is perceived [3,4]. The approach has brought to the forefront of cancer research the concepts of breast cancer heterogeneity – that distinct molecular subtypes of breast cancer are underpinned by distinct genetic and epigenetic aberrations, and that distinct subtypes of breast cancer may have their prognosis and response to therapy governed by distinct molecular pathways and networks [5,6]. It should be noted, however, that

microarray-based expression profiling and comparative genomic hybridisation provide data with important limitations. For instance, microarray-based expression profiling only provides a semiquantitative assessment of gene expression; it is limited by the nature of the probes included in the platform and their sensitivity and specificity. Comparative genomic hybridisation and SNP array analysis have provided a wealth of data on gene copy number aberrations in breast cancer and have helped identify potential therapeutic targets for subgroups of breast cancer patients; however, this technology does not provide any information about structural genomic aberrations and base pair mutations [7].

An ideal tool for the genetic characterisation of cancers is one that could provide information about copy number aberrations, allelic information, somatic rearrangements and base pair mutations in a single experiment [7]. Furthermore, data generated with such technology should be presented in such a way that the presence of cells other than cancer cells in the samples would not constitute an insurmountable hurdle. Such a tool, a few years ago, would belong to the realms of science fiction.

Technology, however, has evolved at an unprecedented pace. We are currently witnessing yet another molecular revolution, one that will most certainly dwarf the paradigm shifts brought about by the introduction of microarrays: the advent of massively parallel sequencing (also known as next-generation sequencing). This technology allows for the accrual of qualitative and quantitative information about any type of nucleic acid in a given sample at an incredible throughput while incurring relatively limited costs (reviewed in [8-13]).

What is massively parallel sequencing and why the fuss?

For the past 15 years, Sanger sequencing and fluorescence-based electrophoresis technologies have been extensively used in somatic and germline genetic studies. Improvements

PCR = polymerase chain reaction; SNP = single nucleotide polymorphism.

in instrumentation coupled with the development of high-performance computing and bioinformatics have reduced the cost of sequencing. However, increases in the throughput of Sanger DNA sequencing are achieved by the use of additional sequencers in parallel, owing to the requirement of gel electrophoresis or additional wells for the capillary sequencing of each reaction.

Using different approaches, massively parallel sequencing methods overcome the limited scalability of traditional Sanger sequencing by either creating micro-reactors and/or attaching the DNA molecules to be sequenced to solid surfaces or beads, allowing for millions of sequencing reactions to happen in parallel. At present, there are four technologies commercially available and several other promising approaches are in various stages of development and implementation (Table 1) (reviewed in [8-13]). The current generation of massively parallel sequencers has led to a quantum leap in our ability to sequence genomes, so much so that 10-fold coverage of the human genome (30 Gb DNA sequence) can be obtained in a single run for no more than US\$15,000 to US\$20,000. (Note that the Human Genome Sequencing Consortium generated 3 Gb at the cost of approximately US\$3 billion and took 13 years!)

Perhaps more important than the sequencing throughput provided by this technology and its relative low cost compared with traditional sequencing methods is the type of data it generates. Instead of long reads generated from a PCR-amplified sample, massively parallel sequencing methods provide much shorter reads (~21 to ~400 base pairs), but millions of them [8-13]. Unlike previous sequencing methods that required DNA amplification (that is, the final sequence was representative of modal population of DNA templates), sequencing can now be performed from single DNA molecules. The short reads generated in the sequencing of each DNA molecule can be counted and quantified, allowing the identification of mutations in nonmodal populations of cells (that is, identification of a somatic mutation in a small subpopulation of cells immersed in a modal population with wild-type sequences) and accurate copy number assessment of each genomic region ([14] and references therein). In addition, with the recent introduction of approaches that allow for the sequences of both ends of a DNA molecule (that is, paired end massively parallel sequencing or mate pair sequencing), it has become possible to detect balanced and unbalanced somatic rearrangements (that is, fusion genes) in a genome-wide fashion [12,14,15].

Not surprisingly, this massive increase in throughput has come at a cost, with the accuracy of each short read being significantly lower than the output generated from Sanger sequencing. Although this is circumvented by the depth of sequencing (that is, multiple reads of the same region), it is accepted that physical validation using traditional sequencing

methods is required. Note that each type of next generation sequencing leads to specific types of artefacts (reviewed in [8-13]); however, as we are writing the book on next-generation sequencing as we go along, one should be aware of unexpected artefacts and new findings should be interpreted with caution.

What can be done with massively parallel sequencing?

Next-generation sequencing has already been applied to re-sequencing studies, which have led to sequencing of complete normal and cancer genomes being performed in a matter of weeks [16-18]. Massively parallel sequencing can be employed for the simultaneous characterisation of cancer genomes in terms of somatic base pair and in-del mutations, balanced and unbalanced rearrangements, and copy number changes in a single experiment [14,18]. Apart from sequencing whole genomes, massively parallel sequencing can be coupled with DNA capturing methods for focused analysis of specific genomic regions, specific genes or the whole exome [19]. In fact, the Breast Cancer International Cancer Genome Consortium has pledged to complete sequencing the genome of 1,500 breast cancers [20]. This study will provide a comprehensive catalogue of the genetic alterations found in breast cancer in general and in the different subtypes of the disease.

Massively parallel sequencing can be applied to germline DNA for gene association studies and for the analysis of cancer genomes [8-14], and may constitute a paradigm shift in the way mutations that cause rare diseases can be identified. In fact, the power of this technology to unravel genes whose germline mutations cause rare mendelian disorders is exemplified by the identification of *MYH3* germline mutations as a cause of Freeman–Sheldon syndrome through the targeted sequencing of all protein-coding regions (exomes) of four individuals with this syndrome and eight unrelated individuals [19]. Although in the interpretation of results from target exome and whole genome sequencing studies of a small number of subjects, investigators will have to deal with the previously underestimated number of private SNPs and copy number DNA polymorphisms, the 1000 Genomes Project will provide a more complete catalogue of SNPs, copy number polymorphisms, and short insertion and deletion polymorphisms in the general population [21], which may facilitate the discovery of pathogenic germline mutations.

In addition to the ability to sequence DNA, massively parallel sequencing can be applied to sequencing RNA [22]. Four main applications have already been developed – namely, digital gene expression, RNA sequencing, paired end RNA sequencing, and small and noncoding RNA sequencing. An in-depth discussion of these methods and their impact on our ability to perform transcriptomic analyses is beyond the scope of this short communication, and readers are referred to excellent reviews on this topic [13,22]. Suffice it is to say

Table 1**Summary of massively parallel sequencing technologies**

Method	Amplification	Read length (base pairs)	Templates per run	Data production/day	Sequence reaction	Reference
Commercially available technologies						
ABI 3730xl	PCR	~900 to 1,100	96	1 Mb/day	Sanger method	www.appliedbiosystems.com
454 FLX Roche	Emulsion PCR	~400	1,000,000	400 Mb/run/ 7.5 to 8 hours	Pyrosequencing	www.rocheapplied-science.com
Illumina (Solexa) Genome Analyzer	Bridge PCR	36 to 175	40,000,000	>17 Gb/run/ 3 to 6 days	Reverse terminator	www.illumina.com
ABI SOLiD	Emulsion PCR	~50	85,000,000	10 to 15 Gb/ run/6 days	Ligation sequencing	www.appliedbiosystems.com
Helicos Heliscope	None	30 to 35	800,000,000	21 to 28 Gb/ run/8 days	Single molecule sequence by synthesis	www.helicosbio.com
Technologies in development						
Pacific Biosciences	None	>1,000	NA	NA	Single molecule real-time DNA sequencing	www.pacificbiosciences.com
Intelligent Biosciences	Yes ^a	NA	NA	NA	Sequence by synthesis	www.intelligentbiosystems.com
Visigen Biotechnologies	None	NA	NA	NA	Base-specific FRET emission	www.visigenbio.com
ZS Genetics	None	NA	NA	NA	ZSG atomic labelling and electron microscopy	www.zsgenetics.com

NA, not available at present. ^aAmplification method not yet standardised. FRET, Förster resonance energy transfer.

these approaches have already led to the identification of multiple novel splice variants [23], novel gene rearrangements [24,25] and novel fusion genes [26,27], and to the identification of read-throughs [27], which are RNA molecules resultant from the co-splicing of two genes that are contiguous in the genome in the absence of a structural genomic aberration. When combined with DNA massively parallel sequencing, RNA sequencing has the potential to unravel RNA editing events, such as the nonsynonymous transcript editing of the *COG3* and *SRP9* genes in a metastatic invasive lobular carcinoma [18]. Furthermore, massively parallel sequencing studies of noncoding and small RNAs coupled with the results of the ENCODE project [28] are likely to reveal a level of transcriptional regulation way beyond our current models.

Modifications of the protocols for massively parallel sequencing also allow for an unbiased assessment of DNA methylation [29,30] and histone acetylation, and are likely to replace microarrays in the analysis of high-throughput immunochromatin precipitation assays [13,31]. Next-generation sequencing is also replacing microarrays in high-throughput RNA interference screens: one can perform genomewide screens to identify genes that interfere with the viability of cancer cells using pools of short hairpin RNAs, and the results can be deconvoluted using next-generation sequencing [32]. This latter approach is likely to provide a wealth of information on genes that are selectively required for cancer cell survival and potential drug targets.

Massive parallel sequencing: opportunities and challenges

The multiple applications and uses of massively parallel sequencing are likely to reshape several aspects of breast cancer research. Given the unprecedented ability to identify mutations, copy number aberrations and somatic rearrangements in cancer genomes, the information accrued by massively parallel sequencing of breast cancers may lead to a paradigm shift in the way breast cancers are classified. In fact, this technology offers a unique opportunity to move from the current descriptive and prognostic classification systems to a functional genomic taxonomy that is based on the molecular aberrations that drive specific subgroups of cancers, in a way akin to the classification system currently used for leukaemias and lymphomas. With the availability of information of the genetic alterations required for the survival of cells of a given cancer, tumours may be classified according to the genetic aberrations they harbour, according to the molecular networks activated or inactivated by these genetic aberrations, and, importantly, according to the agents these tumours are sensitive to.

Studies performing large-scale conventional sequencing of breast cancers [33,34] revealed that there are a relatively low number of genes frequently mutated and a high number of genes rarely mutated in breast cancer. It should be noted,

however, that the number of mutations found in oestrogen receptor-negative breast cancer cell lines [34] was higher than that found in an oestrogen receptor-positive breast cancer [18]. It is therefore plausible that different types of breast cancer are driven by distinct constellations of genetic aberrations. It should be noted, however, that even tumours from the same type may be characterised by mutations of distinct genes in the same or complementary molecular networks, which would result in a similar phenotype.

Recent whole-genome characterisation of M1 leukaemias [35,36] and of a metastatic deposit of an invasive lobular carcinoma of the breast [18] has demonstrated the power of this technology for the identification of novel potential mutations that drive specific subtypes of complex and heterogeneous diseases such as leukaemias and breast cancer, and has demonstrated how the mutational spectrum of a cancer evolves over time. Furthermore, next-generation sequencing analysis of cancer types whose tumours are rather homogeneous in terms of their molecular makeup, such as some special types of breast cancer [37-41], may lead to the identification of pathognomonic genomic alterations, in a way akin to C134Y *FOXL2* mutations in granulosa cell tumours of the ovary [42]. These driver genetic alterations (for example, mutations, amplifications and fusion genes) have the potential of being exploited as therapeutic targets.

Although the presence of non-neoplastic tissues (that is, stroma, inflammatory infiltrate and entrapped normal tissues) represents a challenge for the analysis of the genomes of preinvasive lesions, primary breast cancers and their metastatic deposits, there is evidence to suggest that if a tumour is sequenced at a sufficient depth then accurate sequences at base pair resolution can be obtained and somatic mutations identified [18].

Another important application of massively parallel sequencing due to its ability to deep sequence specific genomic regions is the identification of secondary mutations as mechanisms of resistance to specific agents [43,44]. There are several lines of evidence to demonstrate that *de novo* and acquired resistance to some targeted therapies is driven by secondary mutations in the target genes (for example, the T790M mutation in the *EGFR* gene causing resistance to anti-epidermal growth factor receptor agents [45], and secondary *KIT* mutations leading to resistance to imatinib mesylate and sorafenib [46]) or in genes whose inactivation is synthetically lethal in the presence of the targeted therapy (for example, *BRCA2* and *BRCA1* revertant mutations as a mechanism of resistance to platinum salts and poly(ADP-ribose) polymerase inhibitors [47-49]).

It should be noted, however, that the deluge of data derived from next-generation sequencing studies might take a relatively long time to be translated into information that is clinically relevant. Given that each cancer genome may have

an excess of 10,000 somatic mutations, it is unclear how much validation through the identification of recurrent mutations [14] or by laborious functional studies will be required to separate driver mutations (that is, those that either confer growth/survival advantage for a tumour or are required for the cancer cells for the maintenance of their malignant behaviour) from passenger alterations (that is, genomic noise). Furthermore, next-generation sequencing is likely to unravel a much greater complexity of the normal human genome in terms of SNPs and copy number polymorphisms [50], some of which may be confined to some somatic tissues in the same individual [51,52]. Massively parallel sequencing will require an availability of high-performance computing and bioinformatic support that is way beyond that of most research laboratories. Furthermore, quality control and standardisation of the massively parallel sequencing experiments and data reporting are important issues to consider. Finally, the ethical aspects of next-generation sequencing are by no means trivial, and the readers are referred to excellent reviews covering these aspects [9,11].

Conclusion

One could argue that massively parallel sequencing is not only an end, but also a means for performing experiments that may answer questions that could not even be asked previously. The revolution that is likely to be brought about by massively parallel sequencing methods is akin to the revolution fostered by the introduction of the PCR in the 1980s. It is undeniable that this technology will constitute a quantum leap in breast cancer basic and translational research; however, numerous challenges lie ahead. We ought to learn from our recent experience with microarrays, and avoid any sort of unjustified overoptimism. The greatest danger of using this revolutionary technology is that it comes with new problems; if we move too quickly, the lessons we are beginning to learn from previous high-throughput studies may be forgotten when massively parallel sequencing is applied to clinical and translational questions.

Competing interests

The author declares that they have no competing interests.

Acknowledgments

The author is grateful to the Molecular Pathology Team and Dr Britta Weigelt for critically reading this manuscript. JSR-F is funded in part by Breakthrough Breast Cancer. NHS funding to the NIHR Biomedical Research Centre is also acknowledged.

This article has been published as part of *Breast Cancer Research* Volume 11 Suppl 3 2009: Controversies in Breast Cancer 2009. The full contents of the supplement are available online at <http://breast-cancer-research.com/supplements/11/S3>.

References

- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrum J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, *et al.*: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351.
- Sotiriou C, Pusztai L: **Gene-expression signatures in breast cancer.** *N Engl J Med* 2009, **360**:790-800.
- Correa Geyer F, Reis-Filho JS: **Microarray-based gene expression profiling as a clinical tool for breast cancer management: are we there yet?** *Int J Surg Pathol* 2009, **17**:285-302.
- Geyer FC, Lopez-Garcia MA, Lambros MB, Reis-Filho JS: **Genetic characterisation of breast cancer and implications for clinical management.** *J Cell Mol Med* 2009, Epub ahead of print; doi:10.1111/j.1582-4934.2009.00906.x.
- Pusztai L: **Current status of prognostic profiling in breast cancer.** *Oncologist* 2008, **13**:350-360.
- Tan DS, Lambros MB, Natrajan R, Reis-Filho JS: **Getting it right: designing microarray (and not 'microarray') comparative genomic hybridization studies for cancer research.** *Lab Invest* 2007, **87**:737-754.
- Pettersson E, Lundeberg J, Ahmadian A: **Generations of sequencing technologies.** *Genomics* 2009, **93**:105-111.
- ten Bosch JR, Grody WW: **Keeping up with the next generation: massively parallel sequencing in clinical diagnostics.** *J Mol Diagn* 2008, **10**:484-492.
- Voelkerding KV, Dames SA, Durtschi JD: **Next-generation sequencing: from basic research to diagnostics.** *Clin Chem* 2009, **55**:641-658.
- Tucker T, Marra M, Friedman JM: **Massively parallel sequencing: the next big thing in genetic medicine.** *Am J Hum Genet* 2009, **85**:142-154.
- Fullwood MJ, Wei CL, Liu ET, Ruan Y: **Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses.** *Genome Res* 2009, **19**:521-532.
- Morozova O, Marra MA: **Applications of next-generation sequencing technologies in functional genomics.** *Genomics* 2008, **92**:255-264.
- Stratton MR, Campbell PJ, Futreal PA: **The cancer genome.** *Nature* 2009, **458**:719-724.
- Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, Teague JW, Menzies A, Goodhead I, Turner DJ, Clee CM, Quail MA, Cox A, Brown C, Durbin R, Hurles ME, Edwards PA, Bignell GR, Stratton MR, Futreal PA: **Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing.** *Nat Genet* 2008, **40**:722-729.
- Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, *et al.*: **The diploid genome sequence of an Asian individual.** *Nature* 2008, **456**:60-65.
- Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niaz F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM: **The complete genome of an individual by massively parallel DNA sequencing.** *Nature* 2008, **452**:872-876.
- Shah SP, Morin RD, Khattri J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S: **Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution.** *Nature* 2009, **461**:809-813.
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J: **Targeted capture and massively parallel sequencing of 12 human exomes.** *Nature* 2009, **461**:272-276.
- Breast Cancer International Cancer Genome Consortium** [<http://www.icgc.org/>]
- 1000 Genomes Project** [<http://www.1000genomes.org/>]
- Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool**

- for transcriptomics. *Nat Rev Genet* 2009, **10**:57-63.
23. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keefe S, Haas S, Vingron M, Lehrach H, Yaspo ML: **A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome.** *Science* 2008, **321**: 956-960.
 24. Zhao Q, Caballero OL, Levy S, Stevenson BJ, Iseli C, de Souza SJ, Galante PA, Busam D, Leversha MA, Chadalavada K, Rogers YH, Venter JC, Simpson AJ, Strausberg RL: **Transcriptome-guided characterization of genomic rearrangements in a breast cancer cell line.** *Proc Natl Acad Sci U S A* 2009, **106**: 1886-1891.
 25. Hampton OA, Den Hollander P, Miller CA, Delgado DA, Li J, Coarfa C, Harris RA, Richards S, Scherer SE, Muzny DM, Gibbs RA, Lee AV, Milosavljevic A: **A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome.** *Genome Res* 2009, **19**:167-177.
 26. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM: **Transcriptome sequencing to detect gene fusions in cancer.** *Nature* 2009, **458**:97-101.
 27. Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtkova I, Barrette TR, Grasso C, Yu J, Lonigro RJ, Schroth G, Kumar-Sinha C, Chinnaiyan AM: **Chimeric transcript discovery by paired-end transcriptome sequencing.** *Proc Natl Acad Sci U S A* 2009, **106**:12353-12358.
 28. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, *et al.*: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447**:799-816.
 29. Lister R, Ecker JR: **Finding the fifth base: genome-wide sequencing of cytosine methylation.** *Genome Res* 2009, **19**: 959-966.
 30. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES: **Genome-scale DNA methylation maps of pluripotent and differentiated cells.** *Nature* 2008, **454**:766-770.
 31. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA: **ChIP-seq accurately predicts tissue-specific activity of enhancers.** *Nature* 2009, **457**:854-858.
 32. Iorns E, Lord CJ, Turner N, Ashworth A: **Utilizing RNA interference to enhance cancer drug discovery.** *Nat Rev Drug Discov* 2007, **6**:556-568.
 33. Greenman C, Stephens P, Smith R, Dalgleish GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens S, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, *et al.*: **Patterns of somatic mutation in human cancer genomes.** *Nature* 2007, **446**:153-158.
 34. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, *et al.*: **The genomic landscapes of human breast and colorectal cancers.** *Science* 2007, **318**:1108-1113.
 35. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, *et al.*: **Recurring mutations found by sequencing an acute myeloid leukemia genome.** *N Engl J Med* 2009, **361**:1058-1066.
 36. Walter MJ, Payton JE, Ries RE, Shannon WD, Deshmukh H, Zhao Y, Baty J, Heath S, Westervelt P, Watson MA, Tomasson MH, Nagarajan R, O'Gara BP, Bloomfield CD, Mrozek K, Selzer RR, Richmond TA, Kitzman J, Geoghegan J, Eis PS, Maupin R, Fulton RS, McLellan M, Wilson RK, Mardis ER, Link DC, Graubert TA, DiPersio JF, Ley TJ: **Acquired copy number alterations in adult acute myeloid leukemia genomes.** *Proc Natl Acad Sci U S A* 2009, **106**:12950-12955.
 37. Reis-Filho JS, Lakhani SR: **Breast cancer special types: why bother?** *J Pathol* 2008, **216**:394-398.
 38. Weigelt B, Horlings HM, Kreike B, Hayes MM, Hauptmann M, Wessels LF, de Jong D, Van de Vijver MJ, Van't Veer LJ, Peterse JL: **Refinement of breast cancer classification by molecular characterization of histological special types.** *J Pathol* 2008, **216**:141-150.
 39. Weigelt B, Geyer FC, Horlings HM, Kreike B, Halfwerk H, Reis-Filho JS: **Mucinous and neuroendocrine breast carcinomas are transcriptionally distinct from invasive ductal carcinomas of no special type.** *Mod Pathol* 2009, **22**:1401-1414.
 40. Marchio C, Iravani M, Natrajan R, Lambros MB, Geyer FC, Savage K, Parry S, Tamber N, Fenwick K, Mackay A, Schmitt FC, Busso-lati G, Ellis I, Ashworth A, Sapino A, Reis-Filho JS: **Mixed micropapillary-ductal carcinomas of the breast: a genomic and immunohistochemical analysis of morphologically distinct components.** *J Pathol* 2009, **218**:301-315.
 41. Weigelt B, Reis-Filho JS: **Histological and molecular types of breast cancer: is there a unifying taxonomy?** *Nat Rev Clin Oncol* 2009, Epub ahead of print; doi:10.1038/nrclinonc.2009.1166.
 42. Shah SP, Köbel M, Senz J, Morin RD, Clarke BA, Wiegand KC, Leung G, Zayed A, Mehl E, Kalloger SE, Sun M, Giuliany R, Yorlida E, Jones S, Varhol R, Swenerton KD, Miller D, Clement PB, Crane C, Madore J, Provencher D, Leung P, DeFazio A, Khattra J, Turashvili G, Zhao Y, Zeng T, Glover JN, Vanderhyden B, Zhao C, *et al.*: **Mutation of FOXL2 in granulosa-cell tumors of the ovary.** *N Engl J Med* 2009, **360**:2719-2729.
 43. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, Downing JR: **Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia.** *Science* 2008, **322**: 1377-1380.
 44. Ashworth A: **Drug resistance caused by reversion mutation.** *Cancer Res* 2008, **68**:10021-10023.
 45. Godin-Heymann N, Ulkus L, Brannigan BW, McDermott U, Lamb J, Maheswaran S, Settleman J, Haber DA: **The T790M 'gate-keeper' mutation in EGFR mediates resistance to low concentrations of an irreversible EGFR inhibitor.** *Mol Cancer Ther* 2008, **7**:874-879.
 46. Heinrich MC, Maki RG, Corless CL, Antonescu CR, Harlow A, Griffith D, Town A, McKinley A, Ou WB, Fletcher JA, Fletcher CD, Huang X, Cohen DP, Baum CM, Demetri GD: **Primary and secondary kinase genotypes correlate with the biological and clinical activity of sunitinib in imatinib-resistant gastrointestinal stromal tumor.** *J Clin Oncol* 2008, **26**:5352-5359.
 47. Edwards SL, Brough R, Lord CJ, Natrajan R, Vatcheva R, Levine DA, Boyd J, Reis-Filho JS, Ashworth A: **Resistance to therapy caused by intragenic deletion in BRCA2.** *Nature* 2008, **451**: 1111-1115.
 48. Sakai W, Swisher EM, Karlan BY, Agarwal MK, Higgins J, Friedman C, Villegas E, Jacquemont C, Farrugia DJ, Couch FJ, Urban N, Taniguchi T: **Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers.** *Nature* 2008, **451**:1116-1120.
 49. Swisher EM, Sakai W, Karlan BY, Wurz K, Urban N, Taniguchi T: **Secondary BRCA1 mutations in BRCA1-mutated ovarian carcinomas with platinum resistance.** *Cancer Res* 2008, **68**:2581-2586.
 50. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE: **Personalized copy number and segmental duplication maps using next-generation sequencing.** *Nat Genet* 2009, **41**:1061-1067.
 51. Piotrowski A, Bruder CE, Andersson R, de Stahl TD, Menzel U, Sandgren J, Poplawski A, von Tell D, Crasto C, Bogdan A, Bartoszewski R, Bebek Z, Krzyzanowski M, Jankowski Z, Partridge EC, Komorowski J, Dumanski JP: **Somatic mosaicism for copy number variation in differentiated human tissues.** *Hum Mutat* 2008, **29**:1118-1124.
 52. Bruder CE, Piotrowski A, Gijbsbers AA, Andersson R, Erickson S, de Stahl TD, Menzel U, Sandgren J, von Tell D, Poplawski A, Crowley M, Crasto C, Partridge EC, Tiwari H, Allison DB,

Komorowski J, van Ommen GJ, Boomsma DI, Pedersen NL, den Dunnen JT, Wirdefeldt K, Dumanski JP: **Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles.** *Am J Hum Genet* 2008, **82**: 763-771.