Breast cancer is a heterogeneous disease that comes in several clinical and histological forms. Its clinical progression is difficult to predict using the current prognostic factors and its treatment is therefore not as effective as it should be. Mortality due to breast cancer is decreasing in most western countries, because of mass screening, frequent use of post-operative chemotherapy and/or hormone therapy and the recent introduction of new drugs. However, novel drugs and therapeutic strategies could be more successful if we understood breast cancer heterogeneity better. Two recent papers in *Genome Biology* from the laboratories of Carlos Caldas [1] and Eric Miska [2] use molecular methods to classify breast cancers more precisely.

The first direction of research aims to determine what cell becomes transformed; in other words, the cell of origin of a breast tumor. In the mammary gland, mammary stem cells, which can self-renew and differentiate, generate rapidly dividing progenitors that in turn generate differentiated cells of the mammary gland epithelial lineages: the luminal and myoepithelial lineages. Cancer is thought to originate in these stem cells or in progenitor cells that have acquired self-renewal. Thus, a first degree of heterogeneity comes from whether a tumor comes from a stem cell or a progenitor cell.

The second direction aims to determine what genetic alterations transform a normal breast cell and make it cancerous. The repertoire of genetic alterations can be found by using high-throughput, large-scale methods, such as mass sequencing [3,4] and array comparative genomic hybridization (aCGH) [5,6]. These have revealed a number of alterations - mutations, deletions, amplifications and fusions - that target hundreds of genes, suggesting a high level of heterogeneity. Some tumors can have a high level of genetic instability whereas others can have an apparently normal genome.
The third direction aims to identify breast tumor susceptibility genes. In addition to the BRCA genes, in which mutations confer a high risk of susceptibility to breast cancer, a number of low-risk variants have been recently identified by genome-wide association studies [7,8]. These low-risk susceptibility genes might also introduce some level of heterogeneity, which remains to be evaluated. Susceptibility genes (in the germline) differ from genes changed in the tumor (somatic changes).

The fourth direction aims to classify breast tumors and establish whether all members of a subtype have the same properties. Recently developed high-throughput molecular analyses have provided unprecedented tools for dissecting and understanding cancer heterogeneity. Five subtypes of breast cancer were initially proposed: luminal A and luminal B (both estrogen receptor (ER)-positive); basal (ER-negative); ERBB2 (erythroblastic leukemia viral oncogene homolog 2)-overexpressing; and normal-like [9,10]. This early classification has been useful and has been validated in many further studies, but several issues remain to be clarified. It is not known how the subtypes relate to the cell of origin, how to classify the many samples (about 10-15%) that could not be assigned any subtype, how homogeneous the different subtypes are, and what the molecular alterations specific to each subtype are. Furthermore, it has been suggested that subtypes of breast tumors are part of a continuum [11]. A recent study has shown that genes associated with susceptibility variants are differentially expressed in the major subtypes [12], and this opens up interesting perspectives. The two recent papers in Genome Biology (Chin et al. [1] and Blenkiron et al. [2]) take this further.

**Tumor subtypes**

Chin et al. [1] studied a series of 171 breast tumors using genome-wide, high-resolution aCGH combined with gene expression analysis by DNA microarrays. This is the largest integrated genomic study of breast cancer reported so far. They determined the patterns of gains and losses of the tumor genomes, explored the taxonomy of tumors using gene copy numbers and established lists of genes potentially altered by deletions, copy-number gains and amplifications.

Interestingly, using hierarchical clustering over the common regions of gene alterations they found a subgroup of tumors (about 15% of them) that showed few or no genomic alterations. Basal ER-negative tumors are generally thought to be of high pathological grade (that is, with cells that are highly abnormal in morphology) and genetically unstable [13]. Strikingly, this novel subgroup with low genetic instability included more of the basal and ER-negative high-grade tumor subtypes than other subtypes.

Further characterization showed that the subgroup was associated with specific gene expression, such as increased expression of inflammatory and defense response genes. Survival in this subgroup was not different from the rest of the samples, even when the analysis was restricted to ER-negative tumors. However, the limited size of the series warrants further statistical analyses on a larger number of samples that have been treated homogeneously. The difference in genomic instability was not associated with the presence or absence of a mutation in the tumor suppressor gene TP53, whose alteration is generally associated with genome instability.

The identification of this subgroup of basal breast cancers was made possible in the Chin et al. [1] study by the presence in their panel of tumors of smaller size than those studied by previous studies. This makes the tumors more representative of tumors currently diagnosed. Their study shows that the ER-negative, high-grade basal subtype can be further subdivided in two subclasses of low and high genetic instability, paving the way to further definition of subgroups of cases with similar features within the subtypes.

Breast cancer may show a continuum of features from high proliferation to high differentiation with a few recognizable stages (that is, the subtypes) associated with specific sets of transcribed genes (Figure 1). The Chin et al. study [1] shows that tumors with the same phenotype and the same transcriptional content (ER-negative, basal) can result from different sets of genomic alterations.

Tumors bearing these different alterations are likely not to respond to the same treatment. Thus, determination of phenotype alone may not be enough for therapeutic selection but knowledge of both genotype and phenotype is required. Furthermore, Chin et al. [1] provide lists of copy number alterations and potential cancer genes in breast cancers from which therapeutic targets may be drawn.

**Tumor heterogeneity and microRNAs**

The work by Blenkiron et al. [2] has addressed another very important question: could tumor heterogeneity be sustained, at least in part, by some particular distribution of microRNAs (miRNAs)? In humans, miRNAs are approximately 22-nucleotide-long single-stranded RNAs that have a key role in the post-transcriptional control of up to 30% of protein-coding genes and regulate many cellular processes during development and adult life. MicroRNAs regulate various cell processes, including self-renewal and tumorigenicity in breast cancer cells [14] and also invasion and metastasis [15,16]. They are thought to be important in oncogenesis (see [17,18] for reviews).
Blenkiron et al. [2] analyzed the expression of 309 miRNAs in 93 breast tumor samples using a bead-based flow-cytometric profiling method. To our knowledge, this is the first integrated analysis of miRNA expression, mRNA expression and genomic changes in breast cancer. They showed that many miRNAs have a variable expression across breast tumor samples. As found at the mRNA level, global hierarchical clustering based on miRNA expression levels separated ER-positive from ER-negative tumors relatively well. Moreover, miRNAs were differentially distributed among the five molecular subtypes. The authors identified a miRNA signature that perfectly discriminated between basal and luminal subtypes in their samples as well as in a small independent validation dataset. These results [2] can be compared with recent in situ hybridization data [19] that showed specific expression of some miRNAs in certain mammary epithelial cells and variations in their expression in tumor cells. Expression of some miRNAs correlated with the molecular subtypes and with two major features of breast cancer (grade and ER status) [2]. The authors [2] used an integrated approach to analyze the reasons for the differential expression of miRNAs. By combining these data with aCGH and mRNA expression data, they found that differences in miRNA expression could be explained by combinations of genomic alterations, transcriptional and post-transcriptional regulation and changes in miRNA biosynthesis. By regulating key target genes, miRNAs may contribute to the genetic determination of breast tumor subtypes.

It had already been demonstrated that miRNA profiles can classify tumors of various origins [20] or various clinical outcomes [21]. This suggests that miRNA expression signatures could be used in clinics as diagnostic and prognostic tools. As pointed out by the authors [2], an advantage of miRNAs over mRNAs is that their short size and stability allows their easy detection in paraffin-embedded tumor samples. Determination of miRNA expression in otherwise characterized breast tumor samples is therefore an important step in understanding the role and potential use of miRNAs as disease classifiers, prognostic markers or therapeutic targets. An exciting possibility is that miRNAs could be involved in stem cell biology and the induction and/or maturation of mammary epithelial cell lineages and could thus participate in the control of luminal and/or myoepithelial differentiation. Testing this will require further studies involving modulation of miRNA expression in cell and animal models.

These two studies [1,2] are good examples of the current attempts of the scientific community to understand breast cancer and its heterogeneity better. This understanding will be achieved by integrating clinical and histological definitions with cellular and molecular definitions. They also show that, at the molecular level alone, complexity is important and needs a complex investigation using integrated analyses. Many factors influence epithelial cell fate and behavior and their interrelations must be delineated.

Figure 1
A schematic summary of breast cancer heterogeneity. According to the cancer stem cell hypothesis, breast cancer is driven by a limited number of cancer-initiating cells. The progeny of these cells can either progress along the differentiation pathway or remain blocked in a proliferation state. Molecular subtypes, such as basal and luminal, differ by their degree of proliferation and differentiation (as shown by the red and blue wedges). They represent stages that can be recognized along a continuum (shown by the dashed line) from progenitor-like proliferative tumors (basal subtype) to luminal differentiated tumors. Breast cancer heterogeneity is due both to different cells of origin and to alterations in the genome and epigenome. The demonstration by the Miska and Caldas groups of a different distribution of miRNAs among subtypes [2] and of variations in genome instability and heterogeneity of estrogen receptor (ER)-negative tumors [1] (arrows) contribute to the understanding and classification of breast cancers.
Important questions that we will have to solve are: how much of subtype heterogeneity reflects a different cell of origin? And how much of it is controlled by microRNAs, gene alterations or stromal interactions? Only when we are able to tell which tumor derives from which targeted by which molecular alteration will we be able to treat all breast cancers effectively.

References
1. Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Guha N, Park PJ, Costa JL, Pinder SE, van de Wiel MA, Green AR, Ellis IO, Porter PL, Tavarah S, Brenton JD, Ystrø B, Caldas C. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 2007, 8:R214.

2. Blenkiron C, Goldstein LD, Thorne NP, Sørlie T, Chanock SJ, Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Kristensen VN: Genomic landscape of human breast cancer identifies new markers of tumor subtype. Genome Biol 2007, 8:R214.

3. Greenman C, Stephens P, Smith R, Daigleish GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, 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, Jenkins A, Jones D, et al.: Patterns of somatic mutation in human breast cancer genomes. Nature 2007, 446:153-158.

4. Wood LD, Parsons DW, Jones S, Lin J, Jóblovil T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Desso Z, Ustyanskiy V, Nikolksaya T, Nikolksy Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kamriner JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Ustyanksy

5. Hunter DJ, Ross DT, Johnston H, Akslen LA, Fluge Ø, Perergamenshikov A, Williams C, Zhu SX, Lønning PE, Bertussen-Dale AL, Brown PO, Botstein D: Molecular portraits of human breast tumours. Nature 2000, 406:747-752.

6. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale A-L: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001, 98:10869-10874.

7. Finetti P, Cervera N, Charafe-Jauffret E, Chabannon C, Charpin C, Chaffanet M, Jacquemier J, Viens P, Birnbaum D, Bertucci F: Sixteen-kilobase gene expression identifies luminal breast cancers with poor prognosis. Cancer Res 2008, 68:767-776.

8. Nordgard SH, Johansen FE, Almès GI, Naume B, Børresen-Dale AL, Kristensen VN: Genomic landscapes of breast cancer subtypes. Cancer Res 2007, 67:113-111.

9. Cleator S, Heller W, Coombes RC: Triple-negative breast cancer: therapeutic options. Lancet Oncol 2007, 8:235-244.

10. Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, Huang Y, Hu X, Su F, Lieberman J, Song E: let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell 2007, 131:1109-1123.

11. Ma L, Teruya-Feldstein J, Weinberg RA: Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature 2007, 449:682-688.

12. Tavazoie SF, Alcaron C, Oskarsson T, Padua D, Wang Q, Bos PD, Gerald WL, Massague J: Endogenous human microRNAs that suppress breast cancer metastasis. Nature 2008, 451:147-152.

13. Lun L, Teruya-Feldstein J, Weinberg RA: Molecular mechanisms of breast and colorectal cancer cell invasion and metastasis. Science 2007, 318:1108-1113.