

Viewpoint

The breast cancer somatic 'muta-ome': tackling the complexity

Andrew E Teschendorff¹ and Carlos Caldas^{2,3}

¹Medical Genomics Group, Paul O'Gorman Building, UCL Cancer Institute, University College London, 72 Huntley Street, London, WC1E 6BT, UK

²Breast Cancer Functional Genomics Laboratory, Cancer Research UK Cambridge Research Institute and Department of Oncology University of Cambridge, Li Ka-Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK

³Cambridge Breast Unit and NIHR Biomedical Research Centre, Addenbrookes Hospital, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge, CB2 2QQ, UK

Corresponding author: Carlos Caldas, carlos.caldas@cancer.org.uk

Published: 30 March 2009

This article is online at <http://breast-cancer-research.com/content/11/2/301>
© 2009 BioMed Central Ltd

Breast Cancer Research 2009, **11**:301 (doi:10.1186/bcr2236)

Abstract

Acquired somatic mutations are responsible for approximately 90% of breast tumours. However, only one somatic aberration, amplification of the HER2 locus, is currently used to define a clinical subtype, one that accounts for approximately 10% to 15% of breast tumours. In recent years, a number of mutational profiling studies have attempted to further identify clinically relevant mutations. While these studies have confirmed the oncogenic or tumour suppressor role of many known suspects, they have exposed complexity as a main feature of the breast cancer mutational landscape (the 'muta-ome'). The two defining features of this complexity are (a) a surprising richness of low-frequency mutants contrasting with the relative rarity of high-frequency events and (b) the relatively large number of somatic genomic aberrations (approximately 20 to 50) driving an average tumour. Structural features of this complex landscape have begun to emerge from follow-up studies that have tackled the complexity by integrating the spectrum of genomic mutations with a variety of complementary biological knowledge databases. Among these structural features are the growing links between somatic gene disruptions and those conferring breast cancer risk, mutually exclusive coexistence and synergistic mutational patterns, and a clearly non-random distribution of mutations implicating specific molecular pathways in breast tumour initiation and progression. Recognising that a shift from a gene-centric to a pathway-centric approach is necessary, we envisage that further progress in identifying clinically relevant genomic aberration patterns and associated breast cancer subtypes will require not only multi-dimensional integrative analyses that combine mutational and functional profiles, but also larger profiling studies that use second- and third-generation sequencing technologies in order to fill out the important gaps in the current mutational landscape.

The copy-number 'muta-ome'

The most prominent feature in the breast cancer copy-number muta-ome is amplification of the HER2 locus, present in about 10% to 15% of all breast tumours. It is remarkable that since the discovery of this amplification no further *ERBB2*-like oncogene has been conclusively identified.

Although two recent large-scale (145 and 171 tumours) genome-wide profiling studies combining high-resolution copy-number and matched gene expression data have confirmed candidate oncogenes in well-known regions of recurrent amplification (notably, 8p12, 8q24, 11q13-14, 17q21-24, and 20q13), none of these appears to be as frequently amplified as *ERBB2* and they rarely exhibit amplification profiles that clearly point at a specific genomic location or target [1,2]. Instead, the amplification profiles are complex and multi-modal, suggesting that multiple targets may coexist within these regions. This identification problem is compounded by the fact that a relatively high proportion of variation at the gene expression level (approximately 20%) is driven by copy-number changes; thus, focusing on regions of expression bias that are driven by underlying amplifications (so-called 'hotspots') still leaves an unmanageably large number of targets. Nevertheless, by focusing within these hotspots on genes that are also druggable, Chin and colleagues [1] prioritised a smaller set of eight targets, *FGFR1*, *IKBKB*, *PROSC*, *ADAM9*, *FNTA*, *ACACA*, *PNMT*, and *NR1D1*, including *ERBB2*. Confirming the robustness of these findings, all of these were also found to reside in amplification hotspots in an independent breast cancer cohort [2] (A.E. Teschendorff and C. Caldas, unpublished data). In spite of this agreement, the two studies were discordant when hotspots were associated with clinical outcome, mirroring the disagreements of initial gene expression studies. Thus, whereas in [1] associations with survival and recurrence were restricted to the amplicons on 8p11-12 and 17q11-12, in [2] outcome-associated amplicons were found on 8q22.3, 8q24.3, 8q24.11-13, and 11q14. As explained in [2], this disagreement is most likely due to substantial differences in the clinical characteristics of the two cohorts, yet the lessons learned from mRNA expression microarray studies also suggest that larger studies

ER = estrogen receptor; GI = genomic instability; HD = homozygous deletion; SNP = single-nucleotide polymorphism.

may help to resolve such discrepancies. In fact, given the increased complexity of genomic breast cancer profiles relative to mRNA profiles, much larger sample sizes (approximately 500 to 1,000) might be needed before substantial overlap between prognostic copy-number signatures is found. In addition, given that among copy-number aberrations it is high-level amplifications that seem to carry most of the prognostic significance [1,2], the development of a future prognostic copy-number signature, similar to the expression signatures currently being tested in clinical trials, would benefit from higher-resolution studies such as those using single-nucleotide polymorphism (SNP) arrays or next-generation sequencing technologies, as these have the capability to detect more focal (that is, less than 20 to 200 Kb) amplifications and homozygous deletions (HDs). Indeed, a recent high-resolution 317,503 to 555,351 SNP array study of 45 breast tumours was able to detect a larger repertoire of copy-number amplifications and HDs, some smaller than 250 Kb [3]. Thus, in addition to identifying focal aberrations encompassing known oncogenes (*CCND1*, *CCNE1*, and *FGFR2*) and tumour suppressors (*CDKN2A* and *PTEN*), they uncovered a number of other important genes with likely oncogenic or tumour suppressor roles (*PCDH8*, *MRE11A*, and *HoxA3*) [3]. The higher-resolution SNP array also allowed Leary and colleagues [3] to estimate an average number of 18 copy-number aberrations (7 HDs and 11 high-level amplifications) that were altered significantly above the background mutation rate, implicating about 24 genes that are driving tumour progression.

Larger-scale profiling studies are also needed in order to more fully characterise the genomic landscape of breast cancer. So far, four main genomic subtypes that roughly correlate with the well-known intrinsic subtypes defined from gene expression studies have been identified [1,2]. Specifically, one subtype (called 'simple') consisted almost entirely of estrogen receptor-positive (ER⁺)/luminal-A tumours and is characterised mainly by a relatively simple genomic profile defined by a gain of 1q and 16p and a simultaneous loss of 16q. Another genomic subtype ('simple amplifier') was characterised mainly, but not exclusively, by amplifications on 11q13-14 and 17q11-13 and mapped roughly to the ER⁺/luminal-A and ER⁺/HER2⁺ expression subtypes, respectively. Thus, amplifications on 11q13-14 and 17q11-13 occurred in a largely mutually exclusive fashion, suggesting that these tumours are using distinct oncogenic mechanisms. A third genomic subtype ('complex amplifier') showed the highest degree of genomic instability (GI) and complex rearrangements, including frequent amplifications at the 8q24 and 8p12 loci. This genomic subtype is that of worst prognosis, consisting mainly of ER⁻/basal and ER⁺/luminal-B tumours. A fourth genomic subtype consisted mainly of ER⁻/basal tumours but was characterised, surprisingly, by a low GI profile ('flat') [2]. It is doubtful, however, whether such classifications will have much relevance as it is likely that the clinically relevant phenotypes are determined by specific

combinations of different types of somatic mutations. Thus, to identify breast cancer genomic subtypes that are more relevant, we envisage that it will be necessary to derive and analyse multi-dimensional mutational profiles.

The point 'muta-ome'

Two recent landmark sequencing studies [4,5] have also revealed the remarkable complexity of the point mutational landscape of breast cancer. In addition to the high-frequency somatic point mutations in *TP53* (53%), *PIK3CA* (26%), and *CDH1* (21%) (the so-called gene 'mountains'), these studies have confirmed a surprisingly larger number of other genes that also appear to be more frequently mutated than what can be accounted for by chance, albeit at much lower frequencies than *TP53* or *PIK3CA*. Specifically, for the 11 breast tumours considered in the discovery screen of [5], a total of 1,137 RefSeq genes contained a somatic mutation and 167 of these were also validated in an independent screen of 24 samples. When a number of different estimation procedures for passenger mutation rates were used, it was found that approximately 120 genes were more frequently mutated than the passenger rate, suggesting that these genes are more likely to carry driver mutations (so-called CAN genes). This translates into an average of 101 non-synonymous somatic mutations in RefSeq genes per breast tumour, and approximately 14 of these are thought to be located in CAN genes [5]. It could be argued that this is a gross overestimate due to the highly tumorigenic nature of the breast samples considered (all were ER⁻ cell lines mostly representing metastases); however, it is also plausible that it would increase as larger-scale sequencing studies get completed. Combining the published estimate with the average number of 24 driver amplifications/HDs affecting any given breast tumour and assuming mutual exclusivity of mutation type, this would implicate an average of approximately 40 CAN genes in any given tumour.

Given that estimates of passenger mutation rates are only approximate and that a high mutation rate may reflect mechanisms that are only indirectly related to tumour genesis or progression, it could be expected that many of the identified CAN genes are false-positives. Nevertheless, several of the CAN genes identified in [4,5] (notably, *IKBKB*, *IKBKA*, *CHD5*, *STK11*, *STK6*, and *BRAF*) have been independently implicated in breast or other tumours, whereas other genes (for example, *ATM* and *FGFR2*) have been associated with germline mutations and an increased risk of breast cancer. Moreover, in an independent bioinformatics study, several breast cancer CAN genes (*GAB1*, *NLE1*, and *CNTN6*) were rediscovered as members of protein interaction subnetworks with combined expression levels that correlated with clinical outcome [6], consistent with the idea that mutations in these genes are disrupting important signalling pathways and thus fuelling malignant progression. Taken together, these observations provide strong support for the direct causal role of the identified CAN genes in tumour progression and, for some, probably tumorigenesis also.

Massively parallel sequencing: a fusion 'muta-ome'?

The emerging pattern of a few highly prevalent CAN genes ('mountains') with a much larger number of low-frequency events ('hills') parallels the aberration landscape pattern observed at the copy-number level. Thus, a question posed by these studies is whether the relative scarcity of cancer gene 'mountains' is a genuine feature of the cancer muta-ome or whether it merely reflects the inability of the technology to detect further high-frequency mutations. While the latter scenario seems to be highly unlikely in the case of point mutations [5], more focal (≈ 20 to 50 kB) amplifications and deletions will be discovered using higher-resolution SNP arrays and massively parallel sequencing, as already demonstrated with SNP arrays by Leary and colleagues [3], yet it also seems unlikely that these will constitute novel gene 'mountains'. Another possibility is that the missing high-frequency events involve other types of genomic rearrangements such as balanced translocations or inversions, which are not detectable using copy-number analysis but which may be discovered using massively parallel paired-end sequencing or transcriptome sequencing, as illustrated recently in lung cancer [7], prostate cancer [8], and breast cancer [9,10] cell lines. Using a combined strategy of short- and long-read transcriptomic sequencing of a metastatic prostate cancer cell line, Maher and colleagues [8] identified a variety of mechanistically different and novel gene fusions, including a novel read-through gene fusion chimera *SLC45A3-ELK4*, which interestingly was also found to be aberrantly expressed in 7 out of 20 (35%) metastatic prostate cancer cell lines, 6 of which were negative for ETS gene fusions. Similarly, long-read transcriptomic sequencing of a breast cancer cell line (HCC1954) identified not fewer than seven chimaeric transcripts arising mostly from intragenic to intergenic gene fusion events leading to protein truncation. Notably, among the seven rearrangements, one, t(4;11)(q32;q21), causes truncation of the DNA-binding domain of *MRE11A*, a candidate tumour suppressor involved in DNA damage repair, whereas two other fusions, t(5;8)(q35.3;q24.21) and t(5;8)(p15.33;q24.21), implicate the known cancer genes *NSD1* and *PVT1* and the 8q24.21 breast cancer susceptibility region. Although these findings are very encouraging, large-scale sequencing studies will be needed to evaluate the prevalence of these genomic fusion aberrations in breast cancer. If these studies were to confirm that specific inter- and intra-chromosomal gene fusions are highly frequent events in breast cancer (as already demonstrated in prostate cancer), it would vindicate the long-held belief that this type of rearrangement constitutes the most prevalent category of somatic aberrations in all tumours.

Tackling the complexity: structural features of the 'muta-ome'

The broad picture emerging from the large-scale low-resolution array-based comparative genomic hybridisation studies [1,2] and the smaller-scale higher-resolution sequen-

cing studies [5,9,10] is that of a highly complex genomic landscape, not only in terms of the sheer numbers of genes that may be directly implicated, but also in terms of the different types and flavours of genomic rearrangements that seem to be causally involved. Clearly, building a vast catalogue of somatic mutations in breast cancer [11] constitutes only the first step on the long road to identifying novel drug targets and developing therapies that are more effective. Although the catalogue is still incomplete, it is comprehensive enough that various structural features are starting to emerge.

First, point mutations that are activating or inactivating often occur in genes that are also commonly amplified or deleted [3,12]. This observation is important for it identifies genes that are disrupted by multiple, and possibly equivalent, mechanisms and that therefore are more likely to drive tumour progression. Thus, it may be possible to categorise all somatic mutations into activating or inactivating aberrations and to carry this simplified picture forward when analysing the systems-level implications of these aberrations. A similar approach was taken by Leary and colleagues [3], who combined the point and copy-number mutomes to identify pathways targeted for disruption.

A second structural feature to emerge has been the mutual exclusivity pattern of mutations [13], often involving genes in a common pathway or genes exhibiting high sequence similarity [14], as well as cooperative or synergistic interactions, often involving physically interacting proteins such as the *SMAD2* and *SMAD3* receptors in colorectal cancer [4]. In the case of breast cancer, there are some clear mutual exclusivity patterns emerging (for example, those of *PIK3CA* and *TP53* mutations; A.E. Teschendorff and C. Caldas, unpublished data) and independent hints of further patterns such as those of *PIK3CA* and *RAS* [13,15] as well as hints of cooperation or coexistence such as that involving *HER2* amplification and *PTEN* loss (including loss of *PTEN* expression) [15]. Although in most of these studies sample sizes are still too small for the results to be conclusive, the strong mutual exclusivity and cooperative patterns observed in, for example, a sequencing study of 188 lung tumours [12] suggest that similar mutational patterns will be seen in breast cancer.

Finally, a third feature to emerge is the evidently non-random distribution of mutations and affected genes in relation to specific protein domains, gene functions, and molecular pathways. In the case of somatic mutations, a strong association with spectric repeat and fibronectin domains, GTPase activation, extracellular matrix, and cell-cell adhesion as well as calcium ion-binding functions was observed [14,16]. This reinforces the crucial role that cytoskeletal, including cell-cell adhesion, dysregulation plays in malignant progression of breast cancer, although it remains to be seen how this may depend on the subtype and stage of the specific tumours studied [14,16]. In the context of pathways, somatic mutations were found to be enriched in many pathways

known to be pathogenic in breast cancer, including interferon signalling, cell-cycle checkpoint (G₁-S), BRCA1/2 (breast cancer 1/2, early onset)-related DNA repair, and phosphatidylinositol 3 kinase (PIK3) and v-akt murine thymoma viral oncogene homolog 1 (AKT) and transforming growth factor-beta (TGF- β) signalling [14,16]. In an integrative strategy, incorporating both point and copy-number aberrations in the enrichment analysis, Leary and colleagues [3] identified further important signalling pathways such as those of Notch, epidermal growth factor receptor (EGFR), fibroblast growth factor (FGF), and v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (ERBB2). Thus, these results should help to direct increasingly focused efforts that aim to develop therapies tailored to restoring these disrupted pathways. The study of Lin and colleagues [14] also illustrates nicely the added insights gained by adopting a more systems-level approach, mapping mutations onto well-curated protein interaction networks in order to identify subnetworks more prone to disruption. Specifically, they were able to show that genes targeted by point mutations are more prone to occupy hubs in the corresponding protein interaction networks as opposed to genes that are not. Interestingly, of the 83 somatically mutated proteins in breast cancer [17], over half (59) were shown to be part of a large interaction cluster involving *TP53*, *BRCA1*, *PIK3R1*, and *NFKB*. Thus, a full interpretation and understanding of how the approximately 30 to 40 CAN genes driving a given tumour cooperate to disrupt important signalling pathways, that themselves exhibit significant cross-talk, will undoubtedly benefit from network-based approaches that try to give meaning to the aberrations in the context of the whole interaction network. Some further insights in this direction were recently offered by Taylor and colleagues [18], who showed that tumours are preferentially disrupted at 'intermodular' hubs, representing proteins that mediate signals between different cellular functions.

The structural features of the complex muta-ome are only beginning to emerge, and full elucidation of these features will be crucial for translating the vast and growing catalogue of somatic mutations into improvements for therapy and eventual cures. Given the heterogeneity of breast cancer, accomplishing this goal will require upscaling the current Genome Atlas Research Project [19] to generate multi-dimensional high-resolution mutational and functional profiles of hundreds if not thousands of tumours.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

AET is supported by a Heller Research Fellowship.

References

- Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW: **Genomic and transcriptional aberrations linked to breast cancer pathophysiologies.** *Cancer Cell* 2006, **10**:529-541.
- Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, Costa JL, Pinder SE, van de Wiel MA, Green AR, Ellis IO, Porter PL, Tavaré S, Brenton JD, Ylstra B, Caldas C: **High-resolution aCGH and expression profiling identifies a novel genomic subtype of early negative breast cancer.** *Genome Biol* 2007, **8**:R215.
- Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW, Jones S, Sjöblom T, Park BH, Parsons R, Willis J, Dawson D, Willson JK, Nikolskaya T, Nikolsky Y, Kopelovich L, Papadopoulos N, Pennacchio LA, Wang TL, Markowitz SD, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE: **Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers.** *Proc Natl Acad Sci U S A* 2008, **105**:16224-16229.
- Greenman C, Stephens P, Smith R, Dalgleish 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, Jenkinson A, Jones D, et al.: **Patterns of somatic mutation in human cancer genomes.** *Nature* 2007, **446**:153-158.
- 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.
- Chuang HY, Lee E, Liu YT, Lee D, Ideker T: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol* 2007, **3**:140.
- 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.
- 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.
- 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.
- 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.
- Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR: **The catalogue of somatic mutations in cancer (COSMIC).** *Curr Protoc Hum Genet* 2008, **Chapter 10**:Unit 10.11.
- Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, Fulton RS, Zhang Q, Wendl MC, Lawrence MS, Larson DE, Chen K, Dooling DJ, Sabo A, Hawes AC, Shen H, Jhangiani SN, Lewis LR, Hall O, Zhu Y, Mathew T, Ren Y, Yao J, Scherer SE, Clerc K, Metcalf GA, et al.: **Somatic mutations affect key pathways in lung adenocarcinoma.** *Nature* 2008, **455**:1069-1075.
- Thomas RK, Baker AC, Debiassi RM, Winckler W, Laframboise T, Lin WM, Wang M, Feng W, Zander T, MacConaill L, Lee JC, Nicoletti R, Hatton C, Goyette M, Girard L, Majumdar K, Ziaugra L, Wong KK, Gabriel S, Beroukhim R, Peyton M, Barretina J, Dutt A, Emery C, Greulich H, Shah K, Sasaki H, Gazdar A, Minna J, Armstrong SA, et al.: **High-throughput oncogene mutation profiling in human cancer.** *Nat Genet* 2007, **39**:347-351.
- Lin J, Gan CM, Zhang X, Jones S, Sjöblom T, Wood LD, Parsons DW, Papadopoulos N, Kinzler KW, Vogelstein B, Parmigiani G, Velculescu VE: **A multidimensional analysis of genes mutated**

- in breast and colorectal cancers. *Genome Res* 2007, **17**:1304-1318.
15. Yuan TL, Cantley LC: **Pi3k pathway alterations in cancer: variations on a theme.** *Oncogene* 2008, **27**:5497-5510.
16. Chittenden TW, Howe EA, Culhane AC, Sultana R, Taylor JM, Holmes C, Quackenbush J: **Functional classification analysis of somatically mutated genes in human breast and colorectal cancers.** *Genomics* 2008, **91**:508-511.
17. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE: **The consensus coding sequences of human breast and colorectal cancers.** *Science* 2006, **314**:268-274.
18. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL: **Dynamic modularity in protein interaction networks predicts breast cancer outcome.** *Nat Biotechnol* 2009, **27**:199-204.
19. Cancer Genome Atlas Research Network: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**:1061-1068.