Molecular portraits: the evolution of the concept of transcriptome-based cancer signatures

Angelika Modelska, Alessandro Quattrone and Angela Re

Abstract
Cancer results from dysregulation of multiple steps of gene expression programs. We review how transcriptome profiling has been widely explored for cancer classification and biomarker discovery but resulted in limited clinical impact. Therefore, we discuss alternative and complementary omics approaches.

Key words: transcriptomes; signatures; cancer; heterogeneity; omics

Introduction
Each step of the gene expression program is tightly controlled by regulatory circuits [1]. Cancer is a disease resulting from dysregulation of these circuits, which can occur at any point of the control [2, 3]. Hence, each of the 'molecular portraits' of cancer obtained through high-throughput profiling (omics) technologies (Figure 1) bears fingerprints, which in principle can be exploited to identify biomarker genes for diagnosis and prognosis, and to find suitable therapeutic targets.

In the nearly 15 years from when the first cancer portrait was painted as an early outcome of the microarray technology [4], the great majority of the subsequent analyses have been limited to the first two levels of gene expression, the linear DNA sequence and the steady-state levels of mRNAs (i.e. genome and transcriptome profiling). Despite continuous advances in the affordability of microarrays to quantify copy number variations and mRNA levels, the transition to the massively parallel sequencing technology, which has an improved resolution, a greater dynamic range and an increased sensitivity depending on the depth of sequencing [5–7], has rendered sequencing-based cancer portraits definitely more informative than those relying on hybridization to probes. In the meanwhile, public repositories of high-throughput analysis of biological data have been heavily loaded with profiles of virtually every cancer type, following a rapid pace of accumulation (Figure 2) also owing to multisite and international projects such as TGCA [8] and ICGP [9]. As a result, our understanding of the anatomy of cancer genomes and transcriptomes has substantially improved. We now map precisely point mutations, copy number aberrations, translocations, splicing defects, alterations of transcription start or end and differential expression [2, 7].

Looking retrospectively, despite this increased mapping ability of genomic and transcriptomic aberrations, the usefulness of omics assay in the distinct tasks of classifying tumor subtypes, predicting clinical outcome (clinical validity) as well as treatment response (clinical utility) [10] appears rather
unsatisfactory. At the basis of the poor clinical usefulness of the molecular portraits produced until now could lie the inability to account for the biological complexity of cancer [11, 12]. It is still out of reach to comprehend in an informative way the multiplicity of processes in the path from DNA to protein, which are known to be dysregulated [13], and the underlying intra-tumor variation poses great difficulties [14–16]. Moreover, tumors bear both ‘driver’ and ‘passenger’ mutations that are difficult to distinguish [17], and are characterized by the tendency to undergo a Darwinian evolution and to evolve with time and treatment [18–20]. Admittedly, the major challenge in cancer research is the heterogeneity of this disease, not only between different tissues and cancer types, but also within the same cancer type among the different patients and temporally in the same patient. Additional hurdles related to technical approaches include the analysis of complex biological materials, insufficient sample size and improper composition of study cohorts, lack of clear objectives and thoughtful statistical design upfront of the experiments and an improper subsequent data analysis including the application of arbitrary thresholds [21, 22]. Finally, to understand highly interconnected mechanisms that drive cancer and to impact on oncology practice, we have to evaluate the still largely unknown potential of protein levels and protein posttranslational modifications, as well as integrative views of molecular mechanisms at multiple regulatory levels.

In this review, we examine the possible reasons of the failure to substantially improve clinical practice using the currently available, predominantly transcriptome-based molecular portraits, and identify recent advancements that could lead to more success in the future as well as present current limitations in the field.

Transcriptome-based portraits: the case of breast cancer

Breast cancer is heterogeneous, consisting of cases that are substantially different in their molecular and clinical characteristics. Owing to this inherent diversity, attempts at class discovery, prediction and management of breast cancer patients pose significant challenges [23]. Breast cancer heterogeneity also prevents the establishment of a multistep model of carcinogenesis, which has been possible in other cancer types, such as colorectal or pancreatic cancer [24, 25]. Despite such complexity, a nearly 15 years long examination of breast cancer transcriptome profiles resulted in substantial contributions into subtyping and assessment of the prognostic (providing information on the likely outcome of the cancer disease in an untreated individual) and predictive (providing information on the likely benefit from treatment) value of gene signatures. Therefore, we use this tumor type as a paradigm of how transcriptome profiles have contributed to the clinical management of this disease.

The initial molecular classification of breast cancer was based on transcriptome profiling and resulted in five subgroups,
One of the biggest hurdles for researchers in this field is the sample size that needs to be increased to obtain meaningful results for such a complex disease as breast cancer. The recent METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) study, using 2000 breast cancer patients with long-term clinical outcome information, is the latest wide-scale effort in producing a molecular classification of the tumor [30]. Therein, integrated analysis of genome and transcriptome portraits resulted in the robust identification of 10 subgroups, each being reproducible in the validation cohort and thus representing trustable tumor subtypes. METABRIC findings may help the researchers to better define future studies and plan clinical trials, as each subtype is likely to need a different treatment strategy to overcome the disease [23]. Furthermore, an assessment of the clinical validity of the integrative subtypes derived from METABRIC has been undertaken [31], even if it is worth noting that the METABRIC subtypes were solely conceived with the intention of best representing breast tumor biology and identifying potential molecular drivers. In general, the clinical potential of breast tumor subtyping warrants further study. At the moment, the descriptive power to define breast cancer subtype is being gained with limited translation into the clinic.

Out of the vast number of proposed breast cancer signatures, only a few of them have been licensed for use in the clinic, the most prominent ones listed in Table 1. The evidence base for these assays develops primarily through observational studies of limited sample sizes in diverse settings. As a matter of fact, unmet minimum sample size of study cohorts hampers the assessment of statistical performances, and prediction strength is known to vary considerably depending on which patient populations and what end point the assay was optimized for. Nonetheless, major conclusions were conveyed by previous systematic reviews and meta-analyses [37]. Evidence of analytic validity (technical performance characteristics) and reproducibility was found limited for all assays. Overall, these assays showed clinical validity and utility, albeit to a different extent. OncotypeDX, in particular, positions at the most advanced point in the development pathway, being the only assay that can predict treatment benefit [38, 39]. Recently, the Oncotype DX assay was modified to predict local recurrence risk and to guide individualized selection of treatment after surgical excision for women with newly diagnosed ductal carcinoma in situ of the breast who meet the ECOG E5194 criteria [40]. The cost-effectiveness of the assays remains inconclusive [41]. For all tests, the relationship of predicted to observed risk in different populations and their added value over conventional predictors, optimal implementation and relevance to patients receiving current therapies need further study, in particular in large cohorts. These should prove key to understanding whether we can rely on these assays rather than on standard clinicopathological features for treatment decisions [42, 43]. For now, they remain only complementary to the traditional tests. A comparative study of gene signatures showed that predictions of distinct assays were similar despite the limited overlap between the genes underlying the assays [44]. It is worth noting that observing a low overlap across gene signatures is not relevant from a purely clinical point of view because it can be due to the presence of multiple genes with similar moderate correlation to the clinical outcome. This scenario was highlighted in a study where different lists of genes with similar moderate correlation to the clinical outcome.

| Assay name (producer) | Description | Number of genes assayed | Reference |
|-----------------------|-------------|-------------------------|-----------|
| MammaPrint (Agedia)   | Assesses the risk of metastasis of early-stage breast cancer, and whether a patient will benefit from chemotherapy. | 70 | [32] |
| Oncotype DX (Genomic Health) | Assesses the chances of disease recurrence in women with early-stage estrogen receptor positive breast cancer and the benefit from certain types of chemotherapy. | 21 | [33] |
| MapQuant Dx (Ipsogen) | Assesses the histological grade and thus predicts the benefit from chemotherapy. | 97 | [34, 35] |
| THEROS CancerTYPE ID (Biotheranostics) | Identifies the origin of cancers with unknown or uncertain primary site. | 92 | [36] |

Table 1. Most prominent signatures used in breast cancer clinical practice

Difficulties to acquire robust cancer subtypes as well prognostic and predictive biomarkers could depend on the limitations of transcriptome-based profiling as the technology of choice. One of the reasons could be that transcriptome-based molecular profiles are not true portraits of the diseased state but mere shadows, as mRNA and protein levels tend to be rather poorly correlated (see below) and depend on genomic lesions in a complex way.

The discordance of information contained in different molecular portraits

The expression of many genes predicted to drive cancer phenotypes is weakly correlated with their copy number, and protein

Downloaded from https://academic.oup.com/bib/article-abstract/16/6/1000/226155 by guest on 28 July 2018
levels are poorly predicted by transcript levels, both in tumor samples and cell lines [46–48]. As the proteome is believed to more closely represent the physiological state of the cell than the genome or the transcriptome, the discordance between transcriptome and proteome data should draw attention to the limits of many cancer studies that use transcriptome as a surrogate for proteome to reconstruct molecular portraits and to identify cancer signatures.

The low estimates of correlation between transcriptome and proteome data could be partially owing to technical reasons, such as different scalability between platforms, detection bias, different kinetics of synthesis and degradation between transcripts and proteins or side effects resulting from the use of drugs that inhibit transcription and translation [49, 50]. In addition, the partial correlation between transcript and protein levels was not mitigated when the quantification was addressed in optimal experimental conditions on a genome-wide scale in unperturbed mammalian cells [51]. In this study, transcript levels could explain only 40% of the variance in protein levels, which, albeit higher than in previous reports [49, 50], is limited. An estimate of the correlation between transcript and protein levels was also addressed computationally. Combining steady-state mRNA and protein levels with the analysis of the sequence features related to translation and protein degradation, allowed to demonstrate that the transcript levels and structural features together could explain two-thirds of the variation in protein levels, whereas transcript levels alone could explain only 25–30% of it [52].

The relative importance of variation in the transcriptome and the translome (the latter being a subtraction of the transcriptome composed of mRNAs engaged in translation) on treatment with different stimuli was assessed in another study [53], which showed an extensive uncoupling of mRNA transcriptome and translome fluctuations, with up to 90% of significant changes being limited to the translome. These results, together with others [54, 55], support the notion that the machineries responsible for regulating mRNA levels as well as mRNA translation can remain largely independent of each other. In line with this view, protein levels are more conserved during evolution than the corresponding mRNA levels [56], as assessed line with this view, protein levels are more conserved during translation can remain largely independent of each other. In order to address this, looking at seven different species and identifying a selective pressure to maintain protein abundances during evolution even when mRNA abundances diverge. Taken together, the currently available evidence suggests that the process of translation and the abundance of proteins might provide better molecular portraits of healthy and disease states than the transcriptome.

The use of translome profiling

Cancer characterization and biomarker development could now take advantage from advances in translome profiling, which examines translation by characterizing mRNA transcripts engaged in interaction with active ribosomes. Polysome profiling, relying on the analysis of polysomal mRNAs that are first fractionated by sucrose gradient centrifugation, is commonly used to assess the translational efficiency of transcripts. To date, no studies have been conducted to assess the utility of the translome in the clinical context, with only one report on polysome profiling from solid tissues [57]. More recently, ribosome profiling data have also been proposed as a proxy for protein abundance [58, 59], and this method is proving useful to evaluate translation deregulation in diseased states [60–62]. Nevertheless, the broad application of polysome and ribosome profiling is limited by the lack of high-throughput settings of the procedures. All in all, translation profiling techniques can be helpful to uncover cancer cell behaviors specifically steered by disarranged translation, and therefore to suggest novel therapeutic targets. However, this field certainly needs further advancements and better standardization of protocols to fulfill its full promise in the clinics.

The use of proteome profiling

The ability to understand proteomes brings us a step closer to the physiological state of a cell, and therefore proteomics promises to provide enhanced ability to predict the disease course, to promote insights in cancer deregulated pathways and associated mechanisms of resistance to treatment. Proteomics has now advanced sufficiently to allow us to observe the variation of thousands of proteins [63]. However, proteome determination still lags behind transcriptome assays in terms of sensitivity, accuracy and scalability. Proteomics-based cancer studies are inherently more complicated, owing to large dynamic range of concentrations of molecular species and the absence of means to amplify them, low relative abundance of many biomarkers and the extent of protein variability owning to splicing and protein modifications [64]. Recent advances in technology platforms relying on mass spectrometry (MS), sample handling and bioinformatics solutions for data analysis have delivered high-coverage surveys of the proteome [65]. Over 4 years, a 100-fold increase in throughput has been achieved [66, 67], which brings the analysis of complete proteomes within reach. Several initiatives, such as the Human Proteome Map, ProteomicsDB, the Chromosome-Centric Human Proteome Project, have demonstrated the feasibility of complementing available genome and transcriptome normal data by comprehensively exploring the proteome on progressively larger scale in the near future [68–70]. In addition, the Cancer Proteome Atlas (TCPA) provides ready access to cancer proteome data, which were generated using 3467 TCGA tumors and 439 cell line samples by using reverse-phase protein arrays [71]. As a proof of concept, a preliminary analysis of TCPA data has recently proved the association of several proteins with clinical outcome [72], but the limitation of this assay is that it targets only 181 proteins [71]. Furthermore, the Clinical Proteomic Tumor Analysis Consortium, launched by the National Cancer Institute, has released a public repository of well-characterized, MS-based, targeted proteomic assays, including a growing number of cancer types [73].

The proteomes of human blood serum and plasma contain a vast amount of useful information about the cancer state. Blood collection is simple and minimally invasive, making it a medium of choice for diagnostic tests. Until recently, the exploitation of blood proteome for discovery of protein biomarkers has been challenging because of the large number of unique proteins, their degradation products and the broad range of protein concentrations. Nonetheless, advances in proteomic technologies are possibly turning the serum proteome diagnostics into a clinical reality [74, 75]. For instance, the VeriStrat serum proteomic test status was found predictive of differential benefit in terms of overall survival for non-small-cell lung cancer patients treated with second-line erlotinib versus chemotherapy [76].

Proteome-based cancer portraits unfortunately suffer from some pitfalls that are specific to proteomic approaches. Typical challenges in the proteomics workflow include inconsistent proteolysis of samples and protease bias [77], the limited working concentration range of species detectable by the instrumentation, the need for reduction in sample complexity, which otherwise is not directly compatible with MS analysis, and the
The use of multi-omic approaches

Improvements in the accuracy and precision of high-throughput technologies have set the stage for the integration of multiple data types. In addition to the mostly investigated transcriptomes and proteomes, additional molecular profiles have been shown informative; for instance, metabolic profiles have an important theoretical advantage in that they are sensitive to both genetic and environmental influences, and can be translated into clinical tools for application to personalized medicine [82–84]. The advantages of integrative analyses of multiple assays are arguably manifold: more fundamental understanding of the complex biology of cancer in its multiple and interrelated facets, power and robustness in tumor classification and prediction of outcome, which cannot be reached by each assay in isolation. A drawback of the integrative methods is a considerable dimensionality issue owing to the increasing of the number of variables without increasing the number of samples. However, an array of techniques are amenable to integrative analysis, such as the preliminary identification of principle components, which are defined as latent axes that maximally capture the variance in the data set, or the imposition of variable sparsity through LASSO constraints [85] or elastic net constraints [86]. To enable the access to multi-assyay data and the exchange of integrative analysis methods, storage and informatics platforms such as Rembrandt [87], Synapse [88] and IntOGen [89] have been already established as public resources for open data-driven collaborative research. The methodological choice of the computational analysis depends on the objectives set by the integration of multi-assay data [90, 91]. For example, a number of approaches used different technologies to confirm the findings based on one assay with an additional one [92, 93]. Tumor stratification has been usually addressed by applying clustering-based approaches to multi-assay data types. The most common approach to subtype discovery across multiple data types is to sequentially cluster each type and finally to integrate the results. For instance, integrating six assays by this approach demonstrated the ability to molecularly define the major breast cancer subtypes [13]. Already some attempts to conduct proteogenomic characterization of human cancers are resulting in improved characterization of candidate driver genes and genomic abnormalities [47]. Alternative clustering approaches, performing joint inference from distinct assays and generating a single integrated cluster assignment, were equally useful in characterizing breast cancer and glioblastoma subtypes starting from genome and transcriptome data [30, 94]. Other attempts integrated molecular profiles with previous knowledge captured in gene interaction networks to stratify ovarian, uterine and lung cancer patients [95].

Key Points

- Transcriptome profiling has resulted in limited clinical utility despite tremendous efforts.
- The potential prognostic and predictive value of profiling protein levels of gene regulation is pitifully unexplored and deserves careful assessment.
- Multiple types of ’omic’ data provide complementary information of which integration might be crucial to improve the assessment of tumor prognosis.
- Effective feature selection and mathematical modeling strategies should be developed in the analysis of individual data types and in the integration thereof.

Conclusions

The past 15 years have seen an amazing growth in the ability to generate molecular data at multiple complementary levels of gene expression, therefore enabling the generation of molecular portraits in unprecedented detail in a variety of cancer contexts. However, with a few exceptions, these investments have failed to substantially improve diagnosis, prognosis or treatment of cancer. A fundamental reason for this discrepancy between data generation and clinical improvement is the relatively immature use of experimental designs and analytical methodologies to handle and meaningfully interpret high-throughput data types. The majority of attempts to stratify tumors or to predict clinical outcome to date have used transcriptome-based data without consideration of the effects in protein levels or have postulated linear correlation of data across these levels. Future advances will focus on increasing the information content of each data type, reducing the amount of input biological material, getting relevant data in less analysis time and improving mathematical modeling to uncover relationships between diverse and complementary data types and clinical phenotypes. Continuous attempts to fully describe the functional alterations at the multiple levels of gene regulation in cancer will be instrumental to the effective interpretation of cancer molecular data and to their translation into the clinic.

Funding

Biotechnology Program of the University of Trento.
References

1. Komili S, Silver PA. Coupling and coordination in gene expression processes: a systems biology view. Nat Rev Genet 2008;9:38–48.
2. Garraway LA, Lander ES. Lessons from the cancer genome. Cell 2013;153:17–37.
3. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009;458:719–24.
4. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
5. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63.
6. Mwenifumbo JC, Marra MA. Cancer genome-sequencing projects. Nature 2010;464:993–8.
7. Soon WW, Hariharan M, Snyder MP. High-throughput sequencing for biology and medicine. Mol Syst Biol 2013;9:640.
8. McLendon R, Friedman A, Bigner D, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061–8.
9. International Cancer Genome Consortium, Hudson TJ, Anderson W, et al. International network of cancer genome projects. Nature 2010;464:993–8.
10. McShane LM, Cavenagh MM, Lively TG, et al. Criteria for the use of omics-based predictors in clinical trials. Nature 2013;502:317–20.
11. Bedard PL, Hansen AR, Ratain MJ, et al. Tumour heterogeneity in the clinic. Nature 2013;501:355–64.
12. Kreso A, O’Brien CA, van Galen P, et al. Variable clonal population dynamics influence chemotherapy response in colorectal cancer. Science 2013;339:543–8.
13. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
14. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature 2013;501:328–37.
15. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer 2012;12:323–34.
16. Burrell RA, Swanton C. The evolution of the unstable cancer genome. Curr Opin Genet Dev 2013;24:61–7.
17. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415–21.
18. Burrell RA, Swanton C. Tumour heterogeneity and the evolution of polyclonal drug resistance. Mol Oncol 2014;8:1095–111.
19. Foo J, Michor F. Evolution of resistance to targeted anti-cancer therapies during continuous and pulsed administration strategies. PLoS Comput Biol 2009;5:e1000557.
20. Almendro V, Cheng Y-K, Randles A, et al. Inference of tumour evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. Cell Rep 2014;6:514–27.
21. Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. Cancer Res 2012;72:6097–101.
22. Ferté C, Trister AD, Huang E, et al. Impact of bioinformatic procedures in the development and translation of high-throughput molecular classifiers in oncology. Clin Cancer Res 2013;19:415–25.
23. Dawson S-J, Rueda OM, Aparicio S, et al. A new genomically driven integrated classification of breast cancer and its implications. EMBO J 2013;32:617–28.
24. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759–67.
25. Yachida S, Jones S, Bozic I, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 2010;467:1114–17.
26. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001;98:10869–74.
27. Abba, E Lacunza, M Buttì, et al. Breast cancer biomarker discovery in the functional genomic age: a systematic review of 42 gene expression signatures. Biomark Insights 2010;5:103–18.
28. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput Biol 2011;7:e1002240.
29. Beck AH, Knoblauch NW, Hefti MM, et al. Significance analysis of prognostic signatures. PLoS Comput Biol 2013;9:e1002875.
30. Curtis C, Shah SP, Chin S-F, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–52.
31. Ali HR, Rueda OM, Chin S-F, et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genom Biol 2014;15:431.
32. van ‘t Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.
33. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.
34. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262–72.
35. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 2007;25:1239–46.
36. Ma X-J, Patel R, Wang X, et al. Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay. Arch Pathol Lab Med 2006;130:465–73.
37. Marchionni I, Wilson RF, Wolff AC, et al. Systematic review: gene expression profiling assays in early-stage breast cancer. Ann Intern Med 2008;148:358–69.
38. Holt S, Bertelli G, Humphreys I, et al. A decision impact, decision conflict and economic assessment of routine Oncotype DX testing of 146 women with node-negative or pNImi, ER-positive breast cancer in the U.K. Br J Cancer 2013;108:2250–8.
39. Hassett MJ, Silver SM, Hughes ME, et al. Adoption of gene expression profiling test and association with use of chemotherapy among women with breast cancer. J Clin Oncol 2012;30:2218–26.
40. Duggal S, Julian TB. A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ. J Natl Cancer Inst 2013;105:681–3.
41. Marchionni I, Wilson RF, Marinopoulos SS, et al. Impact of gene expression profiling tests on breast cancer outcomes. Evid Rep Technol Assess (Full Rep) 2007;1–105.
42. Torrisi R, Garcia-Etienne CA, Losurdo A, et al. Potential impact of the 70-gene signature in the choice of adjuvant systemic treatment for ER positive, HER2 negative tumors: a single institution experience. Breast 2013;22:419–24.
43. Nielsen T, Wallden B, Schaper C, et al. Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using
formalin-fixed paraffin-embedded breast tumor specimens. BMC Cancer 2014;14:177.

44. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 2006;355:560–9.

45. Ein-Dor L, Kela I, Getz G, et al. Outcome signature genes in breast cancer: is there a unique set? Bioinformatics 2005;21:171–8.

46. Myhre S, Lingjærde OC, Hennessey BT, et al. Influence of DNA copy number and mRNA levels on the expression of breast cancer related proteins. Mol Oncol 2013;7:704–18.

47. Zhang B, Wang J, Wang X, et al. Proteogenomic characterization of human colon and rectal cancer. Nature 2014;513:382–7.

48. Moghaddas Gholami A, Hahne H, Wu Z, et al. Global proteome analysis of the NCI-60 cell line panel. Cell Rep 2013;4:609–20.

49. de Sousa Abreu R, Penalvo LO, Marcotte EM, et al. Global signatures of protein and mRNA expression levels. Mol Biosyst 2009;5:1512–26.

50. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett 2009;583:3966–73.

51. Schwänhäuser B, Busse D, Li N, et al. Global quantification of mammalian gene expression control. Nature 2011;473:337–42.

52. Vogel C, Abreu R de S, Ko D, et al. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol Syst Biol 2010;6:400.

53. Tebaldi T, Re A, Viero G, et al. Widespread uncoupling between transcriptome and translational variations after a stimulus in mammalian cells. BMC Genomics 2012;13:220.

54. Larsson O, Li S, Issaenko OA, et al. Euraykotic translation initiation factor eIF4E induced progression of primary human mammary epithelial cells along the cancer pathway is associated with targeted translational deregulation of oncogenic drivers and inhibitors. Cancer Res 2007;67:6814–24.

55. Ramírez-Valle F, Braunstein S, Zavadil J, et al. elf4GI links nutrient sensing by mTOR to cell proliferation and inhibition of autophagy. J Cell Biol 2008;181:293–307.

56. Laurent JM, Vogel C, Kwon T, et al. Protein abundances are more conserved than mRNA abundances across diverse taxa. Proteomics 2010;10:4209–12.

57. del Prete MJ, Vernali R, Dolzhin H, et al. Isolation of polypeptide-bound mRNA from solid tissues amenable for RT-PCR and profiling experiments. RNA 2007;13:414–21.

58. Ingolia NT, Ghaemmaghami S, Newman JRS, et al. Genomewide analysis in vivo of translation with nucleotide resolution using ribosome profiling, Science 2009;324:218–23.

59. Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet 2014;15:205–13.

60. Hsieh AC, Liu Y, Edlind MP, et al. The translational landscape of mTOR signalling steers cancer initiation and metastasis. Nature 2012;485:55–61.

61. Rooijers K, Loayza-Puch F, Nijtmans LG, et al. Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. Nat Commun 2013;4:2886.

62. Helmy K, Halliday J, Fomchenko E, et al. Identification of global alteration of translational regulation in glioma in vivo. PLoS One 2012;7:e46965.

63. Wu L, Chandile SI, Choi Y, et al. Variation and genetic control of protein abundance in humans. Nature 2013;499:79–82.

64. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 2006;24:971–83.

65. Altelaar AM, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 2012;14:35–48.

66. de Godoy LMF, Olsen JV, Cox J, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 2008;455:1251–4.

67. Nagaraj N, Kulak NA, Cox J, et al. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. Mol Cell Proteomics 2012;11:M111.013722.

68. Kim M-S, Pinto SM, Getnet D, et al. A draft map of the human proteome. Nature 2014;509:575–81.

69. Wilhelm M, Schlegel J, Hahne H, et al. Mass-spectrometry-based draft of the human proteome. Nature 2014;509:582–7.

70. Paik Y-K, Jeong S-K, Omenn GS, et al. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. Nat Biotechnol 2012;30:221–3.

71. Li J, Lu Y, Akbani R, et al. TCPA: a resource for cancer functional proteomics data. Nat Methods 2013;10:1046–7.

72. Akbani R, Ng PKS, Werner HMJ, et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. Nat Commun 2014;5:3887.

73. Whiteaker JR, Halusa GN, Hoofnagle AN, et al. CPTAC Assay Portal: a repository of targeted proteomic assays. Nat Methods 2014;11:703–4.

74. Wildes D, Wells JA. Sampling the N-terminal proteome of human blood. Proc Natl Acad Sci USA 2010;107:4561–6.

75. Boersema PJ, Geiger T, Wisniewski JR, et al. Global quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. Mol Cell Proteomics 2013;12:158–71.

76. Gregorc V, Novello S, Lazzari C, et al. Predictive value of a proteomic signature in patients with non-small-cell lung cancer treated with second-line erlotinib or chemotherapy (PROSE): a biomarker-stratified, randomised phase 3 trial. Lancet Oncol 2014;15:713–21.

77. Peng M, Taouatas N, Cappadona S, et al. Protease bias in absolute protein quantitation. Nat Methods 2012;9:524–5.

78. Baker ES, Liu T, Petyuk VA, et al. Mass spectrometry for translational proteomics: progress and clinical implications. Genome Med 2012;4:63.

79. Pommier AJC, Shaw R, Spencer SKM, et al. Serum protein profiling reveals baseline and pharmacodynamic biomarker signatures associated with clinical outcome in mCRC patients treated with chemotherapy ± cediranib. Br J Cancer 2014;111:1590–604.

80. Mehan MR, Williams SA, Siegfried JM, et al. Validation of a blood protein signature for non-small cell lung cancer. Clin Proteomics 2014;11:32.

81. Taguchi A, Politi K, Pitteri SJ, et al. Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. Cancer Cell 2011;20:289–99.

82. Tang X, Lin C-C, Spasojevic I, et al. A joint analysis of metabolomics and genetics of breast cancer. Breast Cancer Res 2014;16:415.

83. Clayton TA, Lindon JC, Cloarec O, et al. Pharmacometabolomic phenotyping and personalized drug treatment. Nature 2006;440:1073–7.

84. Farshidfar F, Weljie AM, Kopciuk K, et al. Serum metabolomic profile as a means to distinguish stage of colorectal cancer. Genome Med 2012;4:42.

85. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc B 1996;58:267–88.
86. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc B* 2005;67:301–20.
87. Madhavan S, Zenklusen J-C, Kotliarov Y, et al. Rembrandt: helping personalized medicine become a reality through integrative translational research. *Mol Cancer Res* 2009;7:157–67.
88. Omberg L, Ellrott K, Yuan Y, et al. Enabling transparent and collaborative computational analysis of 12 tumor types within The Cancer Genome Atlas. *Nat Genet* 2013;45:1121–6.
89. Gundem G, Perez-Llamas C, Jene-Sanz A, et al. IntOGen: integration and data mining of multidimensional oncogenic data. *Nat Methods* 2010;7:92–3.
90. Haider S, Pal R. Integrated analysis of transcriptomic and proteomic data. *Curr Genomics* 2013;14:91–110.
91. Kristensen VN, Lingjærde OC, Russnes HG, et al. Principles and methods of integrative genomic analyses in cancer. *Nat Rev Cancer* 2014;14:299–313.
92. Louhimo R, Hautaniemi S. CNAmet: an R package for integrating copy number, methylation and expression data. *Bioinformatics* 2011;27:887–8.
93. Aure MR, Steinfeld I, Baumbusch LO, et al. Identifying in-trans process associated genes in breast cancer by integrated analysis of copy number and expression data. *PLoS One* 2012;7:e53014.
94. Shen R, Mo Q, Schultz N, et al. Integrative subtype discovery in glioblastoma using iCluster. *PLoS One* 2012;7:e35236.
95. Hofree M, Shen JP, Carter H, et al. Network-based stratification of tumor mutations. *Nat Methods* 2013;10:1108–15.
96. Lando M, Holden M, Bergersen LC, et al. Gene dosage, expression, and ontology analysis identifies driver genes in the carcinogenesis and chemoradioresistance of cervical cancer. *PLoS Genet* 2009;5:e1000719.
97. Mankoo PK, Shen R, Schultz N, et al. Time to recurrence and survival in serous ovarian tumors predicted from integrated genomic profiles. *PLoS One* 2011;6:e24709.
98. Cheng W-Y, Ou Yang T-H, Anastassiou D. Biomolecular events in cancer revealed by attractor metagenes. *PLoS Comput Biol* 2013;9:e1002920.
99. Vaske CJ, Benz SC, Sanborn JZ, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* 2010;26:i237–45.
100. Staiger C, Cadot S, Györfy B, et al. Current composite-feature classification methods do not outperform simple single-genes classifiers in breast cancer prognosis. *Front Genet* 2013;4:289.