

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Bioinformatics in Breast Cancer Research

Beyzanur Yigitoglu, Eyyup Uctepe,
Ramazan Yigitoglu, Esra Gunduz and
Mehmet Gunduz

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/59519>

1. Introduction

Developments both in computer hardware and software allowed for storing, distributing, and analyzing data obtained from biological experimentation, the very definition of bioinformatics. From this standpoint, bioinformatics can be narrowly defined as a field at the crossroads of biology and computer engineering, responsible for the storage, distribution, and analysis of biological information.[1] The term of bioinformatics relatively refers to the formation and advancement of algorithms, computational and statistical techniques, and theory to solve formal and practical problems posed by or inspired from the management and analysis of biological data.[2,3]

Since its emergence as an independent discipline in the 1980s, bioinformatics has been rapidly developing, keeping up with the expansion of genome sequence data. Whereas it is safe to say that 20 years ago, publishing computationally-derived results was a challenge and experimental observations were considered the only way of making progress[1]; after the famous Clinton-Blair handshake for the completion of the human genome in April 2003 [4], headlines such as “the laboratory rat is giving way to the computer mouse” arose.[5] The importance of bioinformatics methods has further increased following the technological improvement of large-scale gene expression analysis using DNA microarrays and proteomics experiments. Wet experiments and the use of bioinformatics analyses go hand in hand in today’s biological and clinical research.[6] Undeniably, it is almost inconceivable that a high-impact research publication in biology does not contain some elements of computing.[1]

To date, the genome, transcriptome and proteome are investigated with large-scale and high-throughput techniques to suggest treatment and predict outcomes. With the availability of high-throughput sequencing in hypothesis driven science, various sequence-based techniques

are originated, namely expressed sequence tags (ESTs)[7], serial analysis of gene expression (SAGE)[8], massively parallel signature sequencing (MPSS)[9], the 'HapMap' project proceeding by means of individual SNPs (single nucleotide polymorphisms) to link specific genotypes to diseases.[10,11] Aside from sequencing techniques, microarray technology is one of the high-throughput techniques, possibly the most promising one. As for protein analysis techniques, tissue arrays[12] and proteomics can be named.

On the one hand, microarrays are microscope slides or chips with immobilized probes, usually cDNA (complementary DNA), BAC (bacterial artificial chromosome), or oligo probes.[13] There are very large numbers of spots on an array, each containing a huge number of identical DNA molecules. Two important applications of microarray technology are gene expression monitoring and Single Nucleotide Polymorphisms (SNP) detection.[14] This technique is widely applicable because less RNA is used to analyze thousands of genes. Despite its increasing use around the world, microarray analysis has some limitations if used as a single method for exploring tumor biology. An obvious weakness is that a microarray represents a single snapshot of the patient.[15] But there are a large number of elements leading to disturbed gene function[16], such as large and small deletions or single base substitutions, mutations that affect promoter regions or splice-sites, as well as epigenetic silencing. Those factors may influence the result but may go undetected as well, depending on the exact type of lesion as well as its location with respect to the area hybridizing with the probe.[17] Furthermore, differentially expressed genes do not necessarily translate into varying protein levels with functional implications; so, it does not always show a correlation between the expression of a gene and the amount of translated protein.[18] Also, compared to RT-PCR (reverse transcription polymerase chain reaction), microarray signals are less sensitive, accurate and not able to resolve smaller differences in gene expression.[19] In addition to its comparative simplicity, microarray technology requires better understanding of the limitations and careful attention to experimental design and data analysis for meaningful results.

Bioinformatics applications are used in analysis of entire gene expression profiles to approach the disease at genome level and pose new hypotheses regarding certain mechanisms including but not limited to signaling pathways governing the process of formation, maintenance and expansion of tumor.[20] Bioinformatics analyses can also be applied to miRNA, DNA copy-number, SNPs, sequence, and methylation data[21] along with the field of medical sciences to know the pathways for diagnosing which genomic changes could give rise to each known inherited disease, i.e., identification of the gene causing disease, and also genetic therapies that can reverse disease phenotype.[14] Different Browser and Databases has been developed to analyze and process this huge quantity of data (Table 1.0 and Table 2.0).

Kept in mind that the discovery of complete protein classes is still in progress, e.g., the kinases of the human genome[22], the classification of proteins with related structure and function[23] will preserve its significance in the molecular dissection of human health and disease. In the future, bioinformatics is expected to continue its fascinating interplay with the field of genomics in cancer research, that is cancer bioinformatics and oncogenomics.[24]

2. Bioinformatics in various cancers

Cancer is one of the prevalent diseases that bring about death worldwide. Given that Scientists have sequenced the human genome[25], now it is time to use these genomic data, and the high-throughput technology developed to generate them, to tackle major health problems such as cancer.[24] Cancer molecular mechanisms are more successfully examined considering the genes and proteins interaction and network. Bioinformatics tools are vital for acquiring a more holistic view of cancer and analyzing the intricate data, speeding up the research process including biomarker discovery. Moreover, cancer clinical bioinformatics is critical to reach systems clinical medicine by combining clinical measurements and signs with human cancer tissue-generated bioinformatics, understanding clinical symptoms and signs, disease development and progress, and therapeutic strategy.[26,27,28]

The leading cause of cancer death is lung cancer but still awaits reliable molecular markers. Kim et al.[29] used multiple clinical samples and combined the bioinformatics analysis of the public gene expression data with clinical validation to identify biomarker genes for non-small-cell lung cancer, which shows poor prognosis and recurrence. They meta-analyzed the SAGE and EST data and chose 20 genes for experimental validation through semiquantitative RT-PCR. Then, applied quantitative RT-PCR to 7 genes (CBLC, CYP24A1, ALDH3A1, AKR1B10, S100P, PLUNC, and LOC147166) identified as potential diagnostic markers, leading to 2 highly probable novel biomarkers (CBLC and CYP24A1).

Liver cancer is the most common type, subsequent to lung cancer, responsible for cancer-related deaths. Sawey et al.[30] performed a forward genetic screen, using a mouse hepatoblast model and RNAi, guided by human hepatocellular carcinoma amplification data. They found that the amplification led to the selective sensitivity to FGF19 inhibition. Hence, FGF19 is an equally important driver gene of 11q13.3 amplicon as CCND1 in liver cancer, which means 11q13.3 amplification could be an effective biomarker for patients predicted to respond to anti-FGF19 therapy.

In a recent study[31], an individualized bioinformatics analysis strategy was applied to previously-established transcriptome data for clear cell renal cell carcinoma (ccRCC) to identify and reposition 8 FDA-approved drugs with negative correlation and P-value <0.05 for anticancer therapy. Authors demonstrated that pentamidine is effective against RCC cells in culture, and slows tumor growth in a RCC xenograft mouse model so it might be a new therapeutic agent to be combined with current standard-of-care regimens for patients with metastatic RCC.

With regard to leukemia, diagnosis and subclassification is mostly based on the application of various techniques like cytomorphology, cytogenetics, fluorescence in situ hybridization, multiparameter flow cytometry, and PCR-based methods which are time-consuming and cost-intensive, also require expertise in central reference laboratories. Therefore, microarray analysis represents a novel promising method to be used as a diagnostic tool.[14] A key determinant in the prognosis of chronic lymphocytic leukemia (CLL) is the mutational status of the immunoglobulin heavy chain variable region (IGHV) genes.[32] For the correct delin-

eation of the mutational status, the patient's leukemic cells and closest germline counterpart should be compared. Unfortunately, public web-based databases are commonly used instead of the patient's germline DNA sequence from non-leukemic cells. Several of these reference databases involve VBASE, GenBank/IgBLAST and the international ImMunoGeneTics information systems that employ different software types, amount of natural IGHV polymorphism and criteria used to map the complementarity determining regions and framework regions. As a result, the correct interpretation of the IGHV mutational status in CLL may be affected.[33]

Because of the heterogeneity of many tumors, it is a very challenging work to identify good molecular targets. For instance, resistant subclones of overexpressed and mutated genes may prevent them from being good molecular targets. Therefore, best target is a 'red dot' gene whose mutation occurs early in oncogenesis and dysregulates a key pathway that drives tumor growth in all of the subclones. Examples include mutations in the genes ABL, HER-2, KIT, EGFR and probably BRAF, in chronic myelogenous leukemia, breast cancer, gastrointestinal stromal tumors, non-small-cell lung cancer and melanoma, respectively. For efficacious therapeutics; identification of red-dot targets, development of drugs that inhibit the red-dot targets, and diagnostic classification of the related pathways are a must.[34]

3. Bioinformatics and breast cancer

Breast cancer occurs in both men and women, yet male breast cancer is less common. Although a cure for each stage of breast cancer has not yet been found, identifying the genetic mutations that cause the disease can play an important role and this is said by scientists to be like looking for needles in a haystack, and after finding the needles or coding regions, they must find disease-related sequences within them.[3⁶] Bioinformatics sets the stage for searching 3 billion base pairs to detect genetic defects.

Allinen et al. described the comprehensive gene expression profiles of each cell type composing normal breast tissue and in situ and invasive breast carcinomas performing SAGE (serial analysis of gene expression) and utilizing cell-type specific cell surface markers and magnetic beads for the rapid sequential isolation. Their results suggest that considerable transcriptional alterations happen in all cell populations while genetic changes were detected only in epithelial cells among myoepithelial, endothelial and stromal cells, myofibroblasts and lymphocytes.[35] To continue with another study, based upon a systematic Sanger sequencing analysis of 13,023 genes in 11 human breast cancers, individual tumors accumulate an average of approximately 90 point mutations in gene coding regions, but only a tiny number of these were recurrent and were in significant genes of breast cancer, including p53 and PIK3CA. A much larger number of the genes do not necessarily contribute to the carcinogenesis.[36] Considering the genomic landscape of breast cancer, these more common mutations resemble "mountains" while the vast majority of genes reflect "hills" that are infrequently mutated. We need to elucidate mechanisms involved in the disease to understand the heterogeneity of human cancers and utilize personal genomics for tumor diagnosis and new therapeutic strategies.[37]

As widely accepted, early detection of breast cancer has an enormous impact on patient's survival. Seeing that genome-wide expression patterns of tumors mirror the biology of the tumors, relating gene expression patterns to clinical outcomes sheds light on the biological diversity of the tumors.[38] In the discovery of genes and pathways that are specifically activated or inactivated during tumor progression, high throughput genome-wide array based techniques like array comparative genomic hybridization (aCGH) and transcriptional profiling can be used.[13] A molecular classification of breast cancer, with more than five reproducible subtypes (basal-like, ERBB2, normal-like, luminal A, luminal B) was defined through gene expression profiling and microarray analysis.[38,39,17] In addition, performing the gene set enrichment analysis (GSEA), a gene set linked to the growth factor (GF) signaling was observed to be significantly enriched in the luminal B tumors.[40] Another study states that multiple pathways were identified by mapping gene sets defined in Gene Ontology Biological Process (GOBP) for estrogen receptor positive (ER+) or estrogen receptor negative (ER-); and among them, in a separate set, pathways related to apoptosis and cell division or G-protein coupled receptor signal transduction are associated with the metastatic capability of ER+or ER-tumors, respectively.[41] Additionally, in a study, it is supported that breast cancer is initiated with mutated stem cells/progenitors, also called "breast cancer stem cells" because they are sufficient to sustain oncogenesis and tumor growth.[42] To identify genetic changes in the progression of breast carcinoma, Yao et al. [43] used aCGH and SAGE combined for ductal carcinoma in situ (DCIS), invasive breast carcinomas, and lymph node metastases. They identified 49 minimal commonly amplified regions and reported that the overall frequency of copy number alterations was more in invasive tumors than in DCIS, with several of them present only in invasive cancer. In breast cancer, gene amplification happens recurrently on some chromosomal locations (e.g. 1q, 8p12, 8q24, 11q13, 12p13, 12q13, 17q21-q23, 20q13) [43,44], which points to the activation of some oncogenes at high frequency during the growth of tumor. Amplification is a mechanism causing the gene expression constitutively enhanced above the level of physiologically normal variation, so the significance of oncogene amplification in tumorigenesis had originated from expression profiling of tumor cells by oncogene arrays.[45]

Bioinformatics is also crucial in the realm of pharmacogenomics. There became a need to develop accurate tools for the effective treatment relying on biological characterization of each patient's tumor. Gene-expression profiling of tumors with DNA microarrays is a powerful tool for pharmacogenomics targeting of treatments. Oncotype DX™ assay (Genomic Health) is a good example, which was described for identifying the subset of node-negative estrogen-receptor-positive breast cancer patients who do not require adjuvant chemotherapy.[46,34] A recent research demonstrated that microarray analysis with qRT-PCR validation reveals distinct pathways of resistance to bevacizumab (BEV) in xenograft models of human ER+breast cancer, showing Follistatin (FST) and NOTCH as the top signaling pathways associated with resistance in VEGF-driven tumors ($P < 0.05$). According to the gene expression analysis, the level of VEGF expression affects the response to BEV therapy and gene pathways.[47] Using appropriate bioinformatics tools, such findings may elucidate the matter of resistance to drugs for individual patients and provide a deeper understanding of treatments and risk factors, opening the door from novel targets and disease-related biomarkers to right drugs.

Last but not least, the effect of epigenetic changes on breast cancer etiology is beyond doubt. In spite of quite a number of DNA methylation research manifesting diverse patterns including tumor suppressor genes and oncogenes, only a small fraction of them connect the epigenome data with the transcriptome. In a recent study by Minning and coworkers[48], DNA methylation and gene expression profiling of primary breast tumor tissues and adjacent non-cancerous breast tissues was carried out. They preferred MS-MLPA or MS-qPCR for validation of results. The overlapping genes between DNA methylation and gene expression datasets were further mapped to the KEGG database to identify the molecular pathways linking the used genes together and supervised hierarchical clustering was used for data analysis. The authors found that most of the overlapping genes belong to the focal adhesion and extracellular matrix-receptor interaction that play important roles in breast carcinogenesis. The more gene signature data is acquired by different studies, the better understanding of epigenetic regulation of gene expression and remedial intervention will be possible.

Advances in bioinformatics and its application are much possible by multidisciplinary teams pursuing focused research. The sensitivity, specificity and combination of tools, methodologies, and databases should be evaluated in a complete matter. On top of that, findings must be confirmed with several molecular techniques before translation into clinical practice.

Database Group	Database	Originator	Web Address
Nucleotide Sequence	GenBank	US National Center for Biotechnology Information (NCBI)	www.ncbi.nlm.nih.gov/genbank
	EMBL	European Bioinformatics Institute	www.ebi.ac.uk/
	DDBJ	National Institute of Genetic, Japan	www.ddbj.nig.ac.jp/
	dbEST		www.ncbi.nlm.nih.gov/dbEST
Protein Sequence	SWISS-PROT	Swiss Institute of Bioinformatics, Geneva	web.expasy.org/docs/swissprot_guideline.html
		European Bioinformatics Institute	www.ebi.ac.uk/swissprot/
	TREMBLE	EBI (translation of coding sequences from the EMBL database that have not yet been deposited in SWISS-PROT)	www.ebi.ac.uk/tremble
	UniProt	Bioinformatics Institute (EMBL-EBI), Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR).	www.uniprot.org
	PIR	US National Biomedical Research Foundation (NBRF)	pir.georgetown.edu
		Japan International Protein Information Database (JIPID)	www.ddbj.nig.ac.jp
		Munich Information Center for Protein Sequences (MIPS)	mips.gsf.de

Table 1. Major electronic nucleotide and protein databases

Genome Browser	Originator	Web Address
Ensemble	Wellcome Trust Sanger Institute/ European Bioinformatics Institute(EBI)	www.ensembl.org/
NCBI Map Viewer	US National Center for Biotechnology Information(NCBI)	www.ncbi.nlm.nih.gov/mapview/
UCSC genome browser	Genome Bioinformatics Group of UC Santa Cruz	http://genome.ucsc.edu/
Genomes Compilations		
EBI Genomes	European Bioinformatics Institute(EBI)	www.ebi.ac.uk/genomes
GOLD	Genomes Online Database	www.genomesonline.org/

Table 2. Commonly used genom browser and databases

Author details

Beyzanur Yigitoglu¹, Eyyup Uctepe¹, Ramazan Yigitoglu², Esra Gunduz^{1*} and Mehmet Gunduz¹

*Address all correspondence to: mehmet.gunduz@gmail.com

1 Department of Medical Genetics, Faculty of Medicine, Turgut Ozal University, Ankara, Turkey

2 Medical Biochemistry, Faculty of Medicine, Turgut Ozal University, Ankara, Turkey

References

- [1] Ouzounis, C. The Emergence of Bioinformatics: Historical Perspective, Quick Overview and Future Trends. in *Bioinformatics in Cancer and Cancer Therapy* (ed. Gordon, G.J.) 1-11 (Humana Press, 2009).
- [2] Sims, A.H. Bioinformatics and breast cancer: what can high-throughput genomic approaches actually tell us? *J Clin Pathol* 62, 879-85 (2009).
- [3] Maryam Gholizadeh, S.A.P., Reza Pasandideh. Proteomics and Bioinformatics Approaches for Breast Cancer Researches. *International Journal of Agriculture and Crop Sciences* 5, 1863-1868 (2013).
- [4] Ouzounis, C.A. Rise and demise of bioinformatics? Promise and progress. *PLoS Comput Biol* 8, e1002487 (2012).

- [5] Economist, T. The race to computerise biology. *The Economist* (2002).
- [6] Daisuke Kihara, Y.D.Y., Troy Hawkins. Bioinformatics resources for cancer research with an emphasis on gene function and structure prediction tools. *Cancer Informatics* 2, 25-35 (2006).
- [7] Adams M.D., K.J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Meril C.R., Wu A., Olde B., Moreno R.F., et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252, 1651-1656 (1991).
- [8] Velculescu V.E., Z.L., Vogelstein B., Kinzler K.W. Serial analysis of gene expression. *Science* 270, 484-487 (1995).
- [9] Sydney Brenner, M.J., John Bridgham, George Golda, David H. Lloyd, Davida Johnson, Shujun Luo, Sarah McCurdy, Michael Foy, Mark Ewan, Rithy Roth, Dave George, Sam Eletr, Glenn Albrecht, Eric Vermaas, Steven R. Williams, Keith Moon, Timothy Burcham, Michael Pallas, Robert B. DuBridge, James Kirchner, Karen Fearon, Jen-i Mao, Kevin Corcoran. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* 18, 630-634 (2000).
- [10] Lon R. Cardon, G.R.A. Using haplotype blocks to map human complex trait loci. *Trends in Genetics* 19, 135-140 (2003).
- [11] Andrew G. Clark, R.N., James Signorovitch, Tara C. Matise, Stephen Glanowski, Jeremy Heil, Emily S. Winn-Deen, Arthur L. Holden, Eric Lai. Linkage Disequilibrium and Inference of Ancestral Recombination in 538 Single-Nucleotide Polymorphism Clusters across the Human Genome. *American Journal of Human Genetics* 73, 285-300 (2003).
- [12] Hector, B. The multitumor (sausage) tissue block: novel method for immunohistochemical antibody testing. *Laboratory Investigation* 55, 244-248 (1986).
- [13] Rennstam, K. & Hedenfalk, I. High-throughput genomic technology in research and clinical management of breast cancer. Molecular signatures of progression from benign epithelium to metastatic breast cancer. *Breast Cancer Res* 8, 213 (2006).
- [14] Vidya Vaidya, S.D. A Review of Bioinformatics Application in Breast Cancer Research. *Journal of Advanced Bioinformatics Applications and Research*. 1, 59-68 (2010).
- [15] Yang, X., Ai, X. & Cunningham, J.M. Computational prognostic indicators for breast cancer. *Cancer Manag Res* 6, 301-12 (2014).
- [16] Lonning, P.E., Knappskog, S., Staalesen, V., Chrisanthar, R. & Lillehaug, J.R. Breast cancer prognostication and prediction in the postgenomic era. *Ann Oncol* 18, 1293-306 (2007).
- [17] Per Eystein Lønning, R.C., Vidar Staalesen, Stian Knappskog, Johan Lillehaug. Adjuvant treatment: the contribution of expression microarrays. *Breast Cancer Research* 9, S14 (2007).

- [18] Jaluria, P., Konstantopoulos, K., Betenbaugh, M. & Shiloach, J. A perspective on microarrays: current applications, pitfalls, and potential uses. *Microb Cell Fact* 6, 4 (2007).
- [19] Dominick Sinicropi, M.C., Mei-Lan Liu. Gene Expression Profiling Utilizing Microarray Technology and RT-PCR. in *BioMEMS and Biomedical Nanotechnology* (eds. Ferrari, M., Ozkan, M. & Heller, M.) 23-46 (Springer US, 2007).
- [20] Eroles, P., Bosch, A., Perez-Fidalgo, J.A. & Lluch, A. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev* 38, 698-707 (2012).
- [21] Schiavon, G. *et al.* Heterogeneity of Breast Cancer: Gene Signatures and Beyond. 13-25 (2012).
- [22] Manning G, W.D., Martinez R, Hunter T, Sudarsanam S. The Protein Kinase Complement of the Human Genome. *Science* 298, 1912-1934 (2002).
- [23] Christos A. Ouzounis, R.M.R.C., Anton J. Enright, Victor Kunin & José B. Pereira-Leal. Classification schemes for protein structure and function. *Nature Reviews Genetics* 4, 508-519 (2003).
- [24] Strausberg RL, S.A., Old LJ, Riggins GJ. Oncogenomics and the development of new cancer therapies. *Nature* 429, 469-74 (2004).
- [25] Human Genome Research Institute, The Human Genome Project Completion. <http://www.genome.gov/11006943National>.
- [26] Duoqiao Wu, C.M.R., Xiangdong Wang. Cancer bioinformatics: A new approach to systems clinical medicine. *BMC Bioinformatics* 13(2012).
- [27] Wang, X. & Liotta, L. Clinical bioinformatics: a new emerging science. *J Clin Bioinforma* 1, 1 (2011).
- [28] Wang, X. Role of clinical bioinformatics in the development of network-based Biomarkers. *J Clin Bioinforma* 1, 28 (2011).
- [29] Kim, B. *et al.* Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data. *Cancer Res* 67, 7431-8 (2007).
- [30] Sawey, E.T. *et al.* Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. *Cancer Cell* 19, 347-58 (2011).
- [31] Zerbini LF, B.M., de Vasconcellos JF, Paccez JD, Gu X, Kung AL, Libermann TA. Computational repositioning and preclinical validation of pentamidine for renal cell cancer. *Molecular cancer therapeutics* 13, 1929-1941.
- [32] Ghia, P. *et al.* ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia* 21, 1-3 (2007).

- [33] Davi, F., Rosenquist, R., Ghia, P., Belessi, C. & Stamatopoulos, K. Determination of IGHV gene mutational status in chronic lymphocytic leukemia: bioinformatics advances meet clinical needs. *Leukemia* 22, 212-4 (2008).
- [34] Simon., R. Bioinformatics in cancer therapeutics--hype or hope? *Nature Clinical Practice Oncology* 2, 223 (2005).
- [35] Allinen, M. *et al.* Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* 6, 17-32 (2004).
- [36] Sjöblom T, J.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* 314, 268-274 (2006).
- [37] Wood L.D., P.D.W., Jones S., Lin J., Sjöblom T., Leary R.J., Shen D., Boca S.M., Barber T., Ptak J., Silliman N., Szabo S., Dezso Z., Ustyanksky V., Nikolskaya T., Nikolsky Y., Karchin R., Wilson P.A., Kaminker J.S., Zhang Z., Croshaw R., Willis J., Dawson D., Shipitsin M., Willson J.K., Sukumar S., Polyak K., Park B.H., Pethiyagoda C.L., Pant P.V., Ballinger D.G., Sparks A.B., Hartigan J., Smith D.R., Suh E., Papadopoulos N., Buckhaults P., Markowitz S.D., Parmigiani G., Kinzler K.W., Velculescu V.E., Vogelstein B. The genomic landscapes of human breast and colorectal cancers. *Science* 318, 1108-1113 (2007).
- [38] Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98, 10869-74 (2001).
- [39] Christos Sotiriou, S.-Y.N., Lisa M. McShane, Edward L. Korn, Philip M. Long, Amir Jazaeri, Philippe Martiat, Steve B. Fox, Adrian L. Harris, Edison T. Liu. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 100, 10393-8 (2003).
- [40] Loi, S. *et al.* Gene expression profiling identifies activated growth factor signaling in poor prognosis (Luminal-B) estrogen receptor positive breast cancer. *BMC Med Genomics* 2, 37 (2009).
- [41] Jack X Yu, A.M.S., Yi Zhang, John WM Martens, Marcel Smid, Jan GM Klijn, Yixin Wang, John A Foekens. Pathway analysis of gene signatures predicting metastasis of node-negative primary breast cancer. *BMC Cancer* 7, 182 (2007).
- [42] Behbod, F. & Rosen, J.M. Will cancer stem cells provide new therapeutic targets? *Carcinogenesis* 26, 703-11 (2005).
- [43] Yao, J. *et al.* Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. *Cancer Res* 66, 4065-78 (2006).

- [44] Frank Courjal, M.C., Joelle Simony-Lafontaine, Genevieve Louason, Paul Speiser, Robert Zeillinger, Carmen Rodriguez, and Charles Theillet. Mapping of DNA Amplifications at 15 Chromosomal Localizations in 1875 Breast Tumors: Definition of Phenotypic Groups. *cancer research* 57, 4360-4367 (1997).
- [45] Larissa Savelyeva, M.S. Amplification of oncogenes revisited: from expression profiling to clinical application. *Cancer Letters* 167, 115-123 (2001).
- [46] Soonmyung Paik, M.D., Steven Shak, M.D., Gong Tang, Ph.D., Chungyeul Kim, M.D., Joffre Baker, Ph.D., Maureen Cronin, Ph.D., Frederick L. Baehner, M.D., Michael G. Walker, Ph.D., Drew Watson, Ph.D., Taesung Park, Ph.D., William Hiller, H.T., Edwin R. Fisher, M.D., D. Lawrence Wickerham, M.D., John Bryant, Ph.D., Norman Wolmark, M.D. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *The new england journal of medicine* 351, 2817-2826 (2004).
- [47] Gokmen-Polar, Y. *et al.* Gene Expression Analysis Reveals Distinct Pathways of Resistance to Bevacizumab in Xenograft Models of Human ER-Positive Breast Cancer. *J Cancer* 5, 633-45 (2014).
- [48] Chin Minning, N.M.M., Norlia Abdullah, Rohaizak Muhammad, Nor Aina Emran, Siti Aishah Md Ali, Roslan Harun, Rahman Jamal. Exploring breast carcinogenesis through integrative genomics and epigenomics analyses. *international Journal of Oncology*, 1959-1968 (2014).

