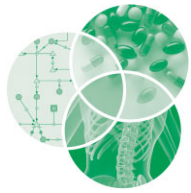
Focus Article



Genome network medicine: innovation to overcome huge challenges in cancer therapy

Dimitrios H. Roukos*

The post-ENCODE era shapes now a new biomedical research direction for understanding transcriptional and signaling networks driving gene expression and core cellular processes such as cell fate, survival, and apoptosis. Over the past half century, the Francis Crick 'central dogma' of single n gene/protein-phenotype (trait/disease) has defined biology, human physiology, disease, diagnostics, and drugs discovery. However, the ENCODE project and several other genomic studies using high-throughput sequencing technologies, computational strategies, and imaging techniques to visualize regulatory networks, provide evidence that transcriptional process and gene expression are regulated by highly complex dynamic molecular and signaling networks. This Focus article describes the linear experimentation-based limitations of diagnostics and therapeutics to cure advanced cancer and the need to move on from reductionist to network-based approaches. With evident a wide genomic heterogeneity, the power and challenges of next-generation sequencing (NGS) technologies to identify a patient's personal mutational landscape for tailoring the best target drugs in the individual patient are discussed. However, the available drugs are not capable of targeting aberrant signaling networks and research on functional transcriptional heterogeneity and functional genome organization is poorly understood. Therefore, the future clinical genome network medicine aiming at overcoming multiple problems in the new fields of regulatory DNA mapping, noncoding RNA, enhancer RNAs, and dynamic complexity of transcriptional circuitry are also discussed expecting in new innovation technology and strong appreciation of clinical data and evidence-based medicine. The problematic and potential solutions in the discovery of next-generation, molecular, and signaling circuitry-based biomarkers and drugs are explored.

How to cite this article:
WIREs Syst Biol Med 2014, 6:201–208. doi: 10.1002/wsbm.1254

INTRODUCTION

Cure of major diseases such as cancer, cardiovascular disorder, diabetes, schizophrenia, and others still remains elusive. Current biology, medicine and drugs discovery over the last 60 years are based on the dogma of reductionism, namely one cause (gene/protein)-one result (trait, phenotype) in a linear relationship. In clinical medicine, clinical symptoms and signs define the start point of diagnostic exploration. Pro-symptomatic disease risk prediction-based prevention or diagnosis at very very early stage can be associated with high cure rates. But despite long-term effort this goal has not been achieved. For example, delay in diagnosis of atherosclerosis with advanced lipids metabolism deregulation and hypertension represent not reversible cellular damage and although available drugs can prolong survival, the rates of coronary heart infarct, stroke, and death in this patients' group are alarmingly high.

*Correspondence to: droukos@uoi.gr

Centre for Biosystems & Genomic Network Medicine, Ioannina University, Ioannina, Greece

Conflict of interest: The authors declare no conflict of interest.

FROM 'CENTRAL DOGMA' TO GENOME NETWORK MEDICINE

Sixty years after the discovery of the DNA double helix by Crick, Watson and colleagues and the gene as the unit of hereditary information flow, we enter into the Genomic Network Medicine (GNM) era.[1] A decade after the completion of the Human Genome Project with the first draft of human genome sequence that rose the major hope for personalized medicine, we know that this goal is misleading suggesting the need to understand gene function itself beyond simple focus on protein-coding sequencing[2].

Since 1958, when Francis Crick[3] published his idea on reductionism-based transcription and translation, it has been the standard approach in biology and Life Science. Even today the 'central dogma' of the protein-coding DNA sequence which is transcribed into messenger RNA (mRNA) and translated into proteins and ultimately into an organism's phenotype represents the basic principle of linear experimentation in biomedical research and clinical medicine. We are now shifting from this simple concept that a single individual mutated gene deregulates transcription, changes protein structure and function leading to a trait, phenotype, or even chronic complex disease such as cancer, diabetes, mental, and other disorders, to a much more complicated approach of nonlinear connectivity, genomic motifs, genes, and transcripts clusters and molecular interactions network-based concept toward network medicine[4].

Indeed, the conventional reductionist approach has driven research in academia, biotechnology, and pharmaceutical industry. Nearly all currently available diagnostics, disease classification, prognostic and predictive markers and modern drugs are based on the 'single-gene transcription dogma'. In a recent catalog of all novel drugs which have been developed and approved by the FDA over the last three decades represent molecular targets of a single gene or protein[5]. However, with the recent publication of the ENCODE data[6] supporting a much more complex polygenic model and regulatory DNA[7] it becomes clear that transcriptional process is much more complex than previously thought[2]. This evidence together with slow progress in developing robust biomarkers and effective drugs for curing common multifactorial diseases, shape now a new avenue in understanding complex and dynamic evolution, biology, disease initiation, and treatment[8]. This focus article discusses latest technological developments in genome sequence[9], computational strategies, and imaging technologies for exploring transcriptional regulatory and signaling networks driving gene expression and cell function[10–15] and the challenges to translate genetic and genomic heterogeneity[16] including transcriptional high diversity[17,18] into clinical medicine advancement.

TARGET THERAPY

With standardization of nonspecific multimodal treatment, the expected further improvement from this treatment is small. Efforts therefore have been focused on targeted drugs. Over the past years an explosion in this therapeutic field has been occurred. More than 35 drugs have been approved by the FDA and many more are in preclinical and clinical staging for the treatment of cancer patients. All these anticancer drugs and other targeted agents are based on the Crick's single gene transcriptional concept. However, the efficacy of these individual signaling pathways inhibitors, which target a mutated gene or its encoded protein, is in most cases modest. The key conclusions, which can currently be drawn about the available target therapy include a temporary antitumor activity, high resistance rates, and a efficacy limited to a small genetic heterogeneity-based subgroup predicted on the basis of biomarkers[19,20].

Despite these substantial limitations, there are some isolated successful paradigms. True overall survival benefit has been rarely observed with a few only drugs. For example, the HER2 signaling pathway inhibitor trastuzumab in patients with HER2-positive breast cancer. Trastuzumab is the single available targeted drug, in a very common solid tumor such as breast cancer, which is effective also in the adjuvant setting and thus can be translated into increased cure rate. But even in this case, more recent evidence suggests high resistance rates[21]. This problem appears to partially be overcome by using trastuzumab emtansine conjugate[20,22].

Another recent paradigm that confirms the genomic cancer heterogeneity has recently published[23] proving the need to move away from 'a size-fits-all' concept. This approach has been widely used till recently by pharmaceutical industry aiming at treating all patients with a major cancer type with the same drug which can dramatically increase the economic gain for the companies. At the same time this study reveals how extremely difficult is to reach personalized clinical medicine. Lung cancer has the highest mortality among all cancers worldwide. Patients with nonsmall-cell lung cancer (NSCLC) and chromosomal rearrangements of the anaplastic lymphoma kinase (*ALK*) gene account for approximately 5% of NSCLC cases and define a distinct molecular subtype of lung cancer. For this small specific subgroup of patients,

a randomized, phase 3 trial comparing crizotinib, an oral tyrosine kinase inhibitor (TKI) targeting ALK, with standard chemotherapy in 347 patients with advanced, previously treated *ALK*-positive NSCLC showed improved clinical response in the crizotinib group. However, this antitumor activity could be translated into progression-free survival only without a true overall survival prolongation and it was associated with significantly higher adverse effects rate[23]. If we consider, in addition to these clinical limitations, also the high cost of both the crizotinib treatment and the ALK test in all NSCLC patients for identifying only ~5% with this rearrangement we can understand the slow progress of this reductionist approach.

GENOMIC HETEROGENEITY AND THE POWER OF NEXT-GENERATION SEQUENCING

Genes are crucial in human physiology. When mutations are accumulated disease can be developed. The unprecedented power of NGS to identify the mutational landscape at low cost, fast and accurately has revolutionized both biomedical research and more recently translational medicine[24]. The ability of applying NGS platforms in biological samples for the inter- and intra-patient assessment of genetic and genomic heterogeneity among patients with the same traditionally defined disease, such as for example cancer, raises the expectation to achieve accurate genomic classification and personalized management of patients[25]. These DNA sequencing machines allow the identification of all classes of mutations including point mutations such as single-nucleotide polymorphisms (SNPs) and insertions-deletions (indels) as well as larger structural changes such as inherited copy-number-variants (CNVs) or somatic copy-number-aberrations (CNAs) and also genomic translocation. All these sequences changes can be detected in both the protein-coding region which accounts for ~1.5% of the genome involving the ~21,000 genes by whole-exome sequencing (WES) and in the large 98.5% of noncoding region by whole-genome sequencing (WGS). Over the past decade ~140 cancer mutated genes have been identified for various cancer types[26] and only recently with the availability of WGS has started the effort to noncoding sequencing for whole-genome mapping of inherited and somatic mutations. This approach may have important clinical implications given the potential functional role of regulatory DNA[27–29].

Table 1 summarizes beyond recently reported cancer mutated genes[26], also cancer susceptibility loci more recently identified by genome-wide association studies GWAS[30–39] as well as new WGS and integrative genomic analysis-based cancer genes identification[40–44]. Despite refinement of genome-wide mapping technologies and evolution from second (NGS) to third-generation sequencing platforms[9] multiple challenges need to be overcome to integrate WES/WGS into routine clinic.

SEQUENCING STRATEGIES AND CLINICAL INTERPRETATION CHALLENGES

Accurate NGS-based genomic classification potentially allows for improved therapeutics of complex

TABLE 1 | Mutations, Genes, and Susceptibility Loci Involved in Some Cancer Types Identified by Genome Sequencing, Genome-Wide Association Studies (GWAS), and Integrative Genomic Analysis-Based Studies

Type of Cancer	Sequencing Number of Genes [ref.]	Number of Mutations	Latest GWAS	Integrative Genomic Analysis	Total Number of Susceptibility Loci	Total Number of Genes	Total Number (Genes–Mutations –Loci)
Breast	36[26]	111[26]	68[30,31]	6[40,41,42]	68	42	221
Ovarian	12[26]	42[26]	4[31]	—	4	12	58
Prostate	17[26]	41[26]	23[31]	—	23	17	81
CRC colorectal cancer	23[26]	66[26]	5[32]	2[43]	5	25	96
Pancreatic	19[26]	45[26]	3[33]	—	3	19	67
Lung	5[26]	310[26]	4[34]	—	4	5	319
Medulloblastoma	4[26]	8[26]	—	6[44]	—	10	18
Glioblastoma	2[26]	49[26]	2[35]	—	2	2	53
Gastric	4[26]	53[26]	7 genes, 11 loci[36–39]	—	11	11	75
Total	122	725	120	14	120	143	988

diseases. Targeting genetic alterations behind the individual patient identified by WES allows the selection among available drugs those which inhibit these mutated genes. However, this personalized approach is still at very initial stage and many hurdles exist to reach clinical decisions. First, the vast majority of mutations identified are passengers with neutral effect on cancer and should be distinguished from much fewer 'drivers' mutations within the genes (intronic) which are causatively involved in cancer[26,40–44]. With an expected dramatic increase in sample size of tumor-normal tissues pairs analysis in sequencing studies and the wide mutational heterogeneity, the false-positive rate of mutated genes responsible for cancer can be grow with currently used analytical methods to distinguish between driver and passenger mutations. Indeed, many large-scale international consortiums are underway including The Cancer Genome Atlas (TCGA)[45] and the International Cancer Genome Consortium[46] for a comprehensive list of cancer mutated genes in many cancer types. Indeed, a latest study assessed a high inaccurate rate by currently used methods and developed and propose the MutSigCV that enable the identification of genes truly associated with cancer[47]. Second, the catalog of genes and mutations is now incomplete. Given the cell-specificity revealed by the ENCODE data[6,7] and the heterogeneity sequencing of hundreds or thousands of cancer-tissues sample pairs for each common cancer type should be analyzed. Third, it appears essential to consider major clinical results from randomized phase 3 trials, meta-analyses and national statistics for example the U.S. cancer morbidity and mortality rates. These findings are crucial for clinical interpretation of sequencing data. For example, for many cancer types early-stage cancer is associated with high cure rates while in advanced and metastatic the disease is currently incurable. The number of driver mutated genes involved in tumorigenesis is small, about two to eight mutated genes and their dysfunction still at initial stage with low metastatic ability of tumor cells enabling high therapeutic response to surgery alone or plus adjuvant therapy. By contrast, in advanced disease not only this number of cancer-mutated genes is much larger but more importantly the deregulation of genome function, of multiple interacting gene expression circuitry and signaling networks is extremely complex, nearly chaotic[48]. This comprehensive aberrant intracellular network is probably associated with high metastatic potential and high death rates[26]. Analyzing genomic studies available, Vogelstein et al.[26] come to the conclusion that given the complexity of advanced cancer, much more effort and investments should be focused on prevention and early detection rather than treatment of advanced disease[26]. Indeed, beyond more complex genome structural landscape in advanced cancer, the understanding, prediction, and restore of whole-genome dysfunction appears currently a daunting challenge[8]. Despite these substantial challenges, emerging biomedical research is increasingly shifted into exploration of gene function. Challenges, innovative approaches, and major clinical expectations are discussed below.

GENE EXPRESSION: WHY IS COMPREHENSIVE UNDERSTANDING OF REGULATORY CIRCUITRY ESSENTIAL?

The need for translating gene expression profiling data into clinic is not new. High-throughput array technologies with the capacity to screen hundreds of genes simultaneously for differentially expressed genes assessment were popular over the past 15 years. Most of these studies evaluating clinical samples generated uncertainty but there are also promising findings. For example, a new molecular classification[49] and prognostic and predictive biomarkers in breast cancer[50] have been suggested. More recently, arrays-based research has moved into miRNAs and CNAs which contribute to gene expression[51]. However, none of these simplified gene expression 'signatures' have proven effective in phase 3 randomized trials and there is now increasing uncertainty whether a simple gene expression profiling or an individual biological system without deep understanding of transcriptional regulatory networks and a comprehensive view of dynamic molecular networks we will be able to reach human biology, physiology, and complex disease cure[52,53].

THE COMPLEXITY OF DYNAMIC CLINICAL GENOME ARCHITECTURE

Confirming the complexity of transcriptional regulatory networks and 3D spatial dynamics of sophisticated whole-genome function, the ENCODE project[6,7,10,11,27] changes the future of both biomedical research and clinical medicine[1]. It is important to note that the ENCODE project launched in 2003 for studying the functional elements of human genome was fully based on standard linear experimentation and reductionist approach at that time[7]. However, with the advent of NGS platforms few years later allowing protein-coding and noncoding sequences and the large number of GWAS disease-associated variants identified falling within the noncoding genome

region affecting transcriptional process and gene expression regulation, it was essential to study molecular networks to understand biology and disease[6,7].

FUTURE PERSPECTIVES: CLINICAL HOPE OR ENDLESS CHALLENGES

Refinement of second and third generation sequencing platforms, increased sequencing accuracy, and particularly the rapidly continuing dropping cost enable the 'driver' mutational landscape assessment, including intra-tumor genomic heterogeneity, for each individual patient. With the completion over the next few years of underway large international cancer genomic projects and many other individual NGS-based studies, there will be a substantial progress in decoding a patient's personal cancer-associated variants.

Will mutational landscape identification alone define optimal clinical therapeutic decision? Although this static knowledge of mutational heterogeneity can improve oncologists's decision on choosing from available drugs, the lacking understanding of how these mutations affect transcriptional, molecular, and signaling networks driving gene expression and cell behavior, raises uncertainty about a true survival benefit. This skepticism comes from: first the functionality of 80% of human genome along with the functional transcriptional role of mutations in the noncoding sequence and that we should consider the RNA transcript and not the gene as the fundamental unit of heredity by the ENCODE data. Second, these data reveal the importance of regulatory DNA[7] suggesting the need for mapping both inherited and somatic variation in regulatory DNA[54,55] to understand deregulated transcriptional activity in cancer cells. Third, latest evidence reveals that beyond genome-wide mutational heterogeneity, highly complex transcriptional heterogeneity drives gene expression[14,17,18]. Fourth, the regulation of transcriptional process is much more complex[2] than the standard linear dogma[3] and it is performed through dynamic networks[10,11]. However, the ENCODE project has completed the characterization of only 10% of human TFs and the transcriptional regulatory networks[1,11] still require validation before clinical implication. Fifth, TFs-binding start and end sites sequences along with RNA polymerase II (Pol II) at the transcription start site are crucial for understanding transcription and gene expression but research in this field is just now is starting[14]. Sixth, beyond the critical role of regulatory DNA, promoters and enhancers, a key contribution in transcriptional process have the ncRNAs including not only the well studied miRNAs but also lncRNA with intensive research only recently to explore their role in regulatory networks[56] along with RNA-binding proteins for determining post-transcriptional regulatory mechanisms[57] as well as on the enhancers RNAs (eRNAs) in breast cancer[58]. Seventh, epigenetic changes across the genome, chromatin state, and 3D shaping of gene networks in nuclear space substantially affect gene expression[59,60]. All these elements and functional processes, just now attract major interest after the publication of the ENCODE data and once all contribute to transcription and gene expression regulation confirming the complexity of global genome function understanding that can be crucial for clinical medicine.

IS CURRENT TECHNOLOGY SUFFICIENT OR IS INNOVATION REQUIRED TO IMPROVE CANCER MANAGEMENT AND HEALTH IN THE POST-ENCODE ERA?

The evidence of the importance of regulatory DNA and ncRNA mapping and dynamic complexity of transcriptional circuitry orchestrating gene expression has resulted in the development of new technologies to study transcription and gene expression in health and disease. However, beyond NGS-based RNA = seq and ChIP-seq for transcriptome analysis, computational strategies and networks visualization techniques in living cells for studying the highly dynamic aspect of regulatory networks, new innovation in technology and scientific methodology is required to, understand, predict, and translate into clinical medicine the dynamic complex circuitry regulating transcriptional process and gene function. In addition, new bioinformatic tools, algorithms, and powerful computers such as, for example, quantum computers[61] and cloud computing for storage, analysis, and transfer a highly tremendous volume of data increasingly generating by studying whole genome structure and function in model organisms and human physiology and disease. Novel software programs are also needed to be developed for translating big data already available by using linear experimentation such as for example those from more than 1200 GWAS and many related databases and the 1000 Genome project into networks principle-based analyses[7]. The future of medicine and cancer therapy is summarized in the terms of Network Medicine[4] and more recently into GNM[1] but it is clear that a long way with multiple hurdles are ahead of us.

CONCLUSIONS

Evidence-based medicine is the standard approach for improving health care. Recent evidence in basic science reveals the crucial role of transcriptional and signaling networks in the regulation of gene expression and cell behavior by the ENCODE project and other individual studies. In addition, current phase 3 trials and meta-analyses suggest slow progress in the management of cancer and other chronic common disorders. Both these data shape now a new epoch of medical research based on genome science and network biology advances. In this new field of GNM, appreciation of clinical data and innovation in technology and science is required. Indeed, as continuously new knowledge arises, there has been uncertainly on the endless complexity of human genome functionality and regulation and the priority which should be given in specific biomedical research areas most likely to achieve translational and clinical implications. For example by comparing traditionally classified homogenous groups of cancer patients with positive (disease-free survival) and negative outcome (therapeutic resistance/recurrence/death) by applying latest technologies in biological samples (NGS, computational, and visualization techniques) it can result in clinically important discoveries.

Exploring and assessing not only the driver mutational landscape but also the aberrant signaling networks and gene expression deregulation behind the phenotype, namely recurrence or disease progression because of drugs resistance, can reveal molecular mechanisms of treatment failure. Thus, beyond a simple WGS, a comprehensive understanding of transcriptional networking process and whole-genome function can be needed to achieve the next-generation of robust biomarkers and novel genome-scale molecular signaling network-based druggable targets.

REFERENCES

1. Roukos DH. Genome network medicine: new diagnostics and predictive tools. *Expert Rev Mol Diagn* 2013, 13:643–646.
2. Ball P. DNA: celebrate the unknowns. *Nature* 2013, 496:419–420.
3. Crick FH. On protein synthesis. *Symp Soc Exp Biol* 1958, 12:138–163.
4. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011, 12:56–68.
5. Rask-Andersen M, Almén MS, Schiöth HB. Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov* 2011, 10:579–590.
6. ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012, 489:57–74.
7. Stamatoyannopoulos JA. What does our genome encode? *Genome Res* 2012, 22:1602–1611.
8. Roukos DH, Baltogiannis GG, Katsouras CS, Bechlioulis A, Naka KK, Batsis C, Liakakos T, Michalis LK. Novel next-generation sequencing and networks-based therapeutic targets: realistic more effective drug design and discovery. Curr Pharm Des (Epub ahead of print; March 19, 2013).
9. Ku CS, Roukos DH. From next-generation sequencing to nanopore sequencing technology: paving the way to personalized genomic medicine. *Expert Rev Med Devices* 2013, 10:1–6.
10. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012, 489:91–100.
11. Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA. Circuitry and dynamics of human transcription factor regulatory networks. *Cell* 2012, 150:1274–1286.
12. Depry C, Mehta S, Zhang J. Multiplexed visualization of dynamic signaling networks using genetically encoded fluorescent protein-based biosensors. *Pflugers Arch* 2013, 465:373–381.
13. Yosef N, Shalek AK, Gaublomme JT, Jin H, Lee Y, Awasthi A, Wu C, Karwacz K, Xiao S, Jorgolli M, et al. Dynamic regulatory network controlling TH17 cell differentiation. *Nature* 2013, 496:461–468.
14. He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. *Nature* 2013, 495:481–486.
15. Chung K, Wallace J, Kim SY, Kalyanasundaram S, Andalman AS, Davidson TJ, Mirzabekov JJ, Zalocusky KA, Mattis J, Denisin AK, et al. Structural and molecular interrogation of intact biological systems. *Nature* 2013, 497:332–337.
16. Roukos D. Integrative deep-sequencing analysis of cancer samples: discoveries and clinical challenges. *Pharmacogenomics J* 2013, 13:205–208.
17. Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 2013, 497:127–131.

18. Pugh BF. Molecular biology: the ends justify the means. *Nature* 2013, 497:48–49.

19. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011, 144:646–674.

20. Cho WC, Roukos DH. Trastuzumab emtansine for advanced HER2-positive breast cancer and beyond: genome landscape-based targets. *Expert Rev Anticancer Ther* 2013, 13:5–8.

21. Martin-Castillo B, Oliveras-Ferraros C, Vazquez-Martin A, Cufí S, Moreno JM, Corominas-Faja B, Urruticoechea A, Martín ÁG, López-Bonet E, Menendez JA. Basal/HER2 breast carcinomas Integrating molecular taxonomy with cancer stem cell dynamics to predict primary resistance to trastuzumab (Herceptin). *Cell Cycle* 2013, 12:225–245.

22. Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, Pegram M, Oh DY, Diéras V, Guardino E, et al, EMILIA Study Group Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med* 2012, 367:1783–1791.

23. Shaw AT, Kim DW, Nakagawa K, Seto T, Crinó L, Ahn MJ, De Pas T, Besse B, Solomon BJ, Blackhall F, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* 2013, 368:2385–2394.

24. Roukos DH, Ku CS. Clinical cancer genome and precision medicine. *Ann Surg Oncol* 2012, 19: 3646–3650.

25. Roukos D, Batsis C, Baltogiannis G. Assessing tumor heterogeneity and emergence mutations using next-generation sequencing for overcoming cancer drugs resistance. *Expert Rev Anticancer Ther* 2012, 12:1245–1248.

26. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science* 2013, 339:1546–1558.

27. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012, 337:1190–1195.

28. Roukos DH, Baltogiannis GG, Baltogiannis G. Mapping inherited and somatic variation in regulatory DNA: new roadmap for common disease clinical discoveries. *Expert Rev Mol Diagn* 2013, 13:519–522.

29. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res* 2012, 22:1748–1759.

30. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013, 45:353–361.

31. Sakoda LC, Jorgenson E, Witte JS. Turning of COGS moves forward findings for hormonally mediated cancers. *Nat Genet* 2013, 45:345–348.

32. Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, Edlund CK, Haile RW, Gallinger S, Zanke BW, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet* 2012, 131:217–234.

33. Parikh H, Jia J, Zhang X, Chung CC, Jacobs KB, Yeager M, Boland J, Hutchinson A, Burdett L, Hoskins J, et al. A resequence analysis of genomic loci on chromosomes 1q32.1, 5p15.33, and 13q22.1 associated with pancreatic cancer risk. *Pancreas* 2013, 42:209–215.

34. Zhang R, Zhao Y, Chu M, Wu C, Jin G, Dai J, Wang C, Hu L, Gou J, Qian C, et al. Pathway analysis for genome-wide association study of lung cancer in Han Chinese population. *PLoS One* 2013, 8(3):e57763.

35. Zhang JX, Zhang J, Yan W, Wang YY, Han L, Yue X, Liu N, You YP, Jiang T, Pu PY, et al. Unique genome-wide map of TCF4 and STAT3 targets using ChIP-seq reveals their association with new molecular subtypes of glioblastoma. *Neuro Oncol* 2013, 15:279–289.

36. Yang XX, Li FX, Zhou CP, Hu NY, Wu YS, Li M. Association of genetic polymorphisms at 1q22 but not 10q23 with gastric cancer in a southern Chinese population. *Asian Pac J Cancer Prev* 2012, 13: 2519–2522.

37. Wang M, Zhang R, He J, Qiu L, Li J, Wang Y, Sun M, Yang Y, Wang J, Yang J, et al. Potentially functional variants of PLCE1 identified by GWASs contribute to gastric adenocarcinoma susceptibility in an eastern Chinese population. *PLoS One* 2012, 7(3): e31932.

38. Zhang H, Jin G, Li H, Ren C, Ding Y, Zhang Q, Deng B, Wang J, Hu Z, Xu Y, et al. Genetic variants at 1q22 and 10q23 reproducibly associated with gastric cancer susceptibility in a Chinese population. *Carcinogenesis* 2011, 32:848–852.

39. Abnet CC, Freedman ND, Hu N, Wang Z, Yu K, Shu XO, Yuan JM, Zheng W, Dawsey SM, Dong LM, et al. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet* 2010, 42:764–767.

40. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012, 486:395–399.

41. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 2012, 486:405–409.

42. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of

2,000 breast tumours reveals novel subgroups. *Nature* 2012, 486:346–352.

43. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, Chaudhuri S, Guan Y, Janakiraman V, Jaiswal BS, et al. Recurrent R-spondin fusions in colon cancer. *Nature* 2012, 488:660–664.
44. Jones DT, Jäger N, Kool M, Zichner T, Hutter B, Sultan M, Cho YJ, Pugh TJ, Hovestadt V, Stütz AM, et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature* 2012, 488:100–105.
45. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012, 487:330–337.
46. International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, Bhan MK, Calvo F, Eerola I, et al. International network of cancer genome projects. *Nature* 2010, 464: 993–998.
47. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013, 499:214–218.
48. Fernández A, Lynch M. Non-adaptive origins of interactome complexity. *Nature* 2011, 474:502–505.
49. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. Molecular portraits of human breast tumours. *Nature* 2000, 406:747–752.
50. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004, 351:2817–2826.
51. Dvinge H, Git A, Gräf S, Salmon-Divon M, Curtis C, Sottoriva A, Zhao Y, Hirst M, Armisen J, Miska EA, et al. The shaping and functional consequences of the microRNA landscape in breast cancer. *Nature* 2013, 497:378–382.
52. Alberts B. Model organisms and human health. *Science* 2010, 330:1724.
53. Anonymous. Time for the epigenome, editorial. *Nature* 2010, 463:587.
54. Thurman RE, Day N, Noble WS, Stamatoyannopoulos JA. Identification of higher-order functional domains in the human ENCODE regions. *Genome Res* 2007, 17:917–927.
55. Vernot B, Stergachis AB, Maurano MT, Vierstra J, Neph S, Thurman RE, Stamatoyannopoulos JA, Akey JM. Personal and population genomics of human regulatory variation. *Genome Res* 2012, 22:1689–1697.
56. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature* 2012, 489:109–113.
57. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 2013, 499:172–177.
58. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 2013, 498:516–520.
59. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. The accessible chromatin landscape of the human genome. *Nature* 2012, 489:75–82.
60. de Wit E, Bouwman BA, Zhu Y, Klous P, Splinter E, Verstegen MJ, Krijger PH, Festuccia N, Nora EP, Welling M, et al. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* 2013, 501:227–231.
61. Powell D. Quark quartet opens fresh vista on matter. *Nature* 2013, 498:280–281.

