Genomic Sequencing in Cancer

Musaffe Tuna and **Christopher I. Amos**
Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX

Abstract

Genomic sequencing has provided critical insights into the etiology of both simple and complex diseases. The enormous reductions in cost for whole genome sequencing have allowed this technology to gain increasing use. Whole genome analysis has impacted research of complex diseases including cancer by allowing the systematic analysis of entire genomes in a single experiment, thereby facilitating the discovery of somatic and germline mutations, and identification of the function and impact of the insertions, deletions, and structural rearrangements, including translocations and inversions, in novel disease genes. Whole-genome sequencing can be used to provide the most comprehensive characterization of the cancer genome, the complexity of which we are only beginning to understand. Hence in this review, we focus on whole-genome sequencing in cancer.

1. Introduction

Genomic alterations, including mutations, copy number changes and structural rearrangements, are the hallmarks of cancer. Whole-genome sequencing (WGS) enables investigators to identify all point mutations and structural rearrangements in place of previous methods that were both costly and inefficient because they could only target specific attributes. Some of alterations in the genome are germline and predispose individuals to cancer, but most alterations in the cancer genome are somatic, and WGS enables researchers to identify all point mutations, indels and structural rearrangements in both germline and somatic tissues. In the past decade, revolutionary advances in genome technology including next-generation sequencing together with advances in analytical tools have led to an improved understanding of the mechanisms underlying cancer pathogenesis. These advances have also enabled researchers to more accurately describe sub-classifications of cancer, predict outcomes in cancer patients, select effective cancer treatments and personalize cancer therapy. The technology of next-generation sequencing is rapidly advancing, and hence in this review manuscript, we give a snapshot overview of next-generation sequencing technologies with current information, and summarize and discuss some of the important findings that have been generated by WGS in cancer.

2. First generation analyses

First-generation sequencing platforms have been used to detect mutations for a limited number of base pairs, therefore, using these platforms to sequence the whole-genome is challenging. The detection of each type of genomic alteration without genomic sequencing requires application of different platforms. For example, DNA microarrays are used to detect DNA copy number alterations in the genome, whereas RNA microarrays are used to identify transcriptomic variation, and Sanger (capillary) sequencing is used to interrogate mutations in a small number of genes.

One of the first-generation sequencing techniques, enzymatic dideoxy DNA sequencing, which is based on chain-terminating dideoxynucleotide analogues, was introduced in 1975 [1,2], and the second technique based on chemical degradation was introduced in 1977 [3]. In the second sequencing technique, terminally labeled DNA fragments are chemically cleaved at specific bases and separated by gel electrophoresis [3]. These two approaches were used to identify cancer-specific somatic mutations in the RAS gene family [4–7], and *RB1* and *DPC4* mutations in human tumors [8,9]. During the two decades following the introduction of first-generation sequencing, remarkable progress was made in the development of automated sequencing instruments [10], which enabled the initial sequencing of the first human genome [11]. Since then, systematic efforts have been made to sequence gene families to identify oncogenic mutations that are targets for cancer therapy, such as *EGFR* and *PIK3CA* [12–16].

3. Next-generation sequencing

In recent years, single-gene sequencing using first-generation approaches has largely been replaced with comprehensive genome-wide sequencing using next-generation sequencing (NGS) techniques, which are classified as either second- or third-generation approaches.

3.1. Second-generation sequencing platforms

The first second-generation sequencing system on the market was developed by 454 Life Sciences (Roche) in 2005. Unlike Sanger sequencing, the Roche system depends on the detection of pyrophosphate release on nucleotide incorporation [17]. In principle, this pyrosequencing technique is based on "sequencing by synthesis". Several other second-generation approaches soon followed the introduction of the Roche approach. These second-generation approaches include SOLID, which depends on the sequential ligation of oligonucleotide probes; and ABI, which is sequencing by reversible dye terminators; and Illumina, which is based on bridge enzymatic amplification. In the Roche and ABI platforms, the amplification method and template preparation are based on emulsion PCR. Several excellent reviews have provided a detailed discussion of the different technologies [18–20]. Each platform generates various base read lengths and is susceptible to various error rate and error profiles relative to those introduced in Sanger sequencing (Table 1) [21]. Improvements in error rates may be reduced with modifications of current second generation sequencing approaches, development third-generation sequencing techniques in which sequencing is determined directly from a single DNA molecule without the need for PCR amplification [22]. Another approach is to develop a hybrid sequencing platform in which the advantages of each sequencing technology are retained and the disadvantages of each technology eliminated, and developing novel algorithms and analysis tools to integrate data across platforms. In addition, increasing depth of coverage reduces the error rate of the assembled sequence and improves sequencing accuracy. Using paired-end reads rather than single-end reads could also reduce errors in the assembly. Longer reads are generally better than shorter ones, as they reduce false positives from mapping ambiguity.

3.2. Third-generation sequencing platforms

Currently, there are four different third–generation sequencing techniques, and there are more under development. One of the first third-generation sequencing techniques was introduced by Heliscope in 2007 [23]. Heliscope sequencing depends on "true single molecule sequencing," which allows DNA sequencing, and "one-base-at-a-time" nucleotide technology, which allows direct RNA sequencing without cDNA library construction [24].

Another third-generation sequencing approach, single-molecule real time (SMRT) sequencing detection, depends on a fluorescence detection system to directly detect each nucleotide, which is phosphor-linked with distinct colors, as they are synthesized without amplification (Pacific Biosciences). During the synthesis process, fluorescence is emitted as the phosphate chain is cleaved, and the nucleotide is incorporated by a polymerase into a single DNA strand. In Oxford nanopore sequencing, sequencing detection relies on the conversion of the electrical signal of nucleotides as they pass through a nanopore, an -hemolysin pore covalently attached to a cyclodextrin molecule [25,26]. Another third-generation sequencing platform is fluorescence resonance energy transfer (FRET) sequencing detection depends on light emission (VisiGen Biotechnologies), which is real-time single-molecule sequencing. This approach does not require cloning and amplification, which eliminates a large part of the cost, relative to current technologies. In addition, read lengths for the instrument are expected to be around 1 kb, longer than the most of current platforms. Solid tumors are often composed of multiple clonal subpopulations, and complex mixtures of cells including non-cancerous fibroblasts, endothelial cells, lymphocytes, and macrophages. It is challenging to define clonal subpopulations and distinct non-cancerous cell from tumor cells with second-generation platforms. Single-molecule sequencing may overcome from this issue. The other third-generation platform is Ion Torrent semiconductor sequencing, which is based on the release of hydrogen ions as a byproduct of nucleotide chain elongation and the detection of pH changes by an ion sensor during DNA synthesis.

3.3. Application of next-generation sequencing

Next-generation sequencing technologies have been successfully applied to whole-genome, transcriptome or whole-exome depending upon the type of input material to cancer samples. If the input material is RNA, it is called whole-transcriptome sequencing. If the input material is DNA the product of the analysis will depend upon the specification. If an initial amplification step is employed, either targeted sequencing or whole genome sequencing can be performed. For targeted sequencing the target can either be a region of interest or all of the exons in the genome, or the exome. When targeted sequencing is requested, an initial capture step is required and the targeted capture can suffer its own biases.

3.3.1. Whole-transcriptome sequencing—Whole-transcriptome sequencing (RNA-seq) is a powerful approach to understanding cancer derived from cDNA (mRNA, total RNA, miRNA, and others). RNA-seq can be used to 1) define known and discover novel splice variants and expressed SNPs; 2) characterize and catalogue the transcripts expressed within a specific cell or tissue—at a particular stage, and quantify the differential expression of transcripts, including chimeric transcripts, generated by somatic structural genome rearrangements; 3) identify allele-specific expression, 4) RNA-seq, 6) can also provide additional insight into the regulation of gene transcription and RNA processing during tumorigenesis. RNA-seq is not limited to known genes, and can be used to detect novel transcripts, alternative splice forms, and non-human transcripts (microbiomes).

3.3.2. Whole-exome sequencing—Whole-exome sequencing (WES) began as exome only, but with subsequent advances, more pieces of the genome including non-coding DNA in exon-flanking regions, promoters and untranslated regions (UTRs) have been added to the

capture kit. WES now covers a much broader range of the genome, including all coding exons; microRNA genes; 5′ UTRs and 3′ UTRs; unannotated transcripts discovered in RNA-seq experiments or the ENCODE project, and all "functional" portions. Sequencing the coding exons is a powerful approach to discovering the activating somatic mutations from specific gene families in cancer. WES is a cost-effective, high-coverage approach to detecting activating and inactivating mutations in known coding genes across the entire genome. Improved exome sequencing technology will be a powerful diagnostic tool to detect known mutations in a large number of cancer samples and discover novel mutations in the cancer genome. One limitation of exome sequencing is that it can only identify those mutations and/or variants that are found in the coding regions of genes that affect protein function. Unlike WGS, WES cannot identify the structural and non-coding variants (intronic variants related to splicing and enhancer variants) associated with a disease. Genome-wide association studies have found that >80% of cancer susceptibility-associated variants are located outside coding regions [27]. These limitations are overcome with the whole genome sequencing approach, which allows us to gain a deeper understanding of genetic variation found in populations and provides information about all types of genetic and genomic alterations found in the cancer genome. The limitation of WGS is the high cost and time associated with sequencing the genome. On the other hand, the advantage of exome sequencing is to the much lower burden of sequencing a very small proportion of the genome, approximately only 1% of the sequencing requirement of WGS. This reduction in sequencing burden allows the investigator to sequence either more samples or at higher depth given limited resources.

3.3.3. Whole-genome sequencing—WGS provides a single approach to identifying the full range of true somatic genomic alterations: nucleotide substitution mutations, indels, rearrangements of repetitive elements, microbial infections, active retrotransposons, copy number alterations, and structural rearrangements including inversions, translocations and complex rearrangements in whole-genome. This includes the intergenic, genic, and regulatory regions of the genome. In WGS of cancer, matched normal samples must be used to distinguish true somatic mutations from inherited changes, and compared to a reference genome. The number of mutations and structural rearrangements in the cancer sequence and their absence in the matched normal sequence must be assessed to identify meaningful differences. In addition, subclonal events can be defined using whole-genome sequencing.

One of the unique strengths of WGS is that it can be used to identify the breakpoints in balanced chromosome translocations and inversions, and also provides information on a genome that is orders of magnitude larger than that provided by the previous genotyping technology, DNA arrays. For humans, DNA arrays currently provide genotypic information on up to five million genetic variants, while WGS provides information on all six billion bases in the human genome, of which ~4 million nucleotides are polymorphic. Of these 3.5 million are single nucleotide polymorphisms and there are additional insertion deletion (indel) and structural variants such as inversions and microsatelites.

Tremendous effort has been put forth to develop next-generation sequencing technologies. Such effort requires the generation of new technologies for the generation of data as well as the development of analytical tools with high sensitivity and specificity to enhance the discovery of mutations, indels and structural rearrangements. NGS technologies have the potential to revolutionize our understanding of cancer as a disease of the genome. Inherited variants that contribute to cancer susceptibility will be discovered using WGS. Thus, this will greatly enhance our understanding of inherited disease, genetic risk factors for cancer, and the somatic changes that initiate cancer and/or metastasis, thereby leading to improved tumor classification and facilitate personalized therapy. NGS will lead to the development of

screening, diagnostic, and prognostic assays, and targeted therapies, and the discovery of predictive and prognostic biomarkers.

4. Applications of WGS

Point mutations are not the only alterations in cancer genome. Therefore tumorigenesis results from more than mutations in genes. One of the major findings of WGS in cancer has been the discovery of many new fusion genes, and chromosomal and complex rearrangements. Fusion genes result from inter-chromosomal translocations and intra-chromosomal inversions, deletions, tandem duplications and aberrant splicing. The breakpoints that cause fusion genes are located not only between open reading frames, but also between known genes and intergenic regions, and between open reading frames and miRNAs.

Until recently, the identification of chromosomal translocations was mostly limited to hematologic malignancies and sarcomas using cytogenetic methods [28]. Thus chromosomal translocations were believed to be rare in epithelial tumors. However, this perception changed following the discoveries of the fusion of transmembrane protease serine 2 (*TMPRSS2*)–*ERG* [(v-ets erythroblastosis virus E26 oncogene homologue (avian)] in prostate carcinoma [29] and the echinoderm microtubule-associated protein like 4 (*EML4*)– anaplastic lymphoma receptor tyrosine kinase (*ALK*) translocations in non-small cell lung [30], breast, and colorectal cancers [31], and rearrangements of the transcription factor *ETS* in prostate cancer [29,32]. A large and growing number of pathogenic chromosomal translocations have since been discovered in multiple solid tumors including breast cancer, lung cancer, prostate cancer, melanoma, hepatocellular carcinoma, pancreatic cancer, medulloblastoma and colorectal adenocarcinoma, and hematologic malignancies including ETP-acute lymphocytic leukemia, relapsed acute myeloid leukemia, acute promyelocytic leukemia, and multiple myeloma [33–65].

Accumulating data reveal that solid tumors have more translocations than previously thought. As of August 2012, of the 340,585 mutations, included in the Catalogue of Somatic Mutations in Cancer, 170,263 are unique variants and 8,004 are fusion genes (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/). However, most of the fused genes are not known to be oncogenes (Supplementary Table 1), furthermore, the role that many of these genes have in tumorigenesis remains unknown. Since the first report of WGS data and analysis of a tumor and matching normal sample in 2008, the number of mutations, fusions, and structural rearrangements has grown exponentially. In 2008, the WGS had only been applied to three primary tumors; one acute myeloid leukemia and two lung cancer cell lines. Since then, WGS data of 487 primary tumor types has been reported.

One of the very first WGS study yielded an improved classification of a hematologic malignancy. The patient initially was diagnosed with acute myeloid leukemia (AML) and received all-trans-retinoic acid chemotherapy treatment for this disease, until defined cryptic fusions (*APL-ATRA*) were identified, that were not consistent with AML. The diagnostic conundrum was solved when WGS detected previously unseen breakpoints that resulted in a cryptic fusion oncogene consistent with non-ATRA-resistant acute promyeolocytic leukemia, which had not been initially detected with cytogenetics [62]. Other findings of fusion genes followed this clinically relevant study. For example, *ETV6-ITPR2* and *NFIA-EHF* fusions have been found in breast cancer [59]. *ETV6* is known to fuse with different genes to form cancer genes in leukemia [66], and ITPR2 encodes inositol-1,4,5-triphosphate receptor type 2 which is involved in signal transduction and the regulation of cellular calcium fluxes. *NFIA* is a transcription factor involved in adenovirus replication. Recurrent *MAGI3* (membrane associated guanylate kinase, WW and PDZ domain containing 3)–

AKT3 (v-akt murine thymoma viral oncogene homologue 3) fusion has been reported in triple-negative breast cancer [58]. AKT3 is one of the members of AKT family (AKT1, AKT2 and AKT3). AKT is a serine/threonine-specific protein kinase that plays a key role in multiple cellular processes such as apoptosis, cell proliferation, cell survival, and cell migration, as well as glycogen synthesis and glucose uptake. Akt1 was originally identified as oncogene, and has been implicated as a major factor in many types of cancer. *AKT3* shares significant homology with *AKT1,* and so is thought to have a major role in phosphorylation relating to cell proliferation. *B2M-TCF12* fusion has been found in lung cancer [43], *PTV1-CHD7* fusion in small cell lung cancer [49], *PREX2-C11orf30* fusion in melanoma [34], and *BCORL1-ELF4* fusion in hepatocellular carcinoma.

WGS facilitates the discovery of other types of gene-disrupting rearrangements, including tandem duplication, inversions and deletions. For example, rearrangements in several known cancer genes may have activated *BRAF, PAX3, PAX5, NSD1, PBX1, MSI2* and *ETV6*, or inactivated *RB1, APC*, and *FBXW7* in breast cancer [59]. Rearrangements in other genes, including *CADM2, PTEN,* and *MAGI2* in prostate cancer [35], *FHIT, WWOX, MACROD2, CSMD1, MAGI2,* and *A2BP1* in melanoma [34] may contribute to cancer development. Structural rearrangements in the *MAGI2* have also been reported in melanoma [34,35]. MAGI2 (membrane associated guanylate kinase, WW and PDZ domain containing 2) gene encodes a PTEN-interacting protein, indicating that *PTEN* modulation may play a role in disrupting different mechanisms in prostate cancer and melanoma that had previously not been recognized.

WGS can also be used to identify complex rearrangements. For example, complex structural rearrangements reported in metastatic melanoma include multiple genes *ETV1, CPA6, TRPA1, C11orf30, STAU2, TTYH3, ICA1, PER4, CARD11* and *PREX2* (phosphatidylinositol-3,4,5-trisphosphate-dependent RAC exchange factor 2), which interacts with the *PTEN* tumor suppressor and modulates its function. Interestingly, *PREX2* is one of the most frequently mutated in melanoma tumor samples [34]. This indicates that multiple mechanisms dysregulate *PREX2* in melanoma, and that *PREX2* may has a distinct function in melanoma. Another complex rearrangement with multiple genes includes *ETV1, RASGRP1, ODZ4, SUSD1, LFNG,* and *PREX2* in melanoma [34]. WGS also facilitates the discovery of cancer predisposition genes and can identify novel rare variants, such as germline mutations in *ATM* gene were identified in two family with hereditary pancreatic cancer [67].

5. Significance of mutated genes in the cancer genome

WGS has revealed mutated genes in multiple tumor types. However, the number and types of driver mutations are highly variable, likely reflecting the differential mutational pressures on individual tumors. For example, *STK11* is highly mutated in lung cancer in smokers but not in other cancer types. On the other hand, some driver mutations are common in multiple tumor types. *TP53* is the only uniformly mutated gene with high frequency in medulloblastoma, pancreatic cancer, breast cancer, lung adenocarcinoma, colorectal adenocarcinoma, hepatocellular cancer, multiple myeloma, chronic lymphocytic leukemia and secondary AML (Supplementary Table 1) [33,37,39–41,46,49,51,54,55,58,61,64,65]. Moreover, high frequencies of *PIK3CA* and/or *PIK3R1* mutations have been found in multiple tumor types including breast cancer, medulloblastoma, colon and rectal cancer and lung adenocarcinoma (Supplementary Table 1).

Interestingly, recent studies have shown that some mutated genes, with either the same or distinct mutations, are common in multiple types of cancer, such as the *IDH1* mutations in gliomas differed from those in AML. In particular, R132C mutation occurred in 50% and

4% of AML and gliomas patients, respectively. On the other hand, R132H mutation is the predominant variant in gliomas (88%) but has a significantly lower incidence in AML (44%) [45,68]. However, both *IDH1* and *IDH2* mutations frequently occur, and predict a poor prognosis in cytogenetically normal AML [69]. The comparative lesion sequencing of primary tumors and their metastases revealed that a relatively small number of additional mutations were needed to transform the precursor primary tumor into metastatic disease [41,52]. It is crucial to define the same mutations that are common in multiple tumors that can be used as target to therapy, additional to tumor specific mutations. For example, *BRAF* is the most commonly mutated oncogene in melanoma, occurring in 50–60% of tumors and the most prevalent mutation is a missense mutation in *BRAF*, which accounts for 90% of all *BRAF* mutations, results in a substitution of glutamic acid to valine at codon 600 (BRAF^{V600E}), and is the target for the BRAF inhibitor vemurafenib. Interestingly, BRAF^{V600E} mutation has been reported in other cancers including colon cancer, serous ovarian cancer, lung cancer and papillary thyroid carcinoma. The question that remains to be answered is: whether BRAF inhibitors can be used to treat other types of BRAF^{V600E} positive tumors and even BRAF translocation positive tumors such as occur in prostate cancer [70]. Thus, the treatment of cancer is likely to move from organ-specific approach to a gene-dependent one [71]. On the other hand, one therapeutic targeted therapy is likely not enough to cure many types of tumor. Hence multiple targets may usually be needed and a combination of therapies may be required for treatment. In addition, a combination of different genes are often mutated in different individuals with the same type of tumor, therefore, therapeutic targets may vary among individuals. Hence, WGS not only allows us to identify significant mutated or driver genes, but also potential therapeutic targets, as well as mutations that are sensitive or resistant to certain therapeutic agents, such as multiple potential therapeutic targets were identified in non-small cell lung cancer, including *EGFR, HGF, MET, JAK2, EPHA3, BRAF, PIK3CG, IGF1R, MET, RET, FGFR1, HDAC1, HDAC2, HDAC6* and *HDAC9* mutations; fused: *KDELR2-ROS1* and *EML4-ALK* fusions [72].

6. Mutation rate

Sequencing studies have revealed the great heterogeneity of somatic mutations and mutation signatures among cancer types, individual tumors of the same cancer lineage type, and intergenic, genic, exon and intron regions, regulatory regions, 5 UTRs and 3 UTRs. The number of mutations in tumor types is very variable. For example, in medulloblastoma, the lowest mutation rate is 0.15–0.6 per Mb [51], in early T-cell precursor acute lymphocytic leukemia, 0.3 per Mb [57], in chronic lymphocytic leukemia, <1 per Mb [50,55], in prostate cancer, 0.9 per Mb [35], in multiple myeloma, 2.9 per Mb [37]. In contrast, most solid tumors have more than one mutation per one million bases: in colorectal adenocarcinoma the frequency is ~5.0 per Mb [33], in hepatocellular carcinoma, it is 4.2 per Mb [40], in breast cancer, it is 1.18–1.66 per Mb [39,58], and the highest rates occur in melanoma, ~ 30 per Mb [34] and in lung cancer, 17.7 per Mb [43], in small-cell lung cancer, 7.4 per Mb [73]. The variations in mutation rate are also correlated with disease subtypes. For example, the mutation rate in chronically ultraviolet radiation-induced melanoma is 111 per Mb whereas the mutation rates in non-ultraviolet radiation-induced melanomas on the hairless skin of the extremities and on hair-bearing skin are only 3–14 and 5–55 per Mb, respectively [34]. In some cancers, the mutation rate is correlated with treatment sensitivity or resistance. For example, the mutation rate in aromatase inhibitor-resistant breast cancers is 1.62 per Mb, whereas the mutation rate in aromatase inhibitor-sensitive breast cancer is 0.824 per Mb. In colon cancer, the mutation rate in the microsatellite stable group (MSS) is 2.8 per Mb, whereas the mutation rate in the microsatellite instable (MSI) group is 47 per Mb [65]. In non-small cell lung cancer, the mutation rate is significantly higher in smokers (median 10.5, range 4.9–17.6) compared to never-smokers (median 0.6, range 0.6–0.9) [72].

Similarly, for analyses restricted to lung adenocarcinomas, mutation rates were significantly higher in smokers (median 9.8, range 0.04–117.4) compared to never-smokers (median 1.7, range 0.07–22.1) [74]. The prevalence of mutation may be associated with age at diagnosis and telomere length as shown in metastatic neuroblastoma [75], and with environmental exposure as discussed in the aforementioned examples, and with transposons. Mutation prevalence may also be associated with other alterations like structural rearrangements in the genome. For example, the number of mutated genes is much higher in smokers (*KRAS, TP53, BRAF, JAK2, JAK3*) compared to never-smokers (*EGFR*), but nonsmoking lung cancer is associated with the prevalence of more fused genes (*ROS1* and *ALK* fusions) [72], indicating that some tumors may be translocation driven while others are affected primarily by mutations. The most important goal for an individual and the clinician is to identify therapeutic targets for each tumor.

The mutation rates in exonic, intronic, intragenic, intergenic, and regulatory regions vary. In general, mutation rates are higher in introns compared to exons, and intragenic regions compared to intergenic regions. In multiple myeloma, for example, the mutation rate in coding regions is significantly lower than that in both intronic and intergenic regions and lower mutation rate in intronic regions compared to intergenic regions, and mutations in regions with non-regulator potential are more common mutations in regions with regulator potential [37]. Significantly fewer mutations occur in the genic (intronic, non-coding exon, and coding exon) regions than in the intergenic regions in hepatocellular cancer [53]. In colorectal carcinoma, the mutation rate in intergenic regions (6.7 per Mb) is higher than that in intronic and exonic sequences (4.8 and 4.2 per Mb, respectively) [33]. In small-cell lung cancer, mutation in coding regions (0.6%) is lower than non-coding, transcribed regions (0.8%), intronic regions (28%) and intergenic regions (70%) [49].

Coding DNA comprises a very small portion (1.5%) of the human genome and the remainder of the genome (98.5%) consists of noncoding DNA (ncDNA). The amount of ncDNA in organisms increases with the organisms' complexities (e.g., 0.25% of the prokaryote). Rather than being junk DNA, ncDNA likely represents the biological complexity of living organisms and is a main force driving diversity among organisms and even individuals. The discovery of endogenous small interfering RNA, microRNA, long interspersed noncoding RNA (lincRNA), promoter-associated small RNA and terminator-associated small RNA, transcription start site-associated RNA, and transcription initiation RNA may be only the tip of the iceberg. These RNAs represent part of interspersed and crosslinking pieces of a complicated transcription puzzle [76] and may correlate with disease predisposition, therapy response, and/or disease outcome. Indeed, mutations have been found to be more prevalent in intergenic regions than genic regions, it would follow that mutations occur more frequently in intergenic regions than in genic regions. The mechanisms underlying this biased distribution of mutation rate phenomena are unknown. Mutation rates in transcribed strands have recently been found be lower than those in non-transcribed strands in a small-cell cancer cell line [43,49], and cutaneous melanoma, but not acral melanoma [34]. A transposable element (TE) is a DNA sequence that can change its relative position within the genome of a single cell. Therefore transposition can create phenotypically significant mutations and insertions [77]. It is not surprise to find TE insertions in genes that are commonly mutated in cancer, and more commonly in introns or UTRs in epithelial tumors [77].

7. Gene-environment interactions in human cancers

In humans, cancer can be caused by environmental factors, including physical and chemical agents, diet and nutrition, lifestyle risk factors such as tobacco and alcohol use, and micro-environmental conditions acting at a systemic, tissue or cellular level, such as chronic

infections, inflammation, or irritation [78,79]. All these factors can generate "mutation signatures" in the genome of human cells. A recent analysis of comprehensive catalogs of mutations of different types of tumors revealed that characteristic mutational patterns can be related to carcinogen exposures, environmental risk factors, and DNA repair processes. For example, G>T/C>A transversions, which are predominant in smoking-associated lung cancer, elicit a pattern compatible with DNA damage induced by tobacco carcinogens such as benzo[a]pyrene diolepoxide, and polycyclic aromatic hydrocarbons [43,49,80,81]. These mutations are enriched at CpG dinucleotides and exhibit a transcriptional strand bias, which is indicative of a past activity of transcription-coupled nucleotide excision repair on bulky adducts of guanine in response to tobacco carcinogens [79]. Similarly, in ultraviolet radiation-associated skin cancers, C>T and CC>TT transitions are the most common nucleotide substitutions and are consistent with damage from ultraviolet irradiation. These transitions occur at dipyrimidines, which are indicative of the formation of pyrimidine dimers following DNA exposure to UV radiation [34,49,82,83], and show transcriptional strand bias due to the action of transcription-coupled repair of UV-induced C>T transitions and pyrimidine dimers. Other examples of exogenous exposures leading to distinctive mutational patterns include G>T transversions in aflatoxin B1-associated hepatocellular carcinomas [84] and A>T transversions in urothelial tumors from patients exposed to aristolochic acid [85].

C>T transitions are the most common mutations in the genome of both primary and relapsed acute myeloid leukemia (AML). The transversion frequency of relapse-specific mutations is significantly higher than that of primary tumor mutations ($P=3.71\times10^{-11}$), indicating that chemotherapy has a substantial effect on the mutational spectrum at relapse [38].

8. Conclusion and future directions

Next-generation sequencing will have an increasingly profound impact on medical research. Accumulated data from next-generation sequencing has been used to characterize novel mutations and structural rearrangements in the genome, leading to the discovery of previously unrecognized genes. Hundreds to thousands of mutations are present in the genome of a tumor; the prospect of assigning relevance to each and distinguishing drivers and passengers is a daunting task. Functional studies using *in vitro* and *in vivo* experimental models are needed to decipher the biological significance of mutations and structural rearrangements, and translational studies are needed to decipher their clinical consequences and to translate this genetic information into the clinic (therapies that benefit patients, classification of tumors, diagnosis and monitor prognosis of disease). The accurate assessment of different tumor types' background mutation rates and mutation signatures is crucial and to the analysis of significantly mutated genes and pathways. Improvements in NGS technology and analysis tools will facilitate the accurate assessment of mutations and structural rearrangements in cancer. Until now, only the tip of the iceberg of the genome has been characterized. To better understand of the cancer genome, more tumor and paired normal samples and even multiple samples from each tumor need to be sequenced. On the other hand, tumor heterogeneity is a challenging issue. Improving the single molecule sequencing and using this technology to sequence whole-genomes may provide clearer views about tumor heterogeneity. Improving the effective diagnosis and treatment of cancer patients requires a better understanding of genic structures, genomic structure, the biology of genes, and tools and strategies for integrating findings from multiple studies. In addition, standards are needed to establish quality to WGS/NGS to use for clinical practice. Such standards should be platform-independent and include 1) gold standards that guide sample collection, storage, and DNA or RNA quality; 2) quality control metrics; 3) definitions of the metrics that will remove the need for a second method of follow up; 4) analytical tools that can easily compare or integrate platforms; 5) analytical tools that quick and easily

provide reliable output data: and 6) standards for developing high-throughput functional methods.

WGS allows researchers to detecting all of the disease-related genetic variants, regardless of the genetic variant's prevalence or frequency, which will enable emerging medical fields of predictive medicine that use genomic information to predict individuals risk to develop diseases and attempt to either minimize the impact of that disease or avoid it,. Next generation sequencing facilitates personalized medicine by identifying the mutations that are associated with the disease and that serve as targets for treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977; 74:5463–5467. [PubMed: 271968]

2. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975; 94:441–448. [PubMed: 1100841]

3. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A. 1977; 74:560–564. [PubMed: 265521]

4. Parada LF, Tabin CJ, Shih C, Weinberg RA. Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene. Nature. 1982; 297:474–478. [PubMed: 6283357]

5. Shimizu K, Birnbaum D, Ruley MA, Fasano O, Suard Y, Edlund L, Taparowsky E, Goldfarb M, Wigler M. Structure of the Ki-ras gene of the human lung carcinoma cell line Calu-1. Nature. 1983; 304:497–500. [PubMed: 6308465]

6. Santos E, Martin-Zanca D, Reddy EP, Pierotti MA, Della Porta G, Barbacid M. Malignant activation of a K-ras oncogene in lung carcinoma but not in normal tissue of the same patient. Science. 1984; 223:661–664. [PubMed: 6695174]

7. Bos JL, Toksoz D, Marshall CJ, Verlaan-de Vries M, Veeneman GH, van der Eb AJ, van Boom JH, Janssen JW, Steenvoorden AC. Amino-acid substitutions at codon 13 of the N-ras oncogene in human acute myeloid leukaemia. Nature. 1985; 315:726–730. [PubMed: 2989702]

8. Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature. 1986; 323:643–646. [PubMed: 2877398]

9. Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science. 1996; 271:350–353. [PubMed: 8553070]

10. Hunkapiller T, Kaiser RJ, Koop BF, Hood L. Large-scale and automated DNA sequence determination. Science. 1991; 254:59–67. [PubMed: 1925562]

11. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki

Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ. C. International Human Genome Sequencing. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]

12. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA, Stratton MR. Patterns of somatic mutation in human cancer genomes. Nature. 2007; 446:153–158. [PubMed: 17344846]

13. Davies H, Hunter C, Smith R, Stephens P, Greenman C, Bignell G, Teague J, Butler A, Edkins S, Stevens C, Parker A, O'Meara S, Avis T, Barthorpe S, Brackenbury L, Buck G, Clements J, Cole J, Dicks E, Edwards K, Forbes S, Gorton M, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jones D, Kosmidou V, Laman R, Lugg R, Menzies A, Perry J, Petty R, Raine K, Shepherd R, Small A, Solomon H, Stephens Y, Tofts C, Varian J, Webb A, West S, Widaa S, Yates A, Brasseur F, Cooper CS, Flanagan AM, Green A, Knowles M, Leung SY, Looijenga LH, Malkowicz B, Pierotti MA, Teh BT, Yuen ST, Lakhani SR, Easton DF, Weber BL, Goldstraw P, Nicholson AG, Wooster R, Stratton MR, Futreal PA. Somatic mutations of the protein kinase gene family in human lung cancer. Cancer Res. 2005; 65:7591–7595. [PubMed: 16140923]

14. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004; 350:2129–2139. [PubMed: 15118073]

15. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004; 304:1497–1500. [PubMed: 15118125]

16. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ, Willson JK, Markowitz S, Kinzler KW, Vogelstein B, Velculescu VE. High frequency of mutations of the PIK3CA gene in human cancers. Science. 2004; 304:554. [PubMed: 15016963]

17. Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 1996; 242:84–89. [PubMed: 8923969]

18. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. 2010; 11:685–696. [PubMed: 20847746]

19. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet. 2011; 52:413–435. [PubMed: 21698376]

20. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2009; 11:31–46. [PubMed: 19997069]

21. Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ, Ji HP, Snyder M. Performance comparison of whole-genome sequencing platforms. Nat Biotechnol. 2012; 30:78–82. [PubMed: 22178993]

22. Schadt EE, Turner S, Kasarskis A. A, window into third-generation sequencing. Hum Mol Genet. 2010; 19:R227–240. [PubMed: 20858600]

23. Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci U S A. 2003; 100:3960–3964. [PubMed: 12651960]

24. Ozsolak F, Milos PM. Single-molecule direct RNA sequencing without cDNA synthesis. Wiley Interdiscip Rev RNA. 2011; 2:565–570. [PubMed: 21957044]

25. Astier Y, Braha O, Bayley H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5 -monophosphates by using an engineered protein nanopore equipped with a molecular adapter. J Am Chem Soc. 2006; 128:1705–1710. [PubMed: 16448145]

26. Rusk N. Focus on next-generation sequencing data analysis. Forward, Nat Methods. 2009; 6:S1.

27. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–753. [PubMed: 19812666]

28. Rowley JD. Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer. 2001; 1:245–250. [PubMed: 11902580]

29. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005; 310:644–648. [PubMed: 16254181]

30. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007; 448:561–566. [PubMed: 17625570]

31. Lin E, Li L, Guan Y, Soriano R, Rivers CS, Mohan S, Pandita A, Tang J, Modrusan Z. Exon array profiling detects EML4-ALK fusion in breast, colorectal, and non-small cell lung cancers. Mol Cancer Res. 2009; 7:1466–1476. [PubMed: 19737969]

32. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, Menon A, Jing X, Cao Q, Han B, Yu J, Wang L, Montie JE, Rubin MA, Pienta KJ, Roulston D, Shah RB, Varambally S, Mehra R, Chinnaiyan AM. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature. 2007; 448:595–599. [PubMed: 17671502]

33. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K, Sougnez C, Voet D, Saksena G, Sivachenko A, Jing R, Parkin M, Pugh T, Verhaak RG, Stransky N, Boutin AT, Barretina J, Solit DB, Vakiani E, Shao W, Mishina Y, Warmuth M, Jimenez J, Chiang DY, Signoretti S, Kaelin WG, Spardy N, Hahn WC, Hoshida Y, Ogino S, Depinho RA, Chin L, Garraway LA, Fuchs CS, Baselga J, Tabernero J, Gabriel S, Lander ES, Getz G, Meyerson M. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. Nat Genet. 2011; 43:964–968. [PubMed: 21892161]

34. Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, Ivanova E, Watson IR, Nickerson E, Ghosh P, Zhang H, Zeid R, Ren X, Cibulskis K, Sivachenko AY, Wagle N, Sucker A, Sougnez C, Onofrio R, Ambrogio L, Auclair D, Fennell T, Carter SL, Drier Y, Stojanov P, Singer MA, Voet D, Jing R, Saksena G, Barretina J, Ramos AH, Pugh TJ, Stransky N, Parkin M, Winckler W, Mahan S, Ardlie K, Baldwin J, Wargo J, Schadendorf D, Meyerson M, Gabriel SB, Golub TR, Wagner SN, Lander ES, Getz G, Chin L, Garraway LA. Melanoma genome sequencing reveals frequent PREX2 mutations. Nature. 2012; 485:502–506. [PubMed: 22622578]

35. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, Onofrio R, Carter SL, Park K, Habegger L, Ambrogio L, Fennell T, Parkin M, Saksena G, Voet D, Ramos AH, Pugh TJ, Wilkinson J, Fisher S, Winckler

W, Mahan S, Ardlie K, Baldwin J, Simons JW, Kitabayashi N, MacDonald TY, Kantoff PW, Chin L, Gabriel SB, Gerstein MB, Golub TR, Meyerson M, Tewari A, Lander ES, Getz G, Rubin MA, Garraway LA. The genomic complexity of primary human prostate cancer. Nature. 2011; 470:214–220. [PubMed: 21307934]

36. Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, Teague JW, Menzies A, Goodhead I, Turner DJ, Clee CM, Quail MA, Cox A, Brown C, Durbin R, Hurles ME, Edwards PA, Bignell GR, Stratton MR, Futreal PA. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet. 2008; 40:722–729. [PubMed: 18438408]

37. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, Harview CL, Brunet JP, Ahmann GJ, Adli M, Anderson KC, Ardlie KG, Auclair D, Baker A, Bergsagel PL, Bernstein BE, Drier Y, Fonseca R, Gabriel SB, Hofmeister CC, Jagannath S, Jakubowiak AJ, Krishnan A, Levy J, Liefeld T, Lonial S, Mahan S, Mfuko B, Monti S, Perkins LM, Onofrio R, Pugh TJ, Rajkumar SV, Ramos AH, Siegel DS, Sivachenko A, Stewart AK, Trudel S, Vij R, Voet D, Winckler W, Zimmerman T, Carpten J, Trent J, Hahn WC, Garraway LA, Meyerson M, Lander ES, Getz G, Golub TR. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011; 471:467–472. [PubMed: 21430775]

38. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012; 481:506–510. [PubMed: 22237025]

39. Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, Van Tine BA, Hoog J, Goiffon RJ, Goldstein TC, Ng S, Lin L, Crowder R, Snider J, Ballman K, Weber J, Chen K, Koboldt DC, Kandoth C, Schierding WS, McMichael JF, Miller CA, Lu C, Harris CC, McLellan MD, Wendl MC, DeSchryver K, Allred DC, Esserman L, Unzeitig G, Margenthaler J, Babiera GV, Marcom PK, Guenther JM, Leitch M, Hunt K, Olson J, Tao Y, Maher CA, Fulton LL, Fulton RS, Harrison M, Oberkfell B, Du F, Demeter R, Vickery TL, Elhammali A, Piwnica-Worms H, McDonald S, Watson M, Dooling DJ, Ota D, Chang LW, Bose R, Ley TJ, Piwnica-Worms D, Stuart JM, Wilson RK, Mardis ER. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature. 2012; 486:353–360. [PubMed: 22722193]

40. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo M, Miya F, Arai Y, Takahashi H, Shirakihara T, Nagasaki M, Shibuya T, Nakano K, Watanabe-Makino K, Tanaka H, Nakamura H, Kusuda J, Ojima H, Shimada K, Okusaka T, Ueno M, Shigekawa Y, Kawakami Y, Arihiro K, Ohdan H, Gotoh K, Ishikawa O, Ariizumi S, Yamamoto M, Yamada T, Chayama K, Kosuge T, Yamaue H, Kamatani N, Miyano S, Nakagama H, Nakamura Y, Tsunoda T, Shibata T, Nakagawa H. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet. 2012; 44:760–764. [PubMed: 22634756]

41. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang Q, Chen L, Lin L, Wendl MC, McMichael JF, Magrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Lin L, Crowder R, Tao Y, Snider JE, Smith SM, Dukes AF, Sanderson GE, Pohl CS, Delehaunty KD, Fronick CC, Pape KA, Reed JS, Robinson JS, Hodges JS, Schierding W, Dees ND, Shen D, Locke DP, Wiechert ME, Eldred JM, Peck JB, Oberkfell BJ, Lolofie JT, Du F, Hawkins AE, O'Laughlin MD, Bernard KE, Cunningham M, Elliott G, Mason MD, Thompson DM Jr, Ivanovich JL, Goodfellow PJ, Perou CM, Weinstock GM, Aft R, Watson M, Ley TJ, Wilson RK, Mardis ER. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010; 464:999–1005. [PubMed: 20393555]

42. Jones DT, Jager N, Kool M, Zichner T, Hutter B, Sultan M, Cho YJ, Pugh TJ, Hovestadt V, Stutz AM, Rausch T, Warnatz HJ, Ryzhova M, Bender S, Sturm D, Pleier S, Cin H, Pfaff E, Sieber L, Wittmann A, Remke M, Witt H, Hutter S, Tzaridis T, Weischenfeldt J, Raeder B, Avci M, Amstislavskiy V, Zapatka M, Weber UD, Wang Q, Lasitschka B, Bartholomae CC, Schmidt M, von Kalle C, Ast V, Lawerenz C, Eils J, Kabbe R, Benes V, van Sluis P, Koster J, Volckmann R,

Shih D, Betts MJ, Russell RB, Coco S, Tonini GP, Schuller U, Hans V, Graf N, Kim YJ, Monoranu C, Roggendorf W, Unterberg A, Herold-Mende C, Milde T, Kulozik AE, von Deimling A, Witt O, Maass E, Rossler J, Ebinger M, Schuhmann MU, Fruhwald MC, Hasselblatt M, Jabado N, Rutkowski S, von Bueren AO, Williamson D, Clifford SC, McCabe MG, Collins VP, Wolf S, Wiemann S, Lehrach H, Brors B, Scheurlen W, Felsberg J, Reifenberger G, Northcott PA, Taylor MD, Meyerson M, Pomeroy SL, Yaspo ML, Korbel JO, Korshunov A, Eils R, Pfister SM, Lichter P. Dissecting the genomic complexity underlying medulloblastoma. Nature. 2012; 488:100–105. [PubMed: 22832583]

43. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, Ha C, Johnson S, Kennemer MI, Mohan S, Nazarenko I, Watanabe C, Sparks AB, Shames DS, Gentleman R, de Sauvage FJ, Stern H, Pandita A, Ballinger DG, Drmanac R, Modrusan Z, Seshagiri S, Zhang Z. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature. 2010; 465:473–477. [PubMed: 20505728]

44. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008; 456:66–72. [PubMed: 18987736]

45. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, Meyer RJ, Schindler JK, Pohl CS, Wallis JW, Shi X, Lin L, Schmidt H, Tang Y, Haipek C, Wiechert ME, Ivy JV, Kalicki J, Elliott G, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson MA, Baty J, Heath S, Shannon WD, Nagarajan R, Link DC, Walter MJ, Graubert TA, DiPersio JF, Wilson RK, Ley TJ. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med. 2009; 361:1058–1066. [PubMed: 19657110]

46. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jonsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerod A, Tutt A, Martens JW, Aparicio SA, Borg A, Salomon AV, Thomas G, Borresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR. C. Breast Cancer Working Group of the International Cancer Genome. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012; 149:979–993. [PubMed: 22608084]

47. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, Menzies A, Stebbings LA, Leroy C, Jia M, Rance R, Mudie LJ, Gamble SJ, Stephens PJ, McLaren S, Tarpey PS, Papaemmanuil E, Davies HR, Varela I, McBride DJ, Bignell GR, Leung K, Butler AP, Teague JW, Martin S, Jonsson G, Mariani O, Boyault S, Miron P, Fatima A, Langerod A, Aparicio SA, Tutt A, Sieuwerts AM, Borg A, Thomas G, Salomon AV, Richardson AL, Borresen-Dale AL, Futreal PA, Stratton MR, Campbell PJ. C. Breast Cancer Working Group of the International Cancer Genome. The life history of 21 breast cancers. Cell. 2012; 149:994–1007. [PubMed: 22608083]

48. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordonez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen L, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia M, Kingsbury Z, Leroy C, Marshall J, Menzies A, Mudie LJ, Ning Z, Royce T, Schulz-Trieglaff OB, Spiridou A, Stebbings LA, Szajkowski L, Teague J, Williamson D, Chin L, Ross MT, Campbell PJ, Bentley DR, Futreal PA, Stratton MR. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010; 463:191–196. [PubMed: 20016485]

49. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordonez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA, Campbell PJ. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010; 463:184–190. [PubMed: 20016488]

50. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, Escaramis G, Jares P, Bea S, Gonzalez-Diaz M, Bassaganyas L, Baumann T, Juan M, Lopez-Guerra M, Colomer D, Tubio JM, Lopez C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernandez JM, Puente DA, Freije JM, Velasco G, Gutierrez-Fernandez A, Costa D, Carrio A, Guijarro S, Enjuanes A, Hernandez L, Yague J, Nicolas P, Romeo-Casabona CM, Himmelbauer H, Castillo E, Dohm JC, de Sanjose S, Piris MA, de Alava E, San Miguel J, Royo R, Gelpi JL, Torrents D, Orozco M, Pisano DG, Valencia A, Guigo R, Bayes M, Heath S, Gut M, Klatt P, Marshall J, Raine K, Stebbings LA, Futreal PA, Stratton MR, Campbell PJ, Gut I, Lopez-Guillermo A, Estivill X, Montserrat E, Lopez-Otin C, Campo E. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011; 475:101–105. [PubMed: 21642962]

51. Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, Phoenix TN, Hedlund E, Wei L, Zhu X, Chalhoub N, Baker SJ, Huether R, Kriwacki R, Curley N, Thiruvenkatam R, Wang J, Wu G, Rusch M, Hong X, Becksfort J, Gupta P, Ma J, Easton J, Vadodaria B, Onar-Thomas A, Lin T, Li S, Pounds S, Paugh S, Zhao D, Kawauchi D, Roussel MF, Finkelstein D, Ellison DW, Lau CC, Bouffet E, Hassall T, Gururangan S, Cohn R, Fulton RS, Fulton LL, Dooling DJ, Ochoa K, Gajjar A, Mardis ER, Wilson RK, Downing JR, Zhang J, Gilbertson RJ. Novel mutations target distinct subgroups of medulloblastoma. Nature. 2012; 488:43–48. [PubMed: 22722829]

52. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009; 461:809–813. [PubMed: 19812674]

53. Totoki Y, Tatsuno K, Yamamoto S, Arai Y, Hosoda F, Ishikawa S, Tsutsumi S, Sonoda K, Totsuka H, Shirakihara T, Sakamoto H, Wang L, Ojima H, Shimada K, Kosuge T, Okusaka T, Kato K, Kusuda J, Yoshida T, Aburatani H, Shibata T. High-resolution characterization of a hepatocellular carcinoma genome. Nat Genet. 2011; 43:464–469. [PubMed: 21499249]

54. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, Larson DE, McLellan MD, Dooling D, Abbott R, Fulton R, Magrini V, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Fan X, Grillot M, Witowski S, Heath S, Frater JL, Eades W, Tomasson M, Westervelt P, DiPersio JF, Link DC, Mardis ER, Ley TJ, Wilson RK, Graubert TA. Clonal architecture of secondary acute myeloid leukemia. N Engl J Med. 2012; 366:1090–1098. [PubMed: 22417201]

55. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, Tesar B, Sievers QL, Shefler E, Gabriel S, Hacohen N, Reed R, Meyerson M, Golub TR, Lander ES, Neuberg D, Brown JR, Getz G, Wu CJ. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med. 2011; 365:2497–2506. [PubMed: 22150006]

56. Weiss GJ, Liang WS, Izatt T, Arora S, Cherni I, Raju RN, Hostetter G, Kurdoglu A, Christoforides A, Sinari S, Baker AS, Metpally R, Tembe WD, Phillips L, Von Hoff DD, Craig DW, Carpten JD. Paired tumor and normal whole genome sequencing of metastatic olfactory neuroblastoma. PLoS One. 2012; 7:e37029. [PubMed: 22649506]

57. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, Lu C, Chen SC, Wei L, Collins-Underwood JR, Ma J, Roberts KG, Pounds SB, Ulyanov A, Becksfort J, Gupta P, Huether R, Kriwacki RW, Parker M, McGoldrick DJ, Zhao D, Alford D, Espy S, Bobba KC, Song G, Pei D, Cheng C, Roberts S, Barbato MI, Campana D, Coustan-Smith E, Shurtleff SA, Raimondi SC, Kleppe M, Cools J, Shimano KA, Hermiston ML, Doulatov S, Eppert K, Laurenti E, Notta F, Dick JE, Basso G, Hunger SP, Loh ML, Devidas M, Wood B, Winter S, Dunsmore KP, Fulton RS, Fulton LL, Hong X, Harris CC, Dooling DJ, Ochoa K, Johnson KJ, Obenauer JC, Evans WE, Pui CH, Naeve CW, Ley TJ, Mardis ER, Wilson RK,

Downing JR, Mullighan CG. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012; 481:157–163. [PubMed: 22237106]

58. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, Cortes ML, Fernandez-Lopez JC, Peng S, Ardlie KG, Auclair D, Bautista-Pina V, Duke F, Francis J, Jung J, Maffuz-Aziz A, Onofrio RC, Parkin M, Pho NH, Quintanar-Jurado V, Ramos AH, Rebollar-Vega R, Rodriguez-Cuevas S, Romero-Cordoba SL, Schumacher SE, Stransky N, Thompson KM, Uribe-Figueroa L, Baselga J, Beroukhim R, Polyak K, Sgroi DC, Richardson AL, Jimenez-Sanchez G, Lander ES, Gabriel SB, Garraway LA, Golub TR, Melendez-Zajgla J, Toker A, Getz G, Hidalgo-Miranda A, Meyerson M. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012; 486:405–409. [PubMed: 22722202]

59. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerod A, Russnes HE, Foekens JA, Reis-Filho JS, van't Veer L, Richardson AL, Borresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature. 2009; 462:1005–1010. [PubMed: 20033038]

60. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK. DNMT3A mutations in acute myeloid leukemia. N Engl J Med. 2010; 363:2424–2433. [PubMed: 21067377]

61. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, Morsberger LA, Latimer C, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal SA, Leroy C, Jia M, Menzies A, Butler AP, Teague JW, Griffin CA, Burton J, Swerdlow H, Quail MA, Stratton MR, Iacobuzio-Donahue C, Futreal PA. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature. 2010; 467:1109–1113. [PubMed: 20981101]

62. Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, Wallis J, Chen K, Payton JE, Fulton RS, Veizer J, Schmidt H, Vickery TL, Heath S, Watson MA, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Ley TJ, Wilson RK. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA. 2011; 305:1577–1584. [PubMed: 21505136]

63. Link DC, Schuettpelz LG, Shen D, Wang J, Walter MJ, Kulkarni S, Payton JE, Ivanovich J, Goodfellow PJ, Le Beau M, Koboldt DC, Dooling DJ, Fulton RS, Bender RH, Fulton LL, Delehaunty KD, Fronick CC, Appelbaum EL, Schmidt H, Abbott R, O'Laughlin M, Chen K, McLellan MD, Varghese N, Nagarajan R, Heath S, Graubert TA, Ding L, Ley TJ, Zambetti GP, Wilson RK, Mardis ER. Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML. JAMA. 2011; 305:1568–1576. [PubMed: 21505135]

64. N. Cancer Genome Atlas. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–337. [PubMed: 22810696]

65. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, Chaudhuri S, Guan Y, Janakiraman V, Jaiswal BS, Guillory J, Ha C, Dijkgraaf GJ, Stinson J, Gnad F, Huntley MA, Degenhardt JD, Haverty PM, Bourgon R, Wang W, Koeppen H, Gentleman R, Starr TK, Zhang Z, Largaespada DA, Wu TD, de Sauvage FJ. Recurrent R-spondin fusions in colon cancer. Nature. 2012

66. Bohlander SK. ETV6: a versatile player in leukemogenesis. Semin Cancer Biol. 2005; 15:162–174. [PubMed: 15826831]

67. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, Gallinger S, Schwartz AG, Syngal S, Cote ML, Axilbund J, Schulick R, Ali SZ, Eshleman JR, Velculescu VE, Goggins M, Vogelstein B, Papadopoulos N, Hruban RH, Kinzler KW, Klein AP. ATM mutations in patients with hereditary pancreatic cancer. Cancer Discov. 2012; 2:41–46. [PubMed: 22585167]

68. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, Kos I, Batinic-Haberle I, Jones S, Riggins GJ, Friedman H, Friedman A, Reardon D, Herndon J, Kinzler KW, Velculescu VE, Vogelstein B, Bigner DD. IDH1 and IDH2 mutations in gliomas. N Engl J Med. 2009; 360:765–773. [PubMed: 19228619]

69. Marcucci G, Maharry K, Wu YZ, Radmacher MD, Mrozek K, Margeson D, Holland KB, Whitman SP, Becker H, Schwind S, Metzeler KH, Powell BL, Carter TH, Kolitz JE, Wetzler M, Carroll AJ, Baer MR, Caligiuri MA, Larson RA, Bloomfield CD. IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol. 2010; 28:2348–2355. [PubMed: 20368543]

70. Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, Han B, Cao Q, Cao X, Suleman K, Kumar-Sinha C, Dhanasekaran SM, Chen YB, Esgueva R, Banerjee S, LaFargue CJ, Siddiqui J, Demichelis F, Moeller P, Bismar TA, Kuefer R, Fullen DR, Johnson TM, Greenson JK, Giordano TJ, Tan P, Tomlins SA, Varambally S, Rubin MA, Maher CA, Chinnaiyan AM. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. Nat Med. 2010; 16:793–798. [PubMed: 20526349]

71. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, Kalyana-Sundaram S, Sam L, Balbin OA, Quist MJ, Barrette T, Everett J, Siddiqui J, Kunju LP, Navone N, Araujo JC, Troncoso P, Logothetis CJ, Innis JW, Smith DC, Lao CD, Kim SY, Roberts JS, Gruber SB, Pienta KJ, Talpaz M, Chinnaiyan AM. Personalized oncology through integrative high-throughput sequencing: a pilot study. Sci Transl Med. 2011; 3:111ra121.

72. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, Chen K, Walker J, McDonald S, Bose R, Ornitz D, Xiong D, You M, Dooling DJ, Watson M, Mardis ER, Wilson RK. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012; 150:1121–1134. [PubMed: 22980976]

73. Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, Menon R, Koker M, Dahmen I, Muller C, Di Cerbo V, Schildhaus HU, Altmuller J, Baessmann I, Becker C, de Wilde B, Vandesompele J, Bohm D, Ansen S, Gabler F, Wilkening I, Heynck S, Heuckmann JM, Lu X, Carter SL, Cibulskis K, Banerji S, Getz G, Park KS, Rauh D, Grutter C, Fischer M, Pasqualucci L, Wright G, Wainer Z, Russell P, Petersen I, Chen Y, Stoelben E, Ludwig C, Schnabel P, Hoffmann H, Muley T, Brockmann M, Engel-Riedel W, Muscarella LA, Fazio VM, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman DA, Snijders PJ, Cappuzzo F, Ligorio C, Damiani S, Field J, Solberg S, Brustugun OT, Lund-Iversen M, Sanger J, Clement JH, Soltermann A, Moch H, Weder W, Solomon B, Soria JC, Validire P, Besse B, Brambilla E, Brambilla C, Lantuejoul S, Lorimier P, Schneider PM, Hallek M, Pao W, Meyerson M, Sage J, Shendure J, Schneider R, Buttner R, Wolf J, Nurnberg P, Perner S, Heukamp LC, Brindle PK, Haas S, Thomas RK. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet. 2012; 44:1104–1110. [PubMed: 22941188]

74. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansen S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Janne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012; 150:1107–1120. [PubMed: 22980975]

75. Cheung NK, Zhang J, Lu C, Parker M, Bahrami A, Tickoo SK, Heguy A, Pappo AS, Federico S, Dalton J, Cheung IY, Ding L, Fulton R, Wang J, Chen X, Becksfort J, Wu J, Billups CA, Ellison D, Mardis ER, Wilson RK, Downing JR, Dyer MA. Association of age at diagnosis and genetic mutations in patients with neuroblastoma. Jama. 2012; 307:1062–1071. [PubMed: 22416102]

76. Jacquier A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet. 2009; 10:833–844. [PubMed: 19920851]

77. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, Wheeler DA, Gibbs RA, Kucherlapati R, Lee C, Kharchenko PV, Park PJ. N. The Cancer Genome Atlas Research. Landscape of Somatic Retrotransposition in Human Cancers. Science. 2012

78. Pfeifer GP, Besaratinia A. Mutational spectra of human cancer. Hum Genet. 2009; 125:493–506. [PubMed: 19308457]

79. Hainaut P, Pfeifer GP. Patterns of p53 G-->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. Carcinogenesis. 2001; 22:367–374. [PubMed: 11238174]

80. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene. 2002; 21:7435–7451. [PubMed: 12379884]

81. Denissenko MF, Chen JX, Tang MS, Pfeifer GP. Cytosine methylation determines hot spots of DNA damage in the human P53 gene. Proc Natl Acad Sci U S A. 1997; 94:3893–3898. [PubMed: 9108075]

82. Tommasi S, Denissenko MF, Pfeifer GP. Sunlight induces pyrimidine dimers preferentially at 5-methylcytosine bases. Cancer Res. 1997; 57:4727–4730. [PubMed: 9354431]

83. Pfeifer GP, You YH, Besaratinia A. Mutations induced by ultraviolet light. Mutat Res. 2005; 571:19–31. [PubMed: 15748635]

84. Mace K, Aguilar F, Wang JS, Vautravers P, Gomez-Lechon M, Gonzalez FJ, Groopman J, Harris CC, Pfeifer AM. Aflatoxin B1-induced DNA adduct formation and p53 mutations in CYP450-expressing human liver cell lines. Carcinogenesis. 1997; 18:1291–1297. [PubMed: 9230270]

85. Nedelko T, Arlt VM, Phillips DH, Hollstein M. TP53 mutation signature supports involvement of aristolochic acid in the aetiology of endemic nephropathy-associated tumours. Int J Cancer. 2009; 124:987–990. [PubMed: 19030178]

Table 1

Next-generation sequencing technologies

	Platforms					
	Roche (454's, GS FLX, Titanium)	**Illumina (Solexa's GA)**	**ABI (SOLID)**	**Pacific Bio (SMRT)**	**Ion Torrent**	**BioSciences (Helicos)**
Sequencing method	Sequencing by synthesis (pyrosequencing)	Sequencing by synthesis (Reversible dye terminators)	Sequencing by ligation (octamers with two-base encoding)	Sequencing by synthesis (real time sequencing without clonal amplification)	Sequencing by synthesis (Hydrogen ion detection)	Polymerase (0ne base-at-a-time, asynchronous extension)
Library	Fragm & MP	Frag & MP	Frag & MP	Fragment only	Frag & MP	Frag & MP
Chemistry	Pyrosequencing	Reversible dye terminator	Cleavable probe SBL	Phosphor-linked fluorescent nucleotides	Ion semiconductor sequencing	Reversible die terminators
Amplification method and template preparation	Emilsion PCR on bead surface	Bridge enzymatic amplification on glass (substrate-based PCR)	Emilsion PCR on bead surface	Single molecule	Emulsion PCR	Single molecule
Sample requirements	1 μg (shotgun) 5 μg (paired-end)	<1 μg (single-end) <1 μg (paired-end)	2 μg (single-end) 5–10 μg (paired-end)	~1 μg (single-end)	100–1000ng	500ng-3μg
Read length	400–700 bp (SE, PE)	36–150 bp (SE, PE)	35–75 bp (SE, PE)	<3000 bp (SE)	<200 bp	35–55 bp
Reads per run	0.4–0.6 Gb/0.035 Gb	1.2–600 Gb	10–155 Gb	<100Gb per hour	100 Mb-1 Gb	35 Gb
Time per run	10–23 h	27h–14 d *	6–8 d	2 h	2–4 h	8 d
Single read	yes	yes	yes	yes	yes	yes
Paired-end Reads	yes	yes	yes	no	no	yes
RNA-seq	cDNA	cDNA	cDNA	Direct RNA	cDNA	Direct RNA
Primary error	Indel	Substitution	A-T bias	Indel (insertion)	Indel	Indel
Advantages	Long read length, fast run time	High sequence yield per run, sequencing accuracy, low cost/Mb	High fidelity	Fast, long read length (>1 kb)	suitable for microbial sequencing	Non-bias representation of templates, smallest samples, and no ligation, amplification, and cDNA requirement
Disadvantages	High reagents cost, high error rates in areas of repeated nucleotides (homolymer), indel errors, cDNA seq.	Shorter reads, cDNA seq.	Long run times, short read length, cDNA seq	Highest error rates compared with other NGS platforms	High error rate (homopolymer), Short read length	High error rates compared with other reversible terminator chemistries

Frag, fragment; GA, genome analyzer; GS, genome sequencer; MP, mate-pair; SE, single-end read; PE, paired-end read; SBL, sequencing by ligation; SOLID, support oligonucleotide ligation detection;

*
depending on sequencer.