Cellular & Molecular Biology
Cell. Mol. Biol. 2015; 61 (5): 91-102
Published online October 30, 2015 (http://www.cellmolbiol.com)
Received on September 29, 2015, Accepted on October 21, 2015.
doi : 10.14715/cmb/2015.61.5.16

Next generation sequencing technologies in cancer diagnostics and therapeutics: A mini review

W. Li[1], K. Zhao[2], M. Kirberger[3], W. Liao[4] and Y. Yan[5]✉

[1] IHRC, inc. 2 Ravinia Drive, Suite 1750, Atlanta, GA 30346-2147, USA
[2] Department of Mathematics and Statistics, Georgia State University, Atlanta, Georgia, 30303, USA
[3] Department of Natural Sciences, Clayton State University, 2000 Clayton State Boulevard, Morrow, GA, 30260, USA
[4] College of Food and Bioengineering, South China University of Technology, Guangzhou, 510640, China
[5] School of Medicine, Hangzhou Normal University, 16 Xuelin St, Hangzhou, Zhejiang, 310036, China

Corresponding author: Yutao Yan, School of Medicine, Hangzhou Normal University, 16 Xuelin St, Hangzhou, Zhejiang, 310036, China. E-mail: yutaoyan@yahoo.com

Abstract

The development of advanced molecular technologies has ushered in the era of 'omics' science, including transcriptomics, proteomics, and genomics. Genomics, or whole genome approach, has become the most comprehensive investigative method to identify new gene mutations, signal pathways and drug targets for cancers. The purpose of this review is to summarize current second generation sequencing techniques in applied genomics, and to analyze the advantages and/or problems associated with each of the various sequencing platforms. Our understanding of molecular factors associated with tumorigenesis is no longer limited to the mutation of well-known cancer related genes, but may involve a broader range of factors involved in tumor development, including novel somatic mutations, gene fusions, long non-coding RNAs, microRNAs, copy number variations, methylation, and genomic structural variations. Furthermore, these new methods are not limited to analyses of single genetic or epigenetic factor, but offer comprehensive molecule profiling as a more critical and powerful approach to decoding the mystery of tumor development and identifying more reliable cancer biomarkers.

Key words: Next generation sequencing, *de novo* assembly, whole genome sequencing, whole exome sequencing, gene mutation, biomarker.

Introduction of Next Generation Sequencing (NGS)

Cancer persistently remains one of the leading causes of death annually, due in part to the complicated nature of cancer development, as well as the limited availability of precise diagnostics and effective treatments in clinical practices. Traditional diagnostic methods include biopsy, Magnetic Resonance Imaging (MRI), endoscopy, and blood tests (1,2). Surgery, chemotherapy, radiotherapy or transplantation are the primary approaches to cancer therapies, but the complexity of the cancer microenvironment and variations in immune-responses require more advanced technologies designed to decode individual disease profiles. This has led to the evolution of high-throughput, next generation methodologies to accelerate sequencing using parallel processing. The resulting, enhanced whole genome and transcriptome sequencing techniques have significantly changed our understanding of genomic variations in cancer diseases, and by narrowing the resolution of genetic variations among cancer patients to single base pairs, it seems likely that future approaches to diagnostics and therapy will incorporate personalized data in the development of clinical applications (3).

Acquiring the information necessary for this type of medical innovation requires the capability of rapidly analyzing large volumes of genetic data, as would be found in an organism's genome. This can be achieved using Whole Genomics Sequencing (WGS) technology, which generates millions of short DNA sequence fragments (i.e., reads) from a genome (4). The range of read lengths may vary depending on the technology used. For example, Illumina reads 25–300 base pairs with 98% accuracy, while Pacbio can read thousands of base pairs with 87% single-read accuracy (5). In top down sequencing projects, computational algorithms are used to merge and organize reads into longer sequences called "contigs" (a.k.a assemblies). These calculated contigs are then reorganized into a scaffold if a complete reference genome is available, thus providing a rapid methodology for genome sequencing. Under ideal circumstances, computer algorithms would process vast quantities of data both accurately and rapidly, but in reality, the application of computational methods in scientific research frequently involve some minor compromises between quality and quantity, where some loss of accuracy (quality) may be acceptable in exchange for rapid analysis (quantity). Typically, some minimum standard is established by consensus. To address this question for genomic data, a definition of "high-quality" assembly has been promulgated by Genome Assembly Gold-Standard Evaluations (GAGE) (6,7) and applied in two competitions to evaluate the quality of contigs generated by popular assembler software: "Assemblathon" (8,9) and the "*De Novo* Genome Assembly Assessment Project" (dnGASP) (10). Thus, recent improvements to sequencing technologies have enhanced our ability to explore the genetic factors of cancer biology in terms of timeliness, accuracy and reduced costs, although these improvements have also come with mathematical, statistical, and computational challenges.

Platforms of Next Generation Sequencing

Frederick Sanger introduced a Sequencing methodology in 1977 that has proven to be effective for small scale sequencing, and has been used extensively for the last 35 years (11,12). However, whole genome sequencing techniques remained unattainable until the year 2000, following successful launches of Roche 454 GS20 (13)and ABI SOLID (14). The preliminary successes of these platforms were soon followed by Ion Torrent (15),PacBio (16) and Illumina (17) platforms, that rapidly permeated the market of genomic sequencing. More recently, third generation sequencing platforms, such as Nanopore (18,19), have been generated for commercial use. A summary timeline of the development of DNA sequencing platforms is presented in Figure 1. Brief introductions of some commonly used NGS platforms are described below.

Illumina

The Illumina sequencing platform launched later than its competitors, but quickly dominated the NGS market. This platform utilized Sequencing by Synthesis (SBS) technology, which allowed for the analysis of millions of base pairs of the genome sequence within the relatively short span of several hours. NGS sequencing on the Illumina instrument is divided into four major steps: Library preparation; Cluster Generation; Pair-end Sequencing and Data analysis.

Library preparation: During this step, genomic DNA is fragmented randomly. Methods for DNA shearing include: physical shearing (e.g., acoustic shearing, sonication, or hydrodynamic shearing) or enzymatic shearing (e.g., shearing by restriction endonuclease or transposase (20)). One commonly used acoustic shearing device is the Covaris® instrument, which shears genomic DNA into 100–1500bp (microTube), 2–5kb (miniTube) or 6–20kb (gTube) fragments (21). Similarly, Megaruptor® from Diagenode was designed to use hydrodynamic shearing to produce libraries between 2–90kb. This system can produce a higher and more reproducible yield of fragments, exhibiting a more narrow size distribution. Conversely, NEBNext®Double-Stranded DNA (dsDNA) Fragmentase uses two enzymes to randomly generate nicks on a single strand of dsDNA, and then cleaves the opposite strand at the same location (22). Another approach, tagmentation, uses a transposase to simultaneously fragment and insert adapters onto ds DNA (23). Subsequent to shearing, adaptors with 6–8 bp specific nucleotide sequences are ligated to the 5′ and 3′ ends of each fragment. These adaptor-ligated fragments are PCR amplified and bead purified. Commercial library preparation kits that are extensively used include Illumina Nextera®, NEBNext Ultra®, Agilent XT®, Nugen Ovation Ultralow®, QiagenGeneread®, LucigenNxseq®, BiooNEXTflex®, and Rapid DNA-seq.

Cluster generation: In this step, the library is diluted to 4–10 pmol, denatured using NaOH, mixed with 1–5% PhiX control, and loaded into the flow cell which contains 1 (MiSeq), 4 (NextSeq) or 8 (HiSeq) lanes (24). The loaded DNA fragments bind the surface-bound oligos complementary to the library adaptors. Each fragment is amplified to the local cluster by bridge amplification.

Pair-end sequencing: Illumina uses sequencing by synthesis method (SBS), where four reversible, terminator-bound dNTPs are naturally incorporated into the synthesized sequence (25). The fluorescence of each bound nucleotide is observed by camera and the dyes with the terminal 3′ blockers are washed away.

Data Analysis: After sequencing, information acquired from base reads, and the associated quality scores, are stored in a Fastq file. Adaptor sequences are then trimmed from each read. The sequences are then either mapped to reference genomes, or a de-novo assembly is constructed if the genome reference is unavailable. Following the *de novo* assembly, segments of the continuous sequences are then aligned, based upon overlapping regions, to form the contigs (26).

Roche 454

The Roche 454 sequencer is based on a pyrosequencing method, where genomic DNA strands are sheared into fragments between 400–600bp, ligated with adaptors, and then separated into individual strands (14). Emulsion oil is then injected into the DNA/enzyme/reagent mixture to form emulsion droplets, and each

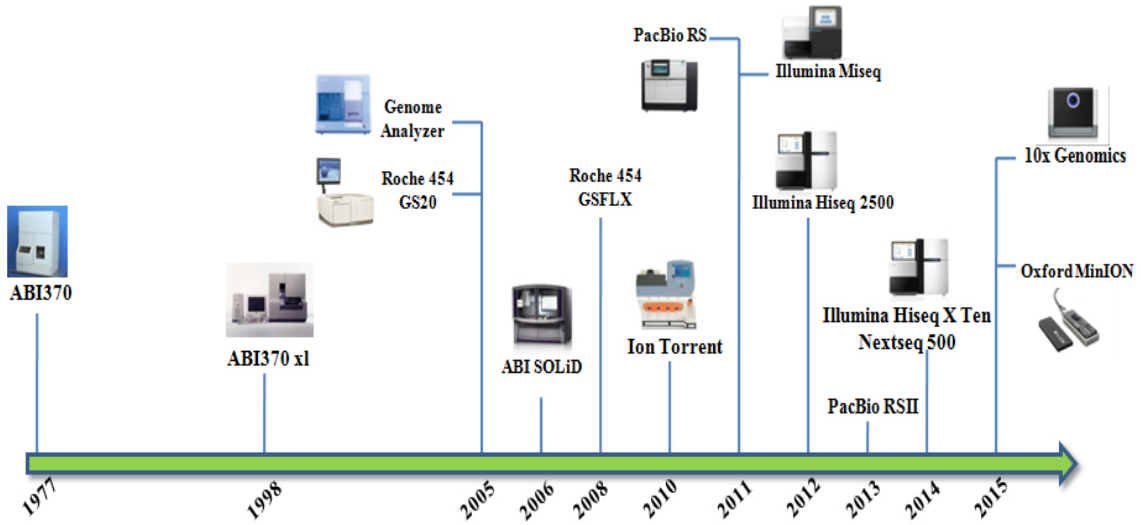

Figure 1. Timeline of the development of DNA sequencing platforms.

droplet contains a single DNA fragment with a synthetic enzyme. Inside each droplet, the single nucleotide sequences bind to the surface of the beads, and an emulsion based PCR (em-PCR) is performed to amplify the single fragments into multiple copies on the bead surface. The DNA-captured beads are then placed onto a PicoTiter Plate containing 1.6 million wells. Each well has a 75 pL volume, and measures 44 μm in diameter. Free nucleotides attached to luciferase are sequentially transferred onto the PicoTiter Plate, and single nucleotide complements to the template then generate light signals, which can be captured by a charge-coupled device (CCD) camera (27).

One main prerequisite of Roche 454 sequencing is that each bead is covered with multiple copies of the same DNA fragments. Two different adaptors can be ligated sequentially onto the 5′ and 3′ termini of the DNA fragments, where one of the adaptors is complementary to the oligonucleotides on the surface of the beads, after DNA fragmentation and bead binding, while the second adaptor bound to the other free side will bind to special capture beads, and the unbound DNA fragments are subsequently washed out (27,28).

ABI SOLiD

The SOLiD sequencing platform, produced by Technologies/Applied Biosystems (ABI), performs sequencing by ligation method. Using methods similar to those used for Roche 454 library preparation, the genomic double stranded DNA molecules are sheared into small pieces, and ligated with P1 and P2 adaptors on the two prime ends. The end carrying the P1 adaptor binds to the surface of the magnetic bead, and emulsion PCR is used to amplify the single nucleotide fragments. The oil is then washed out, and four fluorescent labeled di-base probes are added to the bead mixture. By matching the 1st and 2nd positions of the template using di-base probes, fluorescence can be detected, while the extra tail carrying the fluorescent probe is cleaved out. Following multiple cycles of ligation, detection and tail cleavages, the extended chain will reach the end of the template. The entire extension chain is then removed and a new starting primer is incorporated, which is moved downstream by one nucleotide position, thus binding it onto the template for another cycle of reaction. In total, five rounds of primer binding cycles are performed to complete the sequencing of each fragment (29,30).

Limitation of NGS

Errors may be introduced at every step of the sequencing process, including library preparation, cluster generation, sequence amplification and data analysis. Barcode incompatibility may result in a reading failure of the index due to color imbalance (31). As Illumina uses green laser to read G/T and red laser to read A/C, it is important to maintain the reads associated with the two different colors at each cycle. Therefore, checking the index combination before pooling libraries is essential for successful sequencing. The presence of a highly pure library, prior to loading to Miseq, is also essential for optimal sequencing. Cluster generation is another approach used to obtain high quality sequencing reads. Clustering improves data quality but reduces output, while over–clustering may produce higher sequencing errors with poor Q30 (where Q30 represents 1 error in 1000 bases)(32,33). Therefore, both under-and over-clustering will result in lower quality data output. During PCR amplification, DNA polymerase may naturally generate bias by incorporating the wrong nucleotides into the amplified chain, or it may inefficiently amplify G-C rich regions (34,35). Inverted repeat regions inhibit the elongation of lagging strands by folding the single strand nucleotides (36).

Different sequencing platforms have their own pros and cons. The Illumina sequencing platform exhibits a lower error rate (10^{-3}) than other platforms, but begins to exhibit significant loss of quality after 50 cycles. Additionally, the shorter reads (150–300bp) produced by Illumina are more difficult to assemble into a complete genome using a de novo assembly algorithm, and consequently may not properly align with the reference genome (37). Shorter reads may also inhibit accurate assembly if the reads include indels, or translocations that occur during sequence alignment (38). The SOLiD platform requires significantly more time for sequencing, and the reads produced are not suitable for de novo assembly (39), while the neighbor wells in the Roche 454 platform may produce signal noise which can persistently disrupt the reading quality of the target well (40).

The availability of a genome reference greatly reduces the error rate during genome assembly. However, if no reference genome is available, de novo assembly will be performed to combine the overlapping reads into longer pieces, called contigs. Two classic algorithms used for de novo assembly are overlap graphs and *De Bruijn* graphs (26). Assembly errors may be reduced by adjusting the *de novo* assembly parameters used to map reads back to contigs, including k-mer (word size), bubble size, length fraction and similarity fractions (41,42). Highly repetitive sequences are another challenge for *de novo* assembly, as genomic assemblers are frequently unable to differentiate between two reads exhibiting similar repetitive patterns (26). N50, maximum length and average coverage are also essential factors to be evaluated for assembly quality.

Second generation sequencing may also be limited by a lack of information associated with DNA modifications that occur during sequencing amplification, and this information might be essential for understanding the genetic bases of diseases.

Latest techniques to improve NGS

Currently, 3rd generation sequencing platforms are being developed to improve the accuracy of whole genome sequencing data and acquire a more complete characterization of DNA modification, as well as DNA and protein interactions. These 3rd generation sequencers utilize sensors to directly inspect single nucleotides, and this process eliminates the necessity of sample washing following nucleotide amplification. Nanopore sequencing is one of the most developed of these sequencers. With a diameter of 1 nm, the Nanopore traps a single nucleotide from the fragment, and detects the ion current change, which varies depending on the size, shape and length of the DNA sequence (19). Commercially

available platforms include the Oxford NanoporeMinION® and PromethION®.

The development of libraries with longer inserts, coupled with a more powerful sequencing platform, would reduce the error rates associated with short read pair-end sequencing and improve the accuracy of the genomic data. The PacBio-SMRT® sequencing platform is based on real-time sequencing using a synthesis method, and is designed to sequence longer fragments. After each fluorescence-labeled nucleotide is incorporated into the DNA strand, the fluorescent label is removed, leaving the synthesized strand(5). This approach allows for detection of subsequent DNA modifications, including methylation or acetylation. The technology also provides the capability of sequencing single fragments from 5–20kb, which can close the gaps between contigs.

The Mate Pair library can store insert fragments from 1.5–20kb, but requires a large input of DNA to build the library. For instance, Nxseq® Long Mate Pair Library Kit from Lucigen was able to prepare a long insert library when running a sample on an Illumina sequencer. These libraries can be prepared with fragments from 2–8 kb using bead-based sizing methods, or 10–20 kb using gel-based sizing methods (43). The sheared and end-repaired DNA fragments are ligated to a coupler with a unique Chimera Code sequence, forming a circular DNA. Exonuclease is used to digest the unwanted DNA, following which a junction code is inserted into the sequence. The library is then amplified using PCR, and Mate Pair sequencing is performed using an Illumina sequencer.

Identifying novel cancer biomarkers using advanced genomics technologies

Through the use of advanced molecular techniques, such as Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), Amplicon sequencing, Target resequencing, and Metagenomics, research has identified a number of factors associated with tumorigenesis in addition to mutations of known tumor suppressor genes. These factors include novel gene mutations, gene fusions and re-arrangements, Long non-codling RNAs (IncRNAs), microRNAs, copy number variations (CNVs), methylation, and structural variations, which can contribute to cancer development.

Gene mutations – Traditionally, gene mutations have been identified by PCR amplification and Sanger Sequencing (44-46). However, the more recent development of genomic technologies has led to the discovery of novel genes, biomarkers, or new mutations of previously identified genes. Comparative genomic analyses have improved our ability to construct more complete and systematic gene mutation profiles (47-49). A partial list of genes identified in the last 5 years using advanced genomic approaches is presented in Table 1.

Gene fusions – Gene fusion, caused by chromosome translocation, is the joining of two unrelated exons, producing a chimeric mRNA transcript and protein (50). It has been frequently identified in diverse types of cancers in recent years. In 2011, using high-throughput RNA sequencing, Ha et al. identified 2 previously-characterized, and 3 novel in-frame gene fusions, from 10 breast cancer samples (50). These results suggested a relationship between gene fusions and *BRCA1* breast cancer, but due to the small sample size, it was not possible to identify gene fusion as the major cause of *BRCA1* breast cancer. In another related study, analysis of the structural variations of breast cancer genomes, using a combination of classical molecular cytogenetic approaches with pair-end sequencing, revealed the expression of 12 fusion genes (51). Moreover, 24 novel and 3 previously studied fusions genes in breast cancer were also identified by Edgren et al. using RNA sequencing and improved bioinformatics methods (52). By using pair-end deep sequencing technology, Inaki et al. reported that approximately 50% of genomic rearrangements lead to gene fusions, of which the fusion RPS6KB1-VMP1 is expressed in approximately30% of breast cancers (53). Another recent study of genomic variations of gastrointestinal stromal tumors using transcriptome sequencing identified 328 gene fusions, most of which are related to IGF2 gene (54), while several studies of lung cancer in non-smokers have found that ALK-, ROS1-, RE- and NRG1-related genes fusions are potential therapeutic targets for aggressive lung diseases (55,56). Collectively, these studies illustrate that analyzing a great number of samples by massively parallel sequencing or target sequencing in both genomic and transcriptomic levels is required in order to acquire a more comprehensive understanding of cancer diseases (50). Currently there are several cancer fusion genes databases available, including: ChimerDB (http://biome.ewha.ac.kr:8080/FusionGene/index.jsp), comprised of more than 2700 fusion transcripts with high confidence (57); the Mitelman Database of Chromosome Aberrations and Gene Fusions (http://cgap.nci.nih.gov/Chromosomes/Mitelman) comprised of approximately 10,000 fusions, and the TCGA Fusion gene Data Portal (http://54.84.12.177/PanCanFusV2/) comprised of more than 8700 fusion genes (58).

Long non-coding RNA (lncRNA) – lncRNA is another newly-identified epigenetic factor that plays a significant role in cancer initiation and development. In a very recent study using helicos single molecule sequencing, the expression variations of a subset of Glioblastoma Multiforme (GBM) lncRNA were identified. This study further indicated that the Bromodomain and extraterminal (BET) domain protein, a promising anticancer target, directly regulates lncRNA expression (59). Additionally, analysis of the oncogenic role of a 2455-bp HNF1A-AS1 lncRNA on adenocarcinoma, based on comparison of 40 lung adenocarcinoma and non-tumor samples, revealed that expression levels of HNF1A-AS1 were significantly up-regulated in the cancer tissues (60). The results of these and similar studies, suggest that the development of novel long non-coding RNA-based biomarkers will play a significant role in early cancer diagnostics and cancer stage differentiations.

MicroRNAs – MicroRNAs are small (20–22 nt) non-coding RNAs involved with post-translational regulation of gene expression. Cancers are always closely related to the dysregulation of specific miRNA genes; therefore, discovery and control of dysregulated miRNA could help design future miRNA-based chemotherapies (61). Due to the stability of miRNAs in blood,

Table 1. Selected novel genes mutations in cancer diseases during the last five years.

Cancer Type	Genes or Biomarkers	Mutation rate	Technologies	Function	Ref
Lung Cancer	*PARK2*	2.7%–5.6%	WES	Tumor suppressor	(85)
	U2AF1, RBM10 and ARID1A	3%–7%	WGS, WES	RNA binding protein	(86)
	KIAA1324	78%	RNA seq	Tumor suppressor	(87)
Gastric Cancer	*ARID1A*	73–83%	WES	Chromatin remodeling	(88)
	MLL3, MLL,ARID1A	47%	WES	Chromatin remodeling	(89)
	PIK3CA/B, R-spondin, β-catenin, BRAF/RAF1, APC, ZBTB16/ PLZF	N/A	WES RNA seq	N/A	(90)
Prostate Cancer	*SPOP, FOXA1, MED12*	6–15%	WES	SPOP- encodes the substrate-binding subunit of a Cullin-based E3 ubiquitin ligase;FOXA1-transcriptional activators for liver-specific transcripts;MED12- encodes a subunit of the mediator complex and the Cyclin-dependent kinase 8 (CDK8) sub-complex	(91)
	CHD1, MLL2, FOXA1	3.4%–8.6%	WES	CHD1-Chromatin remodeling MLL2-Chromatin remodeling FOXA1-AR collaborating factor	(92)
	MYCL	8%–20%	WGS& Copy number aberration (CNA)	Amplification of MYCL is associated with TP53 deletion, unique profiles of DNA damage, and transcriptional dysregulation	(93)
	GATA2 and EZH2		WGS RNA seq	**GATA2-**Regulates transcription of genes involved in the development and proliferation of hematopoietic and endocrine cell lineages	(94)
Leukemia	*KRAS and CREBBP*	KRAS-63% CREBBP-24%	Whole-exome and subsequent targeted deep sequencing	KRAS-receptor tyrosine kinase (RTK)/RAS signaling pathway CREBBP-binding protein for the cAMP-response element binding protein	(95)
Renal Cell Carcinoma (RCC)	*TCEB1*	95.4%	WGS WES Copy number and/or methylation analyses	Encodes protein elongin C, a subunit of the transcription factor B (SIII) complex	(96)
	PBRM1(polybromo 1)	41%	WES	Encodes chromatin-targeting subunit of the PBAF SWI/SNF chromatin remodeling complex	(97)
Wilms Tumor	*CASP8, H19, MIR195, RB1, TSPAN32*	N/A	Comprehensive methylome analysis and RNA seq	Tumor suppressors	(98)
Metastatic colorectal carcinoma (mCRC)	INPPL1	4%	WGS	Oncogene, involved in PI3 kinase signaling and promotion of tumor development	(99,100)
Diffuse intrinsic pontine gliomas (DIPG)	*ACVR1*	20–21%	WGS	Encodes a type I activin receptor serine/threonine kinase	(101,102)
Low-grade gliomas (LGGs)	*BRAF, RAF1, FGFR1, MYB, MYBL1,H3F3A and ATRX*	N/A	WGS	N/A	(103)
Glioblastomas	*EGFR, PDGFRA, MET, EGFRvIII, PTEN, 1p/19q*	N/A	OncoCopy and OncoMap	N/A	(104)
Hepatocellular carcinoma (HCC)	*ARID1A, RPS6KA3, NFE2L2, IRF2 RPS6KA3-AXIN1 and NFE2L2-CTNNB1*	N/A	WES Copy number analysis	N/A	(105,106)
Melanoma	*PPP6C, RAC1, SNX31, TACC1, STK19, and ARID2*	N/A	Integration of Exome and copy number data	N/A	(107)
Pancreatic Cancer	*ZIM2, MAP2K4, NALCN, SLC16A4 and MAGEA6*	N/A	WES and copy-number analysis	N/A	(108)

they have been identified as useful biomarkers, and profiling of serum circulating miRNAs has proven to be a useful tool for cancer diagnostics (62,63). In early 2009, Roche 454 pyrosequencing analysis of small RNA from breast cancer samples identified 5 novel miRNAs (64). Later, in 2011, Persson et al. performed extensive next-generation sequencing of paired samples of normal and tumor-adjacent breast tissue, and identified 361 new miRNA precursors. Approximately 10% of these miRNA were located in a highly amplified genomic region of breast cancer cells (65). In another study using Miseq from Illumina and qRT-PCR, samples from 250 bladder cancer patients and 240 controls were compared, and a six-miRNA panel (miR-152, miR-148b-3p, miR-3187-3p, miR-15b-5p, miR-27a-3p and miR-30a-5p) was developed to diagnose bladder cancer (66). Additionally, the analysis of miRNA profiles of gastric cancer using NGS-based whole-transcriptome profiling of 274 gastric tissues resulted in the identification of four gastric cancer-specific miRNAs, among which miR-29c was significantly reduced in cancer tissues, as compared with the controls (67). Therefore, detection of miR-29c

expression levels may represent a sensitive biomarker for early diagnostics.

Cell-free circulating tumor DNA (ctDNA)–CtDNAs have recently been identified as promising non-invasive cancer biomarkers used for the detection, monitoring and treatment of cancers. Several studies have reported the presence of high level of ctDNAs in the serum of patients with lung, colorectal and breast cancers (68-70). To detect these biomarkers, Newman et al. developed CAPP-Seq (Cancer Personalized Profiling by deep Sequencing), an ultrasensitive method to quantify ctDNA (71). In this approach, the library preparation method is improved for low DNA input, and biotinylated DNA probes, specifically targeted to mutated regions of the cancer genome, are used as 'selectors' to identify cancer related mutagenesis. Another reported approach for detection of these biomarkers utilized whole genome sequencing to identify circulating DNA methylation changes as a signature of recurrence risk in cancer patients following surgery (72).

DNA methylation – DNA methylation is a catalytic process by adding methyl groups to the cytosine using DNA methyltransferases (DNMTs) (73). Methylation of DNA is involved in many critical cellular mechanisms, such as suppression of repetitive elements, carcinogeneiss, cell cycle regulation, alteration of gene expression. It has been considered as one of the critical epigenetic elements related to cancer development. Methylation mainly occurred at the cytosine within CpG island, and multiple studies proved that in cancer cells, a large number of genes expression was regulated by DNA methylation at the promoter regions (74).

By using the next generation sequencing technologies, more cancer related DNA methylation patterns, the metastatic or re-occurance status of cancer patient, and the classifications of cancer diseases were detected more accurately. To characterize the status of DNA methylation in the genome, specific techniques were developed, such as whole genome bisulfate-converted sequencing (75) and methylated DNA immunoprecipitation (MeDIP) (76). Bisulfite genomic sequencing was first developed by Frommer et al in 1992 (77). In principle, sodium bisulfate could convert unmethylated cytosine into uracils, which will be recognized as thymines during PCR and sequencing. Methylated cytosines, which are not converted, will be observed as cytosines in the sequencing data. Whole genome bisulfate sequencing has been extensively used in cancer diagnostics, surveillance and treatment. For instance, by using Whole-genome bisulfate sequencing and differential methylation analysis, 21 novel hotspots in the CpG islands were identified as biomarkers to differentiate health individuals and breast cancer free survivors (72). In order to target on single cytosine modification on the whole genome sequence, bisulfate-treated sequence data were aligned to reference genome and single nucleotide variants were detected by *MethylExtract* software (78). Methylated DNA immunoprecipitation (MeDIP) is another widely used technique for reduced regional sequencing. Basically, antibody against 5-methylcytosine (5mC) were used to purify and enrich methylated DNA sequences. These purified DNA pieces will be sequenced by next-generation sequencing. An example is the MeDIP sequencing of whole-genome DNA methylation profiles of 8 human breast cancer cell lines. A massively reduced methylation level were observed in the CpG-poor regions of breast cancer cell lines, compared to normal control (79).

Conclusions

The optimization of bioinformatics tools has further expanded our capacity to develop combinational studies, utilizing multiple genetic factors related to cancer, to perform comprehensive molecular profiling (80,81). In 2013, the Cancer Genome Atlas Network Group combined 5 genomic analysis platforms, including genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, and microRNA sequencing, to reexamine gene expression subtypes and the existence of four main breast cancer classes (82). In addition to the three major genes (TP53, PIK3CA and GATA3) related to breast cancer (>10% occurrence), numerous novel genes, signal pathways and protein-expression-defined subgroups were also identified. Another comprehensive molecular profiling was compiled based on 276 colorectal carcinoma samples. Exome sequencing, copy number variations, methylation, and miRNA analyses were conducted, and 24 significant gene mutations and recurrent copy number variations were identified (83). These examples suggest that high-volume data analysis has the potential to rapidly advance research in cancer genomics.

Whole genome sequencing has also been applied to study the chemotherapeutic drug response of cancer treatment. One example is the Poly ADP ribose polymerase (PARP) inhibitor – olaparib, which has been used to treat patients with germline BRCA mutations. However, evidences have shown that patients with somatic BRCA mutations may also be treated effectively with olaprib, therefore, genetic testing of the *BRCA1* and *BRCA2* variants, using next generation sequencing, would be greatly helpful for cancer diagnostics and proper chemotherapy.

Future Perspective

Accurate genomic profiling of cancer diseases requires both improvement of sequencing platforms and the optimization of bioinformatics tools. It is anticipated that base call errors produced by NGS can be reduced by deep re-sequencing, optimization of clustering and generation, or refinement of data trimming thresholds. Due to the complexity of the tumor microenvironment, intratumor heterogeneity has been observed in different tumor cells within single tumor tissue samples. Single cell sequencing, including single cell genome sequencing and single cell transcriptomics sequencing, has exhibited increased sensitivity and quantitative accuracy (84). Additionally, as an alternative to analyzing single genetic or epigenetic factor in cancer cells, comprehensive molecular profiling is evolving as a more powerful tool to identify novel cancer biomarkers or drug binding targets.

At present, significant challenges exist in the clinical application of NGS for oncology, including the high cost of genome sequencing, long sequencing time and difficulties with data analysis and interpretation. None-

theless, the accumulation of data from these studies support the validity of these approaches as research and diagnostic tools, and further refinements to molecular detection technology will bring us one step closer to the translation of personalized genomic data into medical diagnostics and therapies.

References

1. Soriano, A., Castells, A., Ayuso, C., Ayuso, J. R., de Caralt, M. T., Gines, M. A., Real, M. I., Gilabert, R., Quinto, L., Trilla, A., Feu, F., Montanya, X., Fernandez-Cruz, L., and Navarro, S. Preoperative staging and tumor resectability assessment of pancreatic cancer: prospective study comparing endoscopic ultrasonography, helical computed tomography, magnetic resonance imaging, and angiography. *The American journal of gastroenterology*. 2004, **99**: 492-501. doi: 10.1111/j.1572-0241.2004.04087.x

2. Seitz, M., Shukla-Dave, A., Bjartell, A., Touijer, K., Sciarra, A., Bastian, P. J., Stief, C., Hricak, H., and Graser, A. Functional magnetic resonance imaging in prostate cancer. *European urology*. 2009, **55**: 801-814. doi: 10.1016/j.eururo.2009.01.027

3. Previati, M., Manfrini, M., Galasso, M., Zerbinati, C., Palatini, J., Gasparini, P., and Volinia, S. Next generation analysis of breast cancer genomes for precision medicine. *Cancer letters*. 2013, **339**: 1-7. doi: 10.1016/j.canlet.2013.07.018

4. Ng, P. C., and Kirkness, E. F. Whole genome sequencing. *Methods in molecular biology*. 2010, **628**: 215-226. doi: 10.1007/978-1-60327-367-1_12

5. Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., and Korlach, J. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature methods*. 2013, **10**: 563-569. doi: 10.1038/nmeth.2474

6. Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T. J., Schatz, M. C., Delcher, A. L., Roberts, M., Marcais, G., Pop, M., and Yorke, J. A. GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome research*. 2012, **22**: 557-567. doi: 10.1101/gr.131383.111

7. Magoc, T., Pabinger, S., Canzar, S., Liu, X., Su, Q., Puiu, D., Tallon, L. J., and Salzberg, S. L. GAGE-B: an evaluation of genome assemblers for bacterial organisms. *Bioinformatics*. 2013, **29**: 1718-1725. doi: 10.1093/bioinformatics/btt273

8. Earl, D., Bradnam, K., St John, J., Darling, A., Lin, D., Fass, J., Yu, H. O., Buffalo, V., Zerbino, D. R., Diekhans, M., Nguyen, N., Ariyaratne, P. N., Sung, W. K., Ning, Z., Haimel, M., Simpson, J. T., Fonseca, N. A., Birol, I., Docking, T. R., Ho, I. Y., Rokhsar, D. S., Chikhi, R., Lavenier, D., Chapuis, G., Naquin, D., Maillet, N., Schatz, M. C., Kelley, D. R., Phillippy, A. M., Koren, S., Yang, S. P., Wu, W., Chou, W. C., Srivastava, A., Shaw, T. I., Ruby, J. G., Skewes-Cox, P., Betegon, M., Dimon, M. T., Solovyev, V., Seledtsov, I., Kosarev, P., Vorobyev, D., Ramirez-Gonzalez, R., Leggett, R., MacLean, D., Xia, F., Luo, R., Li, Z., Xie, Y., Liu, B., Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Yin, S., Sharpe, T., Hall, G., Kersey, P. J., Durbin, R., Jackman, S. D., Chapman, J. A., Huang, X., DeRisi, J. L., Caccamo, M., Li, Y., Jaffe, D. B., Green, R. E., Haussler, D., Korf, I., and Paten, B. Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome research*. 2011, **21**: 2224-2241. doi: 10.1101/gr.126599.111

9. Bradnam, K. R., Fass, J. N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J. A., Chapuis, G., Chikhi, R., Chitsaz, H., Chou, W. C., Corbeil, J., Del Fabbro, C., Docking, T. R., Durbin, R., Earl, D., Emrich, S., Fedotov, P., Fonseca, N. A., Ganapathy, G., Gibbs, R. A., Gnerre, S., Godzaridis, E., Goldstein, S., Haimel, M., Hall, G., Haussler, D., Hiatt, J. B., Ho, I. Y., Howard, J., Hunt, M., Jackman, S. D., Jaffe, D. B., Jarvis, E., Jiang, H., Kazakov, S., Kersey, P. J., Kitzman, J. O., Knight, J. R., Koren, S., Lam, T. W., Lavenier, D., Laviolette, F., Li, Y., Li, Z., Liu, B., Liu, Y., Luo, R., Maccallum, I., Macmanes, M. D., Maillet, N., Melnikov, S., Naquin, D., Ning, Z., Otto, T. D., Paten, B., Paulo, O. S., Phillippy, A. M., Pina-Martins, F., Place, M., Przybylski, D., Qin, X., Qu, C., Ribeiro, F. J., Richards, S., Rokhsar, D. S., Ruby, J. G., Scalabrin, S., Schatz, M. C., Schwartz, D. C., Sergushichev, A., Sharpe, T., Shaw, T. I., Shendure, J., Shi, Y., Simpson, J. T., Song, H., Tsarev, F., Vezzi, F., Vicedomini, R., Vieira, B. M., Wang, J., Worley, K. C., Yin, S., Yiu, S. M., Yuan, J., Zhang, G., Zhang, H., Zhou, S., and Korf, I. F. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*. 2013, **2**: 10. doi: 10.1186/2047-217X-2-10

10. Baker, M. De novo genome assembly: what every biologist should know. *Nat Meth*. 2012, **9**: 333-337

11. Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 1977, **74**: 5463-5467

12. Sanger, F., and Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*. 1975, **94**: 441-448

13. Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y. J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X. Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D. M., Margulies, M., Weinstock, G. M., Gibbs, R. A., and Rothberg, J. M. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008, **452**: 872-876. doi: 10.1038/nature06884

14. Voelkerding, K. V., Dames, S. A., and Durtschi, J. D. Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry*. 2009, **55**: 641-658. doi: 10.1373/clinchem.2008.112789

15. Pennisi, E. Genomics. Semiconductors inspire new sequencing technologies. *Science*. 2010, **327**: 1190. doi: 10.1126/science.327.5970.1190

16. Brakmann, S. Single-molecule analysis: A ribosome in action. *Nature*. 2010, **464**: 987-988. doi: 10.1038/464987a

17. Meyer, M., and Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*. 2010, **2010**: pdb prot5448. doi: 10.1101/pdb.prot5448

18. Kasianowicz, J. J., Brandin, E., Branton, D., and Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America*. 1996, **93**: 13770-13773

19. Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature nanotechnology*. 2009, **4**: 265-270. doi: 10.1038/nnano.2009.12

20. Genoscribe. (2014) 6 Methods to Fragment Your DNA / RNA for Next-Gen Sequencing.

21. Covaris. DNA and RNA Shearing. http://covarisinc.com/applications/dnarna-shearing-for-ngs/

22. Knierim, E., Lucke, B., Schwarz, J. M., Schuelke, M., and Seelow, D. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PloS one*. 2011, **6**: e28240. doi: 10.1371/journal.pone.0028240

23. Marine, R., Polson, S. W., Ravel, J., Hatfull, G., Russell, D., Sullivan, M., Syed, F., Dumas, M., and Wommack, K. E. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Applied and environmental microbiology*. 2011, **77**: 8071-8079. doi: 10.1128/AEM.05610-11

24. Illumina. Preparing Libraries for Sequencing on the MiSeq. 2013: 16
25. Illumina.com. Sequencing by Synthesis (SBS) Technology. http://www.illumina.com/technology/next-generation-sequencing/sequencing-technology.html
26. Miller, J. R., Koren, S., and Sutton, G. Assembly algorithms for next-generation sequencing data. *Genomics*. 2010, **95**: 315-327. doi: 10.1016/j.ygeno.2010.03.001
27. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., and Rothberg, J. M. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005, **437**: 376-380. doi: 10.1038/nature03959
28. Quinn, N. L., Levenkova, N., Chow, W., Bouffard, P., Boroevich, K. A., Knight, J. R., Jarvie, T. P., Lubieniecki, K. P., Desany, B. A., Koop, B. F., Harkins, T. T., and Davidson, W. S. Assessing the feasibility of GS FLX Pyrosequencing for sequencing the Atlantic salmon genome. *BMC genomics*. 2008, **9**: 404. doi: 10.1186/1471-2164-9-404
29. Weir, B. A., Woo, M. S., Getz, G., Perner, S., Ding, L., Beroukhim, R., Lin, W. M., Province, M. A., Kraja, A., Johnson, L. A., Shah, K., Sato, M., Thomas, R. K., Barletta, J. A., Borecki, I. B., Broderick, S., Chang, A. C., Chiang, D. Y., Chirieac, L. R., Cho, J., Fujii, Y., Gazdar, A. F., Giordano, T., Greulich, H., Hanna, M., Johnson, B. E., Kris, M. G., Lash, A., Lin, L., Lindeman, N., Mardis, E. R., McPherson, J. D., Minna, J. D., Morgan, M. B., Nadel, M., Orringer, M. B., Osborne, J. R., Ozenberger, B., Ramos, A. H., Robinson, J., Roth, J. A., Rusch, V., Sasaki, H., Shepherd, F., Sougnez, C., Spitz, M. R., Tsao, M. S., Twomey, D., Verhaak, R. G., Weinstock, G. M., Wheeler, D. A., Winckler, W., Yoshizawa, A., Yu, S., Zakowski, M. F., Zhang, Q., Beer, D. G., Wistuba, II, Watson, M. A., Garraway, L. A., Ladanyi, M., Travis, W. D., Pao, W., Rubin, M. A., Gabriel, S. B., Gibbs, R. A., Varmus, H. E., Wilson, R. K., Lander, E. S., and Meyerson, M. Characterizing the cancer genome in lung adenocarcinoma. *Nature*. 2007, **450**: 893-898. doi: 10.1038/nature06358
30. Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J. A., Costa, G., McKernan, K., Sidow, A., Fire, A., and Johnson, S. M. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome research*. 2008, **18**: 1051-1063. doi: 10.1101/gr.076463.108
31. Illumina. (2014) Nextera Low Plex Pooling Guidelines. in *Technical Note: DNA Analysis*, http://www.illumina.com/documents/products/technotes/technote_nextera_low_plex_pooling_guidelines.pdf
32. Krueger, F., Andrews, S. R., and Osborne, C. S. Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by deferred cluster calling. *PloS one*. 2011, **6**: e16607. doi: 10.1371/journal.pone.0016607
33. Kircher, M., Heyn, P., and Kelso, J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC genomics*. 2011, **12**: 382. doi: 10.1186/1471-2164-12-382
34. Chen, Y. C., Liu, T., Yu, C. H., Chiang, T. Y., and Hwang, C. C. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PloS one*. 2013, **8**: e62856. doi: 10.1371/journal.pone.0062856
35. Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C., and Jaffe, D. B. Characterizing and measuring bias in sequence data. *Genome biology*. 2013, **14**: R51. doi: 10.1186/gb-2013-14-5-r51
36. Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M. C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., and Kanaya, S. Sequence-specific error profile of Illumina sequencers. *Nucleic acids research*. 2011, **39**: e90. doi: 10.1093/nar/gkr344
37. Paszkiewicz, K., and Studholme, D. J. De novo assembly of short sequence reads. *Briefings in bioinformatics*. 2010, **11**: 457-472. doi: 10.1093/bib/bbq020
38. Grimm, D., Hagmann, J., Koenig, D., Weigel, D., and Borgwardt, K. Accurate indel prediction using paired-end short reads. *BMC genomics*. 2013, **14**: 132. doi: 10.1186/1471-2164-14-132
39. Huang, Y. F., Chen, S. C., Chiang, Y. S., Chen, T. H., and Chiu, K. P. Palindromic sequence impedes sequencing-by-ligation mechanism. *BMC systems biology*. 2012, **6 Suppl 2**: S10. doi: 10.1186/1752-0509-6-S2-S10
40. Kircher, M., and Kelso, J. High-throughput DNA sequencing--concepts and limitations. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2010, **32**: 524-536. doi: 10.1002/bies.200900181
41. Jayasena, A. S., Secco, D., Bernath-Levin, K., Berkowitz, O., Whelan, J., and Mylne, J. S. Next generation sequencing and de novo transcriptomics to study gene evolution. *Plant methods*. 2014, **10**: 34. doi: 10.1186/1746-4811-10-34
42. Arun-Chinnappa, K. S., and McCurdy, D. W. De novo assembly of a genome-wide transcriptome map of Vicia faba (L.) for transfer cell research. *Frontiers in plant science*. 2015, **6**: 217. doi: 10.3389/fpls.2015.00217
43. Lucigen.com. (2012) Long-span, mate-pair scaffolding and other methods for faster next-generation sequencing library creation. Nature method, http://www.nature.com/nmeth/journal/v9/n9/full/nmeth.f.358.html
44. Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., and Micklem, G. Identification of the breast cancer susceptibility gene BRCA2. *Nature*. 1995, **378**: 789-792. doi: 10.1038/378789a0
45. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., and et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. 1994, **266**: 66-71
46. Wang, K., Li, Y., Jiang, Y. Z., Dai, C. F., Patankar, M. S., Song, J. S., and Zheng, J. An endogenous aryl hydrocarbon receptor ligand inhibits proliferation and migration of human ovarian cancer cells. *Cancer letters*. 2013, **340**: 63-71. doi: 10.1016/j.canlet.2013.06.026
47. Watson, I. R., Takahashi, K., Futreal, P. A., and Chin, L. Emerging patterns of somatic mutations in cancer. *Nature reviews. Genetics*. 2013, **14**: 703-718. doi: 10.1038/nrg3539
48. Toffoli, S., Bar, I., Abdel-Sater, F., Delree, P., Hilbert, P., Cavallin, F., Moreau, F., Van Criekinge, W., Lacroix-Triki, M., Campone, M., Martin, A. L., Roche, H., Machiels, J. P., Carrasco, J., and Canon, J. L. Identification by array comparative genomic hybridization of a new amplicon on chromosome 17q highly recurrent in BRCA1 mutated triple negative breast cancer. *Breast cancer research : BCR*. 2014, **16**: 466. doi: 10.1186/s13058-014-0466-y
49. Li, Y., Wang, K., Jiang, Y. Z., Chang, X. W., Dai, C. F., and Zheng, J. 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) inhibits human ovarian cancer cell proliferation. *Cellular oncology*. 2014, **37**: 429-437. doi: 10.1007/s13402-014-0206-4
50. Ha, K. C., Lalonde, E., Li, L., Cavallone, L., Natrajan, R., Lambros, M. B., Mitsopoulos, C., Hakas, J., Kozarewa, I., Fenwick,

K., Lord, C. J., Ashworth, A., Vincent-Salomon, A., Basik, M., Reis-Filho, J. S., Majewski, J., and Foulkes, W. D. Identification of gene fusion transcripts by transcriptome sequencing in BRCA1-mutated breast cancers and cell lines. *BMC medical genomics*. 2011, **4**: 75. doi: 10.1186/1755-8794-4-75

51. Schulte, I., Batty, E. M., Pole, J. C., Blood, K. A., Mo, S., Cooke, S. L., Ng, C., Howe, K. L., Chin, S. F., Brenton, J. D., Caldas, C., Howarth, K. D., and Edwards, P. A. Structural analysis of the genome of breast cancer cell line ZR-75-30 identifies twelve expressed fusion genes. *BMC genomics*. 2012, **13**: 719. doi: 10.1186/1471-2164-13-719

52. Edgren, H., Murumagi, A., Kangaspeska, S., Nicorici, D., Hongisto, V., Kleivi, K., Rye, I. H., Nyberg, S., Wolf, M., Borresen-Dale, A. L., and Kallioniemi, O. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome biology*. 2011, **12**: R6. doi: 10.1186/gb-2011-12-1-r6

53. Inaki, K., Hillmer, A. M., Ukil, L., Yao, F., Woo, X. Y., Vardy, L. A., Zawack, K. F., Lee, C. W., Ariyaratne, P. N., Chan, Y. S., Desai, K. V., Bergh, J., Hall, P., Putti, T. C., Ong, W. L., Shahab, A., Cacheux-Rataboul, V., Karuturi, R. K., Sung, W. K., Ruan, X., Bourque, G., Ruan, Y., and Liu, E. T. Transcriptional consequences of genomic structural aberrations in breast cancer. *Genome research*. 2011, **21**: 676-687. doi: 10.1101/gr.113225.110

54. Kang, G., Yun, H., Sun, C. H., Park, I., Lee, S., Kwon, J., Do, I., Hong, M. E., Vrancken, M. V., Lee, J., Park, J. O., Cho, J., Kim, K. M., and Sohn, T. S. Integrated genomic analyses identify frequent gene fusion events and VHL inactivation in gastrointestinal stromal tumors. *Oncotarget*. 2015,

55. Fernandez-Cuesta, L., and Thomas, R. K. Molecular Pathways: Targeting NRG1 Fusions in Lung Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2015, **21**: 1989-1994. doi: 10.1158/1078-0432.CCR-14-0854

56. Ha, S. Y., Choi, S. J., Cho, J. H., Choi, H. J., Lee, J., Jung, K., Irwin, D., Liu, X., Lira, M. E., Mao, M., Kim, H. K., Choi, Y. S., Shim, Y. M., Park, W. Y., Choi, Y. L., and Kim, J. Lung cancer in never-smoker Asian females is driven by oncogenic mutations, most often involving EGFR. *Oncotarget*. 2015, **6**: 5465-5474

57. Kim, P., Yoon, S., Kim, N., Lee, S., Ko, M., Lee, H., Kang, H., Kim, J., and Lee, S. ChimerDB 2.0--a knowledgebase for fusion genes updated. *Nucleic acids research*. 2010, **38**: D81-85. doi: 10.1093/nar/gkp982

58. Yoshihara, K., Wang, Q., Torres-Garcia, W., Zheng, S., Vegesna, R., Kim, H., and Verhaak, R. G. The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*. 2014. doi: 10.1038/onc.2014.406

59. Pastori, C., Kapranov, P., Penas, C., Peschansky, V., Volmar, C. H., Sarkaria, J. N., Bregy, A., Komotar, R., St Laurent, G., Ayad, N. G., and Wahlestedt, C. The Bromodomain protein BRD4 controls HOTAIR, a long noncoding RNA essential for glioblastoma proliferation. *Proceedings of the National Academy of Sciences of the United States of America*. 2015. doi: 10.1073/pnas.1424220112

60. Wu, Y., Liu, H., Shi, X., Yao, Y., Yang, W., and Song, Y. The long non-coding RNA HNF1A-AS1 regulates proliferation and metastasis in lung adenocarcinoma. *Oncotarget*. 2015, **6**: 9160-9172

61. Croce, C. M. Causes and consequences of microRNA dysregulation in cancer. *Nature reviews. Genetics*. 2009, **10**: 704-714. doi: 10.1038/nrg2634

62. Moltzahn, F., Olshen, A. B., Baehner, L., Peek, A., Fong, L., Stoppler, H., Simko, J., Hilton, J. F., Carroll, P., and Blelloch, R. Microfluidic-based multiplex qRT-PCR identifies diagnostic and prognostic microRNA signatures in the sera of prostate cancer patients. *Cancer research*. 2011, **71**: 550-560. doi: 10.1158/0008-5472.CAN-10-1229

63. Chen, Z. H., Zhang, G. L., Li, H. R., Luo, J. D., Li, Z. X., Chen, G. M., and Yang, J. A panel of five circulating microRNAs as potential biomarkers for prostate cancer. *The Prostate*. 2012, **72**: 1443-1452. doi: 10.1002/pros.22495

64. Nygaard, S., Jacobsen, A., Lindow, M., Eriksen, J., Balslev, E., Flyger, H., Tolstrup, N., Moller, S., Krogh, A., and Litman, T. Identification and analysis of miRNAs in human breast cancer and teratoma samples using deep sequencing. *BMC medical genomics*. 2009, **2**: 35. doi: 10.1186/1755-8794-2-35

65. Persson, H., Kvist, A., Rego, N., Staaf, J., Vallon-Christersson, J., Luts, L., Loman, N., Jonsson, G., Naya, H., Hoglund, M., Borg, A., and Rovira, C. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. *Cancer research*. 2011, **71**: 78-86. doi: 10.1158/0008-5472.CAN-10-1869

66. Jiang, X., Du, L., Wang, L., Li, J., Liu, Y., Zheng, G., Qu, A., Zhang, X., Pan, H., Yang, Y., and Wang, C. Serum microRNA expression signatures identified from genome-wide microRNA profiling serve as novel noninvasive biomarkers for diagnosis and recurrence of bladder cancer. *International journal of cancer. Journal international du cancer*. 2015, **136**: 854-862. doi: 10.1002/ijc.29041

67. Han, T. S., Hur, K., Xu, G., Choi, B., Okugawa, Y., Toiyama, Y., Oshima, H., Oshima, M., Lee, H. J., Kim, V. N., Chang, A. N., Goel, A., and Yang, H. K. MicroRNA-29c mediates initiation of gastric carcinogenesis by directly targeting ITGB1. *Gut*. 2015, **64**: 203-214. doi: 10.1136/gutjnl-2013-306640

68. Umetani, N., Kim, J., Hiramatsu, S., Reber, H. A., Hines, O. J., Bilchik, A. J., and Hoon, D. S. Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. *Clinical chemistry*. 2006, **52**: 1062-1069. doi: 10.1373/clinchem.2006.068577

69. Holdenrieder, S., Burges, A., Reich, O., Spelsberg, F. W., and Stieber, P. DNA integrity in plasma and serum of patients with malignant and benign diseases. *Annals of the New York Academy of Sciences*. 2008, **1137**: 162-170. doi: 10.1196/annals.1448.013

70. Agostini, M., Enzo, M. V., Bedin, C., Belardinelli, V., Goldin, E., Del Bianco, P., Maschietto, E., D'Angelo, E., Izzi, L., Saccani, A., Zavagno, G., and Nitti, D. Circulating cell-free DNA: a promising marker of regional lymphonode metastasis in breast cancer patients. *Cancer biomarkers : section A of Disease markers*. 2012, **11**: 89-98. doi: 10.3233/CBM-2012-0263

71. Newman, A. M., Bratman, S. V., To, J., Wynne, J. F., Eclov, N. C., Modlin, L. A., Liu, C. L., Neal, J. W., Wakelee, H. A., Merritt, R. E., Shrager, J. B., Loo, B. W., Jr., Alizadeh, A. A., and Diehn, M. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nature medicine*. 2014, **20**: 548-554. doi: 10.1038/nm.3519

72. Legendre, C., Gooden, G. C., Johnson, K., Martinez, R. A., Liang, W. S., and Salhia, B. Whole-genome bisulfite sequencing of cell-free DNA identifies signature associated with metastatic breast cancer. *Clinical epigenetics*. 2015, **7**: 100. doi: 10.1186/s13148-015-0135-8

73. Jeltsch, A. Molecular enzymology of mammalian DNA methyltransferases. *Current topics in microbiology and immunology*. 2006, **301**: 203-225

74. Kulis, M., and Esteller, M. DNA methylation and cancer. *Advances in genetics*. 2010, **70**: 27-56. doi: 10.1016/B978-0-12-380866-0.60002-2

75. Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., and Ecker, J. R. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009, **462**: 315-322. doi: http://www.nature.com/nature/

journal/v462/n7271/suppinfo/nature08514_S1.html
76. Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R. A., Niveleau, A., Cedar, H., and Simon, I. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nature genetics*. 2006, **38**: 149-153. doi: http://www.nature.com/ng/journal/v38/n2/suppinfo/ng1719_S1.html
77. Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*. 1992, **89**: 1827-1831
78. Barturen, G., Rueda, A., Oliver, J. L., and Hackenberg, M. MethylExtract: High-Quality methylation maps and SNV calling from whole genome bisulfite sequencing data. *F1000Research*. 2013, **2**: 217. doi: 10.12688/f1000research.2-217.v2
79. Ruike, Y., Imanaka, Y., Sato, F., Shimizu, K., and Tsujimoto, G. Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC genomics*. 2010, **11**: 137
80. Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014, **511**: 543-550. doi: 10.1038/nature13385
81. Tomczak, K., Czerwinska, P., and Wiznerowicz, M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary oncology*. 2015, **19**: A68-77. doi: 10.5114/wo.2014.47136
82. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012, **490**: 61-70. doi: 10.1038/nature11412
83. Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012, **487**: 330-337. doi: 10.1038/nature11252
84. Ren, S. C., Qu, M., and Sun, Y. H. Investigating intratumour heterogeneity by single-cell sequencing. *Asian journal of andrology*. 2013, **15**: 729-734. doi: 10.1038/aja.2013.106
85. Xiong, D., Wang, Y., Kupert, E., Simpson, C., Pinney, S. M., Gaba, C. R., Mandal, D., Schwartz, A. G., Yang, P., de Andrade, M., Pikielny, C., Byun, J., Li, Y., Stambolian, D., Spitz, M. R., Liu, Y., Amos, C. I., Bailey-Wilson, J. E., Anderson, M., and You, M. A recurrent mutation in PARK2 is associated with familial lung cancer. *American journal of human genetics*. 2015, **96**: 301-308. doi: 10.1016/j.ajhg.2014.12.016
86. Imielinski, M., Berger, A. H., Hammerman, P. S., Hernandez, B., Pugh, T. J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., Sougnez, C., Auclair, D., Lawrence, M. S., 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, B. E., Janne, P. A., Miller, V. A., Pao, W., Travis, W. D., Pass, H. I., Gabriel, S. B., Lander, E. S., Thomas, R. K., Garraway, L. A., Getz, G., and Meyerson, M. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012, **150**: 1107-1120. doi: 10.1016/j.cell.2012.08.029
87. Kang, J. M., Park, S., Kim, S. J., Kim, H., Lee, B., Kim, J., Park, J., Kim, S. T., Yang, H. K., Kim, W. H., and Kim, S. J. KIAA1324 suppresses gastric cancer progression by inhibiting the oncoprotein GRP78. *Cancer research*. 2015. doi: 10.1158/0008-5472.CAN-14-3751
88. Wang, K., Kan, J., Yuen, S. T., Shi, S. T., Chu, K. M., Law, S., Chan, T. L., Kan, Z., Chan, A. S., Tsui, W. Y., Lee, S. P., Ho, S. L., Chan, A. K., Cheng, G. H., Roberts, P. C., Rejto, P. A., Gibson, N. W., Pocalyko, D. J., Mao, M., Xu, J., and Leung, S. Y. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nature genetics*. 2011, **43**: 1219-1223. doi: 10.1038/ng.982
89. Zang, Z. J., Cutcutache, I., Poon, S. L., Zhang, S. L., McPherson, J. R., Tao, J., Rajasegaran, V., Heng, H. L., Deng, N., Gan, A., Lim, K. H., Ong, C. K., Huang, D., Chin, S. Y., Tan, I. B., Ng, C. C., Yu, W., Wu, Y., Lee, M., Wu, J., Poh, D., Wan, W. K., Rha, S. Y., So, J., Salto-Tellez, M., Yeoh, K. G., Wong, W. K., Zhu, Y. J., Futreal, P. A., Pang, B., Ruan, Y., Hillmer, A. M., Bertrand, D., Nagarajan, N., Rozen, S., Teh, B. T., and Tan, P. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nature genetics*. 2012, **44**: 570-574. doi: 10.1038/ng.2246
90. Robinson, D., Van Allen, E. M., Wu, Y. M., Schultz, N., Lonigro, R. J., Mosquera, J. M., Montgomery, B., Taplin, M. E., Pritchard, C. C., Attard, G., Beltran, H., Abida, W., Bradley, R. K., Vinson, J., Cao, X., Vats, P., Kunju, L. P., Hussain, M., Feng, F. Y., Tomlins, S. A., Cooney, K. A., Smith, D. C., Brennan, C., Siddiqui, J., Mehra, R., Chen, Y., Rathkopf, D. E., Morris, M. J., Solomon, S. B., Durack, J. C., Reuter, V. E., Gopalan, A., Gao, J., Loda, M., Lis, R. T., Bowden, M., Balk, S. P., Gaviola, G., Sougnez, C., Gupta, M., Yu, E. Y., Mostaghel, E. A., Cheng, H. H., Mulcahy, H., True, L. D., Plymate, S. R., Dvinge, H., Ferraldeschi, R., Flohr, P., Miranda, S., Zafeiriou, Z., Tunariu, N., Mateo, J., Perez-Lopez, R., Demichelis, F., Robinson, B. D., Schiffman, M., Nanus, D. M., Tagawa, S. T., Sigaras, A., Eng, K. W., Elemento, O., Sboner, A., Heath, E. I., Scher, H. I., Pienta, K. J., Kantoff, P., de Bono, J. S., Rubin, M. A., Nelson, P. S., Garraway, L. A., Sawyers, C. L., and Chinnaiyan, A. M. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015, **161**: 1215-1228. doi: 10.1016/j.cell.2015.05.001
91. Barbieri, C. E., Baca, S. C., Lawrence, M. S., Demichelis, F., Blattner, M., Theurillat, J. P., White, T. A., Stojanov, P., Van Allen, E., Stransky, N., Nickerson, E., Chae, S. S., Boysen, G., Auclair, D., Onofrio, R. C., Park, K., Kitabayashi, N., MacDonald, T. Y., Sheikh, K., Vuong, T., Guiducci, C., Cibulskis, K., Sivachenko, A., Carter, S. L., Saksena, G., Voet, D., Hussain, W. M., Ramos, A. H., Winckler, W., Redman, M. C., Ardlie, K., Tewari, A. K., Mosquera, J. M., Rupp, N., Wild, P. J., Moch, H., Morrissey, C., Nelson, P. S., Kantoff, P. W., Gabriel, S. B., Golub, T. R., Meyerson, M., Lander, E. S., Getz, G., Rubin, M. A., and Garraway, L. A. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature genetics*. 2012, **44**: 685-689. doi: 10.1038/ng.2279
92. Grasso, C. S., Wu, Y. M., Robinson, D. R., Cao, X., Dhanasekaran, S. M., Khan, A. P., Quist, M. J., Jing, X., Lonigro, R. J., Brenner, J. C., Asangani, I. A., Ateeq, B., Chun, S. Y., Siddiqui, J., Sam, L., Anstett, M., Mehra, R., Prensner, J. R., Palanisamy, N., Ryslik, G. A., Vandin, F., Raphael, B. J., Kunju, L. P., Rhodes, D. R., Pienta, K. J., Chinnaiyan, A. M., and Tomlins, S. A. The mutational landscape of lethal castration-resistant prostate cancer. *Nature*. 2012, **487**: 239-243. doi: 10.1038/nature11125
93. Boutros, P. C., Fraser, M., Harding, N. J., de Borja, R., Trudel, D., Lalonde, E., Meng, A., Hennings-Yeomans, P. H., McPherson, A., Sabelnykova, V. Y., Zia, A., Fox, N. S., Livingstone, J., Shiah, Y. J., Wang, J., Beck, T. A., Have, C. L., Chong, T., Sam, M., Johns, J., Timms, L., Buchner, N., Wong, A., Watson, J. D., Simmons, T. T., P'ng, C., Zafarana, G., Nguyen, F., Luo, X., Chu, K. C., Prokopec, S. D., Sykes, J., Dal Pra, A., Berlin, A., Brown, A., Chan-Seng-Yue, M. A., Yousif, F., Denroche, R. E., Chong, L. C., Chen, G. M., Jung, E., Fung, C., Starmans, M. H., Chen, H., Govind, S. K., Hawley, J., D'Costa, A., Pintilie, M., Waggott, D., Hach, F., Lambin, P., Muthuswamy, L. B., Cooper, C., Eeles, R., Neal, D., Tetu, B., Sahinalp, C., Stein, L. D., Fleshner, N., Shah, S. P., Collins, C. C., Hudson, T. J., McPherson, J. D., van der Kwast, T., and Bristow, R. G. Spatial

genomic heterogeneity within localized, multifocal prostate cancer. *Nature genetics*. 2015. doi: 10.1038/ng.3315

94. Gosse, G., Celton, M., Lamontagne, V., Forest, A., and Wilhelm, B. T. Whole genome and transcriptome analysis of a novel AML cell line with a normal karyotype. *Leukemia research*. 2015, **39**: 709-718. doi: 10.1016/j.leukres.2015.03.017

95. Malinowska-Ozdowy, K., Frech, C., Schonegger, A., Eckert, C., Cazzaniga, G., Stanulla, M., Zur Stadt, U., Mecklenbrauker, A., Schuster, M., Kneidinger, D., von Stackelberg, A., Locatelli, F., Schrappe, M., Horstmann, M. A., Attarbaschi, A., Bock, C., Mann, G., Haas, O. A., and Panzer-Grumayer, R. KRAS and CREBBP mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia. *Leukemia*. 2015. doi: 10.1038/leu.2015.107

96. Sato, Y., Yoshizato, T., Shiraishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H., Nagata, Y., Yoshida, K., Kon, A., Suzuki, Y., Chiba, K., Tanaka, H., Niida, A., Fujimoto, A., Tsunoda, T., Morikawa, T., Maeda, D., Kume, H., Sugano, S., Fukayama, M., Aburatani, H., Sanada, M., Miyano, S., Homma, Y., and Ogawa, S. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nature genetics*. 2013, **45**: 860-867. doi: 10.1038/ng.2699

97. Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C. K., Stephens, P., Davies, H., Jones, D., Lin, M. L., Teague, J., Bignell, G., Butler, A., Cho, J., Dalgliesh, G. L., Galappaththige, D., Greenman, C., Hardy, C., Jia, M., Latimer, C., Lau, K. W., Marshall, J., McLaren, S., Menzies, A., Mudie, L., Stebbings, L., Largaespada, D. A., Wessels, L. F., Richard, S., Kahnoski, R. J., Anema, J., Tuveson, D. A., Perez-Mancera, P. A., Mustonen, V., Fischer, A., Adams, D. J., Rust, A., Chan-on, W., Subimerb, C., Dykema, K., Furge, K., Campbell, P. J., Teh, B. T., Stratton, M. R., and Futreal, P. A. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature*. 2011, **469**: 539-542. doi: 10.1038/nature09639

98. Charlton, J., Williams, R. D., Sebire, N. J., Popov, S., Vujanic, G., Chagtai, T., Alcaide-German, M., Morris, T., Butcher, L. M., Guilhamon, P., Beck, S., and Pritchard-Jones, K. Comparative methylome analysis identifies new tumour subtypes and biomarkers for transformation of nephrogenic rests into Wilms tumour. *Genome medicine*. 2015, **7**: 11. doi: 10.1186/s13073-015-0136-4

99. Shanmugam, V., Ramanathan, R. K., Lavender, N. A., Sinari, S., Chadha, M., Liang, W. S., Kurdoglu, A., Izatt, T., Christoforides, A., Benson, H., Phillips, L., Baker, A., Murray, C., Hostetter, G., Von Hoff, D. D., Craig, D. W., and Carpten, J. D. Whole genome sequencing reveals potential targets for therapy in patients with refractory KRAS mutated metastatic colorectal cancer. *BMC medical genomics*. 2014, **7**: 36. doi: 10.1186/1755-8794-7-36

100. Forbes, S. A., Bindal, N., Bamford, S., Cole, C., Kok, C. Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., Teague, J. W., Campbell, P. J., Stratton, M. R., and Futreal, P. A. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic acids research*. 2011, **39**: D945-950. doi: 10.1093/nar/gkq929

101. Buczkowicz, P., Hoeman, C., Rakopoulos, P., Pajovic, S., Letourneau, L., Dzamba, M., Morrison, A., Lewis, P., Bouffet, E., Bartels, U., Zuccaro, J., Agnihotri, S., Ryall, S., Barszczyk, M., Chornenkyy, Y., Bourgey, M., Bourque, G., Montpetit, A., Cordero, F., Castelo-Branco, P., Mangerel, J., Tabori, U., Ho, K. C., Huang, A., Taylor, K. R., Mackay, A., Bendel, A. E., Nazarian, J., Fangusaro, J. R., Karajannis, M. A., Zagzag, D., Foreman, N. K., Donson, A., Hegert, J. V., Smith, A., Chan, J., Lafay-Cousin, L., Dunn, S., Hukin, J., Dunham, C., Scheinemann, K., Michaud, J., Zelcer, S., Ramsay, D., Cain, J., Brennan, C., Souweidane, M. M., Jones, C., Allis, C. D., Brudno, M., Becher, O., and Hawkins, C. Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations. *Nature genetics*. 2014, **46**: 451-456. doi: 10.1038/ng.2936

102. Taylor, K. R., Mackay, A., Truffaux, N., Butterfield, Y. S., Morozova, O., Philippe, C., Castel, D., Grasso, C. S., Vinci, M., Carvalho, D., Carcaboso, A. M., de Torres, C., Cruz, O., Mora, J., Entz-Werle, N., Ingram, W. J., Monje, M., Hargrave, D., Bullock, A. N., Puget, S., Yip, S., Jones, C., and Grill, J. Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma. *Nature genetics*. 2014, **46**: 457-461. doi: 10.1038/ng.2925

103. Zhang, J., Wu, G., Miller, C. P., Tatevossian, R. G., Dalton, J. D., Tang, B., Orisme, W., Punchihewa, C., Parker, M., Qaddoumi, I., Boop, F. A., Lu, C., Kandoth, C., Ding, L., Lee, R., Huether, R., Chen, X., Hedlund, E., Nagahawatte, P., Rusch, M., Boggs, K., Cheng, J., Becksfort, J., Ma, J., Song, G., Li, Y., Wei, L., Wang, J., Shurtleff, S., Easton, J., Zhao, D., Fulton, R. S., Fulton, L. L., Dooling, D. J., Vadodaria, B., Mulder, H. L., Tang, C., Ochoa, K., Mullighan, C. G., Gajjar, A., Kriwacki, R., Sheer, D., Gilbertson, R. J., Mardis, E. R., Wilson, R. K., Downing, J. R., Baker, S. J., Ellison, D. W., and St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome, P. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nature genetics*. 2013, **45**: 602-612. doi: 10.1038/ng.2611

104. Ramkissoon, S. H., Bi, W. L., Schumacher, S. E., Ramkissoon, L. A., Haidar, S., Knoff, D., Dubuc, A., Brown, L., Burns, M., Cryan, J. B., Abedalthagafi, M., Kang, Y. J., Schultz, N., Reardon, D. A., Lee, E. Q., Rinne, M. L., Norden, A. D., Nayak, L., Ruland, S., Doherty, L. M., LaFrankie, D. C., Horvath, M., Aizer, A. A., Russo, A., Arvold, N. D., Claus, E. B., Al-Mefty, O., Johnson, M. D., Golby, A. J., Dunn, I. F., Chiocca, E. A., Trippa, L., Santagata, S., Folkerth, R. D., Kantoff, P., Rollins, B. J., Lindeman, N. I., Wen, P. Y., Ligon, A. H., Beroukhim, R., Alexander, B. M., and Ligon, K. L. Clinical implementation of integrated whole-genome copy number and mutation profiling for glioblastoma. *Neuro-oncology*. 2015. doi: 10.1093/neuonc/nov015

105. Guichard, C., Amaddeo, G., Imbeaud, S., Ladeiro, Y., Pelletier, L., Maad, I. B., Calderaro, J., Bioulac-Sage, P., Letexier, M., Degos, F., Clement, B., Balabaud, C., Chevet, E., Laurent, A., Couchy, G., Letouze, E., Calvo, F., and Zucman-Rossi, J. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nature genetics*. 2012, **44**: 694-698. doi: 10.1038/ng.2256

106. Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K. A., Hosoda, F., Nguyen, H. H., 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., and Nakagawa, H. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature genetics*. 2012, **44**: 760-764. doi: 10.1038/ng.2291

107. Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J. P., Nickerson, E., Auclair, D., Li, L., Place, C., Dicara, D., Ramos, A. H., Lawrence, M. S., Cibulskis, K., Sivachenko, A., Voet, D., Saksena, G., Stransky, N., Onofrio, R. C., Winckler, W., Ardlie, K., Wagle, N., Wargo, J., Chong, K., Morton, D. L., Stemke-Hale, K., Chen, G., Noble, M., Meyerson, M., Ladbury, J. E., Davies, M. A., Gershenwald, J. E., Wagner, S. N., Hoon, D. S., Schadendorf, D., Lander, E. S., Gabriel, S. B., Getz, G., Garraway, L. A., and Chin, L. A landscape of driver mutations in melanoma. *Cell*. 2012, **150**: 251-263. doi: 10.1016/j.cell.2012.06.024

108. Biankin, A. V., Waddell, N., Kassahn, K. S., Gingras, M. C., Muthuswamy, L. B., Johns, A. L., Miller, D. K., Wilson, P. J., Patch, A. M., Wu, J., Chang, D. K., Cowley, M. J., Gardiner, B. B., Song, S., Harliwong, I., Idrisoglu, S., Nourse, C., Nourbakhsh, E., Manning, S., Wani, S., Gongora, M., Pajic, M., Scarlett, C. J., Gill, A. J., Pinho, A. V., Rooman, I., Anderson, M., Holmes, O., Leonard, C., Taylor, D., Wood, S., Xu, Q., Nones, K., Fink, J. L., Christ, A., Bruxner, T., Cloonan, N., Kolle, G., Newell, F., Pinese, M., Mead, R. S., Humphris, J. L., Kaplan, W., Jones, M. D., Colvin, E. K., Nagrial, A. M., Humphrey, E. S., Chou, A., Chin, V. T., Chantrill, L. A., Mawson, A., Samra, J. S., Kench, J. G., Lovell, J. A., Daly, R. J., Merrett, N. D., Toon, C., Epari, K., Nguyen, N. Q., Barbour, A., Zeps, N., Australian Pancreatic Cancer Genome, I., Kakkar, N., Zhao, F., Wu, Y. Q., Wang, M., Muzny, D. M., Fisher, W. E., Brunicardi, F. C., Hodges, S. E., Reid, J. G., Drummond, J., Chang, K., Han, Y., Lewis, L. R., Dinh, H., Buhay, C. J., Beck, T., Timms, L., Sam, M., Begley, K., Brown, A., Pai, D., Panchal, A., Buchner, N., De Borja, R., Denroche, R. E., Yung, C. K., Serra, S., Onetto, N., Mukhopadhyay, D., Tsao, M. S., Shaw, P. A., Petersen, G. M., Gallinger, S., Hruban, R. H., Maitra, A., Iacobuzio-Donahue, C. A., Schulick, R. D., Wolfgang, C. L., Morgan, R. A., Lawlor, R. T., Capelli, P., Corbo, V., Scardoni, M., Tortora, G., Tempero, M. A., Mann, K. M., Jenkins, N. A., Perez-Mancera, P. A., Adams, D. J., Largaespada, D. A., Wessels, L. F., Rust, A. G., Stein, L. D., Tuveson, D. A., Copeland, N. G., Musgrove, E. A., Scarpa, A., Eshleman, J. R., Hudson, T. J., Sutherland, R. L., Wheeler, D. A., Pearson, J. V., McPherson, J. D., Gibbs, R. A., and Grimmond, S. M. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*. 2012, **491**: 399-405. doi: 10.1038/nature11547