

Contents

Cancer Biology	4
Tumor Progression	4
Angiogenesis and Vascularization	5
Metastasis	5
Cancer Mechanisms	7
Gene Fusions	7
Whole-Genome Sequencing to Detect Gene Fusions	7
RNA-Seq to Detect Gene Fusions	9
Targeting Gene Fusions in Pre-Clinical Research	10
Understanding the Functional Impact of Gene Fusions with ChIP-Seq	11
Aneuploidy	11
Gene Expression	12
MicroRNA (miRNA)	12
Discovery of Functional miRNAs in Cancer	13
miRNAs for Cancer Detection and Staging	14
Epigenetics	15
Sequencing-Based Methylation Assays	15
Array-Based Methylation Assays	16
DNA-Protein Interactions (ChIP-Seq)	17
Transcription Factors and DNA-Protein Interactions	17
Chromatin Structure and Histone-DNA Interactions	18
Imprinting	19
Biomarker Discovery	19
Risk Markers	20
Prognostic Markers	21
Cancer Treatment	22
General Reviews of Cancer Genomics	24

Angiogenesis

Angiogenesis is the natural process of blood vessel growth that requires a careful balance between stimulation and inhibition. This process is co-opted during tumor development to generate a network of blood vessels that penetrates into cancerous growths to supply nutrients and oxygen. Blood vessels within tumors are usually formed by sprouting of resident tissue endothelial cells. This process is critical for tumor growth and metastasis.

Elevated AKR1C3 expression promotes prostate cancer cell survival and prostate cell-mediated endothelial cell tube formation: implications for prostate cancer progress

Dozmorov MG, Azzarello JT, Wren JD, Fung KM, Yang Q, et al. (2010) BMC Cancer 10: 672.

In localized and advanced prostate adenocarcinoma, the aldo-keto reductase (AKR) 1C family member 3 (AKR1C3), is usually up-regulated. This process is associated with prostate cancer (PCa) aggressiveness. Microarray analysis of a stably AKR1C3-transformed PC-3 prostate cancer cell line reveals that AKR1C3 overexpression promotes angiogenesis and growth of PC-3 cells. These results suggest that AKR1C3-mediated tumor angiogenesis is regulated by estrogen and androgen metabolism with subsequent IGF-1R and Akt activation followed by VEGF expression.

[Illumina technology: HumanWG-6 v2.0 Expression BeadChip \(discontinued product\)](#)

Additional References

- The miR-15/107 group of microRNA genes: Evolutionary biology, cellular functions, and roles in human diseases
Finnerty JR, Wang WX, Hebert SS, Wilfred BR, Mao G, et al. (2010) J Mol Biol 402: 491–509.
- Inhibition of neovascularization to simultaneously ameliorate graft-vs-host disease and decrease tumor growth
Penack O, Henke E, Suh D, King CG, Smith OM, et al. (2010) J Natl Cancer Inst 102: 894–908.

Metastasis

Metastasis is a complex process by which cancer cells break away from the primary tumor and circulate through the bloodstream or lymphatic system to other sites in the body. At the new sites, the cells continue to multiply and eventually form additional tumors. The ability of pancreatic cancer and uveal melanomas to metastasize contributes greatly to their lethality. Many fundamental questions remain about the clonal structures of metastatic tumors, phylogenetic relationships among metastases, the scale of ongoing parallel evolution in metastatic and primary sites, and how the tumor disseminates.

The patterns and dynamics of genomic instability in metastatic pancreatic cancer

Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, et al. (2010) Nature 467: 1109–13.

Pancreatic cancer has a distinctive pattern of genomic instability, dominated by breakage-fusion-bridge. This paper annotates genomic rearrangements in 13 patients with pancreatic cancer and explores the resulting clonal relationships.

[Illumina technology: TruSeq sequencing by synthesis \(SBS\) chemistry](#)

AAAGAATGATAACAGTAAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AATCAACGTACCGTAAACGAAACGTATCAATTGATCCACTGATTCAACGTACCGTAAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AACGACGAAAGAATGATAACAGTAAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGAGCTACCGTAAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCT
TTAACGTACCATTAAGAGCTACCGTAAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AAAGAATGATAACAGTAAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AAGATTACTTGATCCACTGATTCAACGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA

Cancer Mechanisms

The advent of massively parallel sequencing provides an unprecedented toolbox to untangle the causes and mechanisms of cancer.

Gene Fusions

Gene fusions are hallmarks of some cancer types, formed by the fusion of two previously separate genes. Fusions may lead to a gene product with a new or different function from the two fusion partners. The combination of a strong promoter with a functional gene (proto-oncogene) downstream is common in some cancers. The mechanisms of creating fusion genes are as varied as the functions of the resultant genes. There are several approaches to studying fusion genes through sequencing, such as whole-genome sequencing of the tumor, exome sequencing, and mRNA-Seq.

Whole-Genome Sequencing to Detect Gene Fusions

Sequencing the whole genome is a rigorous approach to finding all variants. Provided the coverage is deep enough, the investigator can be sure no mutation will go undetected and valuable samples will not have to be resequenced in the future. In the following examples, some of the gene fusions may have been missed with more targeted approaches such as exome sequencing or microarrays. Sequencing the whole genome allows the integration with ChIP-Seq data, significantly expanding the data interpretation.

The genomic complexity of primary human prostate cancer

Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, et al. (2011) Nature 470: 214–20.

This paper presents the complete sequence of seven primary human prostate cancers and their paired normal counterparts. Several tumors contained complex chains of balanced (copy-neutral) rearrangements that occurred within, or adjacent to, known cancer genes. Some of the breakpoints occurred in intergenic regions that may have been missed by exon-targeted approaches. In 88% of the cases, the fusion point could be mapped to base pair resolution. The most common type of fusion involved a precise join, with neither overlapping nor intervening sequence at the rearrangement junction. This result differed from the patterns seen in breast tumors, in which the most common junction involved a microhomology of 2–3 bp, indicating that the mechanisms responsible for generating these fusions are different for prostate and breast cancers.

Illumina technology: TruSeq sequencing by synthesis (SBS) chemistry

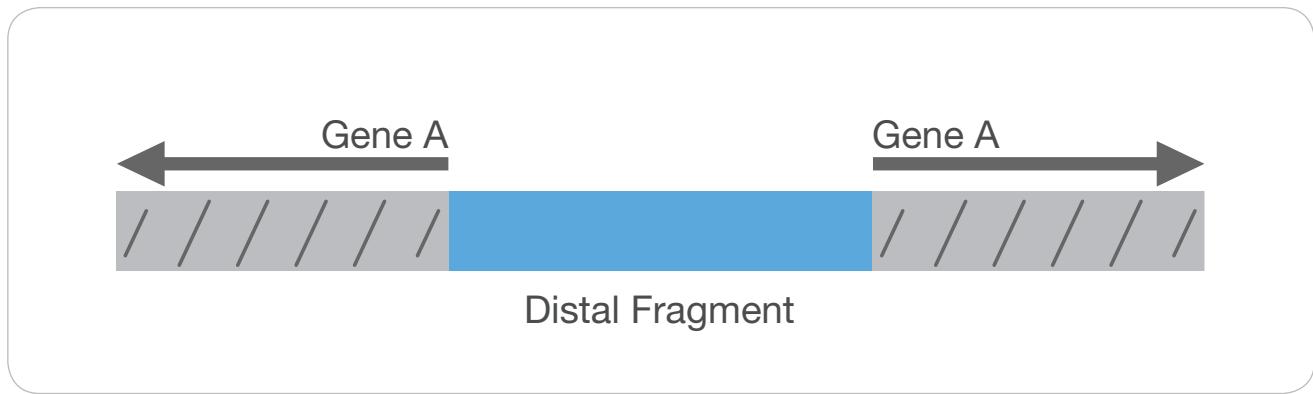


Figure 2. Genomic patterns of fold-back inversions. Some fold-back inversions captured fragments of templated genomic DNA between the two ends. These were often from distant regions of the genome, such as centromeric repeats or adjacent to other dsDNA breaks involved in somatic rearrangements. (Adapted from Campbell et al., (2010) Nature 467: 1109–13.)

Massive genomic rearrangement acquired in a single catastrophic event during cancer development

Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, et al. (2011) Cell 144: 27–40.

This paper describes a phenomenon, called chromothripsis, during which tens to hundreds of genomic rearrangements occur in a one-off cellular crisis. This model of a single catastrophic event during cellular duplications is different from the typical model of cancer progression through the progressive accumulation of mutations. The consequence of the catastrophic rearrangements is local regionalization of complex rearrangements and copy number variants with a limited range, 0–2, indicating that this is the result of a single event. In a cancer progression model where mutations accumulate, there is no upper limit to the copy numbers and it is common to see a wide range of copy numbers. The authors found evidence of chromothripsis in 2–3% of all cancers, across many subtypes, and in ~25% of bone cancers.

[Illumina technology: TruSeq sequencing by synthesis \(SBS\) chemistry](#)

Additional Reference

- Complex landscapes of somatic rearrangement in human breast cancer genomes
Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, et al. (2009) Nature 462: 1005–10.

RNA-Seq to Detect Gene Fusions

Using RNA-Seq to detect gene fusions offers distinct advantages. When compared to whole-genome sequencing, much less sequence needs to be generated. When compared to exome sequencing, it avoids the extra exome enrichment sample preparation step. Only known exomes that can be targeted with primers can be enriched, so *de novo* exomes that result from fusions may be difficult to interpret or missed. The fundamental compromise inherent in this approach is that only gene fusions expressed at the time of sampling can be detected. This is particularly important for cancers such as breast cancer that respond to hormonal stimulation. In studies such as drug target discovery where we assume that only expressed genes will have an impact on the outcome of the disease, this can be cost-effective approach to screening large numbers of patients.

AAAGAATGATAACAGTAACACACCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AATCAACGTACCGTAACGAACGTATCAATTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AACGACGAAAGAATGATAACAGTAACACACCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCT
TTAACGTACCTTAAGAGCTACCGTACGACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCT
AAAGAATGATAACAGTAACACACCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AAGATTACTTGATCCACTGATTCAACGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA
AACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCA

Epigenetics

Epigenetics refers to heritable changes in gene expression that are not accompanied by changes in DNA sequence. There is increasing evidence that aberrant epigenetic events contribute to diseases such as cancer. Epigenetic control is mediated through multiple processes, including DNA methylation, histone modification, and nucleosome remodeling.

Methylation assays can be divided into array- and sequencing-based assays. Sequencing-based analysis can provide highly detailed maps of methylation patterns, while array-based assays provide a cost-effective tool for screening large numbers of samples at lower resolution.

Additional References

- Next generation sequencing based approaches to epigenomics
Hirst M, and MA Marra, Briefings in Functional Genomics, 2010. 9(5-6): p. 455–465.
- Quantitative comparison of genome-wide DNA methylation mapping technologies
Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, et al. (2010) Nat Biotechnol 28(10): 1106–1114.
- Principles and challenges of genome-wide DNA methylation analysis
PW Laird (2010) Nat Rev Genet 11: 191–203.
- The power of NGS technologies to delineate the genome organization in cancer: from mutations to structural variations and epigenetic alterations
Schweiger MR, Kerick M, Timmermann B and M Isau M (2011) Cancer Metastasis Rev [ePub ahead of print].
- Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications
Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, et al. (2010) Nat Biotechnol 28(10): 1097–1105.
- Epigenomics of human colon cancer
Carmona F and M Esteller (2010) Mutat Res 693: 53–60.

Sequencing-Based Methylation Assays

The four most frequently used sequencing-based technologies are the bisulfite-based methods (BS-Seq), reduced representation bisulfite sequencing (RRBS), methylated DNA immunoprecipitation sequencing (MeDIP-Seq), and methylated DNA binding domain sequencing (MBD-Seq). BS-Seq is considered the ultimate methylation assay, covering 91.8% of the human methylation sites (CpGs) at single-base resolution. Under similar conditions, MeDIP-Seq covers 53% of the genome at 150-bp resolution.

Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing

Sun Z, Asmann YW, Kalari KR, Bot B, Eckel-Passow JE, et al. (2011) PLoS ONE 6: e17490.

Reduced representation bisulfite sequencing (RRBS) was used to profile the methylation status of 21,570 CpG islands in eight estrogen receptor positive (ER+) and negative (ER-) breast cancer cell lines. This data was correlated with changes in gene expression and gene copy number changes. Gene expression in cell lines was dominated by ER-associated genes. ER+ and ER- cell lines formed two distinct, stable clusters, and 1,873 genes were differentially expressed in the two groups. The authors identified 149 differentially expressed genes

Additional References

- To infinium, and beyond!
Wright KD and RJ Gilbertson (2010) *Cancer Cell* 17(5): 419–420.
- Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip
Thirlwell C, Eymard M, Feber A, Teschendorff A, Pearce K, et al. (2010) *Methods* 52: 248–54.
- Methylation profiling identifies 2 groups of gliomas according to their tumorigenesis
Laffaire J, Everhard S, Idbaih A, Criniere E, Marie Y, et al. (2010) *Neuro Oncol* 13: 84–98.
- Hepatocellular carcinoma displays distinct DNA methylation signatures with potential as clinical predictors
Hernandez-Vargas H, Lambert PM, Le Calvez-Kelm F, Gouysse G, McKay-Chopin S, et al. (2010) *PLoS ONE* 5(3): e9749.
- Global analysis of CpG methylation reveals epigenetic control of the radiosensitivity in lung cancer cell lines
Kim EH, Park AK, Dong SM, Ahn JH and WY Park (2010) *Oncogene* 29: 4725–4731.

DNA-Protein Interactions (ChIP-Seq)

ChIP-Seq allows researchers to survey DNA-protein binding sites over a complete genome at single-base pair resolution to survey transcription factor binding and chromatin structure. Numerous studies using the technique have been published, making ChIP-Seq a well-established, robust method.

Reviews

- Genomes in three dimensions
M Baker (2011) *Genomics: Nature* 470: 289-294.

Transcription Factors and DNA-Protein Interactions

Transcription factor binding to specific DNA target sequences is the fundamental basis of gene regulation networks. Recent studies show that transcription factors vary greatly in their number of genomic binding sites, and that binding events can significantly exceed the number of known or possible direct gene targets. Defining the relationship between transcription factor binding and target regulation across the entire genome is possible with advances in computing and information processing tools for reconstructing and predicting regulatory networks.

Identification of β -catenin binding regions in colon cancer cells using ChIP-Seq

Bottomly D, Kyler SL, McWeeney SK and GS Yochum (2010) *Nucleic Acids Res* 38: 5735–5745.

The majority of colorectal cancers exhibit mutations in the adenomatous polyposis coli (APC) gene. Cells with mutant APC genes contain elevated levels of the β -catenin transcription coactivator, leading to abnormal expression of genes controlled by β -catenin/T-cell factor 4 (TCF4) complexes. The authors identified 2,168 β -catenin binding regions in HCT116 human colon cancer cells containing core and extended TCF4 motifs and an AP-1 motif.

[Illumina technology: TruSeq sequencing by synthesis \(SBS\) chemistry](#)

Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma

Strub T, Giuliano S, Ye T, Bonet C, Keime C, et al. (2011) *Oncogene* [ePub ahead of print].

The basic helix-loop-helix microphthalmia transcription factor (MITF) is the master regulator determining the identity and properties of the melanocyte lineage. It is regarded as a lineage-specific 'oncogene' with a critical role in the pathogenesis of melanoma. By combining ChIP-Seq and RNA tag sequencing with (siRNA)-mediated MITF knockdown, the authors show that MITF directly regulates a set of genes required for DNA replication, repair, and mitosis.

Illumina technology: [TruSeq sequencing by synthesis \(SBS\) chemistry](#)

Reviews

- Genome-wide transcription factor binding: beyond direct target regulation
Macquarrie KL, Fong AP, Morse RH and SJ Tapscott (2011) *Trends in Genetics* 27: 141–148.

Additional References

- ChIP-Seq and functional analysis of the SOX2 gene in colorectal cancers
Fang X, Yu W, Li L, Shao J, Zhao N, et al. (2010) *OMICS* 14: 369–84.
- Genome-wide assessment of differential roles for p300 and CBP in transcription regulation
Ramos YF, Hestand MS, Verlaan M, Krabbendam E, Ariyurek Y, et al. (2010) *Nucleic Acids Res* 38: 5396–5408.

Chromatin Structure and Histone-DNA Interactions

The role of chromatin remodeling in cancer is complex and relatively poorly understood. In some cases, chromatin remodeling genes are implicated as risk factors, while in many cases it is assumed that the chromatin remodeling is a consequence of the rearrangements of the underlying DNA.

Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma

Varela I, Tarpey P, Raine K, Huang D, Ong CK, et al. (2011) *Nature* 469: 539–42.

The authors sequenced the protein coding exome in primary clear cell renal carcinoma (ccRCC) cases and identified the SWI/SNF chromatin remodeling complex gene PBRM1. PBRM1 truncating mutations were found in 41% (92/227) of cases.

Illumina technology: [TruSeq sequencing by synthesis \(SBS\) chemistry](#)

AGAAATGATAACAGTAAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAAACACACTTCTGTTAACCTT
TCAAAGTACCGTAAACGAACGTATCAATTGAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCG
AACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACG
AGAAATGATAACAGTAAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAAACACACTTCTGTTAACCTT
GATTACTTGATCCACTGATTCAACGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAAACACACTTCTGTTAACCTT
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAAACACACTTCTGTTAACCTT

Prognostic Markers

As cancer progresses, somatic mutations and genomic rearrangements accumulate and profoundly impact patient prognosis. Informative biomarkers are critical for measuring disease progress and response to treatment.

Initial genome sequencing and analysis of multiple myeloma

Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, et al. (2011) *Nature* 471: 467–72.

Whole-genome sequencing (WGS) was performed for 23 multiple myeloma patients and whole-exome sequencing (WES; assessing 164,687 exons) was performed for 16 patients, with one patient analyzed by both approaches. Each tumor was compared to its corresponding normal. The frequency of tumor-specific point mutations was 2.9 per million bases, corresponding to approximately 7,450 point mutations per sample across the genome, including an average of 35 amino acid-changing point mutations plus 21 chromosomal rearrangements that disrupted protein-coding regions. Nearly half the patients showed mutations in genes involved in RNA-processing, protein translation, and the unfolded protein response. The DIS3 (also called RRP44) gene harbored mutations in 4 out of 38 patients and the mutations were clustered within the RNB domain facing the enzyme's catalytic pocket. Based on the sequencing evidence, the authors genotyped an additional 161 multiple myeloma patients for the 12 most common BRAF mutations and found mutations in seven patients (4%). Patients with these mutations may benefit from treatment with BRAF inhibitors.

[Illumina technology: TruSeq sequencing by synthesis \(SBS\) chemistry](#)

Identification of novel SNPs by next-generation sequencing of the genomic region containing the APC gene in colorectal cancer patients in China

Cheng Y, Wang J, Shao J, Chen Q, Mo F, et al. (2010) *OMICS* 14: 315–25.

Through SNP discovery in selected genomic regions of key genes in 27 pairs of colorectal cancers and normal adjacent tissues, 69 novel SNPs in the 123-kb APC genomic region were identified. Eleven SNPs are located in the exonic region, including one novel SNP. Ten of the SNPs are synonymous, but were predicted to affect splicing by creating or removing exonic splicing enhancers or exonic splicing silencers. The authors also identified seven SNPs in the upstream region of the APC gene, three of which were unique to the cancer tissues.

[Illumina technology: TruSeq sequencing by synthesis \(SBS\) chemistry](#)

Genetic variation in glutathione metabolism and DNA repair genes predicts survival of small-cell lung cancer patients

Sun Z, Chen J, Aakre J, Marks R, Garces Y, et al. (2010) *Ann Oncol* 21: 2011–6.

In this study, 248 patients with primary small-cell lung cancer were genotyped for 419 tag SNPs in 49 genes in the glutathione and DNA repair pathways. After analysis at single-SNP, whole-gene, and haplotype levels, 21 SNPs in 11 genes were found to be significantly associated with patient survival. Whole-gene and haplotype analysis identified haplotype combinations and genomic locations underlying the observed SNP associations in 3 of the 11 genes.

[Illumina technology: GoldenGate Genotyping Assay](#)

