Published in final edited form as:

World J Surg. 2009 April ; 33(4): 615–629. doi:10.1007/s00268-008-9851-y.

Basic Principles and Technologies for Deciphering the Genetic Map of Cancer

Georgios Voidonikolas, MD¹, Stephanie S. Kreml, MD¹, Changyi Chen, MD, PhD¹, William E. Fisher, MD, FACS^{1,4}, F. Charles Brunicardi, MD, FACS¹, Richard A. Gibbs, PhD², and Marie-Claude Gingras, PhD^{1,2,*}

¹Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, Texas, USA

²Human Genome Sequencing Center; Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA

³The Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, Texas, USA

⁴The Elkins Pancreas Center, Baylor College of Medicine, Houston, Texas, USA

Abstract

The progress achieved in the field of genomics in recent years is leading medicine to adopt a personalized model in which the knowledge of individual DNA alterations will allow a targeted approach to cancer. Using pancreatic cancer as a model, we discuss herein the fundamentals that need to be considered for the high-throughput and global identification of mutations. These include patient related issues, sample collection, DNA isolation, gene selection, primer design, and sequencing techniques. We also describe the possible applications of the discovery of DNA changes to the approach of this disease and cite preliminary efforts where the knowledge has been translated into the clinical or preclinical setting.

Keywords

Pancreatic cancer; high-throughput sequencing; genotyping; somatic mutations

Introduction

Twenty-eight years after Frederick Sanger made his first attempts to sequence small pieces of DNA, the National Human Genome Research Institute (NHGRI) announced the completion of the Human Genome Project [1-3]. After 13 years of international collaborative effort at a cost of \$3 billion, an essentially complete human genome sequence was finished and the world was ushered into a “genomic” era. In 2007, less than 4 years later, the first genome sequence of a single individual, James Watson, was deciphered in only 2 months at a cost of less than \$2 million. Project “Jim” was initiated by a private company in collaboration with the Baylor College of Medicine-Human Genome Sequencing Center (BCM-HGSC) [4,5]. This project demonstrated the feasibility of characterizing virtually all of the small scale variation in a single individual using a next generation sequencing technology thereby advancing the “personalized” genomic medicine model one step closer to a reality.

*Corresponding author: BCM-Human Genome Sequencing Center One Baylor Plaza, Houston, Texas, 77030 Tel: 713-798-1286 FAX: 713-798-5741 mgingras@bcm.edu .

Cancer has long been recognized as fundamentally driven by genetic mutations [6,7]. Although opinions vary as to whether a comprehensive inventory of the genetic changes in tumor cells can lead to fundamental new insights that will further the diagnosis and treatment of this cancer, pilots for large-scale projects already have been initiated [8-10]. Sjöblom *et al* have shown that the somatic mutations playing a role in the multistep progression of carcinogenesis are far from completely identified, even in the most studied cancer types (breast and colon) [11]. The Tumor Sequencing Project Consortium (TSPC), formed between the NHGRI funded Genome Centers at BCM, Washington University, and the Broad Institute has initiated a pre-pilot project on non-small cell lung cancer (1,000 genes) to demonstrate the potential of a systemic approach to tumor genotyping by DNA sequencing [12]. The benefit of this large-scale project already has been proven by the identification of a novel candidate proto-oncogene [13]. The Cancer Genome Atlas pilot project (TCGA), launched by the National Cancer Institute (NCI) and NHGRI, is now attempting the identification of all genomic alterations significantly associated with cancer. This includes detecting genomic loss or amplification, mutations in coding regions, chromosomal rearrangements, aberrant methylations, and expression profiles. Three tumor types (glioblastoma multiforme, squamous cell carcinoma of the lung, and ovarian carcinoma) have been selected initially by TCGA, with the scope of expanding to all major cancer types [14].

Some success in cancer care has already been obtained with the targeting of specific genetic alterations. Among the first effective applications of targeted therapy was the use of imatinib to inhibit the tyrosine kinase activity of the Bcr-Abl fusion protein formed by the chromosomal t(9;22) translocation in chronic myeloid leukemia [15], and the use of trastuzumab, a recombinant monoclonal antibody against *HER2*, in women with metastatic breast cancer with *HER2* amplification and overexpression [16]. Other breakthroughs in personalized therapies include the treatment of gastrointestinal stromal cell tumors, in which mutations in *KIT* and *PDGFRA* were found to predict a response to imatinib [15,17]. Additionally, mutations in *EGFR* predicted a response to gefitinib and erlotinib in lung adenocarcinoma [18-20]. The fact that specific mutations in multiple loci are a major determinant of the response to targeted therapies suggests that DNA sequencing is likely to provide an effective diagnostic and therapeutic approach to cancer in the future.

This review focuses on the requisites to identify, validate, and confirm mutations, as well as the possible applications the discovery of new DNA changes can have for the characterization and treatment of the disease. Some technical information is provided for those who may be interested in initiating similar projects, and some examples of preliminary attempts to apply current knowledge in a clinical setting are given with the goal of attracting new clinical investigators to bridge the gap between genomics and its application to everyday patient care.

Identification of mutations

Ethical issues

Generating genomic information on a large scale requires strict criteria to protect patient's rights and confidentiality [21]. Institutional Review Board (IRB) oversight and informed consent are unambiguously required. Genome sequencing should be explained to potential participants in the informed consent process. Specific consent should be obtained for the future storage and use of collected samples. For most federally-funded studies, broad data sharing is required. (NHGRI Data sharing policies) Explanation of the anticipated scope of data sharing, along with the risks and benefits of broad data release, should be provided. Under some circumstances (e.g., if data sharing is not required to achieve the primary goals

of the project and data release may impede research participation), it may be appropriate to offer an opt-out provision for public and/or restricted data broadcast [21,22].

Studies involving extensive genome sequencing raise an additional concern about whether individualized research results should be shared with study participants. Convincing arguments have been made that if the research reveals validated data of known clinical relevance, it should be reported to participants [23-27]. The language used to explain return of research results should be carefully crafted to avoid potential legal liability.

Management of clinical information

The patient information collected should include demographics, exposure, family history, symptoms and physical findings at presentation, laboratory values, diagnostic imaging test results, details of the surgical treatment, histology from preoperative and operative specimens, pathologic staging data, details of chemotherapy and radiation treatment, response to treatment in terms of follow-up imaging, disease-free survival and overall survival, and quality of life survey data.

The data should be entered and stored in a password-protected, HIPAA-compliant database. To assure patient confidentiality, the specimen should be logged into the database and then assigned a new serial number for use in the laboratory. Limited access by clinicians and biostatisticians will allow later correlation with clinical data. The tubes containing the samples also should be bar coded to achieve automatic assignment, increase speed of processing, and protection from mixing the samples.

Defining the number of patient samples

The number of patient samples to be sequenced depends on the number of available samples and the size and statistical power of the intended study. We have noticed that the number of modification event detected increases when a larger panel of patients is used in the discovery process (11 patients versus > 40 patients) (Table 1). The issue of statistical power has been raised in reviews of Sjöblom's paper [11,28,29]. Much larger samples than the ones used in that paper are required to detect cancer genes, since in small sample sizes, some candidate genes are expected to display false-positive results. The validation and specificity processes should also be planned with a statistically valid number of patients.

Sample and DNA preparation

It is necessary to collect both tumor and matched normal tissue from each patient for the comparison of germline and tumor genotypes in order to verify that any detected mutations are of somatic origin. Matched normal tissue can consist of surrounding normal tissue and/or blood. Collecting the blood of the patient as the primary source of normal germline DNA is preferable. If the blood is not available, a sample of "normal" tissue taken from an area adjacent to the tumor can be used as an alternative. This choice is less attractive because a "normal" sample may not be entirely normal and may contain some invasive neoplastic cells or normal cells that share alterations with the cancer [30]. In such a case, the sequencing result might be identical between the tumor and "normal" tissue, representing an inherited germline mutation or polymorphism, rather than a somatic mutation.

Samples should be identified as primary tumor tissue or metastasis or abnormal non-cancerous tissue. Strict requirements of quality, quantity, purity, and avoidance of necrotic tissue should be met in order to create an ideal tissue bank. Tissue quality issues encountered by large projects like TCGA can assist in better planning for future projects [31,32]. The degree of homogeneity of the tumor sample to be sequenced is also important. First, it is known that the tumor itself is not homogenous, since it consists of subpopulations

of cells with different phenotypes and genotypes that determine different immunity, aggressiveness and metastatic behavior [33,34]. In addition, the contamination of tumor samples by surrounding tissue must be considered. This has led to the use of a tumor purity threshold of at least 80% as a criterion by TCGA and others [13,35,36] and to the development of Laser Microdissection Technique, especially if the cell population of interest is scant [37].

Fresh tissue is the best source of DNA. However, if the DNA cannot be extracted immediately, freezing remains the best way to preserve tissue. Standard operating procedures for the collection of fresh frozen tissue samples have been developed and used in the European Human Frozen tumor Tissue Bank (TuBaFrost) [38]. Formalin-fixed, paraffin-embedded (FFPE) tissue also can be used, but with limitations [39-41]. In such material DNA is scant, and cross linking and degradation result in sequencing failure. FFPE tissue constitutes a valuable source in terms of the number of available samples, diversity, and global patient information that has been collected over decades. When using such a source, the DNA has to be whole genome amplified and only short amplicons (~100-200 bases) can be sequenced. As a consequence, FFPE tissues can be used in the validation process, but are not preferred for mutation/SNP discovery. Cell lines are not an optimal DNA source since mutations associated with long term *in vitro* culture and unrelated to the *in vivo* development of cancer can be acquired over time.

In cases when the amount of the provided human specimen is limited and the extracted DNA is scarce, whole genome amplification (WGA) is required to overcome the problem. By this method, the original DNA sample is amplified in a specific way from nanogram concentrations to microgram, while the sequence representation of the template is conserved. Of the different methods studied, multiple displacement amplification (MDA) results in DNA products of high molecular weight (up to 12kb) and generates the least bias [42-44]. Original unamplified samples can then be saved to validate the identified mutations.

A process overview is illustrated in Fig. 1 and more detailed and technical information can be found in Table 2.

Candidate gene selection

There is a wide range of genes that can be chosen for a sequencing study: genes with mutations already known to be associated with the tumor by medical literature and data sources [46-49]; genes with mutations associated with familial genetic syndromes that increase the risk of cancer [50,51]; genes known to be differentially expressed in the tumor [52,53]; genes linked to several other cancer types and cellular pathways important in cancer (e.g., oncogenes, tumor suppressor genes, DNA repair genes, cyclins, kinases, phosphatases) [54]. CancerGenes is a valuable source of information for gene selection and prioritization in the elaboration of a gene list [55].

High-throughput DNA sequencing methods have been used on data sets of 200 to 1000 genes in 150-200 patients. In these large scale-up studies, considerable effort is expended in finalizing the gene list [12,14]. In an approach to exon sequencing, Sjöblom *et al* expanded the list to 13,023 genes by limiting the study to a single direction in only 11 patients per tumor type [11]. The power to detect rare mutations was therefore limited in this study, and yet 365 mutations in 236 genes were found. As the ability to comprehensively sequence tens of thousands of exons in a single experiment continually increases, the problem of gene selection diminishes.

This sequencing ability has been further improved by another novelty: the replacement of PCR by capture arrays (Fig. 1). Until now, regions of interest to be sequenced were

selectively sampled through a labor-intensive process whereby each fragment was individually amplified using PCR. This required the parallel design and execution of thousands of reactions. This will now change as NimbleGen (Madison, WI) developed seven custom arrays that contained 204,000 exons, allowing the entire exonic coding genome to be captured, enriched, and sequenced on the 454 or Genome Analyzer platform upon release (Fig. 2) [56,57]. Such enrichment will eliminate the need for large numbers of specific primers, thousands of specific PCR reactions on individual or pooled samples presently associated with the PCR-based large-scale sequencing approaches, thus considerably reducing cost and time expense. This will give superior results, allowing for the identification of polymorphisms/mutations on the whole exonic genome.

Primer design

Currently, when performing an exonic study, specific amplification by PCR of the selected gene exonic regions has to be performed prior to sequencing and specific primer pairs need to be designed (Fig. 1). A pipeline can be established that links several steps for the integrated automating design of primers (Table 3).

Sequencing

The DNA sequencing revolution started in 1977 with Sanger sequencing that soon became by far the most frequently used sequencing technology (Fig. 4) [1]. The Sanger sequencing method, a termination technology, is still widely used to perform individual amplicon sequencing (400-500 bases/read in a 384-well plate = ~150,000 bases/plate). However, the interest in deciphering the whole genome of organisms and large-scale DNA sequencing projects recently prompted the development of other technologies and platforms accommodating high-throughput sequencing.

Discovered in the late 1990s, pyrosequencing technology is based on the real-time monitoring, by bioluminescence (conversion of luciferase into oxyluciferin), of a 4-enzyme sequencing reaction [63]. This technology recently has been adapted to a high throughput setting by 454 Life Sciences in which whole genome or targeted segments can be analyzed in a single run [64]. The technology is based on emulsion PCR of individual DNA fragments captured on 28-micron beads, at a resolution of one DNA molecule per bead, resulting in a 107-fold amplification of the initial DNA copy per bead, followed by pyrosequencing-by-synthesis of each clonally amplified template in a fiber optic slide (Fig. 5). Presently, approximately 500,000 reactions can be performed in parallel in the new FLX system and about 125,000,000 bases are sequenced per run (with each read currently at 200-300 base pairs long, the expectation is that in late 2008, 500 base pair reads will become available). Whereas patient samples and amplicons are pooled making it less expensive and more rapid for discovery, the physical segregation of the DNA-carrying beads in an emulsion during the *in vitro* amplification (clonal amplification) results in the detection of specific mutations in low tumor content samples without the need for tumor cell enrichment by laborious methods. The extreme sensitivity of the 454 technology has been recently demonstrated in a study in which it enabled the detection of mutations in low tumor content samples for which conventional Sanger sequencing has failed to detect any of these mutations [65]. The call for somatic mutation and the percentage of patients carrying a mutation are then estimated during validation by genotyping the matching normal and tumor samples individually for each patient using a different technology.

In addition to the 454 technology, there are several new high-throughput sequencing methods that are being explored. Among these, the Solexa system (recently acquired by Illumina (San Diego, CA) and renamed Genome Analyzer) is a new, massively parallel sequencing platform in which millions of single molecules are covalently attached to a

planar surface and amplified *in situ* by a “bridge amplification” process [66-68]. Sequencing by synthesis is then carried out by adding a mixture of four fluorescently labeled reversible terminators and DNA polymerase to the template on the Solexa flow cell. The nucleotide sequences are determined by the fluorescent signals. After removing the fluorophore and reversing the blockage group and the terminator, the terminator-enzyme mix is added to start a new cycle. The whole process is then repeated until the end of the run (Fig. 6). One nucleotide sequence is read out for a given molecule at each cycle. One Solexa run, on average, generates about 35-45 million reads with read lengths of 36 bases. Between 1.2 and 1.5 billion bases are sequenced per run, about 10 fold higher than the 454 technology throughput. However, the error rate for a single read generated from the Solexa platform is about 1.5% for 36 cycles. As a result, the detection of mutations in samples with low tumor content might be hard to differentiate from the error rate background. The challenges of the technology also include a massive data storage and computational load to manage. Presently, due to the short read length, Solexa has better use in a long-range PCR-based approach (>3 kb) or a whole genome approach.

Other important characteristics associated with tumor development and progression are the variation in gene copy number due to heteroploidy and chromosomal loss. This genomic aspect cannot be detected by sequencing but with techniques such as CGH and SNP arrays [69-70]. The *SKAP2/SCAP2* gene was found to be amplified and associated with the development of pancreatic cancer using this technology [71].

What to look for

The data derived from sequencing can be aligned to identify mutations by parallel comparing of the tumor DNA sequence, the patient normal DNA, and the reference sequence deposited in GenBank (Table 4). This comparison can reveal single nucleotide polymorphisms (SNPs), germline, and somatic mutations. Modifications from the GenBank reference sequence in the patient samples (sequence identical in the normal and tumor samples) are germline mutation or SNP. Somatic mutations are modifications specific to the tumor and not found in the blood or other normal tissue of the patient.

It is now believed that polymorphic variations in the DNA sequence also can be related to population-attributable cancer heritability. SNP is defined as a genomic locus where two or more alternative bases occur with a frequency of at least 1% in a population. SNP accounts for more than 90% of the total variation in the human genome [72]. There are as many as 7 million common SNPs with a minor allele frequency of at least 5% [73,74]. SNPs in close chromosomal proximity can be inherited together on haplotype blocks due to underlining linkage disequilibrium (non random association of alleles from one generation to the next) [75]. SNPs indeed have been associated with predisposition to cancer, prognosis, and response and toxicity to chemotherapy or radiotherapy [76-78]. SNPs in xenobiotic metabolizers, hormone metabolizers, DNA repair genes, genes involved in angiogenesis and cell cycle are under scrutiny in several cancers [79].

Discovery of the germline mutations (mutations found in every cell of the individual) predisposing to hereditary cancer syndromes was the trademark of translational cancer research in 1990s. Genes such as *BRCA1/BRCA2* in breast and ovarian cancer, *APC* in colon cancer with adenomatous polyposis coli and *CDNK2A* in melanoma are examples of mutated genes with high penetrance leading to cancer and studied in family pedigrees [80-82].

Based on observation made by Loeb and others, at least six different metabolic or signaling pathways must be altered to lead to cancer [83-85]. By these alterations cells express insensitivity to growth inhibitory signals, escape apoptosis and acquire limitless replicative

potential and sustained angiogenic, invasive and metastatic abilities [86]. As normal mutation rates cannot by themselves account for the multiple mutations found in cancer cells, it is believed that special mutations called mutator mutations lead to genetic instability and increase the inherent rate of genetic change, thus exhibiting a “mutator phenotype” [85,87].

The whole list of cancer associated genes includes oncogenes, tumor suppressor genes and stability genes. Oncogenes can be abnormally activated by intragenic mutation, chromosomal translocation or gene amplification. Tumor suppressor genes can be inactivated by a missense mutation, nonsense mutation, deletion or insertion of various sizes; by epigenetic silencing; or by amplification of regulatory inhibitors. Finally, the function of stability genes including mismatch repair genes, nucleotide excision repair and base excision repair genes, as well as mitotic recombination and chromosomal segregation genes also can be impaired by a mutation [48].

These DNA modifications can be found in coding (exon) or non-coding (intron and untranslated exonic) regions. Exonic mutations in the coding region can be synonymous (silent) in which the change in base does not affect the amino acid call, or non synonymous resulting in a different amino acid (missense) or protein termination (stop codon, non sense). The non-synonymous mutations can have drastic effects on the protein function and structure. On the other hand, the impact of the silent polymorphisms in the *MDR1* and the Lamin A genes have proved that synonymous mutations should not be overlooked [88,89]. Mutations in non-coding regions located in introns, promoters, splice junctions or untranslated regions of the gene may also contribute to cancer by changing the regulation, exon splicing, mRNA stability, or conformation of the protein [90-93]. Table 5 illustrates the different types of modification events we found in pancreatic adenocarcinoma using Sanger and 454 technology.

Although many somatic mutations can be detected in tumors, it is essential to distinguish between driver mutations which actually contribute to cancer and passenger mutations which randomly happened but are not responsible for cancer [11,94]. In the case where the function of the genes is still not clearly known, the impact of the mutated gene on the development of cancer will have to be evaluated by functional studies.

Validation and confirmation of specificity

Each sequencing technology has its own limitations and pitfalls. After the completion of sequencing and the discovery of mutations, a second technique based on different principles should be performed in order to validate the obtained genotype. There are a number of genotyping methods including TaqMan SNP Genotyping Assay (Applied Biosystem), MIP (ParAllele), SNPstream (Beckman-Coulter), iPLEX Gold Assay (Sequenom), GoldenGate Assay (Illumina) and for genome wide studies GeneChip (Affymetrix) and Infinium (Illumina) [95-101]. Our strategy was to use the Sanger or TaqMan SNP Genotyping Assay to validate 454 pyrosequencing and low throughput pyrosequencing (Biotage) to validate Sanger results. For example, 454 high throughput pyrosequencing detected base modifications, but individual sample validation done by Sanger technique allowed the comparison between tumor and matched normal tissue and thus the differentiation between germline and somatic events (Table 6). We also used differential gel migration of amplicon to validate large deletions.

In order to confirm that the identified mutations are specific to the malignant tissue, other tissues should be genotyped. As an example, we intend to genotype pancreatitis, precancerous lesions, benign cystic tumors, and neuroendocrine tumors to discriminate

between mutations associated with pancreas cancer and non-cancerous or other cancerous conditions.

Present knowledge and applications

The fact that genomics led to deeper understanding and more targeted approach to CML and solid cancers like breast, gastrointestinal and lung cancer is indisputable. In the case of pancreatic cancer, translation of information from the DNA level to the clinical practice is still preliminary and experimental. Effective screening or diagnostic genomic tools have not been established. Targeted therapies based on mutations are just beginning to be tested.

A recurrent pattern of genetic changes associated with pancreatic carcinogenesis has been identified [102]. The genetic changes include the inactivation of *CDKN2A* (p16) and an activating point mutation of *K-RAS* in codon 12 as early events, as well as inactivation of the *SMAD4* (*DPC4*), *TP53*, and *BRCA2* genes as later events [103,104]. *K-RAS* mutation and *CDKN2A* inactivation are detected in over 90% of the tumors; however, *SMAD4* and *TP53* inactivation can only be found in 55%, and 50-75% of the adenocarcinoma, respectively [105].

Mutations and SNPs associated with pancreatic adenocarcinoma risk

While tobacco smoking is a high risk factor for the development of pancreatic adenocarcinoma, numerous associations have been found between smoking behavior and genetic polymorphism in genes responsible for nicotine metabolism [106]. Indirect association of *K-RAS* mutation and activation with occupational exposure to dyes, organic pigments and other agents has been suggested [107,108]. Moreover, a metanalysis evaluating folate intake and genetic polymorphism of 5,10 methylene tetra hydro folate reductase (*MTHFR*) found a *MTHFR* variant associated with an increased risk for pancreatic adenocarcinoma [109]. A polymorphism in glutathione S-transferase gene, affecting detoxification of carcinogens and anticancer agents in the human pancreas, was found to confer a protective effect against pancreatic adenocarcinoma [110].

Other gene polymorphisms that may determine the risk for pancreatic adenocarcinoma include Aurora-A and *CDKN2A* (*p16*) polymorphisms, which were associated with diagnosis of pancreatic adenocarcinoma at an early age [111]. Finally, polymorphisms in DNA repair genes *XPD* and *XRCC2* have been studied as genetic risk modifiers for smoking-related pancreatic adenocarcinoma [112,113], while an *XRCC1* polymorphism had a significant interaction with the *APE1* or *MGMT* polymorphism in modifying pancreatic adenocarcinoma risk [114].

A growing body of evidence suggests that some of the aggregation of pancreatic adenocarcinoma in families has a heritable genetic basis and that as many as 10% of pancreatic adenocarcinoma could be hereditary [51]. Several familial genetic syndromes already have been associated with an increased risk of pancreatic adenocarcinoma such as hereditary pancreatitis, the hereditary nonpolyposis colorectal cancer syndrome (HNPCC), breast cancer, ataxia-telangiectasia, the familial atypical multiple mole-melanoma syndrome (FAMM), and Peutz-Jeghers syndrome. Already, germline mutations in *BRCA2*, *CDKN2A*, *STK11*, *FANCC*, *PRSS1*, and palladin (*PALLD*) have been shown to predispose to pancreatic adenocarcinoma, although with incomplete penetrance [115-117].

Mutations used for Diagnostics of Pancreatic adenocarcinoma

Researchers already have used genetic markers separately or concomitantly to analyze cellular material from pancreatic juice and fine needle aspirates [118-122]. One of these studies has indicated that the analysis of *K-RAS* mutation and *TP53* and *SMAD4* inactivation

can complement traditional cytology and clarify the diagnosis of patients with atypical biopsy samples [119]. EUS-FNA biopsy specimens also are beginning to be examined for some of the mutations associated with pancreatic adenocarcinoma. In one study, analysis of *K-RAS* point mutations improved sensitivity from 44% to 82% [120]. Another similar study examined the utility of immunohistochemistry for *TP53* expression and found that the sensitivity of EUS-FNA was improved from 76% to 90% [121].

Unfortunately, *K-RAS* mutations are not specific to pancreatic adenocarcinoma and also have been detected in 25% of pancreatitis samples, nonneoplastic exocrine pancreatic lesions of smokers, and pancreatic intraepithelial neoplasia (PanIN) [122-128], which proves the need for discovering of a broad panel of mutations that would most likely increase diagnostic specificity.

Mutations and SNPs used as prognostic tools

Some studies have shown an association between patient survival and somatic mutations in *CDKN2A* (*p16*), *TP53*, *SMAD4* (*DPC4*), or germline mutations in *XRCC2*, *XRCC3*, *RecQ1*, *Rad54L*, *ATM*, and *POLB* [129-136]. On the other hand, gene copy number of the epidermal growth factor receptor (*EGFR*) did not have prognostic value in pancreatic adenocarcinoma [137].

Mutations and SNPs used as therapeutic tool

1. Classic chemotherapy—SNPs have been studied regarding the pharmacodynamics and pharmacokinetics of gemcitabine (GEM) in the treatment of lung and other cancers [138], but the knowledge is valuable for pancreatic adenocarcinoma as well. Variants in the promoter region of *ENT1*, the transporter that brings GEM into the cells influence gene expression and probably GEM chemosensitivity [139]. A haplotype of cytidine deaminase (*CDA*), responsible for the detoxification of GEM was found to lead to decreased clearance and a high incidence of neutropenia, while SNPs in the 5' regulatory region of the deoxycytidine kinase gene were found to predict GEM sensitivity [140].

The efficacy and toxicity of other drugs used in treating pancreatic adenocarcinoma like 5-FU and platinum also are affected by polymorphisms. Mutations in dihydropyrimidine dehydrogenase (*DPYD*) gene, and thymidylate synthetase promoter region, involved in the 5-FU catabolism and pharmaceutical effect respectively, may affect drug toxicity and patient survival [141,142]. Moreover, platinum therapy results on survival were found to be affected by polymorphisms in DNA repair genes like *XRCC1*, *ERCC1*, and *ERCC2* [143-145].

Finally, in a preclinical study, sensitivity to cross-linking (mitomycin C, cisplatin, chlorambucil, and melphalan) chemotherapeutic agents was affected by the presence of mutations in *BRCA2*/Fanconi anemia gene [146].

2. Targeted therapy—*K-RAS* mutations, which are found in 70-90% of pancreatic adenocarcinoma tissues, have been the target of several therapeutic approaches. K-ras must be farnesylated to be active. Although no successful clinical trials have been reported, inhibitors of the enzyme farnesyl-transferase have been developed [147,148]. An immunotherapy approach to *K-RAS* mutations has also been proposed. *K-RAS* mutations currently are being studied as a target for immunotherapy with the use of yeast vectors called Tarmogens (Targeted Molecular Immunogens) in a phase II clinical trial, in post-resection, pancreatic cancer patients [149]. One other key downstream target of the Ras family, the phosphoinositol 3-kinase (*PI3K*), may play a role in drug resistance. In a preclinical study, treatment with *PI3K* inhibitors enhanced apoptosis induced by GEM [150].

Another targeted therapeutic approach of pancreatic adenocarcinoma is the inhibition of *EGFR* by tyrosine kinase inhibitors like erlotinib. A phase III study for first-line treatment of advanced pancreatic adenocarcinoma showed that the addition of erlotinib to GEM offered some improvement to survival compared to GEM alone and led to FDA approval [151]. Recently, *EGFR* intron 1 polymorphism was found to influence postoperative patient survival and in vitro erlotinib response [152].

Conclusion

The potential impact of gene sequencing studies on cancer treatment is enormous. Results from large sequencing projects deposited in databases accessible to the scientific community will serve as a referral for mutations associated with cancer. As the different cancer sequencing projects progress, the importance of mutations in cancer development, progression, and metastasis in unsuspected genes will be uncovered. With these discoveries, the list of genes of interest will expand. It will then be possible to orient such knowledge toward epidemiology and familial genetic studies to determine the importance of these mutations in the propensity to develop cancer. Eventually, screening tests and early detection for high-risk relatives and population will follow [153,154]. Functional studies will evaluate the impact on the biologic function of the identified mutated genes. DNA changes will be evaluated as biologic markers for diagnosis, prognosis (disease-free survival after surgery and overall survival), and therapy (including response to or toxicity from chemotherapy or radiation).

Acknowledgments

The study was supported by a grant from the Effie and Wofford Cain Foundation. We would like to acknowledge Mrs. Katie Elsbury for editorial support, Mrs. Sally Hodges for her assistance with patient-related issues and all the people at the Human Genome Sequencing Center who made this work possible.

This work was presented at the Molecular Surgeon Symposium on Personalized Genomic Medicine and Surgery at the Baylor College of Medicine, Houston, Texas, USA, on April 12, 2008. The symposium was supported by a grant from the National Institutes of Health (R13 CA132572 to Changyi Chen).

References

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1997;74:5463–5467. [PubMed: 271968]
2. International Consortium completes Human Genome Project [National Human Genome Research Institute web site]. [April 14, 2003]. Available at: <http://www.genome.gov/11006929>
3. Guttacher AE, Collins FS. Welcome to the genomic era. *N Engl J Med* 2003;349:996–998. [PubMed: 12954750]
4. 454 Life Sciences and Baylor College of Medicine Complete Sequencing of DNA Project. [454 Life Sciences web site]. [May 31, 2007]. Available at <http://www.454.com/news-events/press-releases.asp?display=detail&id=68>
5. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008;452:872–876. [PubMed: 18421352]
6. Kopnin BP. Targets of oncogenes and tumor suppressors: key for understanding basic mechanisms of carcinogenesis. *Biochemistry (Mosc)* 2000;65:2–27. [PubMed: 10702637]
7. Croce CM. Oncogenes and cancer. *N Engl J Med* 2008;358:502–511. [PubMed: 18234754]
8. Elledge SJ, Hannon GJ. An open letter to cancer researchers. *Science* 2005;310:439–441. [PubMed: 16239460]
9. Heng HH. Cancer genome sequencing: the challenges ahead. *Bioessays* 2007;29:783–794. [PubMed: 17621658]
10. Varmus H. The new era in cancer research. *Science* 2006;312:1162–1165. [PubMed: 16728627]

11. Sjöblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006;314:268–274. [PubMed: 16959974]
12. Cancer Sequencing Projects (CSPs) [National Human Genome Research Institute web site]. Available at: <http://www.genome.gov/19517442>
13. Weir BA, Woo MS, Getz G, et al. Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007;450:893–898. [PubMed: 17982442]
14. Collins FS, Barker AD. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci Am* 2007;296:50–57. [PubMed: 17348159]
15. Druker BJ, Talpaz M, Resta DJ, et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* 2001;344:1031–1037. [PubMed: 11287972]
16. Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344:783–792. [PubMed: 11248153]
17. Heinrich MC, Corless CL, Demetri GD, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol* 2003;21:4342–4349. [PubMed: 14645423]
18. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–1500. [PubMed: 15118125]
19. Pao W, Miller VA, Politi KA, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *LoS Med* 2005;2:e73.
20. Kobayashi S, Boggon TJ, Dayaram T, et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2005;352:786–792. [PubMed: 15728811]
21. McGuire AL, Gibbs RA. Genetics. No longer de-identified. *Science* 2006;312:370–371. [PubMed: 16627725]
22. McGuire AL, Gibbs RA. Currents in contemporary ethics: meeting the growing demands of genetic research. *J Law Med Ethics* 2006;34:809–812. [PubMed: 17199822]
23. McGuire AL, Caulfield T, Cho MK. Research ethics and the challenge of whole genome sequencing. *Nat Rev Gen* 2007;9:152–156.
24. Renegar G, Webster CJ, Stuerzebecker S, et al. Returning genetic research results to individuals: points to consider. *Bioethics* 2006;20:24–36. [PubMed: 16680905]
25. Knoppers BM, Joly Y, Simard J, et al. The emergence of an ethical duty to disclose genetic research results: international perspectives. *Eur J Human Gen* 2006;14:1170–1178.
26. MacNeil S, Fernandez C. Informing research participants of research results: analysis of Canadian university based research ethics board policies. *J Med Ethics* 2006;32:49–54. [PubMed: 16373524]
27. Kohane IS, Mandi KD, Taylor PL, et al. Reestablishing the researcher-patient compact. *Science* 2007;316:836–837. [PubMed: 17495156]
28. Rubin AF, Green P. Comment on “The consensus coding sequences of human breast and colorectal cancers”. *Science* 2007;317:1500. [PubMed: 17872429]
29. Getz G, Höfling H, Mesirov JP, et al. Comment on “The consensus coding sequences of human breast and colorectal cancers”. *Science* 2007;317:1500. [PubMed: 17872428]
30. Mahadevan D, Von Hoff DD. Tumor-stroma interactions in pancreatic ductal adenocarcinoma. *Mol Cancer Ther* 2007;6:1186–1197. [PubMed: 17406031]
31. Check E. Cancer atlas maps out sample worries. *Nature* 2007;447:1036–1037. [PubMed: 17597721]
32. Compton C. Getting to personalized cancer medicine: taking out the garbage. *Cancer* 2007;110:1641–1643. [PubMed: 17763370]
33. Tu SM, Lin SH, Logothetis CJ. Stem-cell origin of metastasis and heterogeneity in solid tumours. *Lancet Oncol* 2002;3:508–513. [PubMed: 12147437]
34. Woodruff MF. Cellular heterogeneity in tumours. *Br J Cancer* 1983;47:589–594. [PubMed: 6342645]

35. Kim JH, Tuziak T, Hu L, et al. Alterations in transcription clusters underlie development of bladder cancer along papillary and nonpapillary pathways. *Lab Invest* 2005;85:532–549. [PubMed: 15778693]
36. The Cancer Genome Atlas Biospecimen Selection Process. The Cancer Genome Atlas website. Available at : http://cancergenome.nih.gov/components/hcbr_process.asp
37. Murray GI. An overview of laser microdissection technologies. *Acta Histochem* 2007;109:171–176. [PubMed: 17462720]
38. Mager SR, Oomen MH, Morente MM, et al. Standard operating procedure for the collection of fresh frozen tissue samples. *Eur J Cancer* 2007;43:828–834. [PubMed: 17329097]
39. Ruiz, MI Gallegos; Floor, K.; Rijmen, F., et al. EGFR and K-ras mutation analysis in non-small cell lung cancer: comparison of paraffin embedded versus frozen specimens. *Cell Oncol* 2007;29:257–264. [PubMed: 17452778]
40. Ferrer I, Armstrong J, Capellari S, et al. Effects of formalin fixation, paraffin embedding, and time of storage on DNA preservation in brain tissue: a BrainNet Europe study. *Brain Pathol* 2007;17:297–303. [PubMed: 17465988]
41. Andreassen CN, Sørensen FB, Overgaard J, et al. Optimisation and validation of methods to assess single nucleotide polymorphisms (SNPs) in archival histological material. *Radiother Oncol* 2004;72:351–356. [PubMed: 15450735]
42. Lizardi PM, Huang X, Zhu Z, et al. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* 1998;19:225–232. [PubMed: 9662393]
43. Dean FB, Nelson JR, Giesler TL, et al. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* 2001;11:1095–1099. [PubMed: 11381035]
44. Pinard R, de Winter A, Sarkis GJ, et al. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* 2006;7:216. [PubMed: 16928277]
45. Genomic DNA Preparation from RNeasy™ Preserved Tissues. [Ambion Web site]. 2008. Available at: www.ambion.com/techlib/misc/genomicDNA_rnaseasy.html
46. Catalogue of somatic mutations in cancer [Sanger Institute COSMIC Web site]. Available at <http://www.sanger.ac.uk/genetics/CGP/cosmic/>
47. Pancreatic Cancer Gene Database [PCGD Web site]. Available at <http://www.bioinformatics.org/pcgdb/>
48. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 2004;10:789–799. [PubMed: 15286780]
49. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. *Nat Rev Cancer* 2004;4:177–183. [PubMed: 14993899]
50. Brentnall TA. Management strategies for patients with hereditary pancreatic cancer. *Curr Treat Options Oncol* 2005;6:437–445. [PubMed: 16107246]
51. Berger, DH.; Fisher, WE. Inherited Pancreatic Cancer Syndromes. In: Evans, DB.; Pisters, PWT.; Abbruzzese, JL., editors. *M.D. Anderson Solid Tumor Oncology Series—Pancreatic Cancer*. Springer-Verlag Inc.; New York, NY: 2002. p. 73–82.
52. Logsdon CD, Simeone DM, Binkley C, et al. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. *Cancer Res* 2003;63:2649–2657. [PubMed: 12750293]
53. Buchholz M, Kestler HA, Bauer A, et al. Specialized DNA arrays for the differentiation of pancreatic tumors. *Clin Cancer Res* 2005;11:8048–8054. [PubMed: 16299235]
54. Higgins ME, Claremont M, Major JE, et al. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res* 2007;35:D721–d726. [PubMed: 17088289]
55. Cancer Genes Resequencing Resource. Available at : <http://cbio.mskcc.org/cancergenex/index.php>
56. Albert TJ, Molla MN, Muzny DM, et al. Direct selection of human genomic loci by microarray hybridization. *Nat Methods* 2007;4:903–905. [PubMed: 17934467]
57. Hodges E, Xuan Z, Balija V, et al. Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 2007;39:1522–1527. [PubMed: 17982454]

58. Repeat Masker [computer program] Institute for Systems Biology. 2003. Available at <http://www.repeatmasker.org/>.
59. Single Nucleotide Polymorphism database. [NCBI Website]. [January 17]. 2008 Available at: <http://www.ncbi.nlm.nih.gov/projects/SNP/>
60. Primer3. Howard Hughes Medical Institute. [Primer3Web site]. [February 7, 2007]. Available at: <http://frodo.wi.mit.edu/>.
61. UCSC In-Silico PCR [UCSC Genome Bioinformatics Web site]. Available at: <http://genome.ucsc.edu/cgi-bin/hgPcr?command=start>
62. Parameswaran P, Jalili R, Tao L, et al. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res* 2007;35:e130. [PubMed: 17932070]
63. Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. *Science* 1998;28:363–365. [PubMed: 9705713]
64. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;37:376–380. [PubMed: 16056220]
65. Thomas RK, Nickerson E, Simons JF, et al. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* 2006;12:852–855. [PubMed: 16799556]
66. Solexa Ltd. *Pharmacogenomics* 2004;5:433–438. [PubMed: 15165179]
67. Bennett ST, Barnes C, Cox A, et al. Toward the 1,000 dollars human genome. *Pharmacogenomics* 2005;6:373–382. [PubMed: 16004555]
68. Solexa sequencing Technology Illumina Website. Available at <http://www.illumina.com/pages.ilmn?ID=203>
69. Dutt A, Beroukhi R. Single nucleotide polymorphism array analysis of cancer. *Curr Opin Oncol* 2007;19:43–49. [PubMed: 17133111]
70. Calhoun ES, Hucl T, Gallmeier E, et al. Identifying allelic loss and homozygous deletions in pancreatic cancer without matched normals using high-density single-nucleotide polymorphism arrays. *Cancer Res* 2006;66:7920–7928. [PubMed: 16912165]
71. Harada T, Chelala C, Bhakta V, et al. Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. *Oncogene* 2008;27:1951–1960. [PubMed: 17952125]
72. Cargill M, Altshuler D, Ireland J, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999;22(3):231–8. 1999. [PubMed: 10391209] *Nat Genet* 23:373. Erratum in. [PubMed: 10545957]
73. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 2001;27:234–236. [PubMed: 11242096]
74. Hinds DA, Stuve LL, Nilsen GB, et al. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072–1079. [PubMed: 15718463]
75. Bernig T, Chanock SJ. Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer. *Expert Rev Mol Diagn* 2006;6:319–331. [PubMed: 16706736]
76. Abraham J, Earl HM, Pharoah PD, Caldas C. Pharmacogenetics of cancer chemotherapy. *Biochim Biophys Acta* 2006;1766:168–183. [PubMed: 17141416]
77. Bernig T, Chanock SJ. Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer. *Expert Rev Mol Diagn* 2006;6:319–331. [PubMed: 16706736]
78. West CM, Elliott RM, Burnet NG. The genomics revolution and radiotherapy. *Clin Oncol (R Coll Radiol)* 2007;19:470–480. [PubMed: 17419040]
79. Imanitov EN, Togo AV, Hanson KP. Searching for cancer-associated gene polymorphisms: promises and obstacles. *Cancer Lett* 2004;204:3–14. [PubMed: 14744529]
80. Bodmer WF, Bailey CJ, Bodmer J, et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* 1987;19(32):614–616. [PubMed: 3039373]
81. Cannon-Albright LA, Goldgar DE, Meyer LJ, et al. Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. *Science* 1992;258:1148–1152. [PubMed: 1439824]

82. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* 1994;265:2088–2090. [PubMed: 8091231]
83. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* 1954;8:1–12. [PubMed: 13172380]
84. Renan MJ. How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol Carcinog* 1993;7:139–146. [PubMed: 8489711]
85. Beckman RA, Loeb LA. Genetic instability in cancer: theory and experiment. *Semin Cancer Biol* 2005;15:423–435. [PubMed: 16043359]
86. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100:57–70. [PubMed: 10647931]
87. Venkatesan RN, Bielas JH, Loeb LA. Generation of mutator mutants during carcinogenesis. *DNA Repair* 2006;5:294–302. [PubMed: 16359931]
88. Kimchi-Sarfaty C, Oh JM, Kim IW, et al. A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007;315:525–528. [PubMed: 17185560]
89. McClintock D, Ratner D, Lokuge M, et al. The Mutant Form of Lamin A that Causes Hutchinson-Gilford Progeria Is a Biomarker of Cellular Aging in Human Skin. *PLoS ONE* 2007;2:e1269. [PubMed: 18060063]
90. Conne B, Stutz A, Vassalli JD. The 3′ untranslated region of messenger RNA: A molecular ‘hotspot’ for pathology? *Nat Med* 2000;6:637–641. [PubMed: 10835679]
91. Duan J, Wainwright MS, Comeron JM, et al. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet* 2003;12:205–216. [PubMed: 12554675]
92. Hoogendoorn B, Coleman SL, Guy CA, et al. Functional analysis of human promoter polymorphisms. *Hum Mol Genet* 2003;12:2249–2254. [PubMed: 12915441]
93. Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* 2004;5:389–396. [PubMed: 15168696]
94. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature* 2007;446:153–158. [PubMed: 17344846]
95. Livak KJ. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet Anal* 1999;14:143–149. [PubMed: 10084106]
96. Hardenbol P, Yu F, Belmont J, et al. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res* 2005;15:269–275. [PubMed: 15687290]
97. Bell PA, Chaturvedi S, Gelfand CA, et al. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques* 2002;74:76–77.
99. Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* 2003;68:69–78. [PubMed: 15338605]
100. Genechip Arrays. [Affymetrix Web site]. Available at : <http://www.affymetrix.com/products/arrays/index.affx>
101. Gunderson KL, Steemers FJ, Lee G, et al. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* 2005;37:549–554. [PubMed: 15838508]
102. van der Heijden, MS.; Kern, SE. (200%) Molecular genetic alterations in cancer-associated genes. In: Von Hoff, DD.; Evans, DB.; Hruban, RH., editors. *Pancreatic Cancer*. 1st Edition. Jones and Bartlett Publishers, Inc.; Sandbury, MA: p. 31-41.
103. Moskaluk CA, Hruban RH, Kern SE. p16 and K-ras gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. *Cancer Res* 1997;57:2140–2143. [PubMed: 9187111]
104. Hruban RH, Wilentz RE, Kern SE. Genetic progression in the pancreatic ducts. *Am J Pathol* 2000;156:1821–1825. [PubMed: 10854204]
105. Hruban, RH.; Yeo, CJ.; Kern, SE. Pancreatic cancer. In: Vogelstein, B.; Kinzler, KW., editors. *The Genetic Basis of Human Cancer*. 2nd Edition. McGraw-Hill; NY: 2002. p. 659-673.
106. MacLeod SL, Chowdhury P. The genetics of nicotine dependence: relationship to pancreatic cancer. *World J Gastroenterol* 2006;12:7433–7439. [PubMed: 17167830]

107. Porta M, Malats N, Jarid M, et al. Serum concentrations of organochlorine compounds and K-ras mutations in exocrine pancreatic cancer. PANKRAS II Study Group. *Lancet* 1999;354:2125–2129. [PubMed: 10609819]
108. Alguacil J, Porta M, Kauppinen T, et al. KRAS II Study Group. Occupational exposure to dyes, metals, polycyclic aromatic hydrocarbons and other agents and K-ras activation in human exocrine pancreatic cancer. *Int J Cancer* 2003;107:635–641. [PubMed: 14520703]
109. Larsson SC, Giovannucci E, Wolk A. Folate intake, MTHFR polymorphisms, and risk of esophageal, gastric, and pancreatic cancer: a meta-analysis. *Gastroenterology* 2006;131:1271–1283. [PubMed: 17030196]
110. Jiao L, Bondy ML, Hassan MM, et al. Glutathione S-transferase gene polymorphisms and risk and survival of pancreatic cancer. *Cancer* 2007;109:840–848. [PubMed: 17265526]
111. Chen J, Li D, Wei C, et al. Aurora-A and p16 polymorphisms contribute to an earlier age at diagnosis of pancreatic cancer in Caucasians. *Clin Cancer Res* 2007;13:3100–3104. [PubMed: 17505013]
112. Jiao L, Hassan MM, Bondy ML, et al. The XPD Asp312Asn and Lys751Gln polymorphisms, corresponding haplotype, and pancreatic cancer risk. *Cancer Lett* 2007;245:61–68. [PubMed: 16458430]
113. Jiao L, Hassan MM, Bondy ML, et al. XRCC2 and XRCC3 Gene Polymorphism and Risk of Pancreatic Cancer. *Am J Gastroenterol* 2007;116:2379–2386.
114. Jiao L, Bondy ML, Hassan MM, et al. Selected polymorphisms of DNA repair genes and risk of pancreatic cancer. *Cancer Detect Prev* 2006;30:284–291. [PubMed: 16844323]
115. Hruban, RH.; Yeo, CJ.; Kem, SE. Pancreatic cancer. In: Vogelstein, B.; Kinzler, KW., editors. *The Genetic Basis of Human Cancer*. 2nd Edition. McGraw-Hill; NY: 2002. p. 659–673.
116. Rogers CD, Couch FJ, Brune K, et al. Genetics of the FANCA gene in familial pancreatic cancer. *J Med Genet* 2004;41:e126. [PubMed: 15591268]
117. Pogue-Geile KL, Chen R, Bronner MP, Crnogorac-Jurcevic T, et al. Palladin mutation causes familial pancreatic cancer and suggests a new cancer mechanism. *PLoS Med* 2006;3:e516. [PubMed: 17194196]
118. Berthelemy P, Bouisson M, Escourrou J, et al. Identification of K-ras mutations in pancreatic juice in the early diagnosis of pancreatic cancer. *Ann Intern Med* 1995;123:188–191. [PubMed: 7598300]
119. van Heek T, Rader AE, Offerhaus JA, et al. K-ras, p53, and DPC4 (MAD4) alterations in fine-needle aspirates of the pancreas: a molecular panel correlates with and supplements cytologic diagnosis. *Am J Clin Pathol* 2002;117:755–765. [PubMed: 12090425]
120. Takahashi K, Yamao K, Okubo K, et al. Differential diagnosis of pancreatic cancer and focal pancreatitis by using EUS-guided FNA. *Gastrointest Endosc* 2005;61:76–79. [PubMed: 15672060]
121. Itoi T, Takei K, Sofuni A, et al. Immunohistochemical analysis of p53 and MIB-1 in tissue specimens obtained from endoscopic ultrasonography-guided fine needle aspiration biopsy for the diagnosis of solid pancreatic masses. *Oncol Rep* 2005;13:229–234. [PubMed: 15643503]
122. Lohr M, Muller P, Mora J, et al. p53 and K-ras mutations in pancreatic juice samples from patients with chronic pancreatitis. *Gastrointest Endosc* 2001;53:734–743. [PubMed: 11375580]
123. Pugliese V, Pujic N, Saccomanno S, et al. Pancreatic intraductal sampling during ERCP in patients with chronic pancreatitis and pancreatic cancer: cytologic studies and k-ras-2 codon 12 molecular analysis in 47 cases. *Gastrointest Endosc* 2001;54:595–599. [PubMed: 11677475]
124. Berger DH, Chang H, Wood M, et al. Mutational activation of K-ras in nonneoplastic exocrine pancreatic lesions in relation to cigarette smoking status. *Cancer* 1999;85:326–332. [PubMed: 10023699]
125. Caldas C, Hahn SA, da Costa LT, et al. Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat Genet* 1994;8:27–32. [PubMed: 7726912]
126. Kalthoff H, Schmiegell W, Roeder C, et al. p53 and K-RAS alterations in pancreatic epithelial cell lesions. *Oncogene* 2003;8:289–298. [PubMed: 8426738]

127. Moskaluk CA, Hruban RH, Kern SE. p16 and K-ras mutations in the intraductal precursors of human pancreatic adenocarcinoma. *Cancer Res* 1997;57:2140–2143. [PubMed: 9187111]
128. Tada M, Omata M, Kawai S, et al. Detection of ras gene mutations in pancreatic juice and peripheral blood of patients with pancreatic adenocarcinoma. *Cancer Res* 1993;53:2472–2474. [PubMed: 8495407]
129. Ohtsubo K, Watanabe H, Yamaguchi Y, et al. Abnormalities of tumor suppressor gene p16 in pancreatic carcinoma: immunohistochemical and genetic findings compared with clinicopathological parameters. *J Gastroenterol* 2003;38:663–671. [PubMed: 12898359]
130. Dong M, Ma G, Tu W, et al. Clinicopathological significance of p53 and mdm2 protein expression in human pancreatic cancer. *World J Gastroenterol* 2003;11:2162–2165. [PubMed: 15810085]
131. Yokoyama M, Yamanaka Y, Friess H, et al. p53 expression in human pancreatic cancer correlates with enhanced biological aggressiveness. *Anticancer Res* 1994;14:2477–2483. [PubMed: 7872670]
132. Ghaneh P, Greenhalf W, Humphreys M, et al. Adenovirus-mediated transfer of p53 and p16(INK4a) results in pancreatic cancer regression in vitro and in vivo. *Gene Ther* 2001;8:199–208. [PubMed: 11313791]
133. Tascilar M, Skinner HG, Rosty C, et al. The SMAD4 protein and prognosis of pancreatic ductal adenocarcinoma. *Clin Cancer Res* 2001;7:4115–4121. [PubMed: 11751510]
134. Li D, Li Y, Jiao L, et al. Effects of base excision repair gene polymorphisms on pancreatic cancer survival. *Int J Cancer* 2007;120:1748–1754. [PubMed: 17230526]
135. Li D, Liu H, Jiao L, et al. Significant effect of homologous recombination DNA repair gene polymorphisms on pancreatic cancer survival. *Cancer Res* 2006;66:3323–3330. [PubMed: 16540687]
136. Li D, Frazier M, Evans DB, et al. Single nucleotide polymorphisms of RecQ1, RAD54L, and ATM genes are associated with reduced survival of pancreatic cancer. *J Clin Oncol* 2006;24:1720–1728. [PubMed: 16520463]
137. Lee J, Jang KT, Ki CS, et al. Impact of epidermal growth factor receptor (EGFR) kinase mutations, EGFR gene amplifications, and KRAS mutations on survival of pancreatic adenocarcinoma. *Cancer* 2007;109:1561–1569. [PubMed: 17354229]
138. Ueno H, Kiyosawa K, Kaniwa N. Pharmacogenomics of gemcitabine: can genetic studies lead to tailor-made therapy? *Br J* 2007;97:145–151.
139. Myers SN, Goyal RK, Roy JD, et al. Functional single nucleotide polymorphism haplotypes in the human equilibrative nucleoside transporter 1. *Pharmacogenet Genomics* 2006;16:315–320. [PubMed: 16609362]
140. Kroep JR, Loves WJ, van der Wilt CL, et al. Pretreatment deoxycytidine kinase levels predict in vivo gemcitabine sensitivity. *Mol Cancer Ther* 2002;1:371–376. [PubMed: 12477049]
141. Gross E, Seck K, Neubauer S, et al. High-throughput genotyping by DHPLC of the dihydropyrimidine dehydrogenase gene implicated in (fluoro)pyrimidine catabolism. *Int J Oncol* 2003;22:325–332. [PubMed: 12527930]
142. Pullarkat ST, Stoecklacher J, Ghaderi V, et al. Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy. *Pharmacogenomics J* 2001;1:65–70. [PubMed: 11913730]
143. Stoecklacher J, Ghaderi V, Iobal S, et al. A polymorphism of the XRCC1 gene predicts for response to platinum based treatment in advanced colorectal cancer. *Anticancer Res* 2001;21:3075–3079. [PubMed: 11712813]
144. Ryu JS, Hong YC, Han HS, et al. Association between polymorphisms of ERCC1 and XPD and survival in non-small-cell lung cancer patients treated with cisplatin combination chemotherapy. *Lung Cancer* 2004;44:311–316. [PubMed: 15140544]
145. Park DJ, Stoecklacher J, Zhang W, et al. A Xeroderma pigmentosum group D gene polymorphism predicts clinical outcome to platinum-based chemotherapy in patients with advanced colorectal cancer. *Cancer Res* 2001;61:8654–8658. [PubMed: 11751380]

146. van der Heijden MS, Brody JR, Dezentje DA, et al. In vivo therapeutic responses contingent on Fanconi anemia/BRCA2 status of the tumor. *Clin Cancer Res* 2005;11:7508–7515. [PubMed: 16243825]
147. Sebt SM, Adjei AA. Farnesyltransferase inhibitors. *Semin Oncol* 2004;31:28–39. [PubMed: 14981578]
148. Van Cutsem E, van de Velde H, Karasek P, et al. Phase III trial of gemcitabine plus tipifarnib compared with gemcitabine plus placebo in advanced pancreatic cancer. *J Clin Oncol* 2004;22:1430–1438. [PubMed: 15084616]
149. GI-4000 for mutated-Ras mediated cancers.[GlobeImmune Web site]. Available at: <http://www.globeimmune.com/index.php>
150. Ng SSW, Tsao MS, Chow S, et al. Inhibition of phosphatidylinositol 3-kinase enhances gemcitabine-induced apoptosis in human pancreatic cancer cells. *Cancer Res* 2000;60:5451–5455. [PubMed: 11034087]
151. Moore MJ, Goldstein D, Hamm J, et al. National Cancer Institute of Canada Clinical Trials Group. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol* 2007;25:1960–1966. [PubMed: 17452677]
152. Tzeng CW, Frolov A, Frolova N, et al. Pancreatic cancer epidermal growth factor receptor (EGFR) intron 1 polymorphism influences postoperative patient survival and in vitro erlotinib response. *Ann Surg Oncol* 2007;14:2150–2158. [PubMed: 17453292]
153. Brand RE, Lerch MM, Rubinstein WS, et al. Advances in counselling and surveillance of patients at risk for pancreatic cancer. *Gut* 2007;56:1460–1469. [PubMed: 17872573]
154. Canto MI. Strategies for screening for pancreatic adenocarcinoma in high-risk patients. *Semin Oncol* 2007;34:295–302. [PubMed: 17674957]

Fig. 1.

Sample processing for DNA sequencing. After collection of samples, DNA is extracted and submitted to a whole genome amplification (WGA) step. In traditional exonic studies, gene selection, primer design and specific PCR reactions are required prior to sequencing. The capture of the exonic regions on NimbleGen capture arrays will allow entire exome sequencing in shorter time and at a reduced cost.

Fig. 2. NimbleGen capture arrays. The DNA is fragmented and exons of interest are hybridized to high density oligonucleotide microarrays. The DNA is then eluted from the microarray, universal primers are added. The DNA is amplified and fed into the sequencer. The eluted DNA represents a pool of DNA fragments that are highly enriched for the regions that were targeted by the microarray and set as regions of interest.

Fig. 3.

Primer design. As one can see in the Design Template, there is a padding region of 300 base pairs (bp) before and after the exon where the forward and reverse primers can be designed. The amplicon includes the target region which in turn consists of the region of interest (exon and a padding of 15bp important in the splicing mechanism) and a region of 5-35 bp that is technical requirement of the 454 or Sanger respectively to give accurate sequencing results. Each primer (P1 and 2) contains a specific and a universal region (black and red regions). The specific region attaches to the gene of interest. This primer region is used during the amplification process (PCR) and becomes part of the amplicon. The universal primer used during the sequencing process hybridizes to the universal region allowing the sequencing of any amplicon.

Fig. 4. Technology and achievements. (A) The development of sequencing technology. After the discovery of double helix in 1953, sequencing technology has rapidly developed. The Sanger dideoxy sequencing method set the standard and is still widely used, while modern technologies like 454 and Solexa are competing in cost, throughput, accuracy and completeness. (B) Achievements and milestones. For the past 25+ years, the size of the largest genome sequenced has doubled approximately every 18 months. Indisputably, the largest achievement was the deciphering of the human genome, published in 2001, while the fight against cancer is established by the Tumor sequencing Project and The Cancer

Fig. 5.

The 454 pyrosequencing technology. The DNA from each sample is isolated and quantified, and equal amount of each DNA sample is pooled. Depending on the desired coverage, different amount of samples and amplicons can be pooled. The amount of tumor samples and amplicons illustrated here are given as example. PCRs for each of the genes under study are performed on the pool of DNA samples. Different amplicons are pooled in equal amount and bound to the Capture Beads. The beads are recovered after emulsion-based clonal amplification and the second strand of the amplification products is melted away. The sequencing primer is annealed to the immobilized DNA templates and the beads and the enzymes are loaded to a PicoTiterPlate (1 beads/well). The nucleotides are sequentially flowed over the plate, one at a time, in a cyclical order (TACG) for a total of 42 dNTP flow cycles; a chemiluminescent signal is generated by the release of pyrophosphate (PP_i) with each nucleotide addition, captured by the camera, and processed to determine the base sequence and quality score.

Fig. 6. The Genome Analyzer platform. The DNA is randomly sheared and adaptors are added to each fragment. The linked fragments are enriched with PCR and hybridized to a flow cell. Cluster bridge amplification is then performed followed by sequencing-by-synthesis.

Fig. 7. K-RAS mutations in codon 12 and 13 illustrated in Sanger chromatogram and 454 pyrosequencing Amplicon Variant Analysis software. The different intensity of the superposed bases in the Sanger chromatograms of the different patients illustrates the difficulty for an automatic call with SNP detector. On the other hand, in the 454 pyrosequencing, the mutations are successfully called by AVA. For this reason, the 454 pyrosequencing is a great discovery technology. However, the % given is not representative of the % of patients with the mutations because of factors such as heterogenous sample, heteroploidy and chromosomal loss. Genotyping with another technology then needs to be performed in each sample as well as in the matched normal tissue to determine which mutation is somatic and define the percentage of patients with each mutation.

Table 1

Modification events in function of patient number. When the number of screened patients was increased the number of identified events per gene was substantially increased, irrespective of the technique used.

Sequencing method	Patient number	Gene number	Number of events	Event/gene
Sanger	11	58	304	5.2
	47	4	74	18.5
454	41	21	332	15.8

Table 2**Sample and DNA preparation**

The blood can be collected in any collection tubes normally used in the hospital setting. However, we have found that the PAXgene Blood DNA kit (PreAnalytiX) offers some advantages. The kit has two constituents that can be separately ordered: the blood collection tubes for the clinic and the DNA isolation kit for the laboratory. The PAXgene blood collection tubes contain a proprietary blend of reagents optimized for standardized blood collection, storage, and isolation of high-quality genomic DNA. These reagents stabilize the cellular blood constituents for up to 14 days at room temperature or 28 days at 4°C, allowing pick-up time flexibility. Moreover, we have found that the design of the covering cap minimizes blood spillage upon opening, an important concern for blood pathogen exposure.

In order to study the tumor, a punch biopsy should be taken from the specimen. The punch should be longitudinally dissected. The upper half should be submitted in formalin for paraffin embedding and be used as a mirror image to measure the percentage of viable tumor cells. The lower half should be used for the study. Tissue samples should then be immersed in preservative solutions. We have found that tissues preserved in RNAlater (Ambion) and protease inhibitor (Roche) solution generate high quality DNA, and also allow for parallel expression studies [45]. Tissue preserved in either solution must be washed several times in PBS before DNA isolation to remove any trace of the stabilizing solution that would otherwise interfere with the DNA extraction protocol.

DNA can be isolated from the tissues and WGA can be performed using different commercial kit. When using FFPE material, the REPLI-g FFPE kit (Qiagen, Valencia, CA), which combines DNA isolation, random ligation, and WGA has enabled us to sequence older samples that had previously failed sequencing.

Table 3

Primer design

At BCM-HGSC, the exonic regions of interest are first surrounded with intronic 200-300 base padding sequences in which the primers will be picked (Fig. 3). The DNA interspersed repeated sequences and SNPs in these padding regions are masked using RepeatMasker [58] and the dbSNP database [59] to prevent the design of primers over these bases. The primer set is then designed with Primer3 software [60] using certain basic criteria (primer length between 18-30 bases, melting temperature between 50-60° Celsius and primer pair with similar TM values, GC content ranging between 40% and 60%, 3' end containing a GC clamp but free of secondary structures, repetitive sequences, palindromes, and highly degenerate sequences, no formation of hairpins in the primer and self or cross dimerization with other primers in the reaction). The exclusive specificity of the primers is then checked with PCR software [61] during which, in-silico PCR is used to simulate the PCR process. If the desired sequence to amplify is large, the pipeline breaks apart the sequence and designs multiple primer sets (or amplicon tiles) to cover the large sequence. If the pipeline is unable to come up with some primer designs to amplify a sequence of interest, then manual primer design is done. The pipeline finally attaches forward and reverse universal sequencing primers to the specific primers. These universal regions will be used during the sequencing process to obtain forward and reverse sequence reads. A bar code of few nucleotides also can be added to the primer to segregate the individual sequences if the patient samples are pooled, such as for 454 sequencing [62].

The primers must then be tested on commercial or cell-line human DNA before using them on patient samples. The resulting test amplicons should be run on gel and sequenced to verify the PCR efficacy and specificity, respectively.

Table 4**Mutation call**

The call of the mutations in case of Sanger sequencing can be performed in an automated fashion with SNP Detector (created by Jinhui Zhang at the NCI) using the corresponding sequence deposited in GenBank as reference. Manual verification (visual study of the chromatograms) can be done in software such as Consed and Sequencher (Gene Codes Corp) (Fig. 7) Technologies such as 454 pyrosequencing and Solexa have developed their own analytical software such as Amplicon Variant Analysis (AVA) software (454 Life Sciences) (Fig. 7) Complementary informatic supports also are developed in the different institutions using these technologies for automated high throughput SNP calling and associated function analysis.

Modification events found in pancreatic adenocarcinoma using Sanger and 454 technology. Base shifts and base insertions or deletions (indels) were identified in both the intronic region and the exons of several analyzed genes. Exonic base shifts were either synonymous (syn) or non synonymous (non syn) or detected in the untranslated region (UTR).

Table 5

	intronic events		exonic events			Analyzed genes	TOTAL
	base shift	indel	UTR	syn.	nonsyn.	indel	
Germline	125	8	58	48	17	1	257
Sanger Somatic	52	3	24	11	29	2	121
Total	177	11	82	59	46	3	378
454	106		45	78	103	21	332

Example of validation of the modification events identified with 454 technology. In a total of 2,814 reads of a pooled 41 DNA sample, 454 pyrosequencing detected, among others, two base shift modifications in exon 2 and one in exon 6 of the K-RAS gene. Validation of these events with Sanger sequencing on individual tumor and matched normal tissue samples distinguished between somatic (exon 2) and germline (exon 6) modifications and differentiated homozygous and heterozygous patients.

Table 6

454 result					Sanger validation									
					41 pooled tumors			41 matched individual tumors			41 matched individual normals			
Exon	SNP loc hg18 (chr 12)	codon	Orig. codon	Base shifts	aa	# reads	% read with base shift	homo. major allele	hetero	homo. minor allele	% pt with som. mut.	homo. major allele	hetero	homo. minor allele
2	25289551	12	GGT	G → A	Gly → Asp	2814	2.70	31	10	0	24.4	41	0	0
	25289551	12	GGT	G → T	Gly → Val	2814	2.03	38	3	0	7.3	41	0	0
6	25254044	173	GAT	T → C	Asp	2814	1.14	29	11	1	2.4	29	11	1