Feasibility for Genechron to use of Next Generation Sequencing for Biomarker Identification, with a focus on Oncology

Author: Jonathan Dando PhD
Echino Ltd

Feasibility Study Report

August 2018

Table of contents

1. Introduction

Next-generation sequencing (NGS) has been revolutionary for the clinical diagnostics field; its high throughput sequencing power and decreasing cost, means it is being increasingly used in clinical labs. Instead of assessing genes one at a time by Sanger sequencing, a lab can test a group of genes at the same time. For example, many clinical labs now offer epilepsy gene panel tests that usually sequence 100-500 genes that are known to be causal or have associations with different kinds of epilepsies. Gene panel tests dramatically increases diagnostic efficiency and helps clinicians to focus in on the genetic cause for a certain disease easily and affordably.

NGS technology can also be used to diagnose patients who have been through diagnostic odysseys. Exome sequencing is currently used for this purpose in clinical labs. Exome test sequences an exome that contains all the protein coding regions which comprises 1.5% of the genome but contains 80% of recognized disease-causing mutations.

Numerous examples illustrated that the exome sequencing method can efficiently identify genetic causes for undiagnosed diseases, which not only helps clinicians to obtain accurate diagnoses but also guides clinicians in the personalized care and treatment of their patients.

As an approved (i.e. reimbursable) screening or diagnostic, the use of NGS is still at an early stage; while the establishment of a quality laboratory that can provide such services may seem attractive, there are significant hurdles that could prevent an active investment being made in such an endeavour.

However the question remains, is this commercially feasible and logical, which this study aims to clarify, as it would seem that Genechron would need to establish the platform and service from baseline. As it does not have a manufacturing facility this would also mean that by default the market for Genechron's entrance into this field would be locally and then nationally.

2. The Technology: Next Generation Sequencing (NGS)

NGS is now at its third generation (TGS) of technological development, with the fourth generation anticipated to be obtained by 2020.

Sanger and Maxam-Gilbert sequencing technologies were classified as the First Generation Sequencing Technology who initiated the field of DNA sequencing with their publication in 1977.

Sanger Sequencing is known as the chain termination method or the dideoxynucleotide method or the sequencing by synthesis method. It consists in using one strand of the double stranded DNA as template to be sequenced. This sequencing is made using chemically modified nucleotides called dideoxy-nucleotides (dNTPs). These dNTPs are marked for each DNA bases by ddG, ddA, ddT, and ddC. The dideoxynucleotides are used dNTPs are used for elongation of nucleotide, once incorporated into the DNA strand they prevent the further elongation and the elongation is complete. Then, we obtain DNA fragments ended by a dNTP with different sizes. The fragments are separated according to their size using gel slab where the resultant bands corresponding to DNA fragments can be visualized by an imaging system (X-ray or UV light). The sanger sequencing was widely used for three decades and even today for single or low-throughput DNA sequencing, however, it is difficult to further improve the speed of analysis that does not allow the sequencing of complex genomes such as the plant species genomes and the sequencing was still extremely expensive and time consuming.

The first generation of sequencing was dominant for three decades, however, the cost and time was a major stumbling block. In 2005 and in subsequent years, have marked the emergence of a new generation of sequencers to break the limitations of the first generation.

The basic characteristics of second generation sequencing technology are: (1) The generation of many millions of short reads in parallel, (2) The speed up of sequencing the process compared to the first generation, (3) The low cost of sequencing and (4) The sequencing output is directly detected without the need for electrophoresis.

Short read sequencing approaches divided under two wide approaches: sequencing by ligation (SBL) and sequencing by synthesis (SBS), and are mainly classified into three major sequencing platforms: Roche/454 launched in 2005, Illumina/Solexa in 2006 and in 2007 the ABI/SOLiD.

Roche/454 sequencing

Roche/454 sequencing appeared on the market in 2005, using pyrosequencing technique which is based on the detection of pyrophosphate released after each nucleotide incorporation in the new synthetic DNA strand (http://www.454.com). The pyrosequencing technique is a sequencing-by-synthesis approach.

DNA samples are randomly fragmented and each fragment is attached to a bead whose surface carries primers that have oligonucleotides complementary to the DNA fragments so each bead is associated with a single fragment. Then, each bead is isolated and amplified using PCR emulsion which produces about one million copies of each DNA fragment on the surface of the bead. The beads are then transferred to a plate containing many wells called picotiter plate (PTP) and the pyrosequencing technique is applied which consists in activating of a series of downstream reactions producing light at each incorporation of nucleotide. By detecting the light emission

after each incorporation of nucleotide, the sequence of the DNA fragment is deduced. The use of the picotiter plate allows hundreds of thousands of reactions occur in parallel, considerably increasing sequencing throughput. The latest instrument launched by Roche/454 called GS FLX+ that generates reads with lengths of up to 1000 bp and can produce ~1Million reads per run (454.com GS FLX+Systems http://454.com/products/gs-flxsystem/ index.asp). Other characteristics of Roche/454 instruments are listed in.

Ion torrent sequencing

Life Technologies commercialized the Ion Torrent semiconductor sequencing technology in 2010 (https://www.thermofisher.com/us/en/ home/brands/ion-torrent.html). It is similar to 454 pyrosequencing technology but it does not use fluorescent labeled nucleotides like other second-generation technologies. It is based on the detection of the hydrogen ion released during the sequencing process.
Specifically, Ion Torrent uses a chip that contains a set of micro wells and each has a bead with several identical fragments. The incorporation of each nucleotide with a fragment in the pearl, a hydrogen ion is released which change the pH of the solution. This change is detected by a sensor attached to the bottom of the micro well and converted into a voltage signal which is proportional to the number of nucleotides incorporated. The Ion Torrent sequencers are capable of producing reads lengths of 200 bp, 400 bp and 600 bp with throughput that can reach 10 Gb for ion proton sequencer. The major advantages of this sequencing technology are focused on read lengths which are longer to other SGS sequencers and fast sequencing time between 2 and 8 hours. The major disadvantage is the difficulty of interpreting the homopolymer sequences (more than 6 bp) which causes insertion and deletion (indel) error with a rate about ~1%.

Illumina/Solexa sequencing

The Solexa company has developed a new method of sequencing. Illumina company (http://www.illumina.com) purchased Solexa that started to commercialize the sequencer Ilumina/Solexa Genome Analyzer (GA). Illumina technology is sequencing by synthesis approach and is currently the most used technology in the NGS market.

During the first step, the DNA samples are randomly fragmented into sequences and adapters are ligated to both ends of each sequence. Then, these adapters are fixed themselves to the respective complementary adapters, the latter are hooked on a slide with many variants of adapters (complementary) placed on a solid plate. During the second step, each attached sequence to the solid plate is amplified by “PCR bridge amplification” that creates several identical copies of each sequence; a set of sequences made from the same original sequence is called a cluster. Each cluster contains approximately one million copies of the same original sequence. The last step is to determine each nucleotide in the sequences, Illumina uses the sequencing by synthesis approach that employs reversible terminators in which the four modified nucleotides, sequencing primers and DNA polymerases are added as a mix, and the primers are hybridized to the sequences. Then, polymerases are used to extend the primers using the modified nucleotides. Each type of nucleotide is labeled with a fluorescent specific in order for each type to be unique. The nucleotides have an inactive 3’-hydroxyl group which ensures that only one nucleotide is incorporated. Clusters are excited by laser for emitting a light signal specific to each nucleotide, which will be detected by a coupled-charge device (CCD) camera and Computer programs will translate these signals into a nucleotide sequence. The process continues with the elimination of the terminator with the fluorescent label and the starting of a new cycle with a new incorporation.

One of the main drawbacks of the Illumina/Solexa platform is the high requirement

for sample loading control because overloading can result in overlapping clusters and poor sequencing quality. The overall error rate of this sequencing technology is about 1%. Substitutions of nucleotides are the most common type of errors in this technology, the main source of error is due to the bad identification of the incorporated nucleotide.

ABI/SOLiD sequencing

Supported Oligonucleotide Ligation and Detection (SOLiD) is a NGS sequencer Marketed by Life Technologies (http:// www.lifetechnologies.com). In 2007, Applied Biosystems (ABI) has acquired SOLiD and developed ABI/SOLID sequencing technology that adopts by ligation (SBL) approach.

The ABI/SOLiD process consists of multiple sequencing rounds. It starts by attaching adapters to the DNA fragments, fixed on beads and cloned by PCR emulsion. These beads are then placed on a glass slide and the 8-mer with a fluorescent label at the end are sequentially ligated to DNA fragments, and the color emitted by the label is recorded. Then, the output format is color space which is the encoded form of the nucleotide where four fluorescent colors are used to represent 16 possible combinations of two bases. The sequencer repeats this ligation cycle and each cycle the complementary strand is removed and a new sequencing cycle starts at the position n-1 of the template. The cycle is repeated until each base is sequenced twice. The recovered data from the color space can be translated to letters of DNA bases and the sequence of the DNA fragment can be deduced.

The second-generation of sequencing technologies previously discussed have revolutionized the analysis of DNA and have been the most widely used compared to the first generation of sequencing technologies. However, the SGS technologies generally require PCR amplification step which is a long procedure in execution time and expansive in sequencing price. Also, it became clear that the genomes are very complex with many repetitive areas that SGS technologies are incapable to solve them and the relatively short reads made genome assembly more difficult. To remedy the problems caused by SGS technologies, scientists have developed a new generation of sequencing called “third generation sequencing”. These third generations of sequencing have the ability to offer a low sequencing cost and easy sample preparation without the need PCR amplification in an execution time significantly faster than SGS technologies. In addition, TGS are able to produce long reads exceeding several kilobases for the resolution of the assembly problem and repetitive regions of complex genomes.

There are two main approaches that characterize TGS: The single molecule real time sequencing approach (SMRT) that was developed by Quake laboratory and the synthetic approach that rely on existing short reads technologies used by Illumina (Moleculo) and 10xGenomics (https://www.10xgenomics.com) to construct long reads. The most widely used TGS technology approach is SMRT and the sequencers that have used this approach are Pacific Biosciences and Oxford Nanopore sequencing (specifically the MinION sequencer).

In the following, we present the two most widely used sequencing platforms in TGS; Pacific Biosciences and the MinION sequencing from Oxford Nanopore technology.

Pacific biosciences SMRT sequencing

Pacific Biosciences (http://www.pacificbiosciences.com/) developed the first genomic sequencer using SMRT approach and it's the most widely used third-generation sequencing technology.

Pacific Biosciences uses the same fluorescent labelling as the other technologies, but instead of executing cycles of amplification nucleotide, it detects the signals in real time, as they are emitted when the incorporations occur. It uses a structure composed of many SMRT cells, each cell contains microfabricated nanostructures called zeromode waveguides (ZMWs) which are wells of tens of nanometers in diameter microfabricated in a metal film which is in turn deposited onto a glass substrate. These ZMWs exploit the properties of light passing through openings with a diameter less than its wavelength, so light cannot be propagated. Due to their small diameter, the light intensity decreases along the wells and the bottom of the wells illuminated. Each ZMW contains a DNA polymerase attached to their bottom and the target DNA fragment for sequencing. During the sequencing reaction, the DNA fragment is incorporated by the DNA polymerase with fluorescent labeled nucleotides (with different colors). Whenever a nucleotide is incorporated, it releases a luminous signal that is recorded by sensors. The detection of the labeled nucleotides makes it possible to determine the DNA sequence.

Compared to SGS, Pacific Bioscience technology has several advantages. The preparation of the sample is very fast, it takes 4 to 6 hours instead of days. In addition, the long-read lengths, currently averaging ~10 kbp but individual very long reads can be as long as 60 kbp, which is longer than that of any SGS technology. Pacific Biosciences sequencing platforms have a high error rate of about 13% dominated by insertions and deletions errors. These errors are randomly distributed along the long read.

Oxford nanopore sequencing

The Oxford Nanopore sequencing (ONT) was developed as a technique to determine the order of nucleotides in a DNA sequence. In 2014, Oxford Nanopore Technologies released the MinION device that promises to generate longer reads that will ensure a better resolution structural genomic variants and repeat content. It's a mobile single-molecule Nanopore sequencing measures four inches in length and is connected by a USB 3.0 port of a laptop computer. This device has been released for testing by a community of users as part of the MinION Access Program (MAP) to examine the performance of the MinION sequencer.

In this sequencing technology, the first strand of a DNA molecule is linked by a hairpin to its complementary strand. The DNA fragment is passed through a protein nanopore (a nanopore is a nanoscale hole made of proteins or synthetic materials). When the DNA fragment is translated through the pore by the action of a motor protein attached to the pore, it generates a variation of an ionic current caused by differences in the moving nucleotides occupying the pore. This variation of ionic current is recorded progressively on a graphic model and then interpreted to identify the sequence. The sequencing is made on the direct strand generating the "template read" and then the hairpin structure is read followed by the inverse strand generating the "complement read", these reads is called "1D". If the "temple" and "complement" reads are combined, then we have a resulting consensus sequence called "two direction read" or "2D".

Among the advantages offered by this sequencer: first, it's low cost and small size.

Then, the sample is loaded into a port on the device and data is displayed on the screen and generated without having to wait till the run is complete. And, MinION can provide very long reads exceeding 150 kbp which can improve the contiguity of the denovo assembly. However, MinION produces a high error rate of ~12% distributed about ~3% mismatchs, ~4% insertions and ~5% deletions.

The ONT technology has continued to evolve. Recently, a new instrument has emerged called "PromethION"; it is the bigger brother of the MinION. It is an autonomous worktable sequencer with 48 individual flow cells each with 3000 pores (equivalent to 48 MinIONs) operating at 500 bp per second which is sufficiently powerful to achieve an ultra-high throughput needed for sequencing large genomes such as the human genome. Although the PromethION is not commercially available, the ONT announces that it is capable of producing ~2 to 4 Tb for a duration of 2 days and a length of reads which can attain 200 Kpb which puts this sequencer in competition with the PacBioRSII sequencer from pacific biosciences in terms of read length and HiSeq sequencer from Illumina in cost.

Similarities between different NGS Technologies

Next Generation Sequencing systems have been introduced in the past decade that allow for massively parallel sequencing reactions. These systems are capable of analyzing millions or even billions of sequencing reactions at the same time. Although different machines have been developed with various differing technical details, they all share some common features that are outlined below:

1. Sample Preparation:
All Next Generation Sequencing platforms require a library obtained either by amplification or ligation with custom adapter sequences. These adapter sequences allow for library hybridization to the sequencing chips and provide a universal priming site for sequencing primers. learn more about sample preparation from our Next Generation Sequencing - Experimental Design knowledge base.

2. Sequencing machines:
Each library fragment is amplified on a solid surface (either beads or a flat silicon derived surface) with covalently attached DNA linkers that hybridize the library adapters. This amplification creates clusters of DNA, each originating from a single library fragment; each cluster will act as an individual sequencing reaction.

The sequence of each cluster is optically read (either through the generation of light or fluorescent signal) from repeated cycles of nucleotide incorporation. Each machine has its own unique cycling condition; for example, the Illumina system uses repeated cycles of incorporation of reversibly fluorescent and terminated nucleotides followed by signal acquisition and removal of the fluorescent and terminator groups.

3. Data output:
Each machine provides the raw data at the end of the sequencing run. This raw data is a collection of DNA sequences that were generated at each cluster. This data could be further analysed to provide more meaningful results.

Differences between different NGS Technologies

The differences between the different Next Generation Sequencing platforms lie mainly in the technical details of the sequencing reaction. Below we describe these technical differences briefly. For a full explanation, please visit the manufacturers' webpages at the links provided in each section.

Pyrosequencing

In pyrosequencing, the sequencing reaction is monitored through the release of the pyrophosphate during nucleotide incorporation. A single nucleotide is added to the sequencing chip which will lead to its incorporation in a template dependent manner. This incorporation will result in the release of pyrophosphate which is used in a series of chemical reactions resulting in the generation of light. Light emission is detected by a camera which records the appropriate sequence of the cluster. Any unincorporated bases are degraded by apyrase before the addition of the next nucleotide. This cycle continues until the sequencing reaction is complete.

Disadvantages:
High reagent cost, and high error rate over strings of 6 or more single base nucleotides.

Sequencing by Synthesis

Sequencing by synthesis utilizes the step-by-step incorporation of reversibly fluorescent and terminated nucleotides for DNA sequencing and is used by the Illumina NGS platforms. The nucleotides used in this method have been modified in two ways: 1) each nucleotide is reversibly attached to a single fluorescent molecule with unique emission wavelengths, and 2) each nucleotide is also reversibly terminated ensuring that only a single nucleotide will be incorporated per cycle. All four nucleotides are added to the sequencing chip and after nucleotide incorporation the remaining DNA bases are washed away. The fluorescent signal is read at each cluster and recorded; both the fluorescent molecule and the terminator group are then cleaved and washed away. This process is repeated until the sequencing reaction is complete. This system is able to overcome the disadvantages of the pyrosequencing system by only incorporating a single nucleotide at a time.

Disadvantages:
As the sequencing reaction proceeds, the error rate of the machine also increases. This is due to incomplete removal of the fluorescent signal which leads to higher background noise levels.

Sequencing by Ligation
Sequencing by ligation is different from the other two methods since it does not utilize a DNA polymerase to incorporate nucleotides. Instead, it relies on short oligonucleotide probes that are ligated to one another. These oligonucleotides consist of 8 bases (from 3'-5'): two probe specific bases (there are a total of 16 8-mer probes which all differ at these two base positions) and six degenerate bases; one of four fluorescent dyes are attached at the 5' end of the probe. The sequencing reaction commences by binding of the primer to the adapter sequence and then hybridization of the appropriate probe. This hybridization of the probe is guided by the two probe specific bases and upon annealing, is ligated to the primer sequence through a DNA ligase. Unbound oligonucleotides are washed away, the signal is detected and recorded, the fluorescent signal is cleaved (the last 3 bases), and then

the next cycle commences. After approximately 7 cycles of ligation the DNA strand is denatured and another sequencing primer, offset by one base from the previous primer, is used to repeat these steps - in total 5 sequencing primers are used.

Disadvantages:
This method leads to very short sequencing reads.

Ion Semiconductor Sequencing

Ion semiconductor sequencing utilizes the release of hydrogen ions during the sequencing reaction to detect the sequence of a cluster. Each cluster is located directly above a semiconductor transistor which is capable of detecting changes in the pH of the solution. During nucleotide incorporation, a single H+ is released into the solution and it is detected by the semiconductor. The sequencing reaction itself proceeds similarly to pyrosequencing but at a fraction of the cost.

Disadvantages:
High error rate over homopolymeric stretches of nucleotides.

3. Screening vs Diagnosis

An essential question that the Genechron Executive will need to answer for all products is the kind of product that they would like to sell.

The company does not have any infrastructure or partnership enabling it to manufacture its own CE IVD kits, or Research Use Only (RUO) kits, which means it will always need to buy in the reagents and equipment necessary to perform any work. This prohibits an international dimension to any feasibility study and also means that the market in which they could potentially offer products and/or services will be predominantly local to central Italy and potentially the entire peninsula.

However the company will need to decide whether it wishes to offer diagnostics or screening services: which are very different and require very different infrastructure, staff, regulations and costs.

A **screening test** is done to detect potential health disorders or diseases in people who do not have any symptoms of disease. **Screening tests** are not considered diagnostic, but are used to identify a subset of the population who should have additional **testing** to determine the presence or absence of disease. The primary purpose of **screening tests** is to detect early disease or risk factors for disease **in** large numbers of apparently healthy individuals.

A **diagnostic test** is a procedure performed to confirm, or determine the presence of disease in an individual suspected of having the disease, usually following the report of symptoms, or based on the results of other medical **tests**. Utilizing nuclear medicine techniques to examine a patient having a lymphoma.

A screening, which can be offered to any clinical provider from primary through to quaternary healthcare can be performed within a 'research setting' in which the generated data informs the medical practitioner of further diagnoses to make; in contrast a diagnostic will need to be performed with kits and devices approved for diagnostic application.

	Screening tests	Diagnostic tests
Purpose	To detect potential disease indicators	To establish presence/absence of disease
Target population	Large numbers of asymptomatic, but potentially at risk individuals	Symptomatic individuals to establish diagnosis, or asymptomatic individuals with a positive screening test
Test method	Simple, acceptable to patients and staff	maybe invasive, expensive but justifiable as necessary to establish diagnosis
Positive result threshold	Generally chosen towards high sensitivity not to miss potential disease	Chosen towards high specificity (true negatives). More weight given to accuracy and precision than to patient acceptability
Positive result	Essentially indicates suspicion of disease (often used in combination with other risk factors) that warrants confirmation	Result provides a definite diagnosis
Cost	Cheap, benefits should justify the costs since large numbers of people will need to be screened to identify a small number of potential cases	Higher costs associated with diagnostic test maybe justified to establish diagnosis.

4. Risks of NGS approaches and potential opportunities

Within basic medical research, NGS has rapidly become a well-established technology. Its application in the study of cancer genetics has greatly improved our understanding of the underlying molecular mechanisms and differential responses to interventions. As our knowledge has increased, so also has our appreciation for the potential of NGS for multibiomarker testing in the clinical diagnostic context.

However, there remain many obstacles to the widespread adoption of NGS in routine diagnostic testing, including the inherent complexity of the NGS workflow and the need to handle large amounts of sequencing data.

Although NGS technology has transformed the clinical diagnostics field, it is still not easy for routine clinical molecular labs to adopt this technology. The hurdle is the interpretation of clinical NGS data, which usually has gigabytes or terabytes of data. Normally, a clinical exome sequencing test detects 20,000-30,000 variants in protein coding regions per patient, and identifying a disease-causing variant from this large number of variants poses a serious challenge for routine clinical molecular labs which usually only deal with one gene and several variants at a time. Without a bioinformatics team, which helps with the sequence alignment, variant calling, and variant filtration, it is impossible for the lab directors to make sense of the huge amount of NGS data, let alone to make clinical interpretations out of it. Commercial software is available; however, often there is an annual subscription fee and a fee for analysis per sample which a small clinical lab cannot afford.

In addition, even after bioinformatics data analysis and filtration, NGS data still needs manual interpretation of those identified genes and variants. Big academic labs have an NGS data review board which includes lab directors, physicians, and researchers in the molecular biology and genetics fields to manually review sequencing data and make a final clinical interpretation. This manual review process makes sure that the genetic variant found can explain the clinical presentations of the patient and inheritance pattern of the disease if it is known. Without such a quality review system, other medical practitioners maybe reluctant to consider the use of NGS.

The clinical labs that already offer clinical exome sequencing tests have shown the diagnostic yield of the clinical exome is just around 25%. The remaining 75% of clinical cases still could not be diagnosed, preventing appropriate treatment of these patients. The explanation of this low diagnostic yield could be the following:

• Disease-causing variants are located outside of protein coding regions. Not knowing much about these non-protein-coding regions prevents the interpretation of the functional impact of variations seen in them.

• Disease-causing genes or variants are novel or with few functional studies and associated clinical reports; therefore, it is difficult for clinical labs to perform clinical interpretations on those genes or variants. Often, exome-sequencing-identified variants that were not seen previously, the majority of which are missense mutations, and the pathogenicity of variants remains to be tested functionally.

• Multigenic cause contributes to the disease onset. Currently the clinical labs only analyze genomic data according to the simple Mendelian inheritance pattern. However, the disease etiology could be multigenic.

• For clinical testing, each patient's clinical features have to be considered for data analysis too. With more and more applications of whole genome sequencing and epigenomic sequencing, cutting-edge bioinformatic tools are also being created to integrate various genomic data so that clinical lab directors can process complete genomic profile of patients before final interpretation.

• Genetic cause is not enough for the disease onset; environmental insult also plays an important role in disease manifestation. Up to date, how the genetic and environment factors interact with each other to initiate disease onset is not clear yet; however, active research is being done for diseases like autism, allergy, and cancer.

With current trends suggesting that approximately one-third of us will face a cancer diagnosis at some point in our lifetimes, it is no surprise that many sequencing studies have focused on cancer.

Research has revealed the existence of remarkable genetic complexity within even a single malignant tumor, with each tumor including subpopulations of cells containing distinct genetic alterations. This mosaic of genetic alterations, known as tumor heterogeneity, has implications not only for cancer development and progression, but also for responses to therapy. It is becoming increasingly apparent that in order to have the best chance of successfully combating a cancer, we must first obtain insight into the heterogeneity of an individual tumor.

In many hospitals and clinics, single-biomarker testing has become a common practice for cancer diagnostics. Screening for mutations within the BRCA1, EGFR, or KRAS genes, for example, is routinely used to guide treatment decisions, but single-biomarker assays are unable to capture the full genetic complexity of a tumor.

A targeted NGS approach offers a means to obtain a complete overview of the heterogeneity of a specific cancer, and with it the potential for a personalized treatment approach.

However, as NGS transitions from basic research into applications providing actionable insights for cancer diagnostics, clinical laboratories face a set of challenges different from those experienced in the academic labs where high-throughput sequencing has grown up. Therefore, the transition from PCR-based single biomarker testing to implementing an NGS workflow is not simple.

To explore genomes as part of a search for new discoveries, an academic laboratory may perform lengthy whole-genome sequencing projects lasting several months. By contrast, a clinical laboratory may examine many new samples every day, each different and each requiring a different insight to guide a treatment decision. In this environment, the potential for speed and accuracy is an important attribute of any genomic system.

Differences in infrastructure also pose a challenge to easy adoption of NGS in routine diagnostics. An academic institution may have multiple centralized facilities, including instrumentation and specialized staff for daily operation, bioinformatics support, and centralized computational resources that can be leveraged for the analysis and storage of large NGS datasets. By contrast, a clinical diagnostics laboratory in a small clinic or hospital is unlikely to have such resources at its disposal, and many labs cite bioinformatics as a major bottleneck.

Further barriers standing in the way of an easy transition of the technology include

workflow complexity, ensuring assays of clinical relevance and actionability, and the lengthy process of achieving regulatory clearance for any given assay. These observations are supported by the findings of a survey of clinical laboratories conducted by Qiagen, in which the most common perceived challenge for NGS was identified as complexity of data analysis, followed closely by complexity of the workflow. Mirroring these observations were the findings that labs desire an easy-to-use workflow and a bioinformatics solution as part of their NGS system.

NGS is an inherently complex technology, or set of potentially fragmented technologies. While this can be said of many laboratory workflows, the low volume and precious nature of the starting sample for a clinical laboratory, combined with the importance of each workflow step for achieving a successful outcome, make NGS unique.

The fragmentation of NGS processes is largely the result of vendors specializing in individual sections of an NGS workflow (such as sample preparation, library preparation, sequencing, or bioinformatics), but being unable to offer integrated solutions for the complete process. A laboratory often has no option but to piece together the workflow using components from different vendors.

As a complete process, an NGS workflow typically takes multiple days from start to finish, with each step requiring multiple kits, steps, and operating manuals. Such a complex process offers many opportunities for human error during the handling of precious sample material. Combined with the formidable task of filtering the large and complex datasets generated for meaningful results, the complete process can be daunting for a lab new to NGS.

With many small- to medium-sized clinical labs lacking the resources for in-house NGS technology specialists, there is a gap in the market for a single vendor providing a cohesive, end-to-end system plus technical support.

Within the Lazio and central Italian Geography, this may represent an opportunity for Genechron, providing reimbursement can be obtained.

To benefit from NGS for cancer usage, Genechron should decide among a variety of technologies, and balance the scope of information (plus regulatory based quality constraints) that can be gained against the cost and time required to perform sequencing and the paying customer.

Options in NGS include

- Whole-genome sequencing for the broadest possible coverage of three billion base pairs;
- Whole-exome sequencing for analysis of the protein-coding portions (1% to 2%) of the genome;
- Gene panels targeting ‘hotspots’ of biomarkers and spanning up to 500 genes.

Academic researchers often choose to sequence whole genomes or exomes, because time is not a major constraint and scientists value the ability to identify a more comprehensive dataset of variants. In clinical applications, however, such large NGS datasets present a logistical problem in terms of data storage, and potential ethical issues related to data usage and security.

Evidence suggests that clinical laboratories are more likely to focus on gene panels

to target potentially relevant, actionable information from sequencing. Gene panels are quicker, more targeted, and less expensive than whole-genome or whole-exome sequencing for detection of biomarkers linked to the development of diseases.

In addition to saving time, targeted sequencing with gene panels enables a laboratory to run more samples per instrument cycle and allows more straightforward interpretation of data. Targeted sequencing also provides greater sensitivity for research into variant detection of rare mutations by distributing the sequencing reads in more tightly focused genomic regions.

For 'screening' several predesigned panels exist, such as the Qiagen QIAact Actionable Insight Tumor (AIT) panel. This panel interrogates 773 unique variant positions on 12 genes found in the most common cancers (ALK, BRAF, EGFR, ERBB2, ERBB3, ESR1, KIT, KRAS, NRAS, PDGFRA, PIK3CA and RAF1). The panel was designed to focus research on clinically relevant mutations included in approved therapeutic labels, professional practice guidelines, and active late-stage clinical trials.

Alternatively, if a more costly 'diagnostic' is required, CE-IVD kits do exist which work with several machines, but at present only one machine is CE-IVD marked, which are discussed later in this report.

NGS now offers the capacity to use single genomic markers to predict an individual's likelihood of developing cancer, and even to guide practitioners in making the best treatment decisions for their patient's specific disease.

Moving NGS technologies from basic research into routine diagnostic applications promises to provide even greater diagnostic and prognostic power. However there are still many hurdles that must be overcome to achieve widespread adoption of NGS into clinical practice.

An NGS diagnostic solution must be demonstrated to be capable of highly sensitive and accurate detection of mutations; issues relating to complexity of the technology and data analysis must be resolved; and the requirement for well-designed assays targeted at generating meaningful disease insights needs to be addressed.

With increasing adoption of genomic sequencing technology in clinical labs, the main challenge is to interpret the clinical genomic data accurately in a timely manner so that the lab can give back helpful clinical reports to physicians. The main obstacle for the interpretation of genomic data is due to the unreadiness of processing huge amounts of data, lack of knowledge of genetic variation in normal populations, insufficient clinical and research studies on important disease genes and variants, and unknown functionalities of non-protein-coding regions in the genome.

5. Who needs access to NGS?

Almost every large hospital (tertiary centre) has a molecular diagnostic and/or molecular biology laboratory or department, and ones that do not, either outsource such requirements to collaborators or small service providers.

A critical point is that in the case of 'on site' measurements, because hospitals are by default highly regulated environments, even if a molecular analysis department is not performing 'diagnostics' *per se*, the information they generate is rapidly translated into diagnostic process.

For any entity wishing to perform such molecular analysis as a service a decision on 'screening' vs 'diagnostic' needs to be made.

For all intents and purposes this would argue that a private provider of such a service should focus on 'screening' for Primary Care, following which the patients medical practitioner can then liaise with specialists in Tertiary care settings where further 'diagnostic' analysis is performed.

This is reinforced by recent regulations from the USA FDA: In March of this year, for the first time the FDA implied that each patient at an advanced cancer stage has the right to have her/his cancer genome deciphered at the highest possible level of complexity compatible with current knowledge and technology, linking molecular information to state-of-the-art systemic therapies, as they become available.

For the first time extended NGS testing has become **a reimbursable standard of care in oncology**. Additionally the closest thing to a ' US public healthcare system' the Centers for Medicare and Medicaid Services (CMS) determined that the use of NGS as a diagnostic laboratory test is "reasonable and necessary, and covered nationally, when performed in a CLIA-certified laboratory (certified to perform diagnostics), when ordered by a treating physician, and when all of the following requirements are met":

Patient has:

- Either recurrent, relapsed, refractory, metastatic, or advanced stages III or IV cancer.
- Either not been previously tested using the same NGS test for the same primary diagnosis of cancer or repeat testing using the same NGS test only when a new primary cancer diagnosis is made by the treating physician.
- Decided to seek further cancer treatment (eg, therapeutic chemotherapy).

The diagnostic laboratory test using NGS must have:

- FDA approval or clearance as a companion in vitro diagnostic.
- An FDA approved or cleared indication for use in that patient's cancer.
- Results provided to the treating physician for management of the patient using a report template to specify treatment options.

It is likely that the EMA will follow the FDA and approve such approaches in Europe, but the critical components are that the laboratory must be accredited, it is an approved diagnostic kit and it is for patients with late stage cancer.

Importantly, While the overwhelming majority of current NGS users and applications

remain more attached to basic medical research than to the responsibilities of providing care for individual patients, a growing number of studies have demonstrated that NGS offers great potential for improving diagnostic accuracy and guiding the selection of patient therapies.

In the next future, NGS profiling will likely be requested at progressively earlier stages, leading to a change in the engagement rules.

i.e. the entire tumour mutation test will have to be made available to the medical team as soon as possible after diagnosis and much before any specific therapy becomes applicable. This will give more time to anticipate the best and the worst case for a given patient, come up with a spectrum of therapeutic options, and define a sequence of treatments aimed at optimizing response.

As a result, the crucial turnpike between local and systemic cancer that usually defines the boundaries of intervention amongst surgeons, radiation therapists and medical oncologists will be blurred, multidisciplinary therapy plans being implemented early on during clinical course.

As a result, providing costs are kept down we can anticipate that such screening occurs via a primary care setting in managing the disease and the patient.

As an additional perspective regarding the application of NGS as a diagnostic within Europe, it is worthwhile to consider the guidelines published in 2015 by the European Society of Human Genetics (ESHG), a powerful group in the sector that will undoubtedly influence approval and reimbursement policy. The guidelines are included in their entirety in the section below, but the cautiousness and constant need for quality control, reflects the FDA position issued 3 years later. Given the specialists any 'statement' applied will result in labour and infrastructure needs, plus quality management, all of which will impact pricing.

ESHG Guidelines for diagnostic NGS

State of the art

The available NGS platforms are not stable yet in a sense that the technology and applications change constantly and rapidly. However, this should not prevent the implementation of NGS technology in diagnostics as NGS offers a potential overall benefit for the patient. The one thing that should prevent people from prematurely offering NGS diagnostics is poor quality. Insufficiently validated tests do present a threat to patients, and their use in a clinical diagnostic setting is unacceptable.

STATEMENT 01: NGS should not be transferred to clinical practice without an acceptable validation of the tests according to the emerging guidelines.

Whether the aim of a diagnostic test is to exclude or confirm a diagnosis has to be defined beforehand as the distinction is significant. The distinction mainly depends on the completeness of the test and warrants not only different settings but more importantly a different view on diagnostics.

STATEMENT 02: The laboratory has to make clear whether the test that is being offered may be used to exclude a diagnosis, or to confirm a diagnosis.

Diagnostic/clinical utility

The benefit of implementing NGS in diagnostics is the introduction of testing many genes at once in a relatively short time and at relatively low costs, and thereby yielding more molecular diagnoses.

The limitations of NGS are dependent on the platform and on the enrichment methods (if any) and have to be considered as they will influence the choice of enrichment method and sequencing platform and determine which additional tests (if any) will be necessary to deliver high-quality diagnostics.

STATEMENT 03: The aim and the utility of the test or assay should be discussed at the beginning of the validation and a summary should be included in the validation report.

The 'diagnostic yield' is defined as the chance that a disease-causing variant is identified and molecular diagnosis can be made. The value is calculated per patient cohort. It establishes the performance of NGS primarily from a clinical point of view and may be a good indicator of the efficiency of the test (beyond its analytical aspects) and of its clinical utility.

STATEMENT 04: When a laboratory is considering introducing NGS in diagnostics, it first has to consider the diagnostic yield.

In practice, diagnostic laboratories will preferably offer gene panels. The conditions for including a gene into a panel have to be defined when developing a diagnostic test. Ideally, this is an issue that should be dealt with at the community level, in a multidisciplinary way. The aim is to compile the list of genes that should be included in all diagnostic offers. This is important to harmonize genetic testing. It is definitely important from the standpoint of the patients and medical practitioners who would like to see equal access and uniform services across Europe.

STATEMENT 05: For diagnostic purpose, only genes with a known (ie, published and confirmed) relationship between the aberrant genotype and the pathology should be included in the analysis.

There is a strong opinion that for genes that are responsible for a significant proportion of the defects, referred as 'core genes', the sensitivity should not be compromised by the transition from Sanger to NGS. A strong issue is made about the BRCA1 and BRCA2 genes, where the sensitivity of Sanger sequencing plus deletion/duplication analysis reportedly reaches 99%. The reasoning equally applies to other genes with a high yield in diagnostics. Adding additional genes will of course increase the diagnostic yield, but this should not be at the expense of missing mutations that would previously have been detected. The incremental detection rate is thus the key determining factor in defining the core gene list and in dealing with the gaps.

STATEMENT 06: For the sake of comparison, to avoid irresponsible testing, for the benefit of the patients, 'core disease gene lists' should be established by the clinical and laboratory experts.

Laboratories will apply different (technical and diagnostic) settings for NGS tests, irrespective of guidelines. Indeed, there are too many variables still that cannot be fixed through prescriptive guidelines. Therefore, we propose a simple rating system for NGS diagnostics that will warrant fair scoring and easy comparison between what different labs are offering.

1. Type A test: The lab warrants >99% reliable reference or variant calls of the coding region and flanking intronic sequences, and fills all the gaps with Sanger sequencing (or another complementary sequencing analysis), and, depending on the platform used, performs extra analysis of, for example, the homopolymer stretches.

2. Type B test: The lab describes exactly which regions are sequenced at>99% reliable reference or variant calls, and fills some of the gaps with Sanger (or other) sequencing.

3. Type C test: The type C test solely relies on the quality of NGS sequencing, while no additional Sanger (or other) sequencing is offered.

STATEMENT 07: A simple rating system on the basis of coverage and diagnostic yield, should allow comparison of the diagnostic testing offer between laboratories.

Informed consent and information to the patient and clinician

The implications of a diagnostic test based on NGS depend on the procedures, platforms, filtering processes and data storage used in the laboratory. It is thus crucial that the referring physician is fully informed about the limitations and possible unfortunate effects of a genetic test.

STATEMENT 08: The laboratory has to provide for each NGS test the following: the diseases it targets, the name of the genes tested, their reportable range, the analytical sensitivity and specificity, and, if possible, the diseases not relevant to the clinical phenotype that could be caused by mutations in the tested genes.

The implications of a test based on NGS are mainly based on the chance of unsolicited and secondary findings. Although unsolicited findings are found in the genes linked to the tested disease, secondary findings are found in disease genes not implicated in the etiology of the tested disease.

STATEMENT 09: The analysis pipeline of diagnostic laboratories should focus on the gene panel under investigation in order to avoid the chance of secondary findings, and be validated accordingly.

The chance of unsolicited findings in a gene panel is very low and is mainly dependent on the genes involved. However, heterozygous mutations in recessive conditions might be detected, thereby detecting disease carriers. This will have consequences for counseling, reproductive choices, and so on.

STATEMENT 10: Laboratories should provide information on the chance of unsolicited findings.

Before implementing a NGS-based test, the clinical (genetic) center needs to set up an 'unsolicited and secondary findings protocol' that has to be in accordance with the decisions of an ethical committee. It should be decided – at the laboratory, institute or national level – whether patients are offered opt-in, opt-out options to get additional information besides the initial diagnostic result. The protocol should also specify whether unsolicited findings and carrier status are reported. The laboratory has to make sure that it can manage the different options that are offered.

STATEMENT 11: If a clinical center or a laboratory decides to offer patients an opt-in, opt-out protocol to get carrier status for unrelated diseases and secondary findings all the logistics need to be covered.

Also, pre-test genetic counseling is necessary and should include a discussion on both expected results and the potential for unsolicited and secondary findings. Adequate information should be provided.

STATEMENT 12: The local policy about dissemination of unsolicited and secondary findings should be clear for the patient.

STATEMENT 13: It is recommended to provide a written information leaflet or online available information for patients.

Validation

The quality of a sample is a combination of many parameters such as the amount of data produced, the proportion of PCR duplicates and the coverage. In diagnostic setting, only good-quality samples must be analyzed. It is thus essential to define the criteria to characterize high-quality targeted gene panels, exomes or genomes.

STATEMENT 14: All NGS quality metrics used in diagnostics procedures should be accurately described.

NGS technology requires the monitoring of run-specific and analysis/sample-specific features. Monitoring data do not have to be reported but should be used for continuous validation.

STATEMENT 15: The diagnostic laboratory has to implement a structured database for relevant quality measures for (i) the platform, (ii) all assays, and (iii) all samples processed.

A sample tracking method has to be used as NGS workflows are very complex and comprise multiple processing steps both in the lab and during the computational analysis.

STATEMENT 16: Aspects of sample tracking and the installation of barcoding to identify samples, should be dealt with during the evaluation of the assay, and included in the platform validation.

During platform validation, the laboratory has to make sure that all its devices and reagents satisfy the manufacturers' requirements. The limitations of each technology must be identified and taken into account during test development and data analysis. The laboratories may distinguish features (for validation) that belong to the platform, the specific test, or the analysis pipeline.

STATEMENT 17: Accuracy and precision should be part of the general platform validation, and the work does not have to be repeated for individual methods or tests.

Evidently every sequencing technology harbors its strengths and weaknesses. The bioinformatics tools must reflect these characteristics.

STATEMENT 18: The bioinformatics pipeline must be tailored for the technical platform used.

During pipeline validation the diagnostic specifications must be measured by assessing analytical sensitivity and specificity. For instance, algorithms that are optimized for SNP detection are less accurate for (small) insertions or deletions. The laboratory has to show that it is aware of such peculiarities and that the pipelines for variant detection are adequately tested.

STATEMENT 19: Analytical sensitivity and analytical specificity must be established separately for each type of variant during pipeline validation.

Any changes in chemistry, enrichment protocols, or the bioinformatics analysis platform will warrant re-validation.

STATEMENT 20: The diagnostic laboratory has to validate all parts of the bioinformatic pipeline (public domain tools or commercial software packages) with standard data sets whenever relevant changes (new releases) are implemented.

An in-house database containing all relevant variants provides an important tool in order to identify platform-specific artifacts, keep track of validation results, and provide an exchange proxy for locus-specific databases and meta-analyses. Typically, this database should allow for further annotations (eg, false-positives, published mutations, segregating variants, and so on), which greatly streamline the diagnostic process.

STATEMENT 21: The diagnostic laboratory has to implement/use a structured database for all relevant variants with current annotations.

Data storage should stick to the standard open file formats FASTQ, BAM, and VCF, which should also be used for data exchange with other laboratories. When storing the analysis results, full-log files have to be stored in addition to the analysis results. The log files should be as complete as possible, making the whole analysis from FASTQ data to the diagnostic report reproducible. Unfortunately, there is no (international) consensus yet on what should be stored. However, the storage has to be in line with national requirements and common sense.

STATEMENT 22: The diagnostic laboratory has to take steps for long-term storage of all relevant data sets.

Prior to launching any assay, the clinical target, that is, all coding regions plus the conserved splice sites, has to be defined. The clinical target depends on the diagnostic test and the defined gene panel.

STATEMENT 23: The reportable range, that is, the portion of the clinical target for which reliable calls can be generated, has to be defined during the test development and should be available to the clinician (either in the report or communicated digitally).

STATEMENT 24: The requirements for 'reportable range' depend on the aim of the assay.

For instance, an exome sequencing assay with the aim to achieve a high diagnostic yield does not require additional analysis to achieve high coverage in all genomic regions covered, but needs clear communication to the clinician that the test cannot be used to exclude a particular clinical diagnosis.

The performance of the diagnostic test must be evaluated in terms of accuracy, analytical sensitivity, analytical specificity, and precision. In principle, this is not new but is generally seen as cumbersome. However, the ISO norm is very strict about this.

STATEMENT 25: Whenever major changes are made to the test, quality parameters have to be checked, and samples have to be re-run. The laboratory should define beforehand what kind of samples and the number of cases that have to be assayed whenever the method is updated or upgraded.

Reporting

It is essential that NGS results are reported in a clear and consistent manner, as laboratory reports may be read by both experts and non-experts. From a practical standpoint, the clinically significant conclusions and the relevant test and test quality data should feature on the first page.

STATEMENT 26: The report of a NGS assay should summarize the patient's identification and diagnosis, a brief description of the test, a summary of results, and the major findings on one page.

Four examples of reports, with and without annexes, are included in the supplementary information to the guidelines.

All pathogenic (class 5) and likely pathogenic (class 4) variants have to be reported. Whether or not Unclassified Variants (UVs – class 3) are reported will depend on local practice. The latter has to be clear for the laboratory scientists, as well as for the referring clinicians.

STATEMENT 27: A local policy, in line with international recommendations, for reporting genomic variants should be established and documented by the laboratory prior to providing analysis of this type.

STATEMENT 28: Data on UVs have to be collected, with the aim to eventually classify these variants definitively.

A community activity is needed to collect and share the available information, with the aim to definitely classify the variants into pathogenic (class 5) or benign (class 1).

The policy that has been adopted by the laboratory or institute, with respect to unsolicited and secondary findings, has to be reflected in the laboratory practice and in the report.

STATEMENT 29: Laboratories should have a clearly defined protocol for addressing unsolicited and secondary findings prior to launching the test.

A diagnostic laboratory should not become overloaded with requests to analyze 'old' data in the view of new findings and progress in the fields. A diagnostic request is a contract at a certain point in time. A laboratory will only be able to offer what is known, and validated, at a given point in time.

STATEMENT 30: The laboratory is not expected to re-analyze old data systematically and report novel findings, not even when the core disease gene panel changes.

On the other hand, if at a particular moment, it is decided – by the lab or by the community of experts in the disease – to change a variant from one class to another, the lab is responsible for reanalyzing the available data, re-issuing a report on the basis of the novel evidence, and also re-contacting referring clinicians for the patients that are possibly affected by the new status of the variant. A system effectively linking patients and variants, and allowing for the retrieval of the affected cases when variants are re-classified is necessary in such a situation.

STATEMENT 31: To be able to manage disease variants, the laboratory has to set up a local variant database for the different diseases for which testing is offered on a clinical basis.

Distinction between research and diagnostics

With the increasing possibilities of genome-wide testing in diagnostics and research, the line between diagnostics and research is blurred. It is thus important to describe what can and should be done with diagnostic patient data, and for what type of analyses-specific (additional) research consent is needed. However, this does not exclude a suggestion for further research as a result of a diagnostic investigation. But the distinction between research and diagnostics has to be clear at all times.

STATEMENT 32: A diagnostic test is any test directed toward answering a clinical question related to a medical condition of a patient.

STATEMENT 33: A research test is hypothesis driven and the outcome may have limited clinical relevance for a patient enrolled in the project.

STATEMENT 34: The results of a diagnostic test, particularly by analysis of a whole exome or genome, can be hypothesis generating.

The use of exome or genome data obtained by NGS in a diagnostic setting are acceptable, if the aim is to obtain a genetic diagnosis and the analysis is limited to genes that are known to be linked to (the) disease.

STATEMENT 35: Diagnostic tests that have as their primary aim to search for a diagnosis in a single patient should be performed in an accredited laboratory.

When participating to a research project, patients and families must be aware that such a project may lead to a diagnosis or predictive information about a genetic disease. In research, clinically relevant results should only be transferred into the patient's medical record after confirmation in a diagnostic setting.

STATEMENT 36: Research results have to be confirmed in an accredited laboratory before being transferred to the patient.

Most laboratories set up a database of variant frequencies of all locally sequenced and/or analyzed samples (ideally healthy parents) in order to ease variant interpretation. As such a database does not contain any sensitive information, considerations based on privacy rules do not weigh against the importance of such data for improving diagnosis and healthcare.

STATEMENT 37: The frequency of all variants detected in healthy individuals sequenced in a diagnostics and/or research setting should be shared.

Ideally, all variants detected in disease linked genes should be submitted to databases of pathogenic variants and linked to the clinical data of the patient. The criteria and arguments used for variant classification should also be clearly described.

STATEMENT 38: All reported variants should be shared by submission to federated, regional, national, and/or international databases.

Given the greyness of the area of applicability of NGS it is therefore worthwhile to integrate in logistics into the feasibility study: the implication is clear, at the outset there will need to be a significant capital outlay for assets and staff, followed by annual maintenance and quality management costs.

6. Feasibility logistics - Setting up a lab

Establishing a functioning NGS laboratory poses many challenges particularly in the context of detecting somatic mutations in oncology which range from assay design to sample preparation, data analysis, variant calling, and automation of the entire workflow from samples to results. In terms of the assay design, it is critical to ensure that targets are selected so as to be both of clinical relevance and actionable. In addition, the tests should achieve the required levels of analytical and clinical sensitivity and specificity for diagnostic application.

In addition to the selection of clinically relevant targets, and the requirements of analytical and clinical performance, an automated workflow from sample preparation to data analysis, variant calling and result reporting is critical. The importance of this requirement is well exemplified in the NGS tests, where the workflow involves many steps from DNA extraction from tumor tissue to library construction, template preparation, sequencing and data analysis.

Validation is needed to ensure the reproducibility of the workflow, and automation is a critical component for this. For example, robotic automation should ensure reproducibility in terms of quality and quantity of the DNA extraction from tumor tissue and library preparation as well as minimizing operator errors during the tedious manual steps. Automation also brings other significant advantages i.e. reducing the labour requirement for the work and avoiding the need to repeat samples due to operator errors (thereby increasing cost effectiveness). Other critical requirements in the diagnostic workflow lie in ensuring sample traceability, where automatic sample ID list is being transferred from one part of the workflow to another, this serves to avoid manual entry errors in the ID list.

Another challenge that hinders the widespread application of NGS is data analysis, which requires specialist expertise in bioinformatics. This includes the optimization of analysis parameters, alignment of sequence reads, variant calling and annotation. This is important to ensure high sensitivity and specificity with respect to detecting the target mutations. Building a robust and validated bioinformatic analysis pipeline to meet the strict requirements of analytical and clinical performance is resource intensive. Furthermore, the analysis part should also be easy enough so that laboratory personnel without specialist bioinformatics training can perform it.

The reports generated at the end of the analysis must also be designed specifically to meet requirements in the clinical context, where sufficient QC information must be included to allow assessment of the quality of the runs and samples, whilst the report for medical practitioners must be concise, summarizing only the mutations being sought or detected.

To achieve this final, commercially viable outcome, both now and for the future, how the whole laboratory and process of handling samples and data is structured, plays a critical role in implementation and thereby feasibility of the approach, as this has asset dependent realities. This is detailed below:

Accumulating genomic data and expansion of genome-based clinical trials suggest that many of the future treatments and clinical trials will require comprehensive panels that allow the detection of multiple mutations at the same time. Such tests might involve either probe-based capture or primer-based amplification for the enrichment of genomic regions to be tested.

The number and scope of genes to be tested depend on the purpose of the test. If the purpose is limited to companion diagnostics for current standard care, the number of genes would be very limited. However, if there is a need for clinical trials for which NGS-based tests are required to stratify patients, a broader range of genes should be interrogated.

In this context, the currently used panel tests are mainly focused on clinically actionable genomic alterations at selected protein-coding regions that are defined by the availability of approved drugs and pathogenomic molecular features.

However, there is an increasing need for adding genomic alterations associated with resistance to molecularly targeted therapies or predicted response to investigational drugs.

Thus, commercial laboratories may choose their preferred platforms based on their individual requirements, such as expected sample status (i.e., small biopsies, resected tissue samples, or liquid biopsies, etc.), expected number of samples, and/or types of variants to be analyzed (e.g., copy number analysis might be limited in platforms that use primer-based amplification during target enrichment).

It is important for laboratories to be aware of platform characteristics and perform adequate quality controls depending on the platform characteristics.

Specimen handling

Sample transportation, receipt, and storage

Adequate processing of tissue samples is essential in a reliable NGS testing. Required specimen handling procedures are nearly the same as those required for traditional single-gene tests. Briefly, the quality and the amount of neutral buffered formalin relative to the size of the specimen should be monitored. The time interval from specimen acquisition to fixation should be minimized, and optimal fixation duration should be monitored. The optimal fixation duration depends on the dimension of each sample because formalin penetrates tissue at a rate of approximately 1 mm per hour. The rule of thumb for recommended fixation duration for surgically resected specimen is 24 hours.

Nucleic acid extraction, quantification, and storage

For clinical tests, DNA extraction kits should have a high level of performance specification to obtain DNA of sufficient quality and quantity for intended NGS tests from formalin-fixed paraffin embedded (FFPE) samples. In addition, DNA extraction procedures should have appropriate mechanisms by which sample contamination or misidentification could be avoided. Several commercially available kits that use silica-based or magnetic bead–based extraction protocols meet such requirements.

A major cause of sequence artifacts is deamination of cytosine resulting in C to T transitions during amplification. Formalin fixation and longer storage period contribute to this process. Since those sequence artifacts are usually present at a very low frequency, such artifacts are unlikely to affect test results if sufficient amounts of unique DNA molecules are available. However, if the DNA input amount is too small or if the purpose is to detect variants with low allelic fraction, these artifacts would be a problem. In addition, amplicon-based methods are more susceptible to these artifacts than hybridization capture-based methods.

Thus, laboratories may choose DNA extraction protocols with enzymatic removal of

uracil-containing templates when they interrogate FFPE samples with highly fragmented DNA or with low tumor purity by amplicon-based methods. In the case of decalcified specimens, DNA quality or quantity is inferior to those of the specimens without decalcification.

Quantitation of extracted nucleic acids can be done by Nano-Drop, Qubit, or the Picogreen method. Among them, Nano-Drop is not recommended because it also detects nucleic acids that are not suitable for downstream analyses. Nucleic acids should be stored under highly controlled conditions in order to maintain sample identity and integrity.

Extracted DNA is to be stored at –20˚C and RNA at –80˚C. Sequencing libraries and polymerase chain reaction (PCR) products may be stored in –20˚C but should be separated from pre-amplification materials to prevent them from contaminating pre-amplification materials.

Sample identity tracking
Like other single gene–based tests, verification of sample identity is the most basic and important aspect in clinical NGS test. The NGS panel tests involve many steps, making them inherently subject to sample mix-up or swapping. Thus, test procedures should have an appropriate system to minimize such critical events. For example, panels could be designed to include a number of single nucleotide polymorphism (SNP) markers that allow molecular barcoding of patient samples so that sample identity can be traced. In addition, an electronic laboratory information management system could be useful for this purpose. With any tool in place, it is important to enable end-to-end sample tracking in clinical NGS panel tests.

Library preparation
Library preparation is the step where extracted nucleic acids are prepared for the sequencing reaction. It involves DNA fragmentation, adaptor ligation, and enrichment of the target of interest. Target enrichment can be done either by an amplicon-based approach or by hybridization-based capture. Molecular barcodes are usually introduced to enable sample identification and pooling of multiple samples in a single flow cell lane. To avoid contamination, all steps before amplification should be done in a separate space.

In addition, extra-caution should be in place during sample transportation from pre-PCR area to post-PCR area. After library preparation, appropriate quality controls should be applied to determine whether the rest of the procedures should be continued. Quality controls include quantitation, fragment size analysis, and quantitative PCR using adaptor sequences for priming.

Data analysis of sequence reads
Bioinformatics pipelines used for the analysis of NGS data consists of multiple steps, such as de-multiplexing, read alignment, de-duplication, base calibration, variant calling, filtering, and annotation. Currently, no single “gold-standard” algorithm exists. Therefore, laboratories should choose the most suitable algorithm for the types of variants to be reported and optimize them. In the absence of a “gold standard,” it is important to validate the analytic performance of the bioinformatics pipelines. It is also important to make sure that all versions of algorithms are traceable and properly updated.

Sequencing read
The initial step after sequencing includes converting the base intensities in a sequencer to digital-level nucleotide sequences, called FASTQ. Although the types

of signals differ among various sequencing platforms, most have their internal software for translating base calling into the compressed and de-multiplexed FASTQ files. The nucleotides in FASTQ have corresponding base quality scores that are in a form of logarithmic scales indicating the probabilistic confidence level of the bases. Appropriate quality control should be done to confirm the general integrity of the sequencing data, such as the total number of bases, sequence contents (including GC contents), per base sequence quality, etc. Read trimming is recommended if the base quality or composition does not meet the quality control threshold that each laboratory has empirically set up or that the sequencer manufacturer has recommended.

Alignment

This step is to find where the short read sequences are located. In general, single or pair-end reads are first aligned (mapped) to a human genome reference. It is essential to include the version of the human reference sequence in the clinical report. Since poorly mapped sequences may lead to compromise the reliability of called variants, especially in solid tumors with low tumor content, sequencing reads should be filtered based on mapping quality score so that only confidently mapped reads are processed further.

After the initial mapping, read duplicates should be removed, because unwanted clonal amplification of reads with sequence artifacts may lead to false-positive variant calls. Alignment is often challenging when the sequence reads come from genomic regions having large insertions or deletions, repetitive regions, pseudogenes, or homologous genes because there can be several other similar-looking genomic regions across the reference genome. Realignment using known references that have the suspected insertions or deletions may be necessary in such cases.

Variant calling: single nucleotide variation, insertion/deletion, copy-number variation, and translocation

Somatic variants can be identified by subtracting variants found in non-neoplastic cells from those found in cancer. If a laboratory chooses not to sequence the corresponding normal samples, it should be noted that some germline variations might exist in the result. Laboratories should consider the implementation of modular analysis pipelines, in which different algorithms or settings are used to call different types of variants: single nucleotide variations (SNVs), insertion/deletion (Indels), copy-number variations (CNVs), and translocations.

SNV/Indel

The quality of variant calls is strongly related to the quality of alignment. The key challenge of variant calling is to distinguish real variants from sequencing errors. In general, the more times the variant is sequenced, the more reliable the variant call is. The minimum depth of coverage depends on the required sensitivity of the intended assay, the sequencing platform and the types of mutations to be detected.

Although sensitivity is increased in proportion to sequencing depth, false-positive calls may also be increased especially in cases with low tumor content. There are various reasons for false-positive variant calls; they often result from PCR errors, sequencing errors, mis-mapped reads on repetitive sequences or homologous regions and so forth. Because each variant calling algorithm uses different strategies to filter out false calls, different algorithms sometimes generate discrepant results. Thus, laboratories should find optimal parameter settings during assay development and validation to minimize algorithm-dependent result variability.

Reliable identification of Indels is particularly challenging because sequence reads containing this type of variants are not often accurately mapped. Thus, the sensitivity and specificity for this type of variants is often reduced. With regard to this, laboratories should identify clinically important Indels and validate assay performances to establish reasonable sensitivity and positive predictive value for the identification of those Indels. Due to the high rate of false-positives, manual review by using visualization tools and comparison with the same regions in other samples on the same run is recommended for all Indel calls.

Copy-number variation

The reliable identification of CNV is quite difficult in NGS panel sequencing because of the uneven target coverage related to hybridization capturing steps, the absence of matched normal data, or the lack of coverage uniformity. Although algorithms for detecting CNV in targeted NGS tests are improving, the inherent limitations in cancer panel sequencing of clinical FFPE samples require robust validation of test performance. This type of validation can be done by testing characterized cell lines or clinical samples with known CNV profiles although there is no agreement upon the minimal number of samples for appropriate validation.

Translocation

Translocation can be identified based on the DNA level, using discordant or split sequencing reads. However, the inherent limitations of short-reads in terms of alignment can result in many false positive calls. Therefore, the test performance should be appropriately validated using reference materials with known translocations across targeted genomic regions that have been confirmed by the current gold standard.

Furthermore, it is highly recommended that all translocation calls be manually reviewed by using genome visualization tools such as the Integrative Genomics Viewer. Finally, it should be noted that a translocation can be missed if the breakpoint of fusion (may be somewhere in the introns not covered by the panel) is not included in targeted genomic regions, even though protein-coding regions of the translocation partners are included in the panel.

Variant annotation and filtering

Variant annotation determines if a sequence variant is real and provides predicted resulting amino acid changes. To identify true somatic variants, false variants should be properly filtered. Important sources of false variants include cytosine deamination, amplification errors, and sequencing errors. Cytosine deaminations are introduced *ex vivo*; these variants are not copied to the opposite strand, meaning that the artifacts are only present on one strand.

To facilitate the detection of cytosine deamination artifacts, laboratories may use techniques such as molecular inversion probe and HaloPlex and Duplex sequencing, to enrich and sequence both the sense and antisense DNA strands. Amplification errors can be introduced due to DNA polymerase errors during amplification steps of library enrichment.

These errors might be minimized by the application of unique barcodes to individual DNA molecules during library enrichment. If the same variant is detected in multiple unique molecules, the variant might be real because it is unlikely that individual molecules acquire the same polymerase error during amplification. Sequencing errors are highly dependent on the sequencing platform and sequencing chemistry. The ability to call SNVs and Indels is known to be similarly accurate for data generated on the PGM and Illumina platform, provided that there is sufficient

coverage.

Interpretation of computational output

Depending on the types of variants (e.g., missense, nonsense, etc.) and the types of genes (e.g., hotspots in oncogenes vs randomly distributed mutations in tumor suppressor genes), interpretation of detected variants can be simple or quite complicated.

Basically, it is strongly recommended that interpretation is performed by trained staffs such as clinical molecular geneticists or molecular pathologists. A multi-disciplinary sequencing data analysis team with various scientific backgrounds including clinical oncology, genomics, bioinformatics, and pathology, is recommended for accurate interpretation.

For variants that are not hotspot mutations, germline variants should be excluded first. Since most panel tests do not analyze matched non-diseased tissue, laboratories should prepare mechanisms to filter out potential germline variants based on the genetic polymorphism data on the population to which the tested patient belongs. In most instances, laboratories use public databases on germline polymorphism such as the 1000 Genomes Project, ExAC, or dbSNP), but ideally, these data should be derived from the same ethnic group as the tested patient.
Since the clinical and biological significance of disease-related genomic variants are increasingly characterized, many variants detected in most panel tests have related information in public databases. However, previously uncharacterized variants may also be detected considering the characteristics of NGS tests.

Potential biological significance, or pathogenicity, could be inferred from the archived genotype-phenotype correlation data such as ClinVar, Human Gene Mutation Database, and Leiden Open Variation Database. In addition, the *in silico* prediction of functional impacts is available in dbNSFP or Ensembl Variant Effector. Since no single database is perfect, it is essential to refer to multiple resources for appropriate interpretation. Finally, knowledge about variants will be continuously improved with the accumulation of each lab's experiences and feedbacks from clinicians.

Reporting

General format

Reporting of NGS panel test results should follow the general professional organizations' recommendations and guidelines. There are two major essential parts of a report: proper patient identification and detected actionable variants. The patient identification part is the same as that in the current standard single gene-based tests, but the detected variant part is more complex and sophisticated in clinical NGS panel tests because typical NGS panel tests detect multiple variants at the same time.

The clinical NGS report should include the essential information and the most pertinent information, such as actionable variants and a critical summary of those variants, should be placed in a clearly visible section on the first page. Detected sequence variants should be annotated in concordance with the Human Genome Variation Society mutation nomenclature and the version of the human reference sequence to which sequence reads were aligned should also be included in the report.

Variants, which are not clinically actionable but are potentially useful for future practice, might follow the actionable variants. Then, detailed information about the detected variants (see the "Interpretation" section) and essential technical

information, such as genes or genomic regions included in the panel and key quality control metrics, may be listed. Inclusion of granular details of technical information is not recommended, but a description of how clinicians can obtain the details may be included.

Presentation of detected variants and clinical translation

Since typical panel results include several disease-related genomic variants with different levels of clinical or biological evidences, variants should be classified and reported according to the level of evidences.

Interpretation of detected variants in terms of their clinical impact and pathogenicity is necessary and many information sources such as public databases, published guidelines, and computational prediction algorithms should be integrated for proper interpretation. In addition to the previously mentioned public databases on germline polymorphism, *in silico* prediction of the functional impacts, genotype-phenotype relationship, and several disease-specific mutation databases are available online.

These resources are very useful in the interpretation of variants, specially in the context of cancer, but it is important to make sure that the database is properly curated, referenced and updated in a regular basis. Regarding this, FDA guidance suggested that appropriate databases should implement decision matrices with published details of each variant's interpretation and have documented standard operating procedures (SOPs) for the curation and update of this information.

In cases where only limited interpretation can be made and full quality control standards could not be met, it is essential that pathologists make a professional judgment on whether the result should be reported. Also, any limitation of the analysis should be clarified in the report. Furthermore, clinical NGS laboratories should have clearly defined protocols about when additional confirmatory tests should be advised, as well as performance validation data in cases where those confirmatory tests are not necessary.

Requirements for clinical NGS laboratories

Wet Bench Analytic Process

Documentation

The detailed documentation of SOPs is critical for quality assurance of a complex, multi-step wet bench process. All SOPs of each step of the wet bench process must be documented so that each step can be traced. This includes documentation of all methods, reagents, instruments and controls (if applicable). Most of the documentations should be similar to those of current standard single gene testing, but those specific to NGS testing include detailed information regarding captured regions, such as genomic coordinates of captured probes and lists of genes and target enrichment protocols.

Clinical laboratories that process different types of samples, such as FFPE samples or blood, should establish SOPs for each validated sample type. Metrics for quality control to assess run status must also be documented. Examples include mean target coverage, percentage of reads that map to target regions, and the fraction of bases meeting specified quality and coverage thresholds. Laboratories must define and document acceptance or rejection criteria for each step of the wet bench process, such as DNA extraction, library preparation, and sequencing.

Validation

Before testing patient samples, clinical NGS laboratories must establish the analytical validity of the intended tests. If the intended test is CE approved laboratories can verify the performance specifications established by the manufacturer. If the intended test is not approved by certified bodies, i.e., a laboratory-developed test, laboratories must establish performance characteristics such as accuracy, precision, sensitivity, specificity, and reportable range.

Because NGS panel tests involve complex, multistep processes, each step needs to be empirically optimized to determine optimal assay conditions. Once those optimizations are done, an analytic validation should be performed for a whole test in a “beginning to end” fashion. Test performance should be separately validated for different types of samples, such as FFPE or blood.

For “beginning to end” validation, a number of samples should be analyzed to assess the test performance. There is no general agreement on how many samples are required for this type of validation. With regard to this issue, the CAP concluded that adding a minimum number of samples for validation is premature given the ongoing evolution of NGS technologies and the diversity of diagnostic applications.

It is important to evaluate as many different genomic regions as possible because sequence context can influence sequencing results. In addition, laboratories should determine analytic performances for all variant types relevant to the intended test (e.g., SNV, Indel, CNV, and translocations). Since NGS-based tests interrogate multiple variants at the same time, the validation of test performances involves two parts: method-specific (detection of as many variants as possible in a single sample) and analyte-specific (detection of a certain genomic variant in multiple samples). Laboratories can establish the test performances by combining those two approaches.

Accuracy validation is a “method-specific” way of validation. For this type of validation, laboratories might use well-characterized disease cell line samples to interrogate whether the intended test accurately detects all known variants within the genomic regions covered by the panel. In this case, reference cell line samples should be periodically monitored for identity and passage number to prevent a significant genetic drift.

Analytic sensitivity can be evaluated by comparing NGS test results with the current gold standard test results for known genetic variants in reference materials. To this end, laboratories may use historical controls such as accumulated clinical samples with well-characterized genomic variants by gold standard methods, provided that the gold standard tests were done in an appropriately accredited environment.

Analytic specificity is theoretically calculated by determining the fraction of test negatives (wild type sequence calls by the intended test) per true-negatives (samples that are known to have wild type sequences by the gold standard method). However, this concept often does not work well for NGS-based tests because too many potential variants are included in typical cancer panels. For most laboratories, it might be reasonable to leverage specificity by determining the average number of false-positive calls for the regions tested in a number of well-characterized clinical samples.

Any changes to a clinical NGS test, such as changes in instruments, specimen types,

reagents, or software, require that performance specifications be re-established or be shown to be unchanged. For example, inclusion of new genes to an existing gene panel requires revalidation to make sure that new sequence variations are reliably detected without compromising the quality of the original assay. The extent of re-validation depends on the predicted influence on the test performance related to the change. For example, if only the bioinformatics pipeline is updated, it may not be necessary to re-validate all steps before data analysis.

Quality management

Clinical NGS laboratories should establish and follow a quality management plan. This plan should be integrated within the institution's overall quality assurance program. Components of the NGS quality management program are not much different from the traditional single gene-based tests. Once laboratories establish an initial validation of test performances, laboratories must perform internal quality controls daily and external quality controls periodically.

Quality assurance program

Recently, guidelines for quality control and recommendations for the use of NGS in different applications have been published and are summarized below. This guidance will work as a checklist for each component of the intended NGS test and related quality control metrics that require reviewing in order to feel confident about the quality of results. It is impractical to include multiple positive controls with different variant types during each run due to unacceptable cost and time. Instead, a single characterized external control with known variants in each run may be sufficient to demonstrate that the procedure is successful.

Typically, this can be done by preparing a number of DNA aliquots from a large FFPE sample block whose genome has already been well-characterized genomically and include those aliquots in each run of patient samples.

Recommended items to check prior to releasing NGS results for diagnostic use and QC metrics

Item	**Checklist**	**Consequences of non-conformity**	**Improvement suggestions**
Tissue sample adequacy	Criteria for inadequate specimen	Testing inadequate specimens may lead to a waste of time and money or depletion of available samples.	Check sample adequacy rigorously before testing.
	Minimum diseased tissue content		Request further sampling in case of inadequate samples.
	Appropriate sample handling including fixation and transportation	Inadequate amount or diseased tissue content can lead to false-negative test results.	
Nucleic acid extraction	DNA quantity and quality in terms of amplifiable DNA	DNA with suboptimal quality may inhibit sequencing reaction.	Failed samples should be reported as such and further material might be requested with specified requirement.
		Small amount or fragmentation of DNA may lead to poor quality	

		sequencing data with insufficient or uneven coverage and/or high duplication rate.	Trying another validated extraction method may often helps.
Sample identification	Sample identity tracking throughout all steps	Misidentification of samples could lead to incorrect patient management.	If there is any concern about sample identity, starting over from DNA extraction may be necessary.
			Introduction of polymorphic SNP markers into gene panel and running another genotyping method with the same marker set might be helpful.
Library preparation	Minimum library concentration	Poor sequencing library may lead to insufficient or uneven coverage.	Consider modification of library preparation method or an alternative method to verify any uncertain results.
		Libraries with poor complexity or bias may result in false-negatives. False-positives may also occur due to potential amplification bias.	
Sequencing	Criteria for minimum sequencing depth and other quality metrics (% reads mapped to target regions, % of targets with specified coverage, duplication rate)	Inadequate coverage is associated with higher levels of uncertainty of the test results.	Repeat sequencing with existing library or start over from DNA extraction step.
		Genomic regions with insufficient local coverage may lead to inaccurate results for variants located in those regions.	Verification of uncertain results with another method may be helpful, especially, in case of actionable variants.
Variant detection and review	Variant allele frequency, local sequencing depth and quality score	Failure to filter out sequencing artifacts may lead to false-positive results.	Manually review of clinically important variants even if computational algorithms called no mutation on them.
	Presence of the same variant in forward and reverse strands	Clinically important variants may sometimes be missed.	Any ambiguous or unexpected results should be reviewed by laboratory scientists and pathologists.
	Mapping quality of		Verify variants with

	sequencing reads Potential sequencing artifacts		another method, if applicable.
Bioinformatics	Correct pipeline and version Appropriate version and build of human reference sequence Cross-contamination?	Using outdated or inadequate software can lead to false-positive or false-negative results.	Update software on a regular basis.
Reporting	Endorsed by an authorized competent pathologist?	Variants with clinical significance may be reported erroneously, leading to inappropriate treatment.	Responsible pathologists should be given enough time and opportunities for education and training.

Policies: confirmatory testing, laboratory records, upgrades

As NGS panel tests with appropriate quality controls are reported to meet a clinical grade performance, routine confirmatory testing is not recommended. However, laboratories should have a policy that clearly documents both any indication for confirmatory testing as well as any performance validation data upon which they decided that such confirmatory testing is not necessary.

In addition, the CAP is flexible regarding the methods used for any needed confirmatory testing.

Keeping comprehensive laboratory records is essential in monitoring complex, multi-step NGS cancer panel tests. In this regards, such records should be maintained in such a way that all detailed information about test procedures including reagents, sequencing runs, wet lab, and bioinformatics procedures and responsible technicians is traceable. While all details need not be included in the clinical report, laboratories should maintain a database from which detailed information regarding the analysis of individual specimen can be obtained.

Laboratories must be prepared for upgrades to make sure that they are not using obsolete methods. A policy for the upgrade of instruments, sequencing chemistries, and reagents or kits, as well as subsequent post-upgrade validation of test performances should be in place. The policy may include specified intervals for upgrade and required validation processes, depending on the type or extent of upgrade.

Bioinformatics process

Documentation

Laboratories should document all bioinformatics processes, including all data files, variant caller's parameters, and versions of the bioinformatics algorithms. Sources and versions of all bioinformatics algorithms should be documented and updated properly. Quality control information on bioinformatics analysis, such as the cut-off of read depth, base quality score, and mapping quality, should also be documented.

Validation and quality management

General validation principles were already discussed in the Validation proportion of the "Wet Bench Analytic Process" section. Briefly, for bioinformatics pipelines,

laboratories should iteratively find parameters for optimal performance of computational algorithms before applying it to the lab process. Once the pipeline is initially validated, variations between sequencing runs should be monitored daily. Principles of the quality management were already discussed in the “Wet Bench Analytic Process” section. In short, laboratories should monitor any deviation from established performance characteristics in terms of quality metrics and analysis results. For any deviation, laboratories should document the investigational measures and corrective actions made to resolve the deviation. Essential quality metrics for bioinformatics performance verification include depth of coverage, uniformity of coverage and base call quality scores. In addition, GC bias, proportion of reads that map to nontargeted regions, and percentage of duplicated reads could also be used to monitor performances of sequencing reaction and subsequent bioinformatics analyses

Policies: upgrades, storage, and data management

The bioinformatics pipeline should be revalidated upon any changes in operating systems, software, or overall pipelines, which may otherwise affect its analytic performance. Since a huge amount of data files are generated from the bioinformatics pipeline, it is impractical to store all sorts of files considering the significant cost. Instead, the CAP NGS workgroup recommended that some important file formats, such as FASTQ, BAM, and VCF, should be stored for quality controls or investigational use. There is no general agreement on the required storage period, but it is important for laboratories to set their own storage policies in accordance with local or national requirements (if any) and inform clients of those policies.

7. Feasibility logistics - Equipment costs

Various types of NGS equipment can be purchased with varying degrees of application which is indicated in the table below.

Equipment costs: values indicated below are in 1000's of Euros; additional instruments correspond to computing and storage related devices, while service contracts are the annual maintenance and component exchange costs that will need to be incurred to maintain the quality standard.

Note: only those pieces of equipment marked with a *, have been CE-IVD authorised for 'Diagnostic' purposes, all other equipment can only be used for RUO or screening.

Instrument	Purchase Cost	Additional Instruments	Service Contract
Illumina MiniSeq	49	–	5
Illumina MiSeq	99	–	14
Illumina MiSeq Dx*	125		14
Illumina NextSeq 500	250	–	32
Illumina NextSeq 500 Dx*	350		32
Illumina HiSeq 2500	690	55	74.5 + 6
Illumina HiSeq 4000	900	55	81 + 6
Illumina HiSeq X (whole genome sequencing)	1200	55	92.7 + 6
Ion Torrent – PGM (314/316/318 chips)	49	18/32	4.3-8.5/9.9
Ion Torrent – Proton	224	19	19.9/32.8
Ion Torrent – S5	65	19	4.3-8.5/9.9
Ion Torrent – S5xl	150	19	8.6-17/18.5
Oxford Nanopore MinION	1	0	0
Oxford Nanopore PromethION	75	0	0
PacBio RS	695	–	84
PacBio Sequel	350	–	20
Qiagen Genereader	120	(included)	Unknown
ThermoFisher SOLiD – 5500xl	251	54	44.4

It is important to distinguish between regulations which govern health institutions and private entities: through article 1.5 of the EU IVDD (regulation 33) a hospital can establish a molecular diagnostics lab using non CE-IVD marked equipment providing it meets the safety and quality standard defined by the health authority. All of the above pieces are permitted, to be used in this context: this exemption does not apply to private service providers.

With regard to reagent costs, these are the main driver in the testing costs: The cost of a somatic targeted next-generation sequencing panel is an average €607 according to a new study by researchers in France set out to gauge whether next-generation sequencing tests can be affordable in a clinical setting. While a US team performing a 46-gene hotspot cancer panel assay allowing multiple gene testing from small diagnostic biopsies from non-small-cell lung carcinoma (NSCLC), colorectal carcinoma, and melanoma was performed at a cost of €449 per patient, the panel was less expensive locally than performing more than two or three single gene tests.

Costing should anticipate to include over 3 hours of labour, costing approximately €60, plus overhead and associated company costs and costs of equipment (estimated at an average of €30 per sample), meaning that one patient cancer panel screen costs a maximum of €900.

It is questionable whether a public health system will be willing to pay this price given the absence of proven evidence of real use: a private health system may consider this. This may also explain why the FDA has only approved this approach as a reimbursable diagnostic for late stage cancer.

8. Feasibility logistics - Where the technology is going; trends and applications

It is also important to indicate where the field is going, as any purchase today may not be sufficient in the near future. To date, progress in NGS technology has decreased the cost of sequencing from $10 million to $1000 and further evolution of this technology is aimed at decreasing the cost of sequencing to $100. Though that price point has not yet been reached, and likely won't be for several years, the plummeting cost of sequencing coupled with the rise of personalized medicine will accelerate demand in the applied sector.

The global next generation sequencing market is majorly dominated by the U.S. players. Companies such as Illumina, Inc. and Thermo Fisher Scientific have been reported to contribute more than 60% of the market value, but the emerging technologies such as those based upon Nanopore technology is estimated to witness the highest CAGR.

Illumina are also in the process of launching their new semi conductor based system: the firefly. The tech specs look impressive, but proof will be needed: to quote Illumina

"Illumina also previewed a new sequencing system designed to democratize NGS and truly enable the adoption of genomics worldwide. The highly-reliable, easy-to-use NGS platform will offer customers a low capital cost and plug and play installation. The most integrated sequencer ever developed, the system will take users from purified DNA to answers, making it the ideal tool for virtually any laboratory. The stackable two module system will minimize hands-on time for both library preparation and sequencing. Leveraging Illumina digital fluidics technology, the first module will make library preparation simple and efficient preparing eight normalized samples in parallel on a library preparation cartridge. A separate cartridge for sequencing, loaded into the second module, will deploy a one channel version of Illumina's sequencing-by-synthesis chemistry on a semiconductor chip. Sequencing data will seamlessly move to BaseSpace for analysis.

The new system will be superior to competing semiconductor-based sequencing systems with a raw error rate of less than one percent, data quality comparable to that of a HiSeq X™ Sequencing System. With output of approximately 1.2G per run, the platform will be ideal for numerous markets including academic research, oncology, infectious disease, inherited disease, and reproductive health.

Illumina expects to commercialize this system in the second half of 2017, priced at less than $30,000 USD for both modules. For customers running eight samples at once, the company projects per-sample consumable pricing near $100."

However a paradigm shift in the dynamics market is expected in the year 2020 with the introduction of portable, or Point of Care technologies in full scale. The leading players in the market are taking the necessary steps to tackle the shift in the dynamics and sustain in the competition: at present first movers are Oxford nanopore with their MinION system (https://nanoporetech.com/products/minion)

9. The market for CE-IVD NGS products for tumour diagnostics/ screening

This global cancer diagnostics market is expected to reach $13.1 Billion by 2020 from $7.1 Billion in 2015, and is poised to grow at a CAGR of 12.9% during the forecast period. The total market includes the evolution of molecular diagnostics as well as novel imaging systems, immunoassays and cell analysis.

Cancer molecular diagnostics is usually performed based on platforms such as PCR, leveraging panels. The global market for cancer molecular diagnostics kits was $ 335.9 million in 2016 and is expected to reach $ 6980 million in 2026 with an average annual growth rate of 32.9%.

Limited products (kits) exist on the market, but the number is growing and the competitors global, as non European entities obtain CE marking for their own manufactured kits. Note that the kits are in general not promiscuous i.e. they are designed to be used on specific pieces of equipment, they cannot be used on all sequencers.

Sentosa SQ Melanoma panel: developed by Vela Diagnostics, The *Sentosa* SQ Melanoma Panel is a Next-Generation Sequencing-based *in vitro* diagnostic test that simultaneously detects hot spot mutations in 10 genes from formalin-fixed paraffin-embedded (FFPE) samples that have previously demonstrated clinical relevance in the management of patients with melanoma. http://www.veladx.com/product/ngs-oncology/sentosa-sq-melanoma-panel-95.html

Sentosa SQ Colorectal cancer panel: developed by Vela Diagnostics, The *Sentosa* SQ CRC Panel is a Next-Generation Sequencing (NGS)-based *in vitro*diagnostic test that simultaneously detects hot spot mutations in 11 genes from formalin-fixed paraffin-embedded (FFPE) samples that have previously demonstrated clinical relevance in the management of patients with colorectal cancer (CRC).
http://www.veladx.com/product/ngs-oncology/sentosa-sq-colorectal-cancer-panel-112.html

Sentosa SQ Non-small cell lung cancer panel: developed by Vela Diagnostics, The *Sentosa* SQ NSCLC Panel is a Next-Generation Sequencing (NGS)-based *in vitro*diagnostic test that simultaneously detects hot spot mutations in 11 genes from formalin-fixed paraffin-embedded (FFPE) samples that have previously demonstrated clinical relevance in the management of patients with non-small cell lung cancer (NSCLC). http://www.veladx.com/product/ngs-oncology/sentosa-sq-non-small-cell-lung-cancer-panel.html

Sentosa SQ thyroid cancer panel: developed by Vela Diagnostics, The *Sentosa* SQ Thyroid Cancer Panel is a Next-Generation Sequencing (NGS)-based *in vitro* diagnostic test that simultaneously detects hot spot mutations in 10 genes from formalin-fixed paraffin-embedded (FFPE) samples that have previously demonstrated clinical relevance in the management of patients with thyroid cancer. http://www.veladx.com/product/ngs-oncology/sentosa-sq-thyroid-cancer-panel.html

Sentosa SQ Oncokey core: developed by Vela Diagnostics, is a four panel product that combines the four screens above. http://www.veladx.com/product/ngs-oncology/sentosa-sq-oncokey-core-apac-emea.html

BRCA MASTR Plus Dx: Developed by Mulitplicom (now part of Agilent technologies) it is a next-generation molecular diagnostic solution (NGS library preparation kit, analysis software and quality control) uniquely detects both germline* (inherited) and somatic (acquired) mutations in BRCA1 and BRCA2 genes.Compatible with multiple market leading sequencers, including MiSeq®, MiniSeq® and NextSeq®. Distributor for Eastern European Countries (Accela)

4Bases Next Generation Sequencing (NGS) kits: Developed by Longwood Diagnostics. Allows for greater specificity and sensitivity in oncological and pharmacogenetic diagnostics, helping the characterization of cancer and treatment personalization.

4Bases kits are based on targeted NGS technology for the detection of clinically relevant mutations from tissue samples, or blood in the case of BRCA.

CE-IVD panels available:

- Breast and ovarian cancer. HR1 BRCA 1/2 kit. Identification of somatic and germinal mutations in BRCA1 + BRCA2 + TP53
- Colorectal cancer. CRC NGS panel. Analysis of the KRAS / NRAS / BRAF genes
- Lung cancer. EGFR NGS kit
- Skin cancer. BRAF NGS kit
- Thyroid cancer. Thyroid-ID NGS kit. Analysis of 13 associated genes: KRAS, NRAS, HRAS, BRAF, PIK3CA, TP53, NOTCH1, PTEN, CDKN2A, EGFR, AKT1, CTNNB1, hTERT
- Multi cancer applications. BENKit. Analysis of specific genes such as KRAS, NRAS, EGFR and BRAF in separate tubes.

OncoTRACE: (developed by OncoDNA SA) capable of monitoring up to 15 gene variants in ctDNA obtained from a patient's blood. The variants are initially identified using the company's OncoDEEP theragnostic product, which combines DNA, RNA and protein analysis of biopsy samples. The test can identify drug resistance or potential recurrence earlier than is possible with imaging technologies.

BRCAaccuTest™ and clinical analysis software, NGeneAnalySys™ for hereditary breast/ovarian cancer genetic testing**: Developed by Korean IVD startup NGeneBio** BRCAaccuTest™ is an amplicon-based targeted NGS panel which detects germline mutations in BRCA1 and BRCA2 genes for breast/ovarian cancer patients and genetic predisposition.

Therascreen BRCA 1-2 NGS FFPE Kit: Developed by Qiagen for qualitative detection of variants within the BRCA1 and BRCA2 genes
- Detection of germline and somatic variants
- Allele frequency cut off of 5.75%
- CE-IVD kit for DNA extracted from FFPE ovarian tissue
- QIAGEN workflow for the Illumina MiSeqDx

The therascreen BRCA 1-2 NGS FFPE Kit aids classification of ovarian cancers by identifying variants in human BRCA1 and BRCA2 genes using DNA derived from formalin-fixed paraffin-embedded (FFPE) ovarian tumor tissue.

MammaPrint® BluePrint® Kit: developed by Agendia it is a next-generation sequencing (NGS)-based MammaPrint® BluePrint® Breast Cancer Recurrence and Molecular Subtyping Kit

BRCA Complete™ kit: Developed by Entrogen it is a targeted next generation sequencing (NGS) panel for BRCA1 and BRCA2 exome sequencing on Illumina® MiniSeq, MiSeq, and NextSeq platforms. BRCA Complete™ is a full solution for BRCA1/BRCA2 sequencing that includes reagents for target enrichment, library preparation, and PCR clean-up in one package. The kit is compatible with blood, fresh frozen, and formalin fixed paraffin embedded (FFPE) samples. EntroGen's proprietary enrichment technology enables detection of somatic mutations in BRCA1/BRCA2 genes with approximately 2% limit of detection.

Oncomine Solid Tumor DNA kit: developed by Thermofisher the kit captures regions of human somatic variants (deletions, insertions, inversions, and substitutions) present in selected regions of cancer-related genes (EGFR, ALK, ERBB2, ERBB4, FGFR1, FGFR2, FGFR3, MET, DDR2,KRAS, PIK3C A, BRAF, AKT1, PTEN, NRAS, MAP2K1, STK11, NOTCH1, CTNNB1, SMAD4, FBX W7, TP53) for analysis using NGS technology.

FoundationOne®: Developed by foundation medicine, this is a laboratory service, but not a kit, which has received CE-IVD certification. It is a fully informative genomic profile for solid tumors used by oncologists to identify the molecular alterations in a patient's tumor and match those alterations with relevant targeted therapies and clinical trials. Using next-generation sequencing in routine cancer specimens, FoundationOne interrogates all genes somatically altered in human cancers that are validated targets for therapy or unambiguous drivers of oncogenesis based on current knowledge.

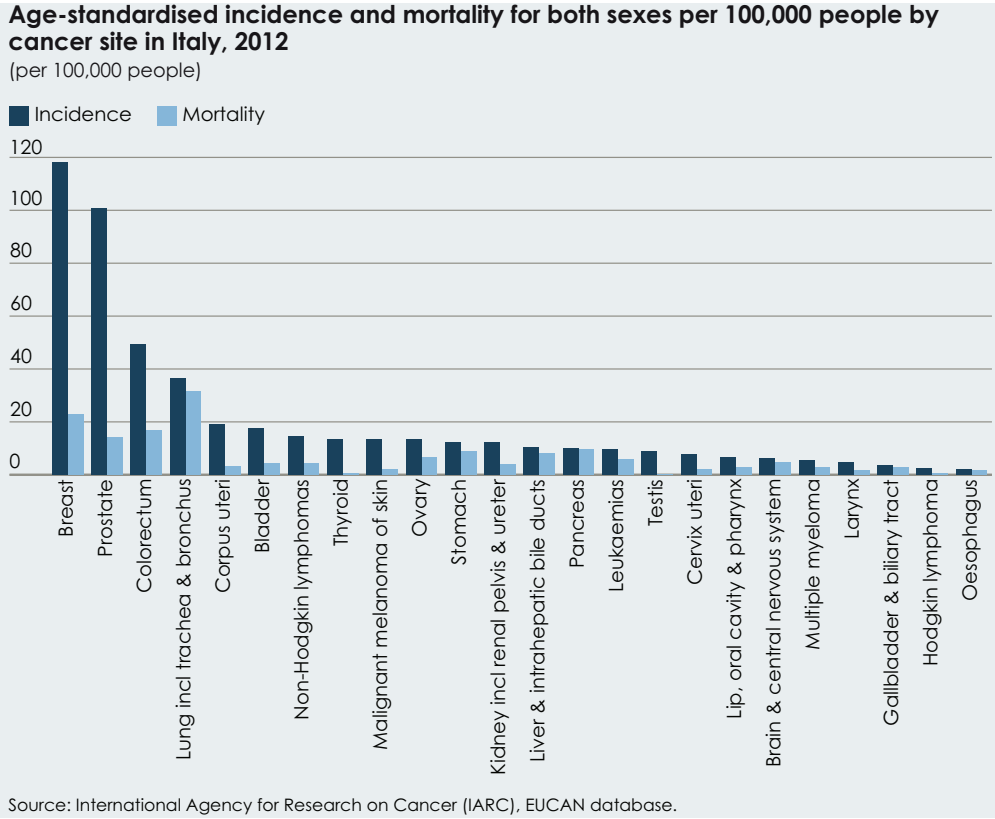
In February 2017, Invivoscribe announced a long-term collaboration with Illumina to develop and commercialize in vitro diagnostic (IVD) assays for the next-generation sequencing MiSeqDx® platform. The announcement reported that under the terms of the agreement, Invivoscribe will work with Illumina to seek FDA clearance or approval of a number of biomarker and immuno-oncology assay kits for use on the MiSeqDx® instrument. Invivoscribe already markets several CE-marked tests outside of the US under the brand name LymphoTrack® Dx Assays, and also Research Use Only (RUO) LymphoTrack® Assays, for use on the MISEQ® platform. This is part of an ongoing initiative of Invivoscribe to develop and provide standardized molecular diagnostic assays to support precision medicine in the oncology field. Invivoscribe also offers CE-IVD LymphoTrack® Dx Assays, and also Research Use Only (RUO) LymphoTrack® Assays for Fisher Scientific's ION PGM next generation sequencing platform.

10. Market trends and market opportunities

10.1 The Italian cancer demographics and value

NGS sensitivity, high speed and less cost per sample make them a highly lucrative platform in comparison to other approaches for screening. Their main application is identification and diagnosis of different types of cancers which occur due to genomic alterations such as mutations. As such, analysis of the Italian demographic for cancer reveals a viable market possibility, if they are willing to pay.

The IARC figures for Italy's reported cancer incidence are below:



A historical analysis of cancer incidence in the Lazio region from 1970-2015 revealed that the most frequent cancer sites were breast, colon-rectum and prostate with 5,529, 5,315 and 4,759 new diagnosed cases, respectively. The cancers with increasing incidence trends were breast cancer, lung cancer and skin melanoma in women, and prostate cancer, colorectal cancer and melanoma in men. The incidence rates of uterine cervix and stomach cancer decreased. The male lung cancer rates increased, reaching a peak in the late 1980s, and then decreased. Prevalence increased for all the considered cancers except cervix cancer. In 2012 breast, colorectal and prostate cancer had the highest prevalence, with 68,239, 36,617 and 33,934 prevalent cases, respectively.

On these figures, if a 20% market uptake is obtained with the €900 price per screen, this would equal a €2 800 000 annual revenue, of which profit would be around €200 000. Modelling scenarios based on if technology evolves (include price drops) and testing can be done for €250 per screen, an anticipated significantly higher uptake (reaching 80%) could result in annual revenues above €3000000, with higher profit margins, as well as the potential to expand into neighbouring.

However if NGS also extends into additional diseases and applications the revenues could be significantly higher.

10.2 The Global market

The key factor driving the growth of the global NGS based diagnostics market is growing incidences of cancers, complicating diseases due to drug-resistance of microbes. On the basis of end user, the market is segmented into research centers and academic & government institutes, hospitals & clinics, pharmaceutical & biotechnology companies, and other end users. Among these end users, the hospitals and clinics segment is expected to register the highest CAGR due to the growing number of agreements between companies and hospitals for the development of advanced tests and products catering to the specific needs of hospitals and clinics; an area where Genechron can have impact in Italy.

Because of the enormous appeal of NGS technology, huge growth for the market has been predicted, which should jump from nearly $3.2 billion in 2017 to $10.5 billion by 2022, with a 27% CAGR. As an example the U.S. clinical oncology next generation sequencing market is anticipated to reach USD 1.53 billion by 2024, while the global market for blood cancer molecular diagnostics is $ 335.9 million as of 2016 and is expected to reach $ 6980 million in 2026 with an average annual growth rate of 32.9%.

Cancer molecular diagnostics kits belonging to the outpatient management or primary care screening market, which is expected to grow globally from $ 60.25 billion in 2016 to $ 78.74 billion by 2021, with an average annual growth rate of 5.5%. Within this market, the molecular diagnostics market is expected to grow from $ 6.54 billion in 2016 to reach $ 10.12 billion in 2021, with an average annual growth rate of 9.1%. This is illustrated by segment in the tables below:

Market size of the global cancer diagnosis by technology platform ($million)

Category	2013	2014	2015	2016	2017	2018	2019	CAGR
RT-PCR	710.5	758.9	998.2	1313.0	1727.1	2271.8	2988.3	31.5%
DNA microarray	357.1	394.4	536.1	728.7	990.4	1346.3	1829.9	35.9%
LOAC	177.9	187.0	240.7	309.8	398.7	513.2	660.5	28.7%
NGS	174.5	240.0	414.3	715.2	1234.6	2131.3	3679.2	72.6%
Multiplex conventional	92.2	98.8	131.9	176.1	235.1	313.9	419.1	33.5%
Next generation capture	68.2	76.7	124.7	202.8	329.7	536.1	871.7	62.6%
Protein microarray	18.0	20.4	29.3	42.2	60.6	87.2	125.3	43.8%
Other	25.3	31.7	38.9	47.8	58.6	71.9	88.3	22.7%
Total	1623.7	1807.9	2514.2	3535.5	5035.0	7271.7	10,662.3	42.6%

Market size of the global cancer diagnosis by cancer type ($million)

Category / Year	2013	2014	2015	2016	2017	2018	2019	CAGR
Breast cancer	375.4	383.8	531.3	735.5	1018.2	1409.5	1951.2	38.4%
Colorectal cancer	346.5	391.6	531.1	720.4	977.1	1325.2	1797.4	35.6%
Cervical cancer	259.0	267.6	304.1	345.5	392.6	446.2	507.0	13.6%
Lung cancer	113.6	128.2	197.0	302.8	465.5	715.4	1099.6	53.7%
Precancer	104.5	124.9	188.3	283.7	427.7	644.6	971.5	50.7%
Prostate cancer	99.2	117.2	170.1	247.0	358.6	520.6	755.8	45.2%
Melanoma	92.0	104.8	134.6	172.8	221.8	284.8	365.7	28.4%
Leukemia	74.5	88.3	130.4	192.6	284.4	419.9	620.1	47.7%
Lymphoma	42.9	51.9	91.0	159.6	279.8	490.7	860.4	75.4%
Pancreatic cancer	26.4	32.7	47.0	67.6	97.1	139.6	200.6	43.7%
Bladder cancer	21.2	25.9	36.4	51.3	72.2	101.6	142.9	40.7%
Chest cancer	19.1	21.8	35.3	57.3	92.9	150.7	244.3	62.1%
Brain cancer	18.4	24.2	34.2	48.3	68.3	96.5	136.4	41.3%
Thyroid cancer	12.6	16.3	28.2	48.9	84.7	146.6	253.9	73.2%
Kidney cancer	11.7	15.6	27.6	49.0	86.8	153.9	272.8	77.2%
Ovarian cancer	6.7	9.5	16.4	28.3	49.0	84.6	146.1	72.7%
Stomach cancer	0.0	3.6	7.9	17.2	37.5	81.9	178.8	118.4%
Other							157.8	
Total	1623.7	1807.9	2511.1	3527.8	5014.1	7212.2	10,662.3	42.6%

However as the technology and applicability takes hold significant usage in neonatal screening, managing complex and rare diseases, cancer biomarkers and pharmacogenetics are anticipated to broaden the market possibilities.

10.3 Managing patients with cancer

Oncologists now having a range of biomarker tests at their disposal to make a more informed and personalized drug choice; treatments which are likely to lead to better response rates and more prolonged responses can be selected based on molecular characteristics exhibited by the patient's tumor.

The market is emerging from the initial exploratory stages of oncology biomarker testing and are on the verge of a more radical change. The central principle at the core of using molecular diagnostics to inform treatment decisions is that the patient's tumor exhibits certain aberrant characteristics that predispose it to interventions at the molecular level.

Unless inherited through the germline, these mutations are not present in non-malignant cells in the patient's body, and therefore the traditional approach for tumor biomarker testing for patients with solid cancers relies heavily on obtaining tissue samples from the actual tumor.

This, comes with several drawbacks: initial biopsies used in diagnosis don't always contain enough viable tissue for testing, biopsy samples cannot be stored indefinitely without degrading, there is considerable intra-tumor genetic heterogeneity, and re-biopsies are often invasive.

The latter point is particularly important when it comes to patients who have experienced a disease progression, which may be indicative of the tumor having acquired novel somatic mutations not present at the time the initial biopsy was taken, which is an inherent limitation.

Solid tumors begin as small groups of localized malignant cells at the primary tumor site, but ultimately spread to distant organs. Cancer metastasizes by shedding cells from the primary tumor, which enter the bloodstream and travel to distant sites. These are called circulating tumor cells (CTCs), cell-free circulating tumor DNA (ctDNA) and RNA (ctRNA), which can be measured via liquid biopsies.

Testing for resistance mutations (as opposed to sensitized ones) requires a new sample to be collected to detect any newly acquired mutations after treatment failure. There is also benefit to regularly test for resistance mutations, to track the evolution of the genetic make-up of the tumor during its exposure to targeted therapies, which now occur much more frequently via outpatient or primary care centres.

These tests only look for mutations/changes in expression levels of single genes or proteins while cancer, is a hugely complex disease and involves the interaction between many different such genes and proteins.

The inherent limitations of single-marker tests are two-fold:

i. Genes do not act in isolation.

Single-marker approaches fail to address the fact that the transformation of healthy cells into cancer cells is often the result of a combination of mutations acting together, affecting intricate signaling pathways within and between cells.

ii. Not all cancers exhibit mutations commonly found in that cancer type.

Clinical trials have focused on alterations that are relatively frequently observed in the target population.

However, just testing for those particular biomarkers would not identify any of the many potential rare abnormalities. By extension, certain mutations are very common in certain cancer types only (for example, KRAS mutations in colorectal carcinoma), but rare in other cancers, and are therefore often bypassed in the testing process in favor of more common mutations in those cancers (for example, BRAF mutations in melanoma). The fact that they are less common, however, does not mean they are non-existent: KRAS mutations can be found in melanomas.

Until relatively recently, the only way to overcome those limitations was to increase the number of single-marker tests performed on individual cancer tissues: in melanoma for example, separate BRAF, KRAS, NRAS and PIK3CA mutation tests could be performed to detect less common aberrations and to gain more insight into the specific molecular characteristics of that patient's cancer. However, this approach requires more viable tumor tissue, significantly complicates the testing workflow, and multiplies the cost of testing.

Hence, using this approach beyond four or five different genes is usually not feasible in practice.

NGS has resulted in a way to approach this problem from a different angle through the above referenced gene panels which aim to sequence a large number of genes simultaneously. The output then provides a mutational status for each of those many genes in one go, greatly increasing the available information regarding that patient's tumor.

One of the major perceived drawbacks of pan-cancer testing panels – whether based on solid tumor samples or liquid biopsies – is that they sometimes deliver an unmanageable amount of information, which is not always actionable.

Identifying rare mutations in cancer types does provide additional insights but if no drugs are available to target those mutations, it leaves both oncologist and patient with more questions than answers.

The market for ctDNA NGS panels is already crowded, and in the USA the reimbursement landscape is not promising, especially for smaller panels (public health reimbursement would pay ~$600 for a 17 gene panel, vs. potentially ~$2900 for >50 gene panels).

Due to the lack of evidence to support therapy selection by ctDNA results alone, the clinical liquid biopsy market is likely to grow at a moderated pace, and tissue testing to remain the clinical gold standard when available as an option. However in the context of long term patient monitoring and care, it does propose to be a first line patient management opportunity.

Indeed the idea of such liquid biopsies has leapt to prominence in just the last year. The biomedical community now expect that the tests will offer a noninvasive way to monitor cancer (screen, not diagnose), find the genetic mutations driving a tumor before symptoms start. Wall Street analysts at JP Morgan expect demand for liquid biopsies to rocket toward $20 billion a year within five years, from about $100 million today.

However, as for the general requirements of a NGS laboratory, liquid biopsy handling also has its own needs: Genechron's prior experience in liquid biopsy handling should well prepare it for the market potential, providing it trains the medical practitioners how to handle liquid biopsies.

Most therapy decisions in the clinic are based on laboratory tests. This testing process can be divided into three phases: the pre-analytical phase, the analytical and the post-analytical phase. While the pre-analytical phase includes the identification and selection of an appropriate test, specimen collection and transport, the analytical phase comprises mainly the laboratory testing itself. The post-analytical phase consists of, for example, data analysis, interpretation of results and reporting, but also - if applicable - archiving of the remaining material. In the whole testing process, 46% to 68% of errors occur in the pre-analytical phase, which adversely influence the quality of the data in the following phases, leading to an increase of diagnostic costs and suboptimal decisions for the patient. The most common mistakes include the use of inappropriate blood collection tube (BCT), poor sample collection procedures (e.g. hemolysis or insufficient volume) or wrong sample storage and transportation but also inaccurate sorting, aliquoting or technical mistakes (e.g. pipetting or centrifugation). As a result, about 10% of patient deaths and 17% of adverse events are reported to be caused by such pre-analytic mistakes. Due to the high relevance of preanalytical sample handling, significant efforts will have to be made to standardize processing and analysis of blood samples for different technologies.

Regarding clinical utility of liquid biopsy as well as usefulness for research, it is very important to have easy-to-use, robust and reproducible workflows. Currently, there are no integrated, multicenter-tested workflows available covering the requirements for the clinical setting. For Genechron to enter the local primary care market, the

primary care practitioners need to be fully integrated into the workflow for the service to be successful.

Such workflows should include Standardized Operating Procedures (SOPs) for all above mentioned phases of laboratory testing starting with specimen collection and ending with result interpretation e.g. via bioinformatics analysis. Blood collection should be performed using certified BCTs, suitable for the respective downstream application (e.g. use of compatible stabilizer or fixative, volume adjusted). Next, fixed established protocols for specimen handling, storage and shipping of the blood sample or how to generate blood plasma (e.g. centrifugation) are indispensable. Depending on the analyte (CTCs, ctDNA, miRNA, etc.), robust extraction, isolation and quantification methods, are necessary. The implementation of an optimized preparation method for each analyte at the analytical sites has to be validated and documented for evaluation of the results.

Although many different liquid biopsy technologies appeared on the market in recent years, there is still a lack of technologies offering reproducible, robust, cost-effective and easy-to-use workflows from the sample to clinically meaningful data. To this day, there is only one FDA approved CTC quantification technology available for three metastatic tumor indications.

In general, the amount of circulating tumor DNA (ctDNA) increases with tumor burden sums up to 1% of total ccfDNA in early-stage disease and up to 40% in late-stage disease and ctDNA allows detection of a relapse relative early. ctDNA is a valuable biomarker, which is already used for treatment response monitoring or the early detection of relapse. Furthermore, the analysis of ctDNA from patients has an impact on therapy decision (e.g. mut EGFR). In addition to improving therapy selection, the analysis of ctDNA can be used to monitor the success of a given therapy.

NGS is already of great value for ctDNA analysis and liquid biopsy (including mRNA and miRNA sequencing). Recently, a 70-gene panel by Foundation Medicine, the FoundationACT® assay, was granted breakthrough device designation by the FDA (http://investors.foundationmedicine.com/news-releases/news-release-details/foundation-medicines-new-liquid-biopsy-assay-granted) , potentially making it the first liquid biopsy NGS panel to achieve regulatory approval.

10.4 Pharmacogenetics

Pharmacogenetics is the study of how genes affect a person's response to drugs. At present there are over 770,000 injuries or deaths due to drug reactions per year in the United States. In Europe, it has been estimated that approximately 5 % of all hospital admissions are caused by ADRs, that 5 % of hospitalized patients will experience an ADR during their hospital stay, and that ADRs cause 197,000 deaths annually throughout the EU. Pharmacogenetics provides the opportunity to lower this number.

NGS provides the opportunity to analyse a persons DNA and see which of their biochemical pathways may be impaired, because they have a genetic polymorphism. They then cross-reference this with all the drugs that require that pathway to break them down.
The insight is a chart that shows us which drugs the person is programmed to be able to break down in the normal fashion and which drugs she or he will not be able

to break down in the normal fashion.

It is a once-in-a-lifetime test; it does not need to be repeated because it's showing which drugs a persons genes can and cannot break down.
Insurance companies who pay for the drugs are starting to recognize these tests. They will often pay for the test, because they realize that if they can prevent the patient from having side effects that require hospitalization, they will save money.

Although the field of pharmacogenetics was established in the 1950s, clinical testing for constitutional pharmacogenetic variants implicated in inter-individual drug response variability has only recently become available to help clinicians guide pharmacotherapy, in part due to US Food and Drug Administration-mediated product insert revisions that include pharmacogenetic information for selected drugs.

Despite pharmacogenetic associations with adverse outcomes, physician uptake of clinical pharmacogenetic testing has been slow. This is due to a lower positive predictive value, which is one reason for underutilization, as well as clinical utility, professional education, and regulatory and reimbursement issues. Additionally, drug efficacy is not influenced solely by genetic variation in drug metabolism genes. Polymorphisms in genes that encode drug transporters and drug targets have also been shown to alter drug responses.

Personalized medicine programs have invested in clinical pharmacogenetics and view it as a logical first step toward incorporating genetics and genomics into more routine and individualized healthcare. However there is a relatively slow clinician uptake of available pharmacogenetic testing.

Insurance coverage for pharmacogenetic testing is currently sporadic, yet the healthcare reimbursement climate is constantly changing. Pharmacogenetic testing needs to demonstrate clinical utility and/or effectiveness before widespread adoption, but for payers, it should also return on its investment. Several reports have attempted to systematically evaluate and review pharmacoeconomic examples without much success.

In the USA, despite the availability of pharmacogenetic testing from regulatory-approved laboratories, physician uptake of clinical pharmacogenetics has been low, in part due to a perceived lack of clinical utility, inadequate professional guidelines for pharmacogenetic based management, and limited insurance reimbursement for testing.

Given that the test is arguably performed once, and costs today around €2000 it is the market exclusively of those with private health insurance. However, prices are decreasing, Illumina wants to do it for €100; nonetheless the largest barriers seem to be medical practitioner uptake, and in the context of one off total genome sequencing, which not only addresses pharmacogenetics but also disease, for many this may be a long term investment worth making.

11. Recommendations

Genomics technology markets are difficult to predict: there is always something new around the corner, and any purchase is a significant capital investment, with sequential high risk.

The clinical market is immature and regulations that apply to tertiary care locations for the use of these systems does not apply elsewhere: any other market provider is going to have to satisfy significant regulatory constraints.

However if the price of testing can be brought down, while the application of NGS increases, specially in pre-tertiary hospital care screening for multiple pathologies, having such a platform and being a Lazio first-mover could be lucrative. As such we think it would be feasible for Genechron to enter this field in Lazio.

Our suggestion would be for the Genechron Executive with their operational and technical staff to review this documentation and establish a “quality work flow” procedure based as closely as possible on the ESHG guidelines and the anticipated necessity to run such a laboratory to fully assess the ease with which samples could be handled and processed, and then wait for the Illumina Firefly system to be released.

A side by side comparison of its capabilities with close-to or equivalent priced devices, the potential panel screen kits that exist for cancer and other applications, to assess total local market possibility and look to start from there, aiming for €250 per screen per patient, providing liquid biopsies can be performed.

Following this, and exclusively for the private healthcare market in Italy, pharmacogenetic whole genome sequencing can be offered at cost. As stated above, a whole genome sequence not only informs pharmacogenetics but also disease management, which for private health insurers will be valuable. A final step would be the extension of the platform to other pathologies, however we would always advise against small companies entering into R&D biomarker fishing expeditions: they are high risk, cost a lot of money and need extensive and multicentric validation to be able to potentially commercialised.

We do not think it would be feasible for a small company to consider purchasing the more expensive or complex systems: in our opinion their shelf life is limited, and likely to be shorter than their amortisation time while the costs per patient will be much higher. While point-of-care small devices look attractive for NGS, their limited scalability could become a bottleneck in high volume markets.


GENECHRON