UNIVERSITÀ DEGLI STUDI DI MILANO – BICOCCA

Facoltà di Scienze Matematiche, Fisiche e Naturali
Dipartimento di Biotecnologie e Bioscienze

Dottorato di ricerca in Biotecnologie Industriali,
XXIII ciclo

Development of New Tools for Genomic Biomarker Investigations

Dott. Valeria Maria Zanda
Matr. 033868

Anno Accademico 2009/2010

Dottorato di Ricerca in Biotecnologie Industriali, XXIII ciclo

Dott. Valeria Maria Zanda
Matr. 033868

Tutor: Marco Vanoni

Il lavoro presentato in questa tesi è stato realizzato presso il dipartimento di Clinical Biomarkers - Biomarker Technologies, MerckSerono, Ivrea, sotto la supervisione del Dott. Francesco Gorreta.

Università degli Studi di Milano-Bicocca
Piazza dell'Ateneo Nuovo 1, 20126, Milano

Dipartimento di Biotecnologie e Bioscienze
P.za della Scienza 2, 20126, Milano

TABLE OF CONTENTS

1 ABSTRACT

During drug research and development, biomarkers are broadly used to improve the understanding of drug mechanism of action, to investigate drug efficacy and safety, to support the selection of target patient population and to optimize treatment schedule. Among different classes of biomarkers, genomic biomarkers are defined as a measurable DNA or RNA characteristics that are indicator of normal biologic processes, pathogenic processes, and/or response to therapeutic or other intervention.

A genomic biomarker can consist, for example, in one or more DNA characteristics such as single nucleotide polymorphisms (SNPs), insertions, deletions or RNA characteristics such as RNA expression levels, RNA processing (splicing and editing) and microRNA levels.

The present research aimed at developing new genomics-based tools using non-conventional biological samples that might support biomarker investigations in clinical settings.

Concerning DNA biomarkers, Single Nucleotide Polymorphisms (SNP) analysis was validated on DNA extracted from Formalin-Fixed Paraffin-Embedded (FFPE) tissue, from Hematoxylin and Eosin (H&E) stained FFPE slides, and from serum samples.

Unconventional samples represent a challenge for genetic analysis due to limitation of biological material and/or the poor quality of the DNA extracted. For example, due to fixation effect and formaldehyde interaction DNA extracted from FFPE samples are characterized by degradation, cross-link, limitation of material, methylol derivatives and PCR inhibitors presence. Therefore before analyzing these samples a method validation is necessary to prove data reliability in accordance with Regulatory Agencies guidelines that encourage the scientific community to perform a fit-for-purpose method validation to support any pharmacogenetic data submission.

SNPs were investigated by Real-Time PCR using TaqMan SNP genotyping assays. Polymorphisms in a panel of genes involved in EGFR pathway, which is directly associated with many type of cancer, were evaluated. In particular, each single assay was first validated for accuracy, intra-assay precision (repeatability under the same operating conditions) and ruggedness (reproducibility with different operators, different batches). These parameters were tested first on good quality DNA such as DNA extracted from cell lines and then on real sample to evaluate the non-conventional matrix.
On this purpose, the impact of fragmented DNA (FFPE samples) and H&E staining on FFPE samples was evaluated. DNA was extracted from different tissues of 10 commercial donors. DNA genotyping results of unstained FFPE and H&E staining were compared with the genotype obtained from high quality genomic DNA extracted from Fresh Frozen (FF) tissues obtained from the same donors (used as reference samples).
Overall, these results demonstrate that SNP genotyping can be performed on archived FFPE tissues providing reliable results.

As additional test serum was used as source of DNA to perform SNPs analyses. Serum is usually used to investigate protein biomarker and is generally collected in most of the clinical trials. It has been demonstrated indeed that free circulating DNA is present in serum: in particular, DNA is present in healthy individual at low concentration while levels are higher in cancer patients, in arthritis, hepatitis (Board et al., 2008; Gahana et al., 2008; Gormally et al., 2007).
To validate SNPs analysis on serum, two aliquots of whole blood were obtained from 35 healthy volunteers. For each subject one aliquot was used to extract good quality DNA, the other was used to prepare serum prior to DNA extraction.
As expected DNA quantity was very low for serum samples. As result, even though DNA was not degraded, genotype analysis was successful only on 70% of the samples.

Overall, the validation conducted showed that serum could be used as source of biological material to conduct genetic analyses. However limitation of DNA does not consent to perform a large panel of analysis. These could be further explored in patients since circulating DNA is present at higher levels in several diseases.

Part of the present thesis focused also on the validation of methods for *KRAS* mutation analysis. This gene encodes for a G-protein which plays a key role in the Ras/mitogen-activated protein kinase (MAPK) signaling pathway and located downstream Epidermal Growth Factor Receptor (EGFR) which is involved in colorectal cancer (CRC).
KRAS status can predict which patients benefit (KRAS wild-type) or do not benefit (KRAS mutated) from anti-EGFR therapy. Since KRAS analysis is also used for diagnostic analysis, an accurate validation of the method was required.
The aim was to compare and validate two different methods for *KRAS* mutation detection on FFPE tumor specimens, and on H&E stained FFPE which represent an unconventional source of samples for this type of analysis.
In particular, DxS ThreraScreen KRAS mutation kit, a Real-Time PCR assay, was compared to the PyroMark KRAS Kit, based on pyrosequencing technology.
The DxS ThreraScreen KRAS mutation test kit is able to detect 7 different mutations present in codons 12 and 13 of the KRAS gene while PyroMark KRAS Kit is able to detect 9 *KRAS* mutations in codon 12-13 and 5 mutations in codon 61.
Results from validation showed that both Pyrosequencing assay both DxS ThreraScreen assay are accurate and reproducible. Moreover no impact of degraded DNA obtained from FFPE or influence due to H&E staining was observed in both methods.
In conclusion PyroMark KRAS Kit showed advantages such as lower amount of DNA needed for analysis, detection of additional mutations in cod.12/13 and codon 61 than DxS TheraScreen KRAS kit; on the

other hand DxS TheraScreen KRAS resulted more sensitive than pyrosequencing assays and less time consuming.

This thesis focused also on establishing a simple method to perform gene expression investigation on hair follicles (HF) and to evaluate its applicability in clinical trials.
Despite 80% of solid cancers arising from epithelial tissues, blood is still one of the most common peripheral tissues used for biomarkers and pharmacogenomic investigations in oncology. Hair follicles may offer a viable alternative since they can reflect biological response in epithelial tissue, they are easy to collect (non-invasive) and available from most individuals.

After the establishment of sample collection and RNA extraction, HFs were collected from 23 health donors to evaluate inter-individual variability of RNA yield and quality.
Gene expression analysis was then conducted on the extracted RNA. First it was evaluated a panel of 16 housekeeping genes to assess the feasibility of the analysis. Then it was shown that in HF a panel of epithelial specific genes were expressed. Indeed, Realtime PCR analyses showed that *EGFR*, Keratin 19 (*KRT19*), Collagen, Melan-A were expressed in HFs but not in RNA derived from blood. On the opposite, *FPR1* and *PRF1* genes were expressed only in blood. These results suggest that HF represents a valuable biological source to study pathways active in epithelial tissue.
Finally, gene expression analysis was conducted on an *in vivo* experiment to evaluate if a response to treatment could be observed in HFs. In particular, PD markers of Interferon treatment were investigated after in vivo subministration of Interferon-beta (IFN-β) in *Macaca fascicularis*. The expression of the known IFN-β responding genes *MxA* was investigated both in blood and in anagen HFs. Results showed that *MxA* induction was observed both in blood and HF: gene induction in blood was observed at 6 hours after subministration while

in HF at 24 hours probably due to a different IFN-β distribution. These data suggest that gene expression analysis can be carried out in HF samples. However, it is important to highlight that in HF the response had a lower degree of induction and higher variability than in blood.
However this preliminary observations need to be further explored in pilot clinical studies to evaluate its applicability.

Overall the validation of different genomic analysis on unconventional sampled opens the possibility to conduct biomarker investigations on several clinical trials conducted in the past or to plan new investigations with non invasive methods.
In addition, from the deep evaluation of the current guidelines from the Regulatory Agencies (and from the open debate in the scientific community) a proper strategy to validate genomic analytical assays was proposed according to fit-for-purpose criteria.

2 RIASSUNTO

Durante il processo di ricerca e sviluppo di un farmaco i biomarkers risultano particolarmente vantaggiosi in quanto possono, ad esempio, migliorare la conoscenza del meccanismo di azione di un farmaco, delle sue caratteristiche di efficacia e sicurezza. Inoltre alcuni marcatori possono essere utilizzati per identificare i pazienti che rispondono meglio al trattamento.
Tra le differenti classi di biomarkers, i marcatori genomici vengono definiti come una caratteristica misurabile del DNA (ad esempio i polimorfismi a singolo nucleotide, inserzioni, delezioni) o dell'RNA (livelli di espressione, splicing o editing, livelli di microRNA) che fornisce indicazioni rispetto a un processo biologico e/o ad una risposta terapeutica.
Lo scopo di questo lavoro è stato quello di sviluppare nuovi metodi in ambito genomico per l'analisi di biomarkers su campioni biologici non convenzionali al fine di supportare studi clinici.

Tra i marcatori a livello del DNA, in questo studio, è stata condotta l'analisi di polimorfismi a singolo nucleotide (SNP) su genomi estratti da tessuti inclusi in paraffina e fissati in formalina (FFPE), slides colorate con eosina ed ematossilina (H&E) e da siero.
I campioni non convenzionali rappresentano una sfida per le analisi genetiche dato che il DNA ottenuto da queste fonti è limitato e di scarsa qualità. Ad esempio il DNA estratto da FFPE risulta degradato e contiene inibitori della PCR a causa del processo di fissazione e della presenza di formaldeide. Di conseguenza prima di analizzare questi campioni è necessario un approfondito metodo di validazione che dimostri l'attendibilità dei dati secondo le linee guida delle principali autorità regolatorie (FDA e EMA). Recentemente questi organismi hanno incoraggiato la comunità scientifica a validare i metodi analitici per supportare la sottomissione di dati riguardanti i biomarkers.

L’analisi sui polimorfismi è stata effettuata mediante Real-Time PCR utilizzando i saggi TaqMan. Per ciascun saggio è stata valutata l’accuratezza, la precisione (ripetibilità nelle stesse condizioni operative) e la robustezza (riproducibilità dell’analisi fra differenti operatori e lotti).
Questi parametri sono stati testati prima su DNA genomico di buona qualità derivante da linee cellulari e poi su campioni reali per valutarne la matrice non convenzionale. Infatti, sono stati valutati l’impatto del DNA frammentato ottenuto dai campioni FFPE e l’effetto della colorazione H&E sui campioni FFPE. Il DNA è stato estratto da diversi tessuti (disponibili commercialmente) di 10 donatori. Il genotipo ottenuto dai campioni FFPE non colorati e H&E FFPE è stato confrontato con quello ottenuto dai tessuti collezionati in azoto liquido (usati come controllo di qualità).
I risultati ottenuti dai differenti tessuti sono attendibili e mostrano che è possibile effettuare analisi di polimorfismi su tessuti inclusi in paraffina di studi clinici.

Un’altra fonte di materiale genomico non convenzionale utilizzata in questo progetto è costituita dal siero. Generalmente il siero viene collezionato nella maggior parte degli studi clinici per l’analisi di marcatori proteici ma non per analisi di genetica. Il DNA presente nel siero viene denominato DNA circolante ed è presente in quantità maggiori in soggetti affetti da tumore o altre patologie (artrite, epatite) rispetto a individui sani (Board et al., 2008; Gahana et al., 2008; Gormally et al., 2007).
Per effettuare la validazione di polimorfismi sul siero, due aliquote di sangue intero sono state collezionate da 35 volontari sani. Per ciascun donatore una aliquota è stata utilizzata per estrarre DNA di buona qualità, l’altra è servita per preparare il siero.
La quantità di DNA ottenuta dal siero è stata limitante, come atteso. Nonostante il DNA non risultasse degradato, è stato possibile assegnare il genotipo solo per il 70% dei campioni. Tuttavia il

genotipo dei campioni di siero assegnati è risultato accurato se confrontato con il corrispettivo ottenuto da sangue.
Le analisi condotte mostrano, quindi, che è possibile utilizzare il siero come fonte per analisi genetiche ma con alcuni accorgimenti dato che il materiale che si ottiene è minimo. Ulteriori analisi potrebbero essere effettuate su pazienti reali in cui la quantità di DNA circolante è maggiore.

Un'altra parte di questo progetto è stata dedicata alla validazione di metodi per l'analisi mutazionale del gene *KRAS*. Questo gene codifica per una proteina ad attività GTPasica che gioca un ruolo chiave nella via di trasduzione del segnale delle MAP chinasi ed è situata a valle del recettore epiteliale di crescita (EGFR) coinvolto nel cancro del colon retto (CRC). Lo stato di *KRAS* può essere di aiuto nella scelta della terapia in pazienti affetti da CRC. Infatti se KRAS risulta wild-type i pazienti potranno ricevere benefici da un trattamento anti-EGFR; nel caso in cui KRAS risulti mutato la terapia anti-EGFR non avrà alcun effetto poiché la via di trasduzione del segnale risulterà sempre attivata.
Lo scopo di questo lavoro è stato confrontare e validare due metodi diversi per l'analisi mutazionale di *KRAS* su DNA ottenuto da tessuti tumorali (FFPE) e slide colorate (H&E FFPE) che rappresentano la fonte non convenzionale di DNA.
E' stato utilizzato un metodo basato sulla tecnologia Real-Time PCR (DxS TheraScreen KRAS mutation kit) e confrontato con un metodo basato sulla tecnologia di pirosequenziamento (PyroMark KRAS kit). Il kit DxS TheraScreen KRAS è in grado di riconoscere 7 differenti mutazioni presenti nei cod. 12 e 13 del gene mentre il kit PyroMark KRAS riconosce 9 mutazioni in questi codoni e in aggiunta 5 mutazioni sul cod. 61. I risultati dimostrano che entrambi i saggi sono accurati e riproducibili; non è stato rilevato alcun effetto dovuto al DNA degradato (FFPE) o alla colorazione dei tessuti (H&E).

In conclusione il kit PyroMark KRAS è vantaggioso in quanto richiede una quantità di DNA minima e consente di analizzare più mutazioni (nei codoni 12-13 e 61) rispetto al kit DxS TheraScreen KRAS. Dall'altra parte quest'ultimo risulta più sensibile e veloce.
Con questa validazione sono state gettate le basi per una valutazioni di un largo pannello di mutazioni utili per gli studi clinici. L'applicazione dei marcatori genetici in campo diagnostico, prognostico e nella scelta del trattamento farmacologico possono essere utili per muovere un ulteriore passo verso una medicina personalizzata migliorando l'outcome dei pazienti.

Un'altra parte della tesi si è focalizzata sulla messa a punto di un metodo semplice per effettuare l'espressione genica sui follicoli dei capelli e valutarne l'applicabilità negli studi clinici. Infatti i follicoli potrebbero essere una fonte vantaggiosa di studio in quanto riflettono la risposta dei tessuti epiteliali che sono all'origine di circa l'80% dei tumori solidi. Inoltre i follicoli sono facilmente collezionabili (e in modo non invasivo) e sono disponibili nella maggior parte degli individui.
Dopo aver stabilito un metodo per il collezionamento e l'estrazione dell'RNA si è proceduto con l'analisi di follicoli su 23 donatori sani per valutare la resa e la qualità dell'RNA tra i vari individui.
Le analisi di espressione genica sono state effettuate sull'RNA estratto. Prima è stato valutato un pannello di 16 geni di controllo (housekeeping genes) in modo da valutare la fattibilità dell'analisi; poi è stato dimostrato che nei follicoli sono espressi geni epiteliali specifici. Infatti le analisi di PCR Real-Time hanno mostrato che i geni *EGFR*, *KRT19*, *COL1A* e *MLANA* sono espressi nei follicoli ma non nel sangue. Al contrario, i geni *FPR1* e *PRF1* sono espressi nel sangue ma non nei follicoli. Da questi dati si evince che i follicoli rappresentano una fonte biologica adatta per studiare pathway presenti nei tessuti epiteliali.

Infine, l'espressione genica è stata verificata in un esperimento *in vivo* per valutare se il follicolo può essere utilizzato per monitorare un trattamento farmacologico.
A questo scopo alcuni marcatori farmacodinamici dell'interferone sono stati analizzati dopo la somministrazione di IFN-β in scimmie del tipo *Macaca fascicularis*. L'espressione del gene *MxA*, noto per essere un gene di risposta dell'interferone, è stata verificata nel sangue e nei follicoli. Dai risultati emerge che l'induzione di *MxA* avviene in entrambi i tessuti ma a tempi diversi: per i follicoli il picco massimo registrato è visibile dopo 24 ore dalla somministrazione mentre per il sangue dopo 6 ore. La causa è da ricercarsi probabilmente in una differente distribuzione nei tessuti dell'interferone. I dati suggeriscono comunque che analisi di espressione genica possono essere condotte in questo tipo di tessuto anche se è importante sottolineare che il livello di risposta nei follicoli è più basso e più variabile rispetto al sangue.
I follicoli, quindi, potrebbero essere utilizzati come promettente fonte non invasiva per studiare i meccanismi molecolari dei tessuti epiteliali. Tuttavia le informazioni preliminari ottenute necessitano di ulteriore approfondimento nel caso che il metodo debba essere applicato a studi clinici.

La validazione di diversi metodi per marcatori genetici su campioni non convenzionali di questo lavoro apre la possibilità di condurre analisi di marcatori su diversi studi clinici condotti in passato o di pianificarne di nuovi attraverso la raccolta di materiale biologico non invasivo.
Inoltre dalla valutazione delle attuali linee guida delle autorità regolatorie e dall'analisi del dibattito in corso all'interno della comunità scientifica si è proposta una strategia per la validazione di metodi analitici per marcatori genomici.

3 INTRODUCTION

3.1 Biomarkers

According to a consensus definition a **biomarker** is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Biomarkers have many valuable applications in disease detection and monitoring of health status. These applications include use as a diagnostic tool for the identification of those patients with a disease or abnormal condition (i.e., elevated blood glucose concentration for the diagnosis of diabetes mellitus), as an indicator prognosis or staging of disease, for prediction and monitoring of clinical response to an intervention (i.e., blood cholesterol concentrations for determination of the risk of heart disease).

Two others concepts that are related to biomarker are the following (Biomarkers Definitions Working Group, 2001):

- **Clinical endpoint** is a characteristic or variable that reflects how a patient feels, functions, or survives. Clinical endpoints are distinct measurements or analyses of disease characteristics observed in a study or a clinical trial that reflect the effect of a therapeutic intervention. Clinical endpoints are the most credible characteristics used in the assessment of the benefits and risks of a therapeutic intervention in randomized clinical trials.
- **Surrogate endpoint** is a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence. Surrogate endpoints are a subset of biomarkers. Although all surrogate endpoints can be considered

biomarkers, it is likely that only a few biomarkers will achieve surrogate endpoint status.

Biomarkers play an important role in drug discovery and development from target identification and validation to clinical application in order to make the overall process a more rational approach.
Biomarkers could be useful in decisions regarding optimal dose selection, benefit–risk assessment, and regulatory approvals related to new formulations, populations, and/or indications for drugs with demonstrated effectiveness using clinical end points. In fact, the general principles of using biomarkers for regulatory decisions are described clearly in several regulatory guidances and publications. Moreover, currently many efforts have been done from pharmaceutical companies to discover and develop biomarkers in the various phases of drug development. Biomarker develop process includes discovery, qualification, verification, assay optimization, clinical validation and at the end commercialization.

Among biomarkers that have been used extensively over the past decades there are for example an elevated level of urine human chorionic gonadotropin (hCG) as a measure of normal pregnancy, a disproportional body weight as a primary determinant of obesity, and increased blood pressure and cholesterol levels as parameters of cardiovascular risks. Other examples of biomarkers are reported in Table 1 and include bone density in osteoporosis and viral load in infectious diseases.

Table 1. Example of biomarkers and their related clinical endpoints.

Therapeutic class	Biomarkers	Clinical endpoint
Antihypertensive drugs	↓ Blood pressure	↓ Stroke
Drugs for glaucoma	↓ Intraocular pressure	Preservation of vision
Drugs for osteoporosis	↑ Bone density	↓ Fracture rate
Antiarhythmic drug	↓ Arrhythmias	↑ Survival
Antiretroviral drugs	↓ Viral RNA ↑CD4	↑ Survival
Antidiabetic drugs	↓ Blood glucose	↓ Morbidity
Lipid-lowering drugs	↓ Cholesterol	↓ Coronary disease
Drugs for prostate cancer	↓ PSA	Tumor response

Advances in technology have allowed the incorporation of measurements of many biomarkers in patient care without increased risk to patients. In fact, almost all types of cancer are nowadays treated with testing biomarkers. Some biomarkers can discriminate normal from diseased states (i.e. diagnostic biomarkers), while others predict the likely course of disease progression (i.e. prognostic biomarkers) or response to therapy (i.e. stratification biomarkers) or drug effect (pharmacodynamic or PD biomarkers).

A more exhaustive overview of biomarker characterization is reported in Table 2 (from Hu et al., 2005).

Table 2. Some categories of biomarkers.

Type	Characteristics	Sources of Biomarkers	Key Validation Requirements
Diagnostic Biomarkers	Discriminate normal from a specific disease	Expression analysis Disease mechanism studies Epidemiology studies	Correlation with a specific disease or disease state
Prognostic Biomarkers	Predict the likely course of disease progression independent of a specific therapy	Expression analysis Animal models Epidemiology studies	Correlation with a clinical outcome
Stratification Biomarkers	Identify patients that respond to a specific drug or suffer from its side-effects prior to administration of the drug	Pre-clinical studies Clinical trials Epidemiology studies	Correlation with a clinical response upon administration of the drug in controlled trials
PD/PK Biomarkers	Correlate response to a specific drug with concentrations of the drug or its metabolites	Drugs or metabolites Animal models	Correlation with the concentration or activity of a specific drug in animal and human studies
Efficacy Biomarkers	Monitor the beneficial effects of a specific drug on the intended drug target or medical condition	Molecular targets or downstream molecules Clinical trials	Correlation with the concentration or activity of a specific drug in clinical trials with placebo controls
Toxicity Biomarkers	Monitor the adverse effects of a specific drug on any unintended cellular processes, cells, tissues or organs	Histopathology Clinical chemistry Toxicology studies Clinical trials	Correlation with the concentration or activity of a specific drug in clinical trials

The ideal clinical biomarker should include:

- **Clinical relevance** meaning that the biomarker should have a rational basis for drug (influenced by exposure to a drug and is assumed to be related to the drug pharmacologic action or intended clinical effect).
- **Sensitivity** defined as the ability to detect the intended measurement and that change of biomarker reflects the change of clinical endpoint.

- **Specificity** meaning a biomarker which is able to identify person with a target disease or responding to certain therapeutics intervention.
- **Reliability** that refers to the ability to measure analytically the biomarker or change in biomarker with acceptable accuracy, precision, robustness, and reproducibility. These parameters are used for biomarker measurement and in particular they refer to the quality and variability of the assay.
- **Practicality** which means a non-invasive procedure or at least less invasive as possible for healthy volunteers or patients.
- **Simplicity** in order to avoid sophisticated equipment or operator skills and decrease time and cost. This is an important feature for applying biomarkers in drug development and even more for clinical practice.

3.1.1 Genomic Biomarkers

Genomic biomarkers (PGBM) are defined as a measurable DNA or RNA characteristics that are indicator of normal biologic processes, pathogenic processes, and/or response to therapeutic or other intervention (EMA - ICH Topic E15, 2007). A genomic biomarker could, for example, reflect the expression, function and regulation of a gene, which can be used as predictors of disease progression or response to treatment at the molecular level. A genomic biomarker can consist of one or more DNA and/or RNA characteristics.

DNA characteristics include for example single nucleotide polymorphisms (SNPs), variability of short sequence repeats, haplotypes, DNA modifications (methylation), deletions or insertions of one or more nucleotides, copy number variations , Cytogenetic rearrangements such as translocations, duplications, deletions, or inversions.

RNA characteristics include RNA sequences, RNA expression levels, RNA processing, such as splicing and editing, microRNA levels.

The definition of a genomic biomarker is not limited to human samples, but includes samples from viruses and infectious agents as well as animal samples. For example genomic biomarkers could be used also for non clinical and/or toxicological studies.
Genomic biomarker definition does not include the measurement and characterization of proteins or low molecular weight metabolites that are features of proteomics and metabolomics.
Genomic biomarker is the main features of pharmacogenomics (PGx). Indeed, pharmacogenomics is defined as the study of variations of DNA and RNA characteristics as related to drug response while pharmacogenetics (PGt) is a subset of pharmacogenomics and is defined as the study of variations in DNA sequence as related to drug response. This latest includes the processes of drug absorption and disposition (pharmacokinetics, PK), and drug effects (pharmacodynamics, PD; drug efficacy, and adverse effects of drugs).
Pharmacogenomics and pharmacogenetics have the potential to improve the discovery, development, and use of drugs. Each of the International Conference on Harmonisation (ICH) regions has published specific pharmacogenomic and pharmacogenetic guidelines, or concept papers, and is in the process of developing others.
In the following sections are reported some examples of genomic biomarkers and their clinical application.

3.1.1.1 SNP

Single nucleotide polymorphisms (SNPs) are the simplest and the most abundant type of sequence variation in the human genome. They occur throughout DNA with a frequency of about 1 in 1,000 bp in coding or noncoding regions (Shastry, 2007). Some polymorphisms have no effect, others are located in coding regions or in regulatory regions and might lead to several biological consequences. For example, SNPs might influence gene expression, protein stability and function and therefore they may be directly involved in different

pathologies such as diabetes, inflammatory diseases, cancer and coronary heart diseases (Anderson et al., 2006; Shastry, 2009; Zhernakova et al., 2009). The interest in identifying new SNPs and their role in disease progression or in drug response has increased over recent years. This investigation was also facilitated by advanced technology, indeed many genome-wide association studies identified an increasing number of loci that might be used as genetic markers (Dutt et al., 2007; Huang et al., 2009; Kirk et al., 2002; LaFramboise et al., 2009; McCarthy et al., 2008).

Clinical pharmacogenetic investigations study the correlation between variation in DNA sequences and drug response with the aim of developing a personalized medicine that allows treatment efficacy to be improved according to one's personal genotype. For example, polymorphisms involved in drug metabolism, distribution, absorption and excretion, such as cytochrome CYP2C9 and CYP2C19 variants or vitamin K epoxide reductase (VKORC1) isoforms (Kirchheiner et al. 2005) have been largely studied.

In clinical trials SNPs investigation may lead to population stratification able to identify subgroups which respond better to drug treatment or present adverse drug reaction.

In the present study the following panel of SNPs was employed:

- Polymorphisms on the Fc fragment of IgG, low affinity IIa (*FCGR2A*) and IIIa (*FCGR3A*) genes, that are members of a family of immunoglobulin Fc receptor found on the surface of many immune response cells. The proteins encoded by these genes are a cell surface receptor found on phagocytic cells such as macrophages and neutrophils. They are involved in antibody-dependent cellular cytotoxicity (ADCC) that might be linked to response after monoclonal antibody treatment (Alizadeh et al., 2007; Bibeau et al., 2009; Bruhns et al., 2009; Cañete et al., 2009).

- Polymorphisms of genes involved in the epidermal growth factor (*EGF*) pathway. In the EGF pathway, variations in ligands and receptor might influence the response to monoclonal antibody therapy. For instance, a polymorphism in *EGFR* in codon 421 results in an amino acid substitution of an arginine (R) with a lysine (K) and could make the cell even more sensitive to targeted receptor inhibition through cetuximab as the polymorphism is located next to the interaction site between EGFR and the monoclonal antybody (Saridaki et al., 2010). Another example is a polymorphism in the natural ligand EGF. Indeed, *EGF* 61 A/G gene polymorphisms have been analyzed in colorectal cancer (CRC) and other cancers but the functional effects of these polymorphisms has to be still completely elucidated because overall survival (OS) and progression free survival (PSF) in some cases are not concordant (Garm Spindler et al., 2009; Wu et al., 2009).
- Polymorphism of cyclin *D1* gene that was reported to be implicated in susceptibility to and early onset of colorectal cancers (Hong et al., 2005). Cyclin *D1* is a key cell cycle regulator that is upregulated by the b-catenin/Tcf pathway in colorectal tumourigenesis and its proliferation is associated with a poor clinical outcome. As a consequence the gene could be used as predictive molecular marker of cetuximab therapy (Saridaki et al., 2010).

3.1.1.2 Mutation

A mutation is any change in the sequence of DNA encoding for a gene.

Mutation effects are different according to the position in which the mutation occurs and if the protein function is altered. In most of cases the mutation is has damaging effect but in others could be advantageous or at least neutral meaning that it does not affect the amino acid sequence of the protein and no appreciable changes in organism phenotype are detected. A subgroup of neutral mutation is silence mutation that codify for the same amino acid even if the sequence is different.

A mutation is defined as somatic mutation when it does not occur in cells that give rise to gametes and it is not passed along to the next generation by sexual means. On the contrary if the mutation occurs in the germ line cells, the mutation will be passed on to the next generation and it will be called germinal mutation.

Furthermore mutation can occur spontaneously for example when a cell divides, it makes a copy of its DNA and sometimes the copy is not quite perfect; in other case mutation could be induced by different factors such as radiations, chemical agents which might be similar to bases or that are able to modify them.

There are many different ways by which changes can occur in DNA and as a consequence there are many classes of mutations. For instance, substitution when a base is exchanged with another, deletion when a sequence of DNA is lost or deleted and as opposite insertion when extra bases are inserted in the sequence. Others types of mutation include frameshift, caused by base insertion or deletion and altering the mRNA message and base inversion.

An example of somatic mutation used as biomarker is the *KRAS* gene, which is reported by Food and Drug Administration to be a Valid Genomic Biomarkers in the Context of Approved Drug Labels.

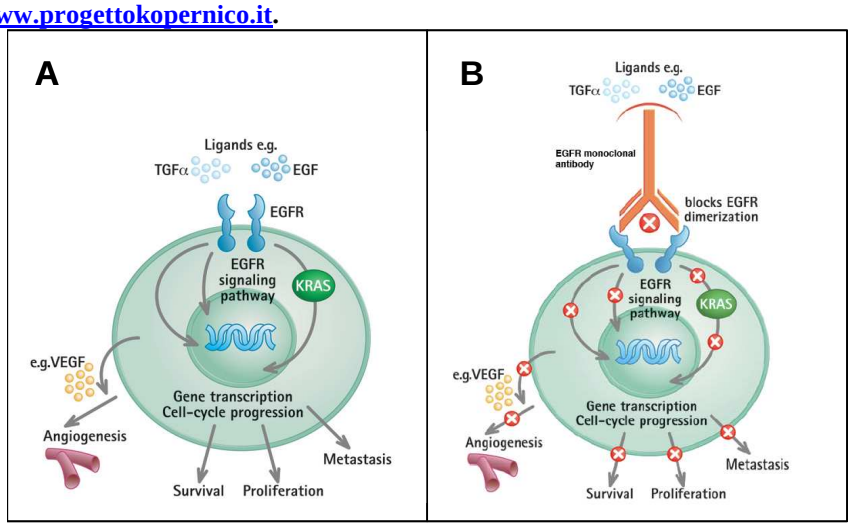
KRAS gene, a Kirsten ras oncogene homolog from the mammalian ras gene family, encodes a protein that is a member of the small GTPase superfamily involved in the Epidermal Growth Factor Receptor (EGFR) signaling pathway. Gene is located on chromosome 12 with a total of 6 exons (Van Krieken et al., 2008). A single amino acid substitution is responsible for an activating mutation. The transforming protein is implicated in various malignancies, including colorectal carcinoma, the most important, lung adenocarcinoma, mucinous adenoma, ductal carcinoma of the pancreas (Roberts et al., 2010).

KRAS mutation status is a useful and important biomarker especially in Colorectal Cancer (CRC) treatment decision (Silvestris et al., 2009). Indeed, mutation of the *KRAS* gene yields to a constitutive activated protein.

EGFR forms a heterodimer with other receptors from HER family and becomes able to receive signals from growth factor. This leads to EGFR activation and as a consequence to auto-phosphorylation of the heterodimer and transcription in the nucleus. In particular, in the cell nucleus specific genes are up or down regulated leading to cell response like proliferation, survival or invasion and metastasis.

Monoclonal antibodies against EGFR block signaling by the receptor and inhibit downstream events, including effects mediated by KRAS (Figure 1). When KRAS is mutated and permanently switched on, blocking EGFR will not prevent events downstream of KRAS and in this case, anti-EGFR therapy will not prevent the tumor from growing and spreading. Therefore, blocking EGFR with a monoclonal antibody will only be effective in KRAS wild-type tumors (Bardelli et al., 2010; Lièvre et al., 2008).

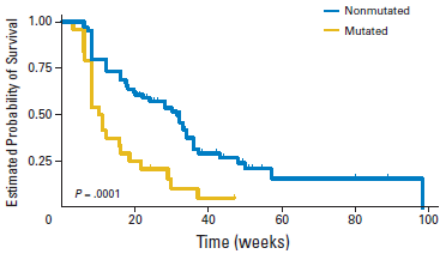
Figure 1. Schematic representation of EGFR pathway (A) and action of monoclonal antibody against EGFR receptor (B). Figure taken from www.progettokopernico.it.



The mutation in *KRAS* gene was present in over that 30% colorectal cancer and the most frequent mutation is in codon 12 and 13 located in exon 2 of the gene as reported in the Catalogue of Somatic Mutations In Cancer (COSMIC: www.sanger.ac.uk/genetics/CGP/cosmic/). Others minor proportion mutations are in codon 61 which clinical relevance are still unclear.

In 2008 the American Society of Clinical Oncology (ASCO) announced that CRC patients that have a mutation in the *KRAS* gene do not respond to therapies that inhibit EGFR; as Cetuximab and Panitumumab (Allegra et al., 2009). However many Phase III studies demonstrated that patients treated with anti-EGFR drugs show a significant survival advantage, but only if the tumor has a non-mutated *KRAS* gene. *KRAS* mutations are associated with poor responses to therapy, reduced PFS and shorter OS in colorectal cancer patients treated with cetuximab alone (De Hertogh et al., 2010; De Roock et al., 2007,) or in combination with chemotherapy (Figure 2).

Figure 2. Example of Progression-free survival of patients according to the presence or absence of *KRAS* mutation (from Lièvre, et al., 2008).



Following recommendations by ASCO, patients are today tested for the *KRAS* gene mutation, in order to determine which kind of therapy is most effective for their treatment.

KRAS testing mutations comprises referral for *KRAS* mutation testing, selection of the tissue block containing the tumor area of interest and finally DNA extraction and *KRAS* mutation analysis.

However, there is a lack of validated testing methods and standardized operating procedures for the detection of *KRAS* mutations. The concordance between different diagnostic methods is also largely unknown. Therefore, there is an urgent need to establish and implement clinical practice guidelines and standardized procedures for *KRAS* mutation testing in patients with colorectal cancer.

3.1.1.3 Gene expression

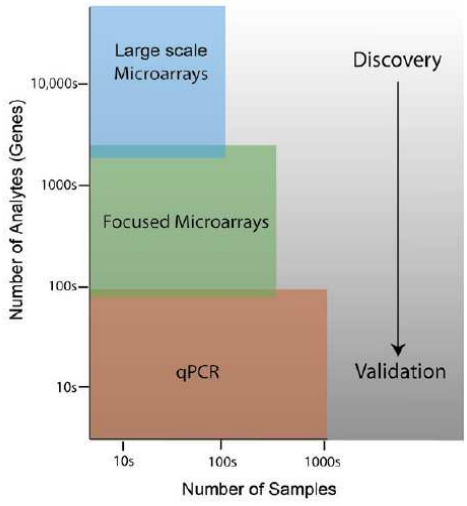
Cells are able to regulate gene expression according to different *stimuli* that they received or in response to various experimental conditions. Indeed, some genes are only expressed in particular cell types or cell cycle phase but their expression could change at developmental stage or in response to extra cellular signals. Gene expression can be regulated both at the transcriptional and post-transcriptional levels. Gene expression is an example of PD/PK biomarkers. Indeed, there are examples of gene expression studies that include validation of the extent of transcription of a gene in order to study the difference in expression of a gene in the diseased state

compared to the normal state; change in expression for cells that are exposed to a chemical substance (e.g., drug, toxin, hormone or cytokine); quantification of non-coding RNA gene expression.

Quantifying gene expression levels has become an important analysis in molecular laboratories because it allows measuring the amount of cellular RNA and determining what extent that particular gene is being expressed (VanGuilder et al., 2008).

Depending on the number of genes and samples being examined there is a multitiered technological approach (Figure 3). For large-scale discovery experiments the common method used is gene expression microarrays (Lesko et al., 2003), however in some case some specific studies focus on microarrays to study few genes in a large number of samples. Due to the costs of whole-genome micorarrays, RT-qPCR is used to validate discovery obtained from microarrays or in that case in which gene to investigate is already known.

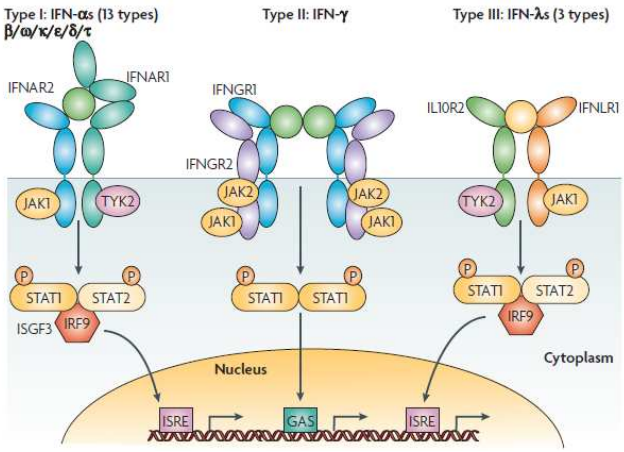
Figure 3. Types of technologies for gene expression analyses.
The plot shows the relation between number of genes and number of sample, the technologies go from being discovery tools to tools for validation and implementation (from VanGuilder et al., 2008).



Among genes studied for their variation of expression there are Interferon (IFN) responding genes which are activated by transcellular signal transduction (Der et al., 1998; Rani et al., 2007).

Briefly, IFN-β is part of IFNs family, releases from lymphocytes and belongs to Type I together with IFN-α and IFN-γ (Borden et al., 2007). IFNs genes are normally transcriptionally silent and IFNs is produced only in presence of pathogens or in response of tumor cells (Williams, 1991). In these situations molecules are induced by TLR-activated transcription factors that bind to their promoter. IFN- β binds to specific cell surface complex including IFN- α receptor known as IFNAR1 and IFNAR2. This trimeric high affinity complex initiates a signaling cascade that results in the activation of the IFN-stimulated gene factor 3 (ISGF3) heterotrimeric complex, which is composed of STAT1, STAT2, and IFN regulatory factor 9 (IRF9). Active ISGF3 translocates to the nucleus, binds to specific sequences (termed IFN-stimulated response elements or ISRE) present in the promoter of interferon regulated genes (O'Doherty et al., 2007), and induces transcription of target genes (Figure 4).

Figure 4. Schematic representations of the three classes of Interferon.
Different signal transduction of IFN classes: the binding complex between ligand and receptor changes according to Type I, II and II (from Borden et al., 2007).



Interferon-Beta (IFN-β) is used as the most common treatment for Multiple Sclerosis (MS) but its effects are complex and its pharmacodynamics at the genomic level in humans is poorly understood (Weinstock-Guttman et al., 2008). As MS is a chronic heterogeneous inflammatory disease of the central nervous system and in rate of progression, patient stratification is difficult. However it might be possible only with defined group of biomarkers so that relative contribution of each of different pathogenetic mechanisms can be determined (Habek et al., 2010; Scagnolari et al., 2007).

There are many studies that investigate gene expression in patient affected by MS versus healthy volunteers to identify new biomarkers (Booth et al., 2005; Brynedal et al., 2010; Riveros et al., 2010). What emerged from these analyses is that Inteferon regulates a high number of genes; more than 300 genes mediate the biological and therapeutic effects of IFN stimulation (Serana et al., 2008). Examples are 2',5'-oligoadenylate synthetases which inhibit a broad range of RNA viruses (corresponding to *OAS1* and *OAS2* genes) and the antiviral protein Mx that was identified because of the resistance of mouse strain A2G to influenza A viruses. The human homologous gene is *MxA* which codifies for a cytoplasmatic protein associated with the intracellular membranes (Pachner et al., 2003).

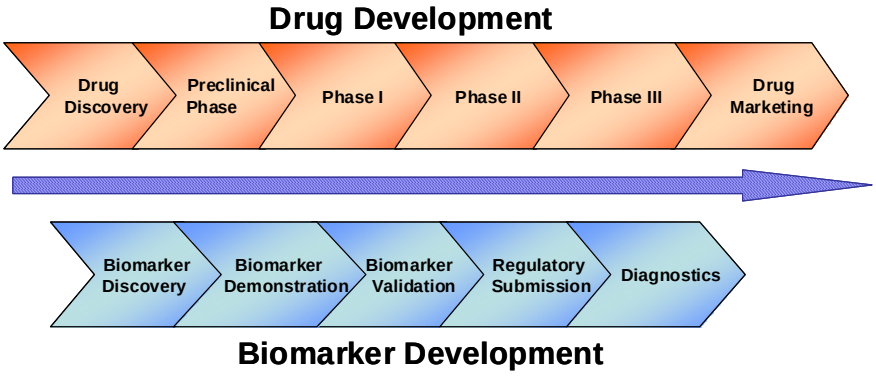
3.2 Biomarkers in drug development

Biomarkers are widely used in drug development, contributing to increased productivity and improved patient care across a spectrum of drug development activities. The increase in biomarker research is due to the widespread belief that appropriate application of biomarkers to preclinical and clinical drug development will accelerate the process (i.e., speeding time to market), increase efficiency (by providing early indications of efficacy or toxicity), and facilitate dose selection prior to expensive phase III clinical trials (Lee et al., 2005).

Given the interest in biomarkers in the pharmaceutical and diagnostics companies, also regulatory and government agencies has begun to focus on them. Indeed, biomarkers are included in the National Institute of Health Road Map which aims to facilitate a more efficient and productive system of medical research (Curry, 2008) and the FDA has published different guidelines about generation and submission of genomic data obtained from trials investigating new biomarkers (FDA, CDRH - Pharmacogenomic Data Submissions - Companion Guidance. 2007; Pharmacogenetic Tests and Genetic Tests for Heritable Markers. 2007).

Regulatory agencies encourage development of biomarkers since the first phases of drug discovery. In particular in PGBM/assay the co-development with a drug should be seen as a continuous process that goes through analytical validity of the PGBM assay at early stage of the drug development, clinical validity studies of PGBM (to ensure that the PGBM assay is able to select/stratify patients) and ultimately clinical usefulness. This latest refers to the ability of the PGBM test to ensure that the patient selected will have an improved benefit-risk profile when treated with the drug after PGBM testing in addition to conventional clinical features (EMA, CHMP. Reflection paper on co-development of pharmacogenomic biomarkers and Assays in the context of drug development. 2010).

Figure 5. Parallel paths of drug development and biomarker development.



3.2.1 Drug Discovery and Preclinical Development

It is important that the development of the PGBM assay is initiated early in the drug development process to be able to bridge data obtained later during clinical PGBM qualification. Biomarkers also play an important role in the preclinical assessment of potentially beneficial and harmful effects of a new drug candidate. Moreover new approaches for target discovery such as epidemiologic studies and microarrays could help to identify links between biomarker and pathophysiology. In particular, screening tests in animals using biomarkers provide important demonstration that a compound is likely to have the intended therapeutic activity in patients. This is the first step of a sequential process to establish a biomarker as surrogate endpoint to predict clinical outcome.

For example, HIV-1 RNA levels are measured in plasma as prognostic indicators to help guide therapy directed at suppressing viral replication and preventing opportunistic infections (Lathia et al., 2009; Lesko et al., 2001). Another example of biomarker was HER2/neu protein that was initially identified as an oncogene by *in vitro* studies and then its overexpression or gene amplification was linked to lymph node involvement by correlation with clinical studies. Her2/neu is now one of the most important targets in human breast cancer and is indicative of poor patient prognosis. (Harris et al., 2007; Hu et al., 2005).

At this step not only the biomarkers identification is important but it is also important to develop a prototype assay to monitor the biomarker. At this stage, however, the prototype should have the essential characteristics permitting its evaluation: acceptable methods/assays, tissue and disease specificities. Parameters associated with sensitivity and specificity of the biomarker shall be proportionate to supporting the rationale for the non-clinical and clinical applications of the biomarker (BM), and according to the development phase precision,

accuracy, reproducibility, detection levels justified for the intended use. The prototype of the assay could also rely on simplified procedures, it could be intuitive and easy-to use, with dynamic visualization, lowered consumption of samples, reagents and energy - to keep up with standard analytical capabilities. Therefore the assay should have a flexible design environment for intuitive layout (Colburn et al., 2003).

Moreover, for future clinical utilization of the platform chosen, appropriate plans are also required at the stage of prototype assay development for an efficient control and the ability to generate sufficient and appropriate data to validate the assay, to facilitate incorporation of quality assurance and control measures and, as appropriate, to meet the legal requirements for IVDs.

3.2.2 Early phase Clinical Development

During this phase of drug development, the biomarker assay is used to confirm that the results from *in vitro* and animal studies (or the initial observation about the behavior of the PGBM for example possible safety marker identified n preclinical data) are relevant to the human *in vivo* context.

The pre-requisite of a translational BM is therefore that the same entity is identified and measured both in animal and man and that the functionality associated to the BM is consistent in the two species. Therefore the PGBM assay needs to be consistent enough to support the extrapolation of PGBM effects observed in animal studies in humans, among a complex set of variables and situations.

It should be demonstrated that the assay identifies the same BM as in animal studies; the analytical performance characteristics of the assay, such as the limit of detection appropriate for the intended use, accuracy, and repeatability need to be confirmed at this stage. In some situations the assay of animal studies has to be adapted to the human situation to cope with respective differences as nucleic acid sequence

differences of the gene of interest. Moreover basic assay performances generated has to be compared between clinical and preclinical specimens (EMA, CHMP. Reflection paper on co-development of pharmacogenomic biomarkers and Assays in the context of drug development. 2010).

In phase I clinical trial new drugs investigation is performed on a restricted number of patient or healthy volunteers to evaluate its tolerance at different doses and to define pharmacological effect. In the following step which is phase II of clinical studies, the number of patients involved is no more than several hundred subjects.
In the past the biomarkers used during these steps were safety parameters, rate of absorption and excretion. These safety and PK parameters establish the duration of exposure to the drug and serve as a guide for safe dose selection and escalation. However the recent trend is to develop targeted therapies that highlighted biomarker therapeutic relevance (Marrer et al., 2010). For example, members of the human epidermal growth factor receptor (EGFR) family of transmembrane tyrosine kinases has been discovered to play important roles in a series of new targets and in different solid tumors such as breast cancer (Abd El-Rehim et al., 2004) and non-small cell lung cancer (Socinski, 2010). Biomarkers of potential use for stratification of patients include the target receptor or enzyme for drug intervention and modalities that can confer resistance to the drug. However at this exploratory level there are a list of possible candidate biomarkers obtained from safety data and patients monitoring drug response.
Regulatory agencies encourage clinical endpoints and surrogate endpoints at this early clinical stages. However, this is not always possible because biomarkers that show a correlation with drug could be used for facilitate future large-scale clinical trials and even in failed clinical trials biomarker testing might provide an explanation for failure and help future drug design.

To summarize, clinical PGBM qualification includes stability of the PGBMs over time and in relation to different phases/status of the clinical phenotype, biological rationale and preliminary data to especially differentiate (disease) prognostic PGBM from predictive PGBM (drug response), performance of the test and PGBM in predicting response/outcome. All these aspects of PGBM use should be GCP-compliant especially if the PGBM is intended to be used in a clinical context for the purpose of patients selection as diagnostic, prognostic, predictor of dose-exposure, predictor of drug response (safety, efficacy).

3.2.3 Late Clinical Development

Phase III trials are required to collect the additional information about effectiveness and safety that is needed to evaluate the overall benefit-risk relationship of the drug for specific and expanded indications and provide an adequate basis for extrapolating the results to the general population.

At this stage, incorporation of validated surrogate biomarkers for the stratification and monitoring of patients might be useful to cost-savings in drug development (Hu et al., 2005). At this stage, biomarker is tested on large cohort studies, therefore the specific assay need to be robust. Indeed, it should be ensured that the candidate companion diagnostic used in clinical trials is suitable for clinical validation and clinical usefulness (e.g. determination of the cut-off, determination of threshold values for quantitative assays). Moreover a full documentation of all aspects of analytical performance shall be available to prepare and facilitate the transition of the testing methodology in the post approval clinical use.

3.3 Biomarker Validation Approach

In biomarker validation it is important to distinguish between analytic method validation and clinical qualification. The first is defined as the process of assessing the biomarker assay, its performance characteristics and the optimal conditions that will generate the reproducibility and accuracy of the assay; the second refers to a linking process between the biological process and clinical endpoints (Taylor et al., 2008).

Clinical qualification is observed in the co-development of biomarkers and drugs with the use of these biomarkers limited to the application of the drug. But this co-development imposes the necessity to generate specific guidelines for analytic test validation in order to facilitate the rapid transfer in clinical diagnostic laboratories.

The assay used has to include some specific attributes such as analytical sensitivity, diagnostic sensitivity, analytical specificity, diagnostic specificity, accuracy, repeatability, reproducibility, linear range of quantitative assays and limits of detection. Analytical performance criteria should be defined and justified in a pragmatic way so as to be proportionate to the stage of development and to risks and benefits of its intended use ("fit-for-purpose"; Lee et al., 2006). Method validation is thus a continuous and iterative process of assay refinement with validation criteria that is driven by the application of the biomarkers with increasing rigor at each successive validation step and focusing on method robustness, cross-validation, and documentation control.

The first step of the biomarker method validation is the choice of the right assay followed by developing of the assay into a validated method.

When an assay is selected there are some attributes that have been considered. At the beginning it is important to evaluate sample type (including any matrix factors), conditions and preparation for sampling, sample handling such as storage conditions and extraction

procedure, patient or subject selection criteria. Furthermore many analytical factors have to be investigated. For example assay protocol, assay platform including amplification efficiency, linearity, precision, dynamic range as well as limit of detection, calibrators and quality control, software and algorithms used for the interpretation of results and performance variables (Chau et al., 2008).

Biomarker method validation is more complicated than a typical bioanalytic assay because it is an endogenous characteristic and also includes key variable assay elements that range from the choice of the matrix to maintaining sample integrity, to assay standardization and accuracy.

Biomarker production, physiology and distribution of the biomarkers are the main parameters to select the biological matrix that should be readily accessible, such as whole blood, plasma, serum, or urine. This choice is important because matrix can influence the validation process as evidenced by feasibility in the acquisition of biological material during the study, such as in the collection of noninvasive (sputum, urine, feces, and saliva) versus minimally invasive (blood or plasma) samples (Chau et al., 2008).

If the method has sufficient sensitivity, then the preferred matrix choice is based on ease of sample collection and analysis. However, if sensitivity is a factor and measurement of the biomarker in the specified matrix poses as a challenge, then the preferred matrix is chosen based on sample concentration even if this presents as a greater challenge for sample collection and preparation. In addition to the influence of sample sources, material collection and processing should be examined to maintain sample integrity.

It is important for researchers to realize that biospecimen collection varies across populations; how they are handled and differences in sample processing variables can dramatically affect the results of a trial. Thus, appropriate conditions for collecting, handling, and storing study samples need to be standardized along with adequate training of

the clinical trial management personnel to preserve the stability and integrity of the analyte. Sample integrity can be affected by repeated cycles of freeze-thawing specimens or by long-term storage, and hence, the stability of the sample becomes compromised. Depending on the type of samples (biological fluids or tissues), minimization in variability at each step of this procedure from collection to processing is critical to ensure consistent and valid analyte measurement at subsequent biomarker assays.

Biomarker analysis requires a systematic review of the analyte stability in calibration standards, quality controls (QCs), and study samples as in bioanalytic method validation. In general, QCs are prepared to evaluate limited ranges of standard curve: the lower, middle, and upper. Whereas QC samples are used during study sample analysis to judge the acceptability of assay runs, validation samples (VS) are used in assay validation experiments to estimate intra-run and inter-run accuracy/ precision and stability. Whereas only three VS concentrations are required in GLP bioanalytic assays at least five different concentrations of VS should be analyzed in duplicate on at least six different runs during the prestudy validation because quantitative biomarker assays often exhibit nonlinear calibration curves; thus, more VS are required (Smolec et al., 2005).

Other important components of biomarker assay validation include the reference materials, precision and accuracy, dynamic range, sample recovery, sample volumes, and instrument validation. Additionally, the variability in method validation can also be affected by assay results at different locations and the correct calibration of the assay at different test sites. Because biomarkers are endogenous substances, difficulty may arise in obtaining biomarker/analyte-free matrices either to do specificity studies on or to prepare for the calibration curve. Most of the time, the target biomarker molecule is not available to act as a certified calibration standard. Researchers then may rely on the use of a non-certified standard, a recombinant protein, or a

surrogate matrix to construct the calibration curve. If assay standards are prepared in a nonauthentic matrix, QC samples should be prepared and tested in the same matrix as the study samples to show that the assay performance is similar between authentic and nonauthentic matrices. Parallelism studies should be conducted when surrogate standards and matrices are used for calibration purposes (Chau et al., 2008).

3.4 Unconventional samples for biomarker analysis

Genetic biomarker investigations are often performed on biological material of good quality especially from blood samples. However, as pharmacogenetics is a recent science, biological source is not always available in retrospective clinical studies. The advantage of use these samples is that all retrospective clinical data are already available; therefore genetic data generated from these studies can be easily evaluated for association to response to treatment. In these cases only biopsies such as formalin-fixed paraffin-embedded (FFPE) tissue samples or Hematoxylin and Eosin (H&E) stained FFPE slides are stored and can be used to perform genetic investigations. These tissue types are a potential source of information because they are often collected for different diseases and clinical trials and are then stored in biological specimen banks (Andreassen et al., 2004).

Interest in new biological source is not limited to sample collected in archived clinical studies. Recently, some methods have been developed to extract DNA or RNA from unconventional samples that allow also a non invasive collection. For example DNA extracted from saliva samples could be used to perform Human Leukocyte Antigen (HLA) typing (McWhinnie et al., 2010).

In the present study, it was evaluated the possibility to carry out mRNA expressions analysis on single plucked human hairs during clinical trials.

3.4.1 Hair Follicle

Despite 80% of cancer arising from epithelial tissues, peripheral blood is the most common surrogate tissue used for biomarkers and pharmacogenomic investigation in oncology. Hair follicles (HF) may offer a viable alternative since they are easy to collect, available from most individual and they can reflect biological response in epithelial tissue. Hair follicles are widely study in order to better understand problems related growth and loss of hairs. However, recently interest in biomarker investigation in Hair follicles is increasing for example to study gene expression in subject with hirsutism (Kim et al., 2006; Oliveira et al., 2003) or to use it in monitoring biological response in epithelial tissue (Brady et al., 2007).

The hair is a complex structure consisting of epithelial and mesenchymal tissue which interacted in utero to form it. During this time, the precise distribution of hair follicles over the surface of the body was established and the future phenotype of each hair (e.g., long scalp hair and short eyebrow hair) was determined. No additional hair follicles are formed after birth even if the size of the follicles and hairs can change with time. The precise spacing and distribution of the follicles are established by genes that are expressed very early in the morphogenesis of the follicles (Krause et al., 2006; Paus et al., 1999).

Hair is composed of two main structures: the hair shaft which growth out of dermis and hair follicle that is located inside it.

The hair shaft serves as an instrument of social communication, a protective device, and as a container for sequestering and excreting unwanted compounds. It consists of dead cells that have turned into keratins.

Hair shaft is composed of different parts:

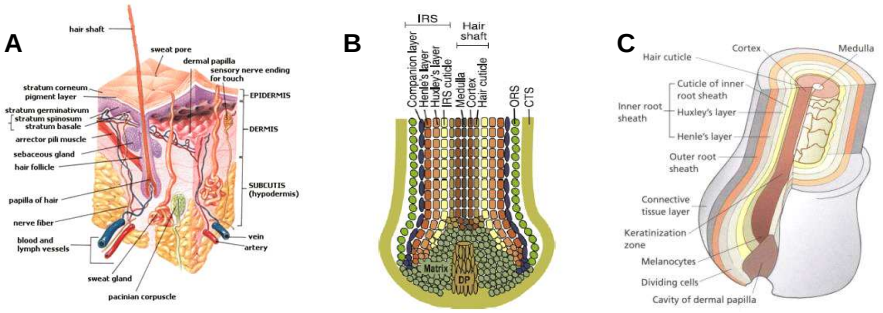
- **Cortex** which is a packed with strands of keratin and contents granules of hair pigment melanin and also a central core called central Medulla.
- **Matrix** which is a mass of sulphur-rich keratins which held together cortex keratin strands.
- **Cuticle** which is the outer layer and it is formed of overlapping layers of long cells (from 6 to 10 layers, 0.3µm thick, 100µm long)

The active part of the hair is hair follicle which is characterized by the bulb, composed by melanocytes which produce melanin granules and by highly proliferating keratinocytes which can differentiate into trichocytes, or cells of the inner root sheath (IRS). This ability allows it to completely reform itself over the cycle.

The hair bulb is composed of:

- **Dermal papilla (DP)** which is the "command center" and determines the size of the anagen hair bulb, the duration of anagen, and the hair shaft diameter.
- **The bulge stem cells** which form the secondary hair germ, and is involved in the generation of the new hair. They can even be reconstituted by dedifferentiating keratinocytes in response to wounding of the bulge area.

Figure 6. Hair composition.
Panel **A** shows a complex overview of the hair composition while panel **B** shows details of hair shaft and panel **C** details of the bulb. Picture from Schneider et al., 2009.



Each individual displays an estimated total number of 5 million hair follicles, of which 80,000 to 150,000 are located on the scalp (Bernard, 2006). It has been estimated that an adult has a surface scalp between 550 and 650 cm^2. Age influences follicle numbers for cm^2 : 1135 at birth, 635 at 30 years, 415 at 60 years.

Each hair follicle perpetually goes through three stages which have different duration and that compose the hair follicle cycle:

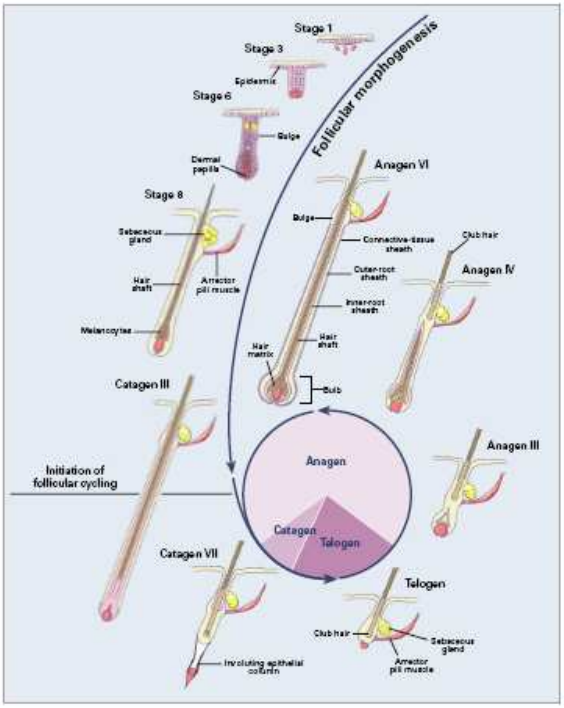
- **Anagen phase**: from 2 to 6 years
- **Catagen phase**: few weeks
- **Telogen phase**: from 2 to 4 months

Approximately 85% to 90% of all scalp hairs are anagen follicles, while 10-14% are in the telogen phase, and 1-2% are in the catagen phase (Alonso et al., 2006).

Numerous key factors such as proteins, hormone, growing factors and growing factor receptors are important for hair follicle development and cycling.

The purpose of hair cycling in mammals with individual (asynchronous) follicle waves (i.e. humans) is not as obvious, but may include cleaning the skin surface of debris and parasites, and excretion of deleterious chemicals by encapsulation within trichocytes. In addition, follicle cycling might serve as a regulator of paracrine or even endocrine secretion of hormones and growth modulators produced within the follicle and secreted into the skin or circulation (Ito et al., 2005; Rendl et al., 2009; Stenn et al., 2001). Finally, hair follicle cycling may act as a safe-guarding system against malignant degeneration by protecting rapidly dividing keratinocytes from oxidative damage by deletion during catagen phase.

Figure 7. Scheme of development and cycling of hair follicles (from Paus et al., 1999).



The **Anagen** is the phase characterized by active growth and it is further subdivided into six sub-stages according to morphological criteria (Stenn et al., 2001). These are:

- **Stage I** -growth of the dermal papilla and on-set of mitotic activity in the germ-like overlying epithelium.
- **Stage II** -bulb matrix cells envelop the dermal papilla and begin differentiation, evolving bulb begins descent along the fibrous streamer.
- **Stage III**-bulb matrix cells show differentiation into all follicular components.
- **Stage IV**-matrix melanocytes reactivate.
- **Stage V**-hair shaft emerges and dislodges telogen hair.
- **Stage VI**-new hair shaft emerges from skin surface.

The **Catagen** phase is characterized by a highly controlled process of involution (like apoptosis). In this phase:

- Follicular melanogenesis ceases
- Toward the end of the catagen stage, the dermal papilla condenses and moves upward, coming to rest underneath the hair-follicle bulge.
- If the dermal papilla fails to reach the bulge during the catagen stage, the follicle stops cycling and the hair is lost

The **Telogen** phase is characterized by quiescence. The follicle remains in this stage until it is reactivated by intra and extra follicular signal. In this stage:

- The hair shaft matures into a club hair, which is eventually shed from the follicle, usually during combing or washing.
- Most people lose 50 to 150 scalp hairs per day.
- An increase in the percentage of scalp follicles in the telogen stage leads to excessive shedding. Therefore, drugs that maintained or reduced the percentage of follicles in this stage would be valuable in treating hair loss.

3.4.2 FFPE and H&E FFPE

Formalin-fixation and paraffin-embedding is the standard method for tissue handling and stored clinical tissue biopsies. According to this procedure the fresh biopsy is first fixed in formalin to stop vital activity of cells, then water is removed by washing steps in alcohol at increasing concentrations. Tissue is treated with xylene which allows the substitution of alcohol with paraffin and finally with the paraffin (embedded step). The result is a tissue block called formalin-fixed paraffin-embedded (FFPE) that could be maintained entire or cut by microtome in section of different thickness. Slides obtained are often used for histological and immunohistochemical staining for medical diagnosis.

For example, hematoxylin and eosin (H&E) staining is the main procedure used by pathologists to identify tumor. Basophilic structures are stained in blue with hematoxylin while eosinophilic structure such as cytoplasm and connective tissues are counterstained in red with eosin.

FFPE are the widest resource for retrospective discovery studies and genomic analyses such as polymorphisms or mutations; indeed in recent years the identification and validation of molecular markers in these tissues has been a field of intensive research to aid in cancer diagnosis and to guide cancer treatment. Where H&E slides are the only source available it is possible to modify analytical methods to use them.

The advantage of using FFPE is that they are stable at room temperature and so they are easy to store. However nucleic acid obtained from FFPE is characterized by degradation, cross-linking, shortage of material, methylol derivatives and PCR inhibitors presence (Srinivasan et al., 2002). Damage is caused by the effect of formaldehyde which denatures the AT-rich region producing sites for chemical interactions, which affect DNA quality and integrity.

Nevertheless increasing efforts have been made to investigate genetic germline variations to study their influence on phenotype, cancer susceptibility and disease outcome (Chen et al. 2008).

Recently some studies focused on evaluate the possibility to extract RNA from FFPE to perform gene expression analyses (Farragher et al., 2008; Penland et al., 2007) and to study microRNA profiling (Doleshal et al., 2008; Goswami et al., 2010; Lehmann, 2010; Liu et al., 2009).

4 AIM OF THE STUDY

During drug research and development, biomarkers are broadly used to improve the understanding of drug mechanism of action, to investigate drug efficacy and safety, to support the selection of target patient population and to optimize treatment schedule.
Among different classes of biomarkers, genomic biomarkers are defined as a measurable DNA or RNA characteristics that are indicator of normal biologic processes, pathogenic processes, and/or response to therapeutic or other intervention. Genomic biomarkers are the main subject of pharmacogenomics. As PGx is a recent science, conventional biological sources to perform genomic analysis are not always available especially in retrospective clinical studies.

The PhD research aimed at developing new genomics-based tools using non-conventional biological samples that might support biomarker investigations in clinical settings.
In particular the focus was on establishing and validating method to analyze genomic biomarker on non-conventional samples such as Formalin-Fixed Paraffin-Embedded (FFPE) tissue, Hematoxylin and Eosin (H&E) stained FFPE, hair follicles.
Among different DNA and RNA characteristics, single nucleotide polymorphisms (SNPs), mutations and gene expression analyses were evaluated in different unconventional tissues.
This study was also useful to identify the proper strategy to validate analytical methods as required by Regulatory Agencies in the field of pharmacogenomics. Methods developed will be used to support biomarker investigation in clinical trials and to produce data for regulatory decision making. The possibility to carry out genomic biomarker investigation in clinical trials might allow stratifying the population in order to identify patients who better benefit from therapeutic treatment. This is a further step towards a personalized medicine.

5 MATERIALS AND METHOD

5.1 Nucleic acid Extraction

5.1.1 DNA Extraction Procedures

DNA extraction was performed from different tissue types and using different technologies. In particular Fresh Frozen biopsy samples, Formalin-Fixed Paraffin-Embedded (FFPE) tissues and Haematoxylin & Eosin stained FFPE (H&E FFPE) were extracted using affinity columns method. DNA obtained from these sources was used in single nucleotide polymorphism or mutation analysis.

Whole blood and serum derived from the same donors were also used for DNA extraction. One aliquot of whole blood was used for automatic DNA extraction using paramagnetic particles; the other aliquot of blood was collected to obtained serum. DNA from serum was then extracted by affinity columns. Blood and serum DNA were used for polymorphism analysis.

5.1.1.1 DNA extraction from Fresh Frozen biopsies

DNA extraction was performed with affinity columns using the DNeasy® Blood & Tissue kit (QIAGEN®) according to manufacturer's instructions and working instructions. This method foresees a first step of lysis, followed by the binding of the DNA to the column membrane. Remaining contaminants and enzyme inhibitors are then removed by washing steps before the DNA is eluted in water or buffer.

Briefly, about 25mg of tissues were placed in a 1.5ml microcentrifuge tube. 180µl of Buffer ATL and 20 µl proteinase K were added at each sample. Samples were mixed thoroughly by vortexing, and incubated at 56°C until tissues were completely lysed. 200µl Buffer AL was added to each sample, and mix thoroughly by vortexing. Then 200µl ethanol (96–100%) was added and the mixture was transferred

(including any precipitate) into the DNeasy Mini spin column placed in a 2 ml collection tube. Samples were washed with two different washing buffers provided by the kit. Then, samples were centrifuged for a minute at 20000 x *g* to remove any residual ethanol. Elution was performed in 100µl Buffer AE which was added directly onto the DNeasy membrane.

5.1.1.2 DNA extraction from Formalin-Fixed Paraffin- Embedded (FFPE)

DNA extraction from FFPE slides was performed with affinity columns using the QIAamp® DNA FFPE Tissue kit (QIAGEN®) according to manufacturer's instructions and internal working instructions.

This procedure removes first the paraffin from each sample. Then samples are incubated for the lysis step and heated to partially reverse formaldehyde modification of nucleic acids. DNA is then bound to membrane while remaining contaminants and enzyme inhibitors are removed by two washing steps. DNA is then eluted in water or buffer.

Briefly, slide was placed for 10 min in a bath of xylene, then in a bath of ethanol (96–100%) for two minute. Slide was allowed drying at room temperature for few minutes. The tissue was scraped with scapel and placed in a 1.5 ml microcentrifuge tube in presence of 180 µl of Buffer ATL and 20 µl of proteinase K. Sample was incubated at 56°C for 1 h under mixing (or until the sample has been completely lysed) and then at 90°C for 1h . 200 µl of Buffer AL and 200 µl of ethanol (96–100%) were added to the sample, and then, mixed thoroughly by vortexing. The entire lysate was transferred to the QIAamp MinElute column (in a 2 ml collection tube) and centrifuged at 6000 x *g* for 1 min. QIAamp MinElute column was washed with 500 µl of Buffer AW1 and 500 µl of Buffer AW2. At each step of centrifugation (8000 x *g* for 1 min) the flow-trough was discarded.

A centrifugation step at 20,000 x *g* for 3 min was required to dry the membrane completely. Elution step was performed in 20 µl of Buffer ATE.

5.1.1.3 DNA extraction from H&E FFPE

DNA extraction from H&E FFPE sample slides was performed with affinity columns using the QIAamp® DNA FFPE Tissue kit (QIAGEN®) according to manufacturer's instructions and internal working instructions.

Compared to the extraction of unstained FFPE, H&E FFPE samples required an additional step to allow the detachment of the cover glass before DNA extraction. For this purpose slides were incubated in a xylene bath for about 24h. Time of incubation depends on preparation protocol and storage. Once the cover glass was detached, samples were incubated in a bath of 100% of ethanol for 10min then they were left to dry at room temperature. Samples were washed with 1ml of 1X PBS and tissues remove by a scalpel. Sample were then immediately transferred in 1.5 ml tube and extracted as described in the previous section.

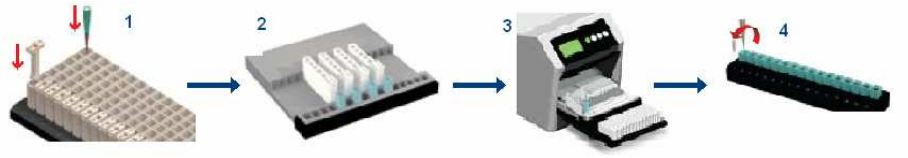
5.1.1.4 DNA extraction from peripheral blood

Whole blood (5ml) was collected into sodium heparin tubes from healthy volunteers. DNA was extracted from 400µl of peripheral blood by automated purification using Maxwell®16 system (Promega). This method is based on paramagnetic particles (PMPs) which provides a mobile solid phase that allows capturing, washing and eluting DNA. The instrument is a magnetic particle-handler that transports the PMPs through purification reagents in the prefilled cartridges.

Briefly, samples were loaded into the cartridge together with plungers in a lysis buffer. For each sample an empty elution tube was placed into cartridge rack and then filled with 100µl of elution buffer. The

cartridge rack was placed into the instrument and the extraction started. First of all plungers were fixed into the magnetic rod assembly then particles were captured by plungers and transferred into the well contained sample and lysis buffer. DNA was bound by particles and transferred into the others wells of the cartridge for washing step. At the end of the extraction program samples were released into elution tubes and transferred by pipetting into new clean tubes. A scheme of the procedure is shown in Figure 8.

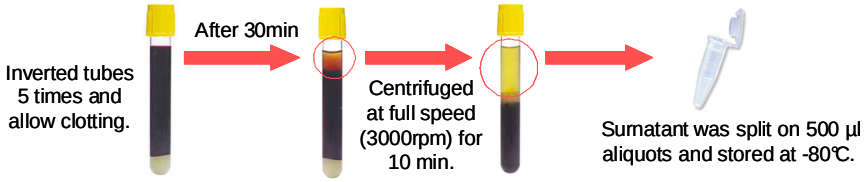
Figure 8. **DNA extraction from peripheral blood using Maxwell® procedure.** (1) Samples and plungers are loaded into the cartridge (2) Elution tubes are placed into the cartridge rack. (3) The cartridge rack is loaded into the instrument to allow DNA extraction. The program of the instrument is run (4) Elution tubes are removed from the cartridge rack and DNA eluted is transferred into a new tube.



5.1.1.5 DNA extraction from serum

Serum was obtained from separation from peripheral blood. 5ml of blood were collected from healthy volunteers into BD Vacutainer™ SST™ Tubes. Samples were left 30 minutes at room temperature to allow the clotting before centrifugation and collection. A schematic representation of the procedure is reported in Figure 9.

Figure 9. Schematic representation of serum preparation.



Serum was split on 500µl aliquots and stored at -80°C until extraction. DNA extraction from serum was performed with affinity columns by QIAamp® DNA Investigator kit (QIAGEN®) according to manufacturer's instructions and internal working instructions.

The QIAamp DNA Investigator procedure consists of 4 steps: sample lysis: under denaturing conditions with proteinase K, DNA binding to the membrane, washing away of residual contaminants and elution.

Serum aliquot was allowed thawing at room temperature and added with 500 µl of Buffer AL and 25 µl of proteinase K. Sample was incubated at 56°C for 10 min under mixing and then briefly centrifuged. 200 µl of ethanol (96–100%) were added to the sample and mixed thoroughly by vortexing. The entire lysate was transferred to the column (in a 2 ml collection tube) and centrifuged at 6000 x *g* for 1 min. Column was washed with 500 µl of Buffer AW1, 500 µl of Buffer AW2 and 500 µl of ethanol (96–100%). At each step of centrifugation the flow-trough was discarded.

A centrifugation step at 20,000 x *g* for 3 min was required to dry the membrane completely. Elution step was performed in 20 µl of Buffer ATE.

5.1.2 RNA Extraction Procedures

Total RNA was extracted from Hair Follicles (HFs), which were used as alternative source of biological material for studying gene expression. For RNA extraction two different methods were compared: RNeasy® Micro Kit (QIAGEN®) and RNAqueous Micro Kit (Ambion). Both kits are based on affinity column binding.

Since HFs in a non-conventional biological source for genomic analysis and that RNA is affected by degradation, a first attempt for the extraction was performed on fresh samples. To test the applicability of the method in clinical studies, during which samples need to be stored and shipped to the analytical laboratory, a collection

and a stabilization method was also evaluated. In particular, follicles were stabilized in RNA*later*® solution before total RNA extraction.
Moreover to study tissue specific gene expression, RNA was also extracted from a pool of peripheral blood obtained from healthy volunteers. This analysis was performed in order to compare gene expression from different tissues.
RNA integrity was monitored after each extraction method.

5.1.2.1 RNA extraction from Hair Follicles: QIAGEN® RNeasy® Micro Kit

RNA extraction from HFs samples was performed with affinity columns by RNeasy® Micro Kit (QIAGEN®) according to manufacturer's instructions and internal working instructions.
This technology is based on selective binding properties of a silica-based membrane of the columns. Guanidine-thiocyanate–containing lysis buffer and ethanol is added to the sample to promote selective binding of RNA to the membrane. Traces of DNA that may copurify are removed by DNase treatment on the RNeasy MinElute spin column. DNase and any contaminants are then washed before eluting the RNA.
Briefly, the lysis step was performed adding an equal volume of lysis buffer and ethanol (350µl of Lysis Buffer and 350µl of 70% ethanol). The lysate was centrifuged at 20000 x *g* for 3 min and supernatant was transferred to the column and again briefly centrifuged. Total RNA bound to the membrane was first washed with 350µl of Buffer RW1 and then incubated with DNase I to remove any DNA residual. Sample was washed with 350µl of Buffer RW1, 500 µl of Buffer RPE and 500 µl of 80% ethanol. At each step of centrifugation (8000 x *g* for 1 min) the flow-trough was discarded.
Total RNA was eluted in 12µl RNase-free water by centrifugation for 1 min at 20000 x *g*.

5.1.2.2 RNA extraction from Hair Follicles: Ambion RNAqueous®-Micro Kit

RNA extraction from HFs samples was performed with affinity columns by RNAqueous®-Micro Kit (Ambion) according to manufacturer's instructions and internal working instructions.
The procedure requires the lysing the sample in the presence of guanidinium thiocyanate. The lysate is then mixed with ethanol and apply to a silica-based filter that selectively bound RNA. A post-elution DNase treatment is performed to remove genomic DNA that could interfere with RT-PCR assays. The DNase is then removed after digestion using a specific resin.
Briefly, the lysis step was performed adding an equal volume of lysis buffer and ethanol (100µl of Lysis solution and 100µl of 100% ethanol). The mixture was loaded into a Micro Filter Cartridge Assembly and centrifuged 10s at 20000 x *g* to bind RNA to the filter. Sequential washing step with 180µl of Wash Solution 1 and Wash Solution 2/3 were required. The flow-trough was discarded and the cartridge replaced in the same collection tube. An additional 1min centrifugation at 20000 x *g* was required to remove residual fluid and dry the filter. The filter was transferred in the elution tube and 10µl of preheated to 75°C Elution Solution was added (in two subsequent step of 5µl to increase the yield). A DNase treatment was performed adding 1µl of DNase I Buffer and 1µl of DNase I. Sample was incubated at 37°C for 20 min and then treated with 2 µl of DNase Inactivation Reagent and left at room temperature for2 min. To pellet the DNase Inactivation Reagent sample was centrifuged at 20000 x *g* for 1.5 min and total RNA transferred in a new RNase-free tube.

5.1.2.3 RNA extraction from peripheral blood

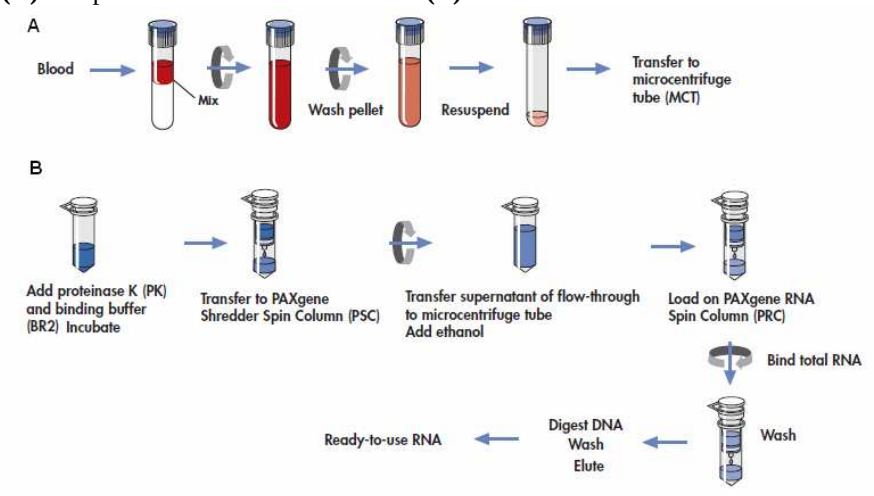
RNA extraction from peripheral blood was performed with the PAXgene Blood RNA System which consists of a blood collection tube (PAXgene Blood RNA Tube) and nucleic acid purification kit

(PAXgene Blood RNA Kit). The first part of the kit allows sample collection and avoids RNA degradation limiting changes in gene expression. The second allows RNA isolation and purification trough affinity columns.

Briefly, 2.5ml of blood from healthy volunteers was collected into PAXgene Blood RNA Tube then before extraction sample were maintained for at least two hours at room temperature to complete lysis of blood cell. A pellet was obtained by centrifugation of sample at 3000 x *g* for 10 min. Then pellet was resuspended and transferred to a microcentrifuge tube (as shown in Figure 10A). A schematic representation of RNA purification trough affinity column is reported in Figure 10B; elution was performed in 50 µl of elution buffer.

Details of the procedure followed were reported in manufacturer's instruction and internal working instruction.

Figure 10. Schematic representation of the PAXgene Blood RNA Procedure.
(A) Sample collection and stabilization **(B)** Purification of total RNA.

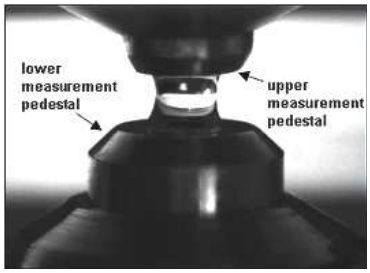


5.2 Nucleic acid quantification

5.2.1 RNA and DNA Quantification

After DNA or total RNA extraction, samples were quantified with NanoDrop®ND-1000 a full-spectrum (220-750nm) spectrophotometer. The instrument is able to quantify the sample using minimal amount; indeed 1 µl sample was pipetted onto the end of a fiber optic cable. A second fiber optic cable was then brought into contact with the liquid sample causing the liquid to bridge the gap between the fiber optic ends. A pulsed xenon flash lamp provided the light source and a spectrometer using a linear Charge Coupled Deviced (CCD) array analyzed the light after passing through the sample (Figure 11).

Figure 11. NanoDrop® principle.
A liquid sample bridges the gap between the receiving fiber and the source fiber to enable spectrophotometrical measurement using a pulsed xenon flash lamp. (From NanoDrop® ND-1000 Spectrophotometer User's Manual).



When the measurement of a sample is taken, the intensity of light that has transmitted through the sample is recorded. The sample intensities along with the blank intensities were used to calculate the sample absorbance according to the following equation:
Absorbance = -log (Intensitysample/Intensityblank)

The sample concentration was calculated according to a modified Beer-Lambert equation: **c = (A * e)/b**

Where **c** is the sample concentration in ng/µL
A is the absorbance (A 260) represented in absorbance units
e is the wavelength-dependent extinction coefficient in ng*cm/µL (for double-stranded DNA: 50 ng*cm/µl, single-stranded DNA: 33 ng*cm/µl, and RNA: 40 ng*cm/µl)
b is the path length in cm

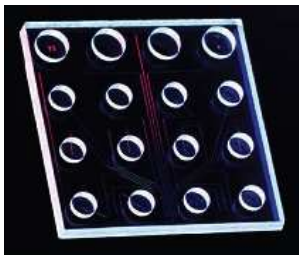
The ratio of absorbance at 260 and 280 nm (A260/A280) was used to assess the purity of nucleic acid. A ratio included between 1.8 and 2 was generally accepted for pure RNA and DNA. If the ratio was lower, it indicated the presence of protein, phenol or other contaminants that absorbed strongly at or near 280 nm. The ratio of sample absorbance at 260 and 230 nm (A260/A230) was a secondary parameter of nucleic acid purity. This value for pure nucleic acid was commonly in the range of 1.8-2.2 but if it was appreciably lower, it may indicate the presence of co-purified contaminants.

5.3 Nucleic Acid Quality Control

5.3.1 Capillary Electrophoresis for RNA analysis

To determine RNA quality, capillary electrophoresis was performed using the RNA 6000 Pico Chip (RNA 6000 Pico Kit) on an Agilent 2100 Bioanalyzer (Agilent Technologies). The electrophoretic assays from Agilent are based on the traditional electrophoresis principles, but have been transferred to a multi-well chip format (Figure 12).

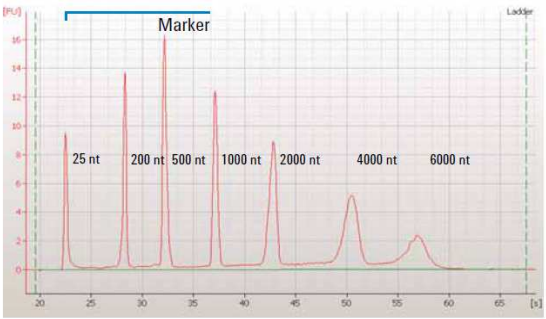
Figure 12. Example of Bioanalyzer chip.
The picture shows the internal part of Bioanalyzer chip, in particular wells and micro-channel.



The chip contains sample wells, gel wells and a well for an external standard (ladder). Wells are interconnected by a network of glass micro-channels which are filled, during chip preparation, with a sieving polymer and fluorescence dye. Once the wells and channels are filled, the chip becomes an integrated electrical circuit. The samples are injected into the micro-channels and the dye molecules intercalate into RNA strands. Sample components are electrophoretically separated by size and components are detected by laser-induced fluorescence.

Data is translated into gel-like images (bands) and electropherograms (peaks). With the help of a ladder that contains six components of known sizes (0.2, 0.5, 1.0, 2.0, 4.0 and 6.0 kb) a standard curve of migration time versus fragments size is plotted. An example is reported in Figure 13.

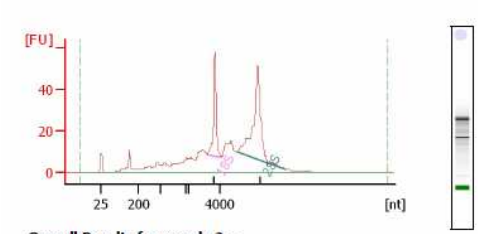
Figure 13. Electropherogram of a Pico RNA Ladder.
The first pick corresponds to the markers while the others to the ladder. For each peak nucleotide length is reported.



From the migration times measured for each fragment in the sample, the size is calculated. Two marker fragments (for RNA only one marker fragment) are run with each of the samples bracketing the overall sizing range. The “lower” and “upper” markers are internal standards used to align the ladder data with data from the sample wells. This is necessary to compensate for drift effects that may occur during the course of a chip run. The area under the upper marker peak is compared with the sample peak areas. Because the concentration of the upper marker is known, the concentration for each sample can be calculated. Besides this relative quantification, an absolute quantification is available for protein assays, using external standard proteins.

For RNA assays, quantification is done considering the ladder area. The area under the ladder is compared with the sum of the sample peak areas. The area under the “lower” marker is not taken into consideration. For total RNA assays, the ratio of ribosomal subunits (28S/18S) is determined as a quality parameter (see Figure 14).

Figure 14. Example of a report after run.
In the electropherogram marker peak and the ribosomal subunit peaks (18S and 28S) are showed.



Overall Results for sample 3 :

RNA Area:	591.4
RNA Concentration:	3,018 pg/µl
rRNA Ratio [28s / 18s]:	2.0
RNA Integrity Number (RIN):	7.5 (B.02.03)
Result Flagging Color:	
Result Flagging Label:	RIN: 7.50

Fragment table for sample 3 :

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	3,174	4,093	66.2	11.2
28S	5,114	8,280	129.7	21.9

Additionally, the RNA integrity number (RIN) can be utilized to estimate the integrity of total RNA samples based on the entire electrophoretic trace, including the presence or absence of degradation products.

Before starting to prepare the chip, all reagents were equilibrated to room temperature for 30 minutes. Sample aliquots and ladder were heat denatured for two minutes at 70°C. For gel preparation 550 µL RNA 6000 Pico gel matrix was pipetted into a spin filter and centrifuged at 1500 x *g* for 10 minutes. 1 µL Vortexed RNA 6000 Pico dye concentrate was diluted in 65 µL filtered gel. The Gel-Dye Mix was vortexed and centrifuged for ten minutes at 13000 x *g*. 9 µL Gel-Dye Mix were added into the designed chip positions on a RNA 6000 Pico chip and dispensed into the microchannel system using the chip priming station for exactly 30 seconds. 9 µL Gel-Dye mix were added into two other designed chip positions. 9 µl of the RNA 6000 Pico conditioning solution were pipetted into the design well. 5 µL RNA Pico marker were pipetted into all twelve sample wells and into the ladder well. 1 µL ladder or 1 µL of each sample were added into the appropriate wells. The chip was vortexed for one minute at 2400 rpm and then run in the Agilent 2100 Bioanalyzer.

5.3.2 Agarose Gel Electrophoresis for DNA analysis

DNA integrity and quality was evaluated by agarose gel electrophoresis. For this purpose samples were analyzed with ReadyAgarose Precast Gel with Etidium Bromide (BioRad®).

Gel concentration differs according to DNA size: genome DNA was analyzed by 1% agarose gel while PCR products by 3% gel. Gel loading solution (Sigma-Aldrich®) was added to each sample in order to facilitate the loading. According to sample analyzed different ladders were used. In particular about 500ng of 1 Kb DNA Ladder (Invitrogen) was used for genomic DNA because it was suitable for sizing linear double-stranded DNA fragments from 500 bp to 12 kb.

For smaller DNA fragment about 500ng of 50 bp DNA Ladder as used because it was suitable for sizing double-stranded DNA from 50 to 800 bp. The electrophoresis gel analyses were performed using ReadySub-Cell GT horizontal electrophoresis cells systems (BioRad) in TAE buffer at 80V for about 35 min. To acquire gel image GelDoc XR (BioRad) was used. The instrument included an UV-tray which was able to detect DNA staining with ethidium bromide.

5.4 DNA amplification: whole genome amplification

Limiting amounts of human biological material often yield insufficient quantity and of poor quality genomic DNA (gDNA) for various assays and reliable analyses. In these situations, whole genome amplification (WGA) methods can be applied to obtain enough material for molecular analysis (Barker et al., 2004; Bredel et al., 2005; Lee CI et al., 2006).

Whole Genome Amplification was applied on DNA extracted from blood and serum samples using the REPLI-g® Whole Genome Amplification (Qiagen®) kit. This method is based on multiple displacement amplification (MDA). MDA is an isothermal (30°C), linear amplification using random hexamer primers and Phi29 DNA polymerase.

Random primers attach randomly to the template adding complementary base pairs to one strand, while displacing the other DNA strand. Primary products initiate secondary priming on the displaced DNA strand and continue displacing to create multiple branches. If the initial sample of genomic DNA is high quality, it can be amplified one million-fold while maintaining accurate loci and allele representation.

Phi29 DNA is a high processivity polymerase where processivity refers to the average number of bases the polymerase adds to a DNA strand being synthesized before the polymerase detaches from the

template nucleic acid. Phi29 DNA polymerase does not dissociate from the genomic DNA template allowing the generation of DNA fragments up to 100 kb without sequence bias. The enzyme has a 3'→5' exonuclease proofreading activity and provides error rates 100 times lower than Taq DNA polymerase-based methods. A schematic representation of REPLI-g® principle is shown in Figure 15.

Figure 15. Schematic representation of REPLI-g DNA amplification.
After annealing of the primers, Phi 29 DNA polymerase moves along the DNA template strand displacing the complementary strand. The displaced strand becomes a template for replication allowing high yields of high-molecular-weight DNA to be generated.

REPLI-g® Midi kit was used for WGA of DNA extracted from whole blood and serum. In particular for blood samples DNA was previously extracted according to procedure explained in section 5.1.1.4 and then amplified. WGA reaction was performed using 25ng of DNA. The procedure of buffer and mix preparation was carried out according to manufacturer's instructions. Then samples were incubated at 30°C for 16 hours into a GeneAmp® PCR System 9700 (Applied Biosystems). A briefly step at 65°C is required to stop the reaction.

For serum amplification a different protocol was used. 500 µl were directly used for WGA. Sample was centrifuged at 8000rpm for 10 min and the supernatant discarded. Pellet was resuspended adding 50 µl TE buffer and vortexing for 5s. Sample was centrifuged 8000rpm for 10 min and the supernatant discarded. Pellet was resuspended adding 10 µl TE buffer and vortexing for 5 s. 10 µl Buffer D2 (prepared before amplification with DTT, 1 M and reconstituted Buffer DLB) was added to each sample and maintained on ice for 10 min. 10 µl Stop Solution were added to each microcentrifuge tube.

The master mix preparation was the same used for blood as temperature-time amplification profile.
The obtained DNA was used for single nucleotide polymorphism analysis in order to increase the number of gene analyzable.

5.5 RealTime PCR Analysis

Real-time Polymerase Chain Reaction (RT-PCR) is the ability to monitor the progress of the PCR as it occurs (i.e., in real time) and it has a great application potential.
In the present work RT-PCR was used for different applications such as gene expression analysis, genotyping or detection of mutations.

5.5.1 Reverse Transcription of Total RNA

This procedure was performed for gene expression analysis on total RNA. Indeed, reverse transcription step was used to prepare first-strand cDNA prior to conventional PCR amplification.
Several methods of reverse transcription were evaluated in order to select the most appropriate.

First, it was used the TaqMan® High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to manufacturer's instruction and internal working procedure.
In this method random primers, as the most non-specific priming method, were used to synthesize cDNA and as starting point for reverse transcriptase. Random primers bound to various sites throughout the length of the RNA and also RNA without a Poly-A tail can be reverse-transcribed.
About 500ng of total RNA was brought to a total volume of 10 µL with nuclease-free water. The diluted RNA was added to 10 µL master mix prepared as indicated in Table 3.

Table 3. Volume reagents for for a total of 20µL/reaction.

Reagents	Volume
RT Buffer 10X	2.0µl
dNTP Mix 25X (100mM)	0.8µl
Random Primers 10X	2.0µl
MultiScribe Reverse Trascriptase	1.0µl
RNA	500ng (variable volume)

Reaction was performed into a GeneAmp® PCR System 9700 (Applied Biosystems) according the temperature-time profile reported in Table 4.

Table 4. Reverse transcription temperature-time profile of High Capacity cDNA Reverse Transcription Kit.

Step	Temperature	Time
Incubation	25°C	10 min
Reverse transcription	37°C	120 min
Enzyme inactivation	85°C	5 min
Storage	4°C	∞

The resulting cDNA was stored frozen (-80°C) until assayed by Real-Time qPCR.

As alternative method, it was used the High Capacity RNA-to-cDNA Master Mix (Applied Biosystems) kit.
The main characteristic of this method was a ready Master Mix which provided all the components needed for first-strand synthesis including the enzyme. Procedure details were followed according to manufacturer's instruction.
Briefly, RNA was diluted into 16µl of H_2O (50ng-500ng RNA/reaction) and added to 4µl of Master Mix (final concentration 1X). Sample was gently mixed, centrifuged and then placed into the thermocycler GeneAmp® PCR System 9700 (Applied Biosystems).
The temperature-time profile is shown in Table 5.

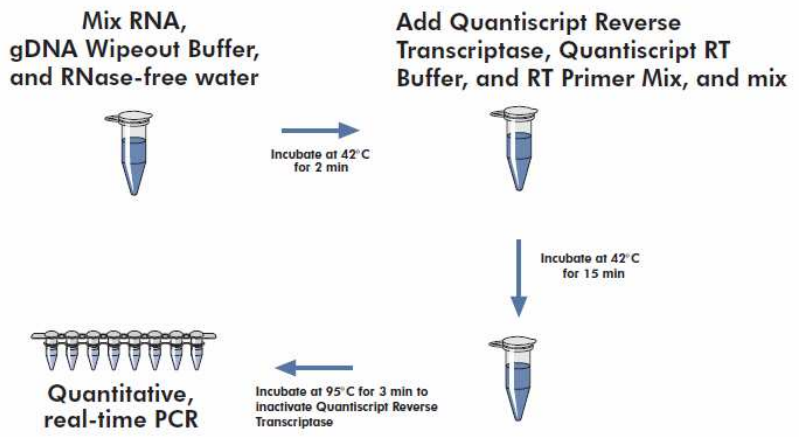
Table 5. Reverse transcription temperature-time profile of High Capacity RNA-to-cDNA Master Mix.

	Step 1	Step 2	Step 3	Step 4
Temperature	25°C	42°C	85°C	4°C
Time	5 min	30 min	5 min	∞

Moreover it was compared the Quantitect Reverse Transcription (Qiagen) kit

The method was based first on elimination of residual genomic DNA and then on reverse transcription. Procedure details were followed according to manufacturer's instruction. A scheme of the procedure is reported in Figure 16.

Figure 16. Quantitect Reverse Transcription (Qiagen®) procedure.
The method is based on a first step in which residual DNA is removed and then RNA is reverse transcribed in cDNA.



Briefly, the purified RNA sample (diluted in 12µl H_2O for 50ng-500ng RNA/reaction) was incubated in gDNA Wipeout Buffer (2µl) at 42°C for 2 minutes to effectively remove contaminating genomic DNA.

After genomic DNA elimination, the RNA sample was ready for reverse transcription using a master mix prepared from Quantiscript

Reverse Transcriptase (1µl), Quantiscript RT Buffer (4µl) and RT Primer Mix (1µl). The entire reaction taked place at 42°C and is then inactivated at 95°C.

- Finally the Super Script® Reverse Transcriptase (Invitrogen) kit was evaluated

The distinctive feature of this method is the SuperScript® III Reverse Transcriptase which is genetically engineered by the introduction of point mutations that increases thermal stability. Procedure details were followed according to manufacturer's instruction.

The first step of the procedure was to prepare a mix of random primers (50–250 ng), total RNA (50–500 ng) and 1 µl 10 mM dNTP Mix (10 mM each dATP, dGTP, dCTP and dTTP at neutral pH). Sterile distilled water was added to get 13 µl volume. Then, the mixture was heated to 65°C for 5 minutes and incubated on ice for at least 1 minute.

The second step was the addition of 5X First-Strand Buffer (4 µl), 0.1 M DTT (1 µl), RNaseOUT™ Recombinant RNase Inhibitor (1 µl) and SuperScript™ III RT (200 units/µl, 1 µl). The reaction was incubated at 25°C for 5 minutes and after this at 50°C for 60 minutes. To inactivate the reaction tube was heated at 70°C for 15 minutes.

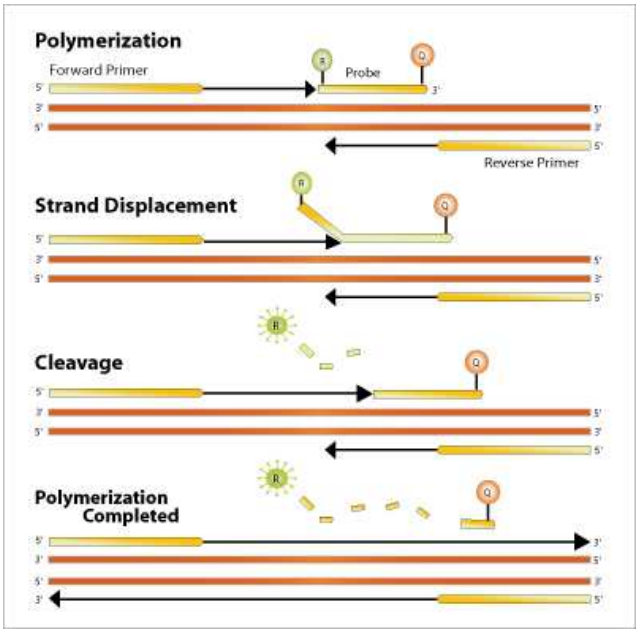
5.5.2 Principals of the TaqMan® Assay for Gene Expression

TaqMan® is a real time quantitative PCR chemistry first described by Heid et al. (1996). This assay measures the amplification of PCR product, which allows gene expression analysis. In particular in the present thesis this application was used to evaluate expression of tissue specific genes, interferon responder genes and some human housekeeping genes.

The method is based on the 5' nuclease activity of the Taq-polymerase and a non-extendible hybridization probe which is labeled at the

extremities. The probe consists of a oligonucleotide, specific for the gene under investigation, linked with a reporter dye to the 5'-end (FAM, 6-carboxyfluorescein) of the probe and a non-fluorescent quencher at the 3'-end of the probe (TAMRA, 6-carboxy-tetramethylrhodamine) shown in Figure 17.

Figure 17. Scheme of TaqMan® technology.
The peculiar characteristic of the technology is the presence of the probe which is specific for the gene target. The probe contains at the extremities a reporter and a quencer. If the probe is intact there is no florescence signal. During reaction, if the probe is annealed to gene target, the DNA polymerase displace the probe and cleave reporter from quencer causing a fluorescence increase during the PCR reaction.



During PCR forward and reverse primers are used to amplify the gene target while the probe is used for detection. When the probe is intact, fluorescence energy transfer occurs and the reporter dye fluorescent emission is absorbed by the quenching dye. If the probe has bound to the target, the fluorescent hybridization probe is cleaved by the 5'-3' nucleolytic activity of the DNA polymerase during the extension phase of the PCR cycle. The reporter dye emission is no longer transferred efficiently to the quencher dye, resulting in an increase of

reporter dye fluorescent emission spectra at 518 nm. A charge-coupled device (CCD camera) in the sequence detector measures the fluorescent emission spectra from 500 to 650 nm continuously during the amplification. Therefore, the reactions are monitored in real time. Fluorescence signal is proportional to DNA copy number produced during PCR and it is correlated to the target starting amount
It is possible to detect a wide spectrum of emission wavelengths and probes containing different reporter fluorescent dyes.

During the present study, a single 35µl reaction was prepared for each sample in a 96-wells plate according to reagent quantities listed in Table 6. Then, plate was mixed and centrifuged briefly (1 min at 1200rpm); samples were transferred in a 384-plate and each reaction was analyzed in triplicate (10µl for well).

Table 6. TaqMan® reaction mix for a single reaction in 35 µl.
The table reports the volume for a single reaction. For multiple DNA samples an extra volume should be consider.

Reagents	Volume
TaqMan Gene Expression Master Mix 2X	17.5µl
TaqMan assay primers 20X	1.75µl
cDNA	3.5µl
H_2O Nuclease-free	12.25µl

In this work TaqMan gene expression assays were used to analyze a panel of genes on RNA extracted from different tissue such as blood and hair follicles. A complete list of the gene analyzed is reported in Table 7.

Table 7. Genes used in this study analysis.
The table reports the complete name of the gene, its symbol and the TaqMan® assay identification number according to manufacture producer (Applied Biosystems).

Gene Name	Gene Symbol	Assay ID
RNA, 18S ribosomal 1	18S	Hs99999901_s1
Interferon alpha, beta and omega receptor 1	IFNAR1	Hs00265057_m1
Interferon alpha, beta and omega receptor 2	IFNAR2	Hs01022059_m1
Epidermal Growth Factor receptor	EGFR	Hs01076092_m1
Keratin 19	KRT19	Hs0151611_gH
Collagen, type I, alpha 1	COL1A1	Hs00164004_m1
Melan-A codes for a melanocyte differentiation antigen	MLANA	Hs00194133_m1
Formyl peptide receptor 1, involved in neutrophils activation	FPR1	Hs00181830_m1
Perforin 1, pore forming protein, one of the main cytolytic proteins of cytolytic granules	PRF1	Hs00169473_m1

The TaqMan Gene Expression Master Mix contains buffer and AmpliTaq Gold® DNA Polymerase.
TaqMan real-time PCR reactions were then performed on ABI Real-Time® 7900HT with the following thermal cycle profiling (Table 8).

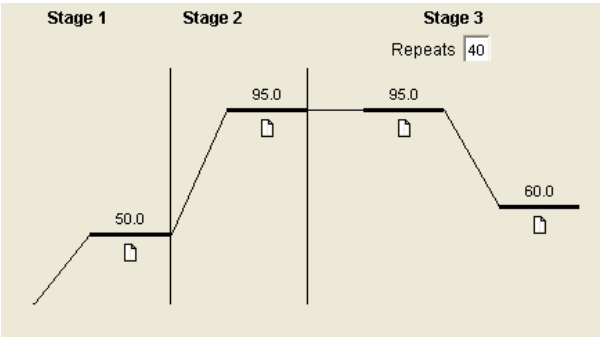
Table 8. PCR-reaction steps of the TaqMan® assay.
The temperature-time profile is reported in the table and includes 40 cycles of denaturation, annealing and extension.

	Temperature	Time
Enzyme activation	50°C	2 min
Initial Denaturation	95°C	10 min
Denaturation (40 cycles)	95°C	15 sec
Annealing and extension (40 cycles)	60°C	1 min

Real-time PCR monitors the fluorescence emitted during the reaction as an indicator of amplicon production at each PCR cycle (in realtime).

In this reaction to have a completed and precise quantification, detection was performed in all PCR phases as shown in Figure 18.

Figure 18. Detection points during PCR.
The plot shows PCR steps and stages in which detection occurs. 🗋 This symbol indicates the phase of the detection.



5.5.3 TaqMan® Low Density Array (LDA)

The TaqMan® Array is a customizable 384-well micro fluidic card (array) that enables to perform hundreds of real-time PCR reactions simultaneously.
TaqMan® Arrays require minimal amounts of sample and allow for 1 to 8 samples to be run in parallel against 12 to 384 TaqMan® Gene Expression Assay targets (including manufacturing controls). The primers are pre-spotted into the plate and this allows an easy sample loaded. Indeed, once the sample is loaded into the port, the plate is centrifuged and sealed and it can be run (see Figure 19).

Figure 19. Scheme of TaqMan® Arrays procedure.
(1) cDNA sample and TaqMan® Universal PCR Master Mix are loaded into each of the eight loading ports of a TaqMan Array. **(2)** The plate is then centrifuged to dispense the sample mix into the individual wells. **(3)** The plate is sealed **(4)** PCR is run on an Applied Biosystems® 7900HT Real-Time PCR System.

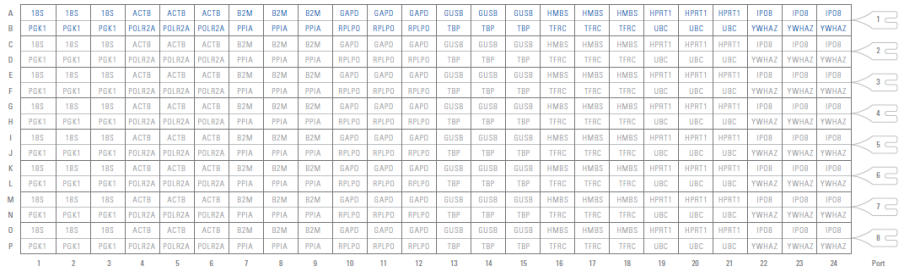
The advantage of this method is that the plate can be spotted with customizable primers according to the application. Moreover, some panels are already commercially available for the evaluation of genes belonging to different pathways or for housekeeping genes. Housekeepings are constitutive genes required for the maintenance of basal cellular function and expressed at relatively constant levels. They are expressed in at least 25 copies per cell and sometimes number in the thousands. Housekeeping genes are broadly used as internal control for gene expression analysis.
TaqMan® Endogenous Control Arrays are LDA containing a panel of housekeeping genes designed to simplify the selection of endogenous controls for gene expression studies in different species such as human, mouse, or rat.
For this study Human TaqMan® Endogenous Control Arrays was used to investigate gene expression in a non-conventional tissue such as

hair follicles. The Array included 16 genes which are listed in Table 9 while a layout of the plate is shown in Figure 20.

Table 9. List of genes in TaqMan® Endogenous Control Arrays.

Gene Name	Gene Symbol	Assay ID
RNA, 18S ribosomal 1	18S	Hs99999901_s1
Phosphoglycerate kinase 1	PGK1	Hs99999906_m1
Actin beta	ACTB	Hs99999903_m1
Polymerase (RNA) II (DNA directed) polypeptide A	POLR2A	Hs00172187_m1
Beta-2-microglobulin	B2M	Hs99999907_m1
Peptidylprolyl isomerase A (cyclophilin A)	PPIA	Hs99999904_m1
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	Hs99999905_m1
Ribosomal protein, large, P0	RPLP0	Hs99999902_m1
Glucuronidase, beta	GUSB	Hs99999908_m1
TATA box binding protein	TBP	Hs99999910_m1
Hydroxymethylbilane synthase	HMBS	Hs00609297_m1
Transferrin receptor (p90, CD71)	TFRC	Hs99999911_m1
Hypoxanthine phosphoribosyltransferase 1	HPRT1	Hs99999909_m1
Ubiquitin C	UBC	Hs00824723_m1
Importin 8	IPO8	Hs00183533_m1
Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	YWHAZ	Hs00237047_m1

Figure 20. Human TaqMan® Endogenous Control Arrays Layout.
The array allows analyzing eight samples for 16 genes. The plate is designed to run each gene in triplicate.

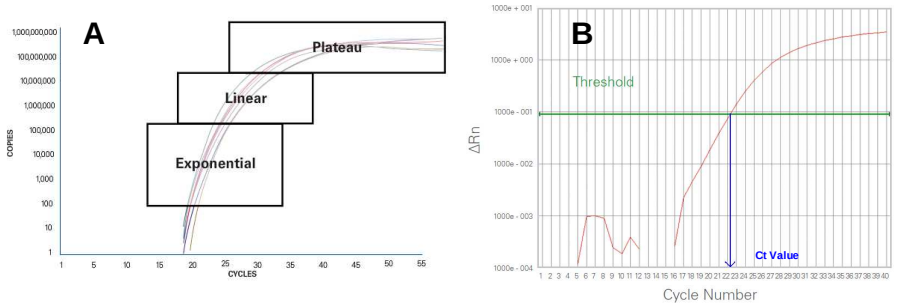


5.5.4 TaqMan gene expression Data Analysis

At the end of PCR reaction, data were analyzed using ABI SDS software 2.3v. The analysis results with a plot as reported in Figure 21A in which three phases could be distinguish: exponential (DNA theoretically doubles every cycle), linear (components are being consumed, the reaction is slowing) and finally a plateau phase (the

reaction stopped, no more products are being made and the PCR products begin to degrade).

Figure 21. Example of Real-Time PCR plots.
(A) The plot reports a general Real-Time PCR curve. X-axis represents the number of cycles while the Y-axis DNA copy number in logarithmic scale. Three main phases could be distinguished exponential, linear and plateau. **(B)** Amplification plot shows the Log (ΔRn) graphed versus cycle. Ct Value is the fractional cycle number at which the fluorescence passes the threshold. ΔRn is the normalization of the Rn obtained by subtracting the baseline (ΔRn = Rn - baseline) where Rn is the ratio of the fluorescence emission intensity of the reporter dye to the fluorescence emission intensity of the passive reference dye.



Data gathered from these phase were used to calculate some parameters useful for analysis such as background signal and Cycle Threshold. C_t (threshold cycle) is the intersection between an amplification curve and a threshold line (see Figure 21B). It is a relative measure of the concentration of target in the PCR reaction and must be set in the linear phase of the amplification. The C_t value increases with a decreasing amount of template.

After defining a baseline and threshold, the resulting values for were exported into Excel. For further analysis, the average Ct values of the triplicates were calculated.

Fold change were determined by the DDCt method as described below:

- for each replicate reaction calculate first the average Ct value
- calculate the DCt as the difference between the average Ct value of the marker gene M and the average Ct value of an

housekeeping gene (HK, *i.e.* 18S ribosomal RNA) used for normalization (18S)

DCt = (Ct_gene-M – Ct_HK)

- then, fold change between one sample and its control (i.e. post treatment vs. pretreatment) is calculated as:

  fold change= 2^ - (DCt post treatment - DCt pre treatment)

  which is also equal to:

  fold change= 2^-DDCt

5.5.5 Principals of the TaqMan® Assay for SNP Analysis

Another application of TaqMan® chemistry is allelic discrimination which allows the determination of sample genotyping combining PCR amplification and allele detection.

The main principle is the same used for gene expression but in this case there are two probes in the reaction mix that compete for hybridization to the target.

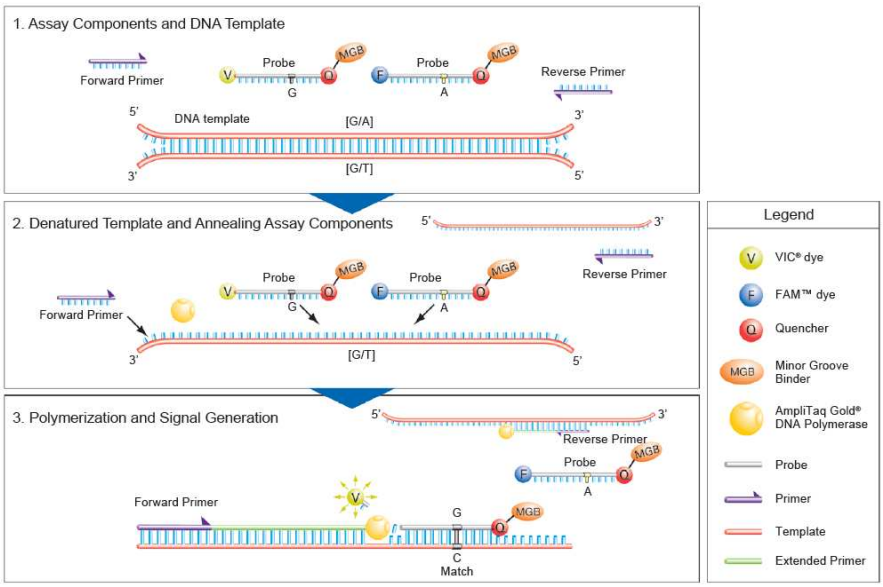
This type of assay is composed by:

- Sequence-specific forward and reverse primers to amplify the polymorphic sequence of interest.
- Two TaqMan® MGB probes: one probe labeled with VIC® dye detects the Allele 1 sequence; one probe labeled with FAM™ dye detects the Allele 2. Probes are different only for one nucleotide corresponding to the polymorphism.

During PCR, the 5'-nuclease activity of Taq DNA polymerase cleaves the reporter dye (FAM or VIC) from an MGB probe that is completely hybridized to the DNA strand (Figure 22).

When separated from the quencher, the reporter dye fluoresces. However, if a single point mismatch is present between the probe and the target DNA strand due to a SNP, the binding of the probe to DNA is destabilized during strand displacement in PCR, which will reduce the efficiency of probe cleavage and quenching of the fluorescent reporter dye.

Figure 22. Scheme of TaqMan® SNP detection.
Allelic Discrimination is achieved by competitive hybridization of TaqMan® MGB Probes.



Therefore, an increase in either FAM™ or VIC® dye fluorescence indicates homozygosity for FAM™- or VIC®-specific alleles (X:X or Y:Y), and an increase in the fluorescence of both dyes indicates heterozygosity (X:Y).

The assay is based on a pre-read run records the background fluorescence of each sample before PCR. Then samples are amplified by real-Time PCR and at the end of the run a post-read run were performed. The pre-read fluorescence is subtracted from the post-read fluorescence to account for pre-amplification background fluorescence.

SDS software measures these parameters and it is able to identify the dye components, to determine the contribution of each dye in the raw data using the multi-component algorithm. The result of the allelic discrimination run is a scatter plot of Allele X versus Allele Y in which three clusters are observable: one corresponding to Allele X, one to Allele Y and one to heterozygote group that shows both alleles.

PCR reactions were prepared according to Table 10. A single reaction for each sample was prepared and then split in triplicate.
The analysis was performed by ABI Prism® 7900HT system in a final volume of 10 μL and temperature-time profile for PCR amplification was the same used for gene expression (see Table 8).

Table 10. TaqMan® reaction mix for a single reaction in 35 μl.

Reagents	Volume
TaqMan Genotyping Master Mix 2X	17.5μl
TaqMan SNP assay Primer Mix 40X	0.88
DNA	from 50ng/ μl to 10ng/ μl
H_2O Nuclease-free	12.25μl

5.6 *KRAS* mutation analysis

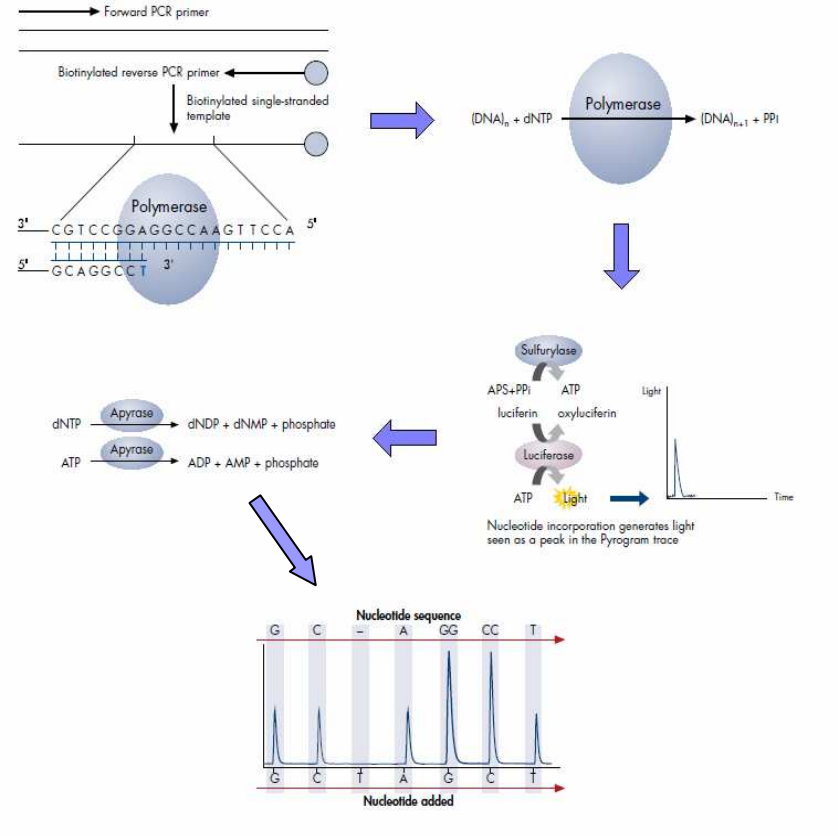
KRAS mutation analysis was performed using two different technologies. The first method was based on pyrosequencing using PyroMark KRAS kit while the second was based on Real-Time PCR using DxS TheraScreen KRAS mutations kit.
Both kits detect *KRAS* mutation in a background of wild-type genomic DNA in codon 12 and 13. Pyrosequencing technology is also able to detect mutations in codon 61.
Both kits are CE- marked diagnostic tests in accordance with the European Union *in vitro* Diagnostic (IVD) Medical Device Directive 98/79/EC.

5.6.1 Pyrosequencing Technology

Pyrosequencing is a technology able to sequence DNA fragment by chemical synthesis. The first step of the method is the DNA target amplification by PCR with a biotinylated primer. The biotinylated

single-stranded PCR amplicons are purified through sepharose beads covered of streptavidin. Then a sequencing primer is hybridized to the amplicon that serves as a template for sequencing. The sequencing reaction contains enzymes (DNA polymerase, ATP sulfurylase, luciferase, and apyrase) as well as their substrates (adenosine 5' phosphosulfate, APS and luciferin).

Figure 23. Pyrosequencing Tecnology scheme.
Target is amplify by PCR using a biotinylated primer. Single biotinylated strand target is purified and placed into PyroMark instrument for sequencing. Polymerase enzyme catalyzes the reaction of sequencing by nucleotides incorporation. Each incorporations event starts some enzymatic reactions that at the end produce a profile called pyrogram. In this profile each peak corresponds to a nucleotide and high of peak is proportional to number of nucleotides incorporated.



The instrument adds deoxyribonucleotide triphosphate (dNTP) to the reaction; DNA polymerase catalyzes the incorporation of the deoxyribonucleotide triphosphate into the DNA strand, if it is complementary to the base in the template strand.
Each incorporation event is accompanied by release of pyrophosphate (PPi) in an equimolar quantity to the amount of incorporated nucleotide. ATP sulfurylase converts PPi to ATP in the presence of adenosine 5' phosphosulfate (APS). This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) chip and seen as a peak in the raw data output (Pyrogram). The height of each peak (light signal) is proportional to the number of nucleotides incorporated.
Apyrase, a nucleotide-degrading enzyme, continuously degrades unincorporated nucleotides and ATP. When degradation is complete, another nucleotide is added.
A schematic representation of the pyrosequencing technology is reported in Figure 23.

5.6.1.1 PyroMark KRAS Mutation kit

The kit is based on three main steps:

- Amplification of the interest region by PCR
- Purification of single strand DNA template
- Sequence analysis by Pyrosequencing instrument

The kit is able to detect and quantify KRAS mutation in codon 12, 13 and 61. The assay is design to detect the more frequent mutations which are in position 35 (second base of codon 12) or 38 (second base of codon 13). However a list of possible mutation detected for codon 12-13 is reported in Table 11 and for codon 61 in Table 12.

Table 11. Summary of possible mutation in codon 12-13 detected by the kit.

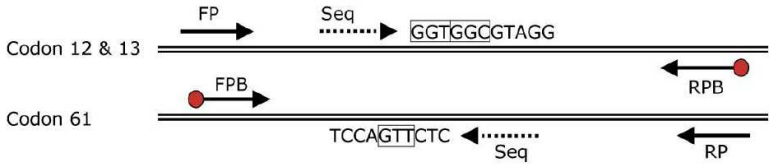
Codon	AA mutation	Nucleotide changed	Position
12	Gly12Asp	GGT>GAT	35
12	Gly12Val	GGT>GTT	35
12	Gly12Ala	GGT>GCT	35
12	Gly12Cys	GGT>TGT	34
12	Gly12Ser	GGT>AGT	34
12	Gly12Arg	GGT>CGT	34
13	Gly13Asp	GGC>GAC	38
13	Gly13Cys	GGC>TGC	37
13	Gly13Ser	GGC>AGC	37

Table 12. Summary of possible mutation in codon 61 detected by the kit.

Codon	AA mutation	Nucleotide changed		Position
		Forward	Reverse	
61	Gln61His	CAA>CAC	TTG>GTG	183
61	Gln61His	CAA>CAT	TTG>ATG	183
61	Gln61Leu	CAA>CTA	TTG>TAG	182
61	Gln61Arg	CAA>CGA	TTG>TCG	182
61	Gln61Pro	CAA>CCA	TTG>TGG	182

The kit includes two assay: one for codon 12-13 (which are sequential) and one for codon 61. Codons 12-13 are sequenced in forward direction while codon 61 in reverse direction as shown in the figure below.

Figure 24. Scheme of PyroMark KRAS assay.
Codons 12 and 13 are sequenced in the forward direction; codon 61 is sequenced in the reverse direction. FP: Forward primer; FPB: Forward biotinylated primer; RP: Reverse primer; RPB: Reverse biotinylated primer; Seq: Sequencing primer.



As first step, PCR reactions were carried out using 10ng of genomic DNA in a final volume of 20 µl using a biotinylated primer. Amplicon length of both assays is about 120bp which allow the analysis on degraded DNA deriving from FFPE samples.
Amplicons were immobilized on Streptavidin Sepharose High Performance beads. Single-stranded DNA was prepared, and the corresponding sequencing primers were added to the DNA. As the assays are independent two different sequencing primers were used. Then samples were analyzed by PyroMark Q24 Instrument.
Details regarding PCR and sequence preparation were followed as described in manufacturer's instructions and internal working instructions.

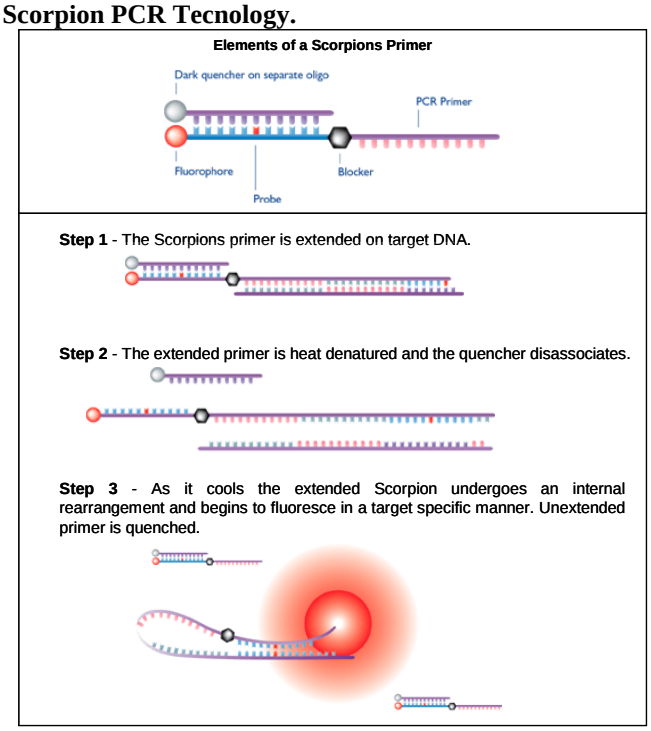
5.6.2 ARMS and Scorpion Technology

The second method that was validated for *KRAS* mutation was based on Real-Time PCR.
The kit used combined two different type of PCR technologies:

- **Scorpion Technology**
  Scorpions are bi-functional molecules containing a PCR primer covalently linked to a probe. The fluorophore in the probe interacts with a quencher which reduces fluorescence. During a PCR reaction the fluorophore and quencher are separated which leads to an increase in light output from the reaction tube. There are two formats for Scorpions. The figure below shows the bi-

molecular Scorpion format. The alternative is known as the unimolecular format in which an integral stem loop sequence is used to bring the quencher close to the fluorophore.

Figure 25. Scorpion PCR Tecnology.



- **A**mplification **R**efractory **M**utation **S**ystem or **ARMS** PCR.
  A common application of this method is the detection of individually point mutation. Taq DNA polymerase is extremely effective at distinguishing between a match and a mismatch at the 3'-end of a PCR primer. Specific mutated sequences can be selectively amplified, even in samples where the majority of the sequences do not carry the mutation as:
  • When the primer is fully matched, the amplification proceeds with full efficiency.
  • When the 3'-base is mismatched, only low level background amplification occurs.

5.6.2.1 DxS Therascreen for KRAS Mutation

DxS ThreraScreen KRAS Mutation Test Kit (CE marked) is based on a panel of 8 assays: one control reaction and 7 reactions to detect mutations present in codons 12 and 13 of the KRAS gene (mutations are listed in Table 13).

Table 13. Summary of mutations detected by DxS Therascreen kit.

Codon	AA mutation	Nucleotide changed
12	Gly12Asp	GGT>GAT
12	Gly12Val	GGT>GTT
12	Gly12Ala	GGT>GCT
12	Gly12Cys	GGT>TGT
12	Gly12Ser	GGT>AGT
12	Gly12Arg	GGT>CGT
13	Gly13Asp	GGC>GAC

All reaction mixes contain an exogenous control (internal control) assay labelled with JOE fluorophore which is detected by 7500 RealTime PCR System, Applied Biosystems. The control assay, labelled with FAM, is used to assess the total DNA in a sample. The control assay amplifies a region of exon 4 of the *KRAS* gene. The primers and probe were designed to avoid any known KRAS polymorphisms. Each mutation assay, labelled with FAM, contains one Scorpion plus one ARMS primer, for discrimination between the wild-type DNA and the mutant DNA detected by a real-time PCR assay.

For each sample one control reaction and seven primer specific reactions were prepared. In particular total volume for each reaction was 25 µl: 20 µl of mater mix plus 5 µl of sample (DNA concentration: 4-10 ng/ µl).

PCR profile included an initial hold step at 95°C for 4 minutes, a 2-step amplification for 45 cycles with a denaturation at 95°C for 30 seconds and annealing at 60°C for 1 minute.
The fluorescence was acquired at the 60°C step.
After amplification, for each sample, the presence of a mutation was assessed by calculating the DeltaCt value between each mutation assay and the control reaction assay as follow:
DeltaCt = [mutation Ct] – [control Ct]
A mutation is detected when the DeltaCt value is lower than the cut-off values provided by the kit.

6 RESULT AND DISCUSSION

6.1 SNP analysis on FFPE, H&E FFPE and serum

A panel of TaqMan SNP genotyping assays was validated on FFPE, H&E FFPE, and on serum samples. The validation program was carried out performing the following tests:

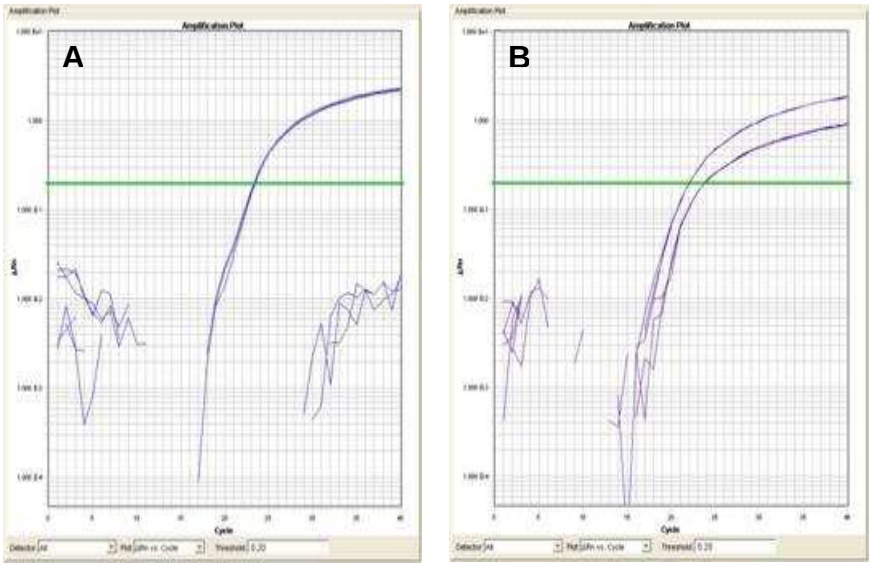
- Accuracy and Precision on good quality DNA positive control
- Impact of fragmented DNA and H&E staining
- Impact of limited quantity of material in FFPE samples
- Impact of serum
- Whole genome amplification and allelic drop-out

6.1.1 Accuracy and Precision on good quality DNA positive control

As a first step, the performance of each SNP assay was tested using good quality DNA extracted from cell cultures.
Samples were obtained from the Coriell Institute for Medical Research (http://www.coriell.org). These DNA samples were used as positive controls since their genotype is known for all gene tested.
In this work a panel of SNPs was investigated including polimorphisms in Fc fragment of IgG, low affinity (*FCGR2A* and *FCGR3A*), in EGFR pathway (*EGF* and *EGFR*) and finally in a key cell cycle regulator gene (*CCND1*). Assay accuracy and precision (intra and inter-run, different operator, different assay batches) were evaluated. According to FDA guidelines, it is important to demonstrate that assays are able to distinguish between a heterozygote and homozygote. To satisfy this condition at least two different homozygote and one heterozygote positive control sample were chosen for each assay, and accuracy and intra-assay precision were evaluated. TaqMan® technology was used to genotype positive control

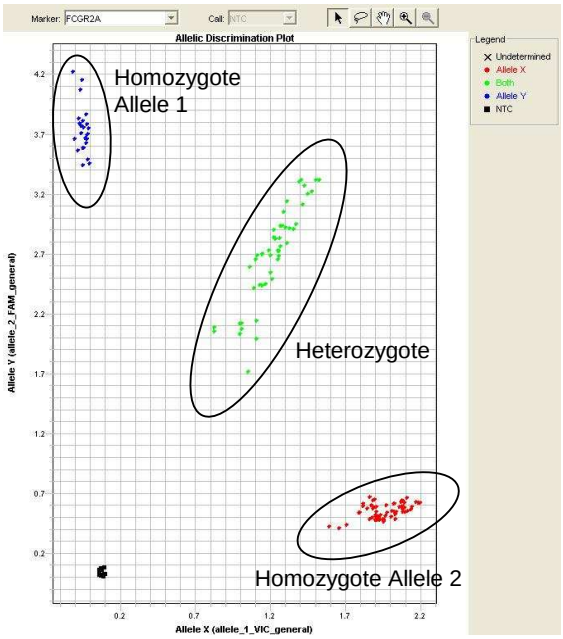
samples. During PCR amplification, a competitive hybridization of the two allelic-specific probes occurs. The amplification plot showed the presence of only a fluorochrome fluorescence increase in homozygote samples or fluorescence increase of both fluorochromes in heterozygote samples. An example of PCR amplification plot is shown in Figure 26.

Figure 26. Example of amplification plots of two different genotypes.
(A) The homozygote sample shows one amplification curve due to detection of one fluorochrome. **(B)** The heterozygote sample shows two amplification curves corresponding to the presence in the sample of the two alleles.



To assign the genotype of one sample, the software calculates the ratio between the end point fluorescence of the two fluorochromes. The result is the allelic discrimination plot as shown in Figure 27 in which samples are distributed in three clusters: two homozygotes (one for Allele 1 and one for Allele 2) and the central corresponding to heterozygote.

Figure 27. Example of allelic discrimination plot.
The X-axis shows the VIC fluorescence associated with the ancestral allele while the Y-axis shows the FAM fluorescence associated with the polymorphic allele. Samples were attributed as homozygote for the polymorphic allele (blue), heterozygote (green), and homozygote for the ancestral allele (red). Negative controls (no DNA) are given in black.



For each gene, results obtained testing control samples are summarized in Table 14.

Table 14. Validation of TaqMan® genotyping assays using Coriell genomic DNA.

Genomic DNA samples with known genotype are listed according to Coriell identification number. The expected genotype and validation results are reported for each assay.

FCGR2A rs1801274 assay C___9077561_20					
Control	**Expected genotype**	**Accuracy: Concordance of genotype**	**Intra-assay precision: Concordance of replicates**	**Ruggedness: Concordance of operators**	**Ruggedness: Concordance of batches**
NA10835	AA	Passed	Passed	Passed	Passed
NA18620	AA	Passed	Passed	Passed	Passed
NA10831	GA	Passed	Passed	Passed	Passed
NA19005	GA	Passed	Passed	Passed	Passed
NA10830	GG	Passed	Passed	Passed	Passed
NA18854	GG	Passed	Passed	Passed	Passed
FCGR3A rs396991 assay C__25815666_10					
Control	**Expected genotype**	**Accuracy: Concordance of genotype**	**Intra-assay precision: Concordance of replicates**	**Ruggedness: Concordance of operators**	**Ruggedness: Concordance of batches**
NA18532	AA	Passed	Passed	Passed	Passed
NA18949	AA	Passed	Passed	Passed	Passed
NA10831	CA	Passed	Passed	Passed	Passed
NA10835	CA	Passed	Passed	Passed	Passed
NA17237	CC	Passed	Passed	Passed	Passed
NA18561	CC	Passed	Passed	Passed	Passed
EGFR rs2227983 assay custom					
Control	**Expected genotype**	**Accuracy: Concordance of genotype**	**Intra-assay precision: Concordance of replicates**	**Ruggedness: Concordance of operators**	**Ruggedness: Concordance of batches**
NA18620	AA	Passed	Passed	Passed	Passed
NA18561	AA	Passed	Passed	Passed	Passed
NA10830	GA	Passed	Passed	Passed	Passed
NA19005	GA	Passed	Passed	Passed	Passed
NA18532	GG	Passed	Passed	Passed	Passed
NA07019	GG	Passed	Passed	Passed	Passed

EGF rs4444903 assay C__27031637_10					
Control	**Expected genotype**	**Accuracy: Concordance of genotype**	**Intra-assay precision: Concordance of replicates**	**Ruggedness: Concordance of operators**	**Ruggedness: Concordance of batches**
NA18532	AA	Passed	Passed	Passed	Passed
NA18563	GA	Passed	Passed	Passed	Passed
NA19005	GG	Passed	Passed	Passed	Passed
CCND1 rs9344 assay C____744725_1_					
Control	**Expected genotype**	**Accuracy: Concordance of genotype**	**Intra-assay precision: Concordance of replicates**	**Ruggedness: Concordance of operators**	**Ruggedness: Concordance of batches**
NA10830	AA	Passed	Passed	Passed	Passed
NA10831	GA	Passed	Passed	Passed	Passed
NA18561	GG	Passed	Passed	Passed	Passed

Concordance with the expected genotype, among replicates and among runs was the acceptance criteria selected for validation. As shown in Table 14, results for all assays satisfied the required conditions.

Additional tests were carried out to evaluate precision under different conditions. In particular, variability between two operators and concordance among TaqMan® assay batches were assessed. As reported in Table 14, it was observed that these conditions had no impact on results and each assay was accurate and reproducible.

6.1.2 Impact of fragmented DNA and H&E staining

To evaluate the performance of each SNP assay on degraded DNA extracted from FFPE and H&E FFPE samples, a panel of tissues from 10 donors was purchased from ProteoGenex, Inc. In particular, for each donor it was obtained:

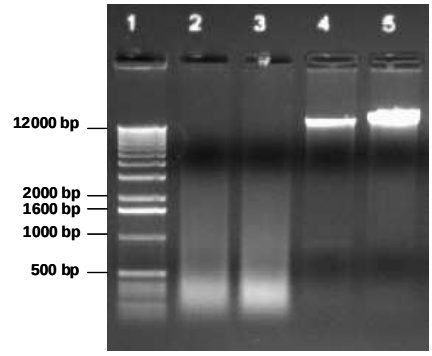
- Fresh Frozen Tissue (**FF**) in order to extract good quality genomic DNA used as reference;

- Unstained Formalin-Fixed Paraffin-Embedded (**FFPE**) tissue slices as source of degraded FFPE DNA;
- Hematoxylin and Eosin stained FFPE slices (**H&E FFPE**) to evaluate the impact of H&E staining on FFPE DNA quality.

All samples derived from gastric cancer donors.

As this validation is performed on unconventional samples it was not possible to obtain DNA extracted from FFPE slides of known genotype commercially. Therefore genotypes obtained from fragmented DNA extracted from FFPE samples were compared to the genotypes obtained from DNA extracted from FF, which was used as high quality genomic reference source. After extraction and quantification, electrophoresis analysis was performed to evaluate DNA quality. Comparison between FF and FFPE samples showed different DNA integrity: genomic DNA from FF is intact while DNA from FFPE is degraded into small fragments (Figure 28).

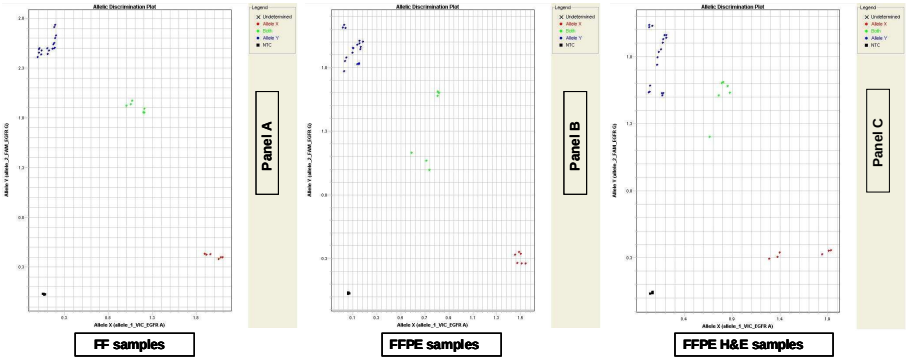
Figure 28. Quality DNA comparison between fresh frozen and FFPE samples. Agarose gel electrophoresis (1%) shows a well-defined band for FF samples and a smear for FFPE samples. Lane 1: DNA Ladder 1kb (Invitrogen); Lanes 2-3: DNA extracted from two FFPE samples while Lanes 4-5: DNA from the matching fresh frozen samples.



Given the poor DNA quality, assay performance was evaluated according to selected criteria. In particular, real samples validation was carried out by testing accuracy and repeatability, evaluating concordance between FFPE and FF, among replicates, different runs

and operators. For each assay, control DNA samples with known genotype were used as positive control and negative controls were also included. Results showed that fragmented DNA did not impact genotyping analyses as all acceptance criteria were met (Figure 29 Panel A-B, Table 15).

Figure 29. Examples of allelic discrimination analysis - *EGFR* polymorphism (rs2227893).
The X-axis shows the VIC fluorescence associated with the ancestral allele while the Y-axis shows the FAM fluorescence associated with the polymorphic allele. Samples were attributed as homozygote for the polymorphic allele (blue), heterozygote (green), and homozygote for the ancestral allele (red). Negative controls (no DNA) are given in black. Panel (**A**) shows FF samples, panel (**B**) FFPE and panel (**C**) H&E FFPE of the same DNA samples. DNA samples were tested at 10 ng.



Another source of genomic DNA evaluated for these assays was Haematoxylin and Eosin staining FFPE (H&E). Unlike unstained FFPE, histological slides are often used for pathology evaluation. For this reason H&E FFPE are archived for many years and might be the only biological sources of clinical studies carried out in the past. The disadvantage is that the tissue is damaged by formaldehyde effects and analysis might be affected by staining with H&E (Banaschak et al., 2000). To evaluate the impact of the staining on results reliability, SNP analysis was conducted on DNA extracted from H&E stained slides. Results were then compared to the genotype obtained from unstained slides of the same donor (Table 15). As shown in Figure 29,

no impact on genotype analysis was due to H&E staining. For these samples accuracy and repeatability were tested and all the acceptance criteria were met. Samples are listed in Table 15.

Table 15. EGFR polymorphism (rs2227893) genotyping results on commercial samples.
Samples are listed according to Proteogenex, Inc. reference number. For each sample different sources of tissue were available: Fresh Frozen (FF), Formalin-Fixed Paraffin-Embedded (FFPE) and Hematoxylin and Eosin (H&E) stained FFPE.

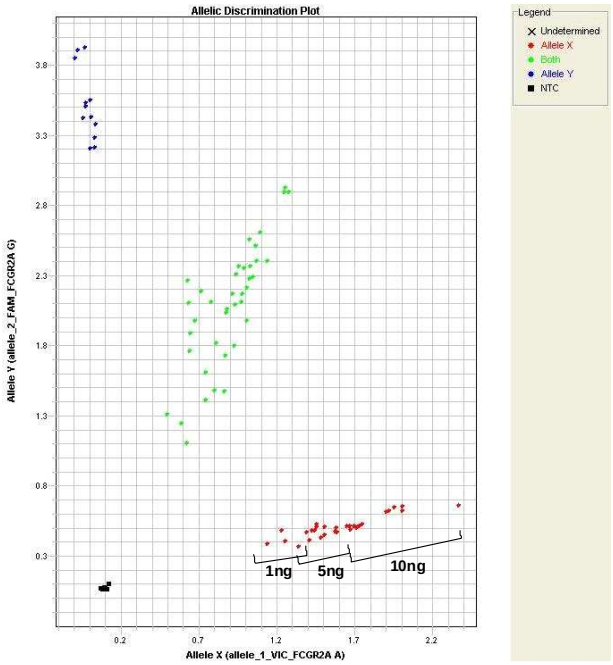
Donor ID	FF	FFPE	H&E FFPE	Concordance
8312	GG	GG	GG	Passed
8315	AG	AG	AG	Passed
8323	GG	GG	GG	Passed
8324	GG	GG	GG	Passed
8325	GG	GG	GG	Passed
8340	GG	GG	GG	Passed
8341	AA	AA	AA	Passed
8346	AA	AA	AA	Passed
8352	AG	AG	AG	Passed
8356	GG	GG	GG	Passed

6.1.3 Impact of limited quantity of material in FFPE samples

Additional analyses were also performed to validate each SNPs assay with lower amounts of DNA (1, 5 and 10 ng). This was aimed at validating the assay for samples with a limited amount of material. The evaluation of the method robustness using a low amount of DNA is extremely important to avoid allelic drop effects, which may cause the detection of only one allele in heterozygote samples (Lips et al., 2005).
As shown in Figure 30, the limited amount of DNA (down to 1 ng) did not impact the genotype call since all samples were correctly assigned in all analytical runs.

However, since late amplification curves could potentially have a negative impact on the assay performance, it was recommended to perform the SNP assay with no less than 10 ng of DNA sample.

Figure 30. Examples of analysis on decreasing amounts of DNA – *FCGR2A* polymorphism (rs1801274).
Allelic discrimination of FFPE samples analyzed at 1, 5 and 10 ng. Coriell positive controls and negative controls (no DNA, in black) were included in the panel. Due to the decreasing amount of DNA, clusters were spread with fluorescence levels closer to negative controls as shown for the allele_1 homozygote samples.



6.1.4 Impact of serum

In the present work serum was also evaluated as an alternative source of DNA. Serum is routinely used in clinical diagnostics for protein biomarkers (i.e. oncology) and it could be used for genomic analysis when whole blood is not available.
However, using serum as source of DNA requires the following considerations.

First of all DNA detectable in serum is free circulating DNA (CFDNA) which is present in healthy individual at low concentration (ng/ml) while levels are higher in cancer patients, in arthritis, hepatitis. Moreover the mechanism of DNA released is not fully understood. The two major hypotheses are apoptosis/necrosis and the release of intact cells in the bloodstream and their subsequent lysis.

Second the DNA must be separated from large amounts of serum protein, many of which are nucleolytic and must be neutralize to avoid degradation.

To test the feasibility of genomic analysis from serum, first whole blood was collected from 35 healthy donors. One aliquot of whole blood was kept for DNA extraction; another aliquot was used for serum preparation. Whole blood DNA was used as positive control.

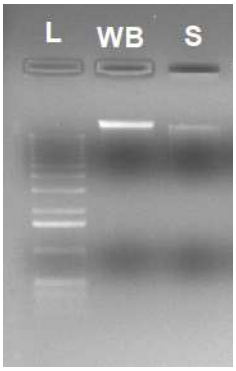
DNA was extracted from whole blood thought Maxwell 16 (Promega). Average yield on 35 samples was 50 ng/µl and ratio 260/280 > 1.8 meaning that genomic DNA has a good quality.

DNA was extracted from serum using QIAamp® DNA Investigator kit (QIAGEN®) and quantified with Nanodrop spectrophotometer.

For all serum samples total yield and quality were very low (yield < 100ng and ratio 260/280<1.6). However, the quality control performed on serum samples showed that the DNA is not degraded. Indeed, in the gel electrophoresis analysis, there was only a band in whole blood as well as in serum corresponding to the genomic DNA (

Figure 31). The band was stronger for WB due to the quantity of amount of DNA.

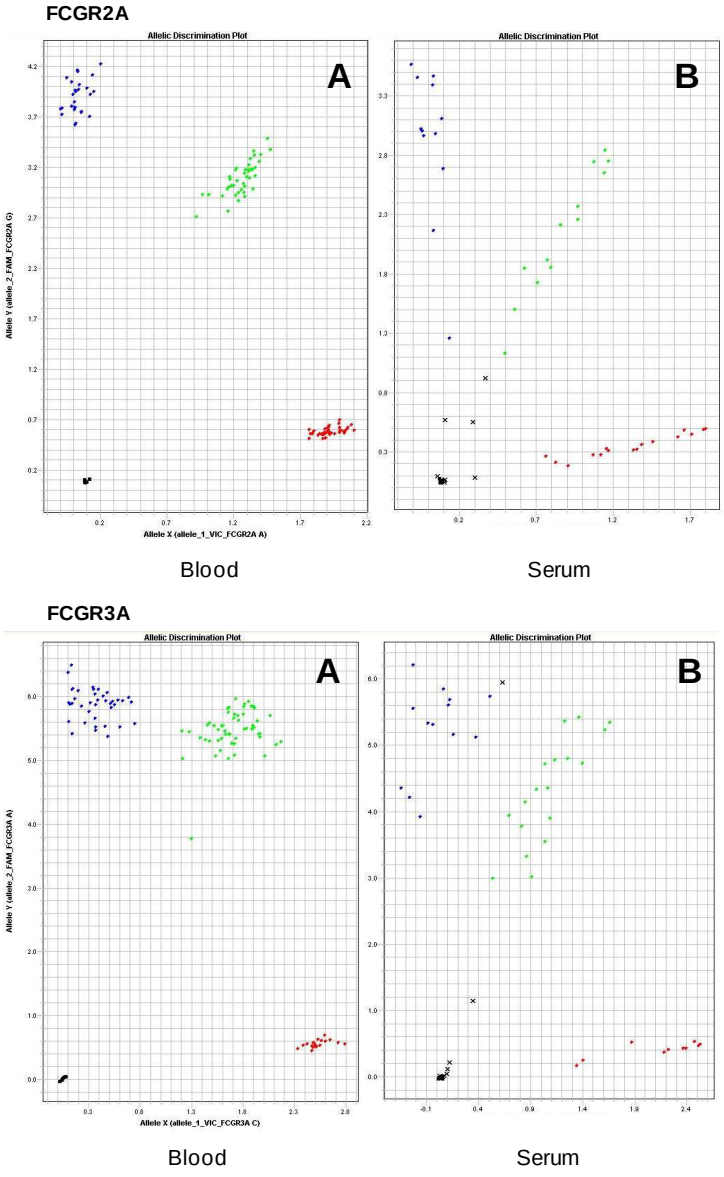
Figure 31. Quality DNA comparison between WB and serum.
Agarose gel electrophoresis (1%) shows a well-defined band for blood as well as serum sample. Lane L: DNA Ladder 1kb (Invitrogen); Lanes WB: DNA extracted from blood sample while Lanes S: DNA from the matching serum sample.



Genotyping analysis on two genes (*FCGR2A* and *FCGR3A*) was performed on both types of samples. Blood was diluted and analyzed at 25 ng for reaction while for each sample of serum 3.5µl were used regardless of the concentration. This discrepancy was due to the low serum extraction yield.

As shown in Figure 32 blood showed a good cluster separation for both assay while for serum clusters were less defined.

Figure 32. *FCGR2A* (above) and *FCGR3A* (belove) allelic discrimination comparison between DNA from blood and serum.
All pictures showed that while blood DNA (A) is a good source for genotyping serum (B) has low quality. While for blood clusters are well-distinguish for serum samples are spread and in some case near to negative control.



For each assay, all 35 samples analyzed for blood were assigned, while for serum 26 were assigned for *FCGR2A* and 27 for *FCGR3A* (about 70%); the remaining samples resulted undetermined.

A complete list of samples and genotype is shown in Table 16.

Table 16. List of sample from blood and serum and their genotype.
On the left are listed samples and genotype for *FCGR2A*, on the right those for *FCGR3A*.

	FCGR2A		**FCGR3A**	
Sample ID	**Call WB**	**Call SERUM**	**Call WB**	**Call SERUM**
ID1	Both	Undetermined	allele_2_FCGR3A A	Undetermined
ID3	allele_1_FCGR2A A	Undetermined	Both	Undetermined
ID4	Both	Undetermined	Both	Undetermined
ID5	allele_2_FCGR2A G	allele_2_FCGR2A G	Both	Both
ID6	Both	Undetermined	Both	Undetermined
ID7	allele_2_FCGR2A G	allele_2_FCGR2A G	allele_2_FCGR3A A	allele_2_FCGR3A A
ID8	allele_2_FCGR2A G	allele_2_FCGR2A G	allele_2_FCGR3A A	allele_2_FCGR3A A
ID9	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_1_FCGR3A C	allele_1_FCGR3A C
ID10	Both	Both	Both	Both
ID11	Both	Undetermined	allele_2_FCGR3A A	allele_2_FCGR3A A
ID12	Both	Both	allele_1_FCGR3A C	allele_1_FCGR3A C
ID13	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_2_FCGR3A A	allele_2_FCGR3A A
ID14	Both	Both	Both	Both
ID15	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID16	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID17	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_1_FCGR3A C	allele_1_FCGR3A C
ID18	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID19	allele_1_FCGR2A A	Undetermined	Both	Undetermined
ID20	Both	Both	Both	Both
ID21	allele_1_FCGR2A A	Undetermined	Both	Undetermined
ID22	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID23	allele_2_FCGR2A G	allele_2_FCGR2A G	allele_2_FCGR3A A	allele_2_FCGR3A A
ID24	Both	Both	Both	Both
ID25	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_1_FCGR3A C	allele_1_FCGR3A C
ID26	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_2_FCGR3A A	allele_2_FCGR3A A
ID27	Both	Undetermined	allele_2_FCGR3A A	Undetermined
ID28	Both	Both	allele_2_FCGR3A A	allele_2_FCGR3A A
ID29	allele_2_FCGR2A G	allele_2_FCGR2A G	allele_2_FCGR3A A	allele_2_FCGR3A A
ID30	Both	Both	Both	Both
ID31	allele_2_FCGR2A G	allele_2_FCGR2A G	Both	Both
ID32	allele_2_FCGR2A G	Undetermined	Both	Undetermined
ID33	Both	Both	Both	Both
ID34	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID35	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_1_FCGR3A C	allele_1_FCGR3A C
ID36	Both	Both	allele_2_FCGR3A A	allele_2_FCGR3A A

Concordance of genotype between DNA sources was evaluated. Agreement was observed for all assigned samples. However, SNP analysis from serum did not result robust since many samples were not assigned.

The low percentage of call in serum could be attributed to low quality and quantity DNA. In the attempt to overcome this problem, whole genome amplification was applied to serum samples. As control the same approach was performed for whole blood DNA.

6.1.5 Whole genome amplification and allelic drop-out

Whole genome amplification was used on DNA extracted from blood and from serum using the REPLI-g® kit from QIAGEN. This method is based on a linear multiple displacement amplification (MDA).

For each blood DNA samples 25 ng were used as starting material before amplification. For serum samples, 10 μL were directly used after extraction for the amplification procedure.

As the amplification produced a great DNA quantity, a sample dilution was necessary before performing genotyping analysis. During preliminary phase, amplified DNA was functionally tested and it was established for both blood and serum a 1:100 dilution before analysis.

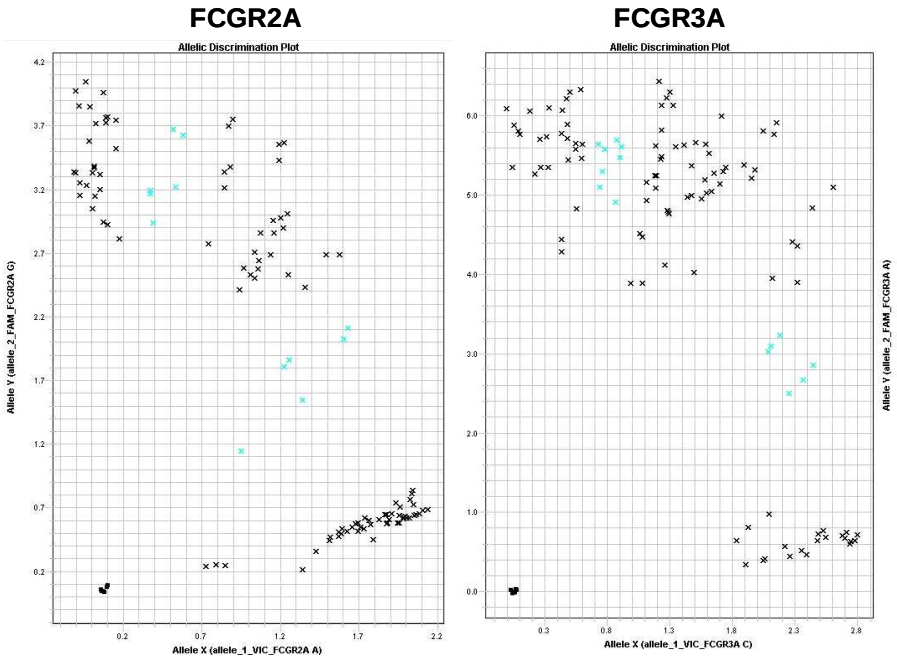
Amplified samples were then diluted and analyzed for *FCGR2A* and *FCGR3A* genotyping.

Results showed that all blood samples were correctly assigned since concordance between amplified and not amplified blood samples was observed.

On the contrary, genotyping of amplified serum samples showed a more complex situation. First of all for both assays the software was not able to assign automatically the genotype. As show in Figure 33 clusters were spread and not well-defined. In this situation it is not possible to distinguish between homozygote and heterozygote classes.

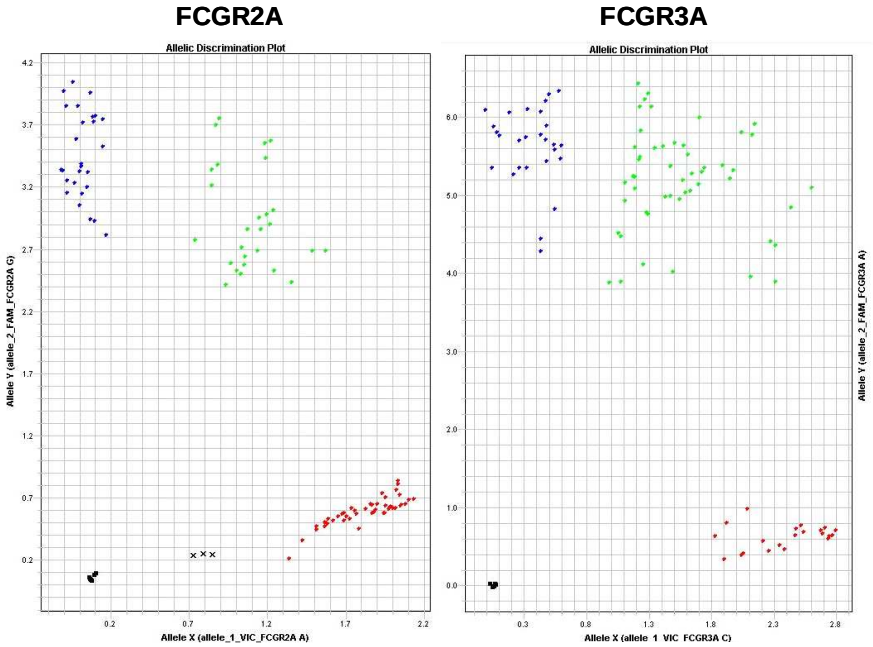
In order to assign the genotype, borderline samples were excluded from the analysis (highlight in light blue in Figure 33).

Figure 33. Allelic discrimination plots of amplified serum.
On the left it is shown the plot for *FCGR2A* while on the right for *FCGR3A*. Samples appeared spread and it is not possible distinguish clusters. In both panel borderline samples excluded from the analysis are highlight in light blue.



Excluding four samples for *FCGR2A* and five samples for *FCGR3A* (run in triplicates), it was possible to assign the genotype as shown in Figure 34.

Figure 34. Allelic discrimination plots from serum after exclusion of borderline samples
On the left it is shown the plot for *FCGR2A* while on the right for *FCGR3A*. Samples were attributed as homozygote for the FAM fluorophore (blue), heterozygote (green), and homozygote for VIC fluorophore (red). Negative controls (no DNA) are given in black square while undetermined samples in black cross-shaped .



With this approach, one sample still resulted undetermined for *FCGR2A*, while the others were assigned as reported in Table 17.
Comparing genotype obtained from blood and those obtained from serum there were some discrepancies. Discordance of genotype was observed in one sample for *FCGR2A* and four samples for *FCGR3A*. The cause of this problem could be attributed to allellic drop-out effect in which allele imbalance produce amplification of only one of the two alleles as seen in the ID3 and ID14 sample for *FCGR3A* analysis. For other samples it has been observed an allele switch or a heterozygote gain. The possible explanation of this change in genotype may have been an inefficient amplification during WGA due

to the variable presence of DNA fragments or protein presence in the serum samples (Li et al., 2006; Lips et al., 2005; Lu et al., 2005).

All together, these data showed that whole genome amplification is not a reliable approach for determining the genotype of limiting DNA material obtained from serum. For this reason it is recommended to genotype serum sample without any amplification.

Regarding the high percentage of failed samples, it has to be considered that in the present study, circulating DNA was obtained from serum of healthy volunteers. In clinical study conducted on patients (i.e. oncology, arthritis or hepatitis) it is possible that higher amount of DNA will be extracted reducing the failure rate.

Table 17. List of sample and their genotype obtained from whole genome amplification.
On the left are listed samples and genotype for *FCGR2A*, on the right those for *FCGR3A*. Samples underlined in red are excluded from the analysis while samples highlight in yellow showed a different genotype between amplified serum and amplified blood.

	FCGR2A		FCGR3A	
Sample ID	**Amplified Call SERUM**	**Amplified Call WB**	**Amplified Call SERUM**	**Amplified Call WB**
ID1	Excluded	Both	allele_2_FCGR3A A	allele_2_FCGR3A A
ID3	allele_2_FCGR2A G	allele_1_FCGR2A A	allele_2_FCGR3A A	Both
ID4	Both	Both	Both	Both
ID5	allele_2_FCGR2A G	allele_2_FCGR2A G	Both	Both
ID6	Excluded	Both	Both	Both
ID7	allele_2_FCGR2A G	allele_2_FCGR2A G	allele_2_FCGR3A A	allele_2_FCGR3A A
ID8	allele_2_FCGR2A G	allele_2_FCGR2A G	Excluded	allele_2_FCGR3A A
ID9	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_1_FCGR3A C	allele_1_FCGR3A C
ID10	Both	Both	Both	Both
ID11	Undetermined	Both	allele_1_FCGR3A C	allele_2_FCGR3A A
ID12	Both	Both	allele_1_FCGR3A C	allele_1_FCGR3A C
ID13	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	allele_2_FCGR3A A
ID14	Excluded	Both	allele_2_FCGR3A A	Both
ID15	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID16	allele_1_FCGR2A A	allele_1_FCGR2A A	Excluded	Both
ID17	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_1_FCGR3A C	allele_1_FCGR3A C
ID18	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID19	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID20	Excluded	Both	Both	Both
ID21	allele_1_FCGR2A A	allele_1_FCGR2A A	Excluded	Both
ID22	allele_1_FCGR2A A	allele_1_FCGR2A A	Excluded	Both
ID23	allele_2_FCGR2A G	allele_2_FCGR2A G	Excluded	allele_2_FCGR3A A
ID24	Both	Both	Both	Both
ID25	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_1_FCGR3A C	allele_1_FCGR3A C
ID26	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_2_FCGR3A A	allele_2_FCGR3A A
ID27	Both	Both	allele_2_FCGR3A A	allele_2_FCGR3A A
ID28	Both	Both	allele_2_FCGR3A A	allele_2_FCGR3A A
ID29	allele_2_FCGR2A G	allele_2_FCGR2A G	Both	allele_2_FCGR3A A
ID30	Both	Both	allele_1_FCGR3A C	Both
ID31	allele_2_FCGR2A G	allele_2_FCGR2A G	Both	Both
ID32	allele_2_FCGR2A G	allele_2_FCGR2A G	Both	Both
ID33	Both	Both	Both	Both
ID34	allele_1_FCGR2A A	allele_1_FCGR2A A	Both	Both
ID35	allele_1_FCGR2A A	allele_1_FCGR2A A	allele_1_FCGR3A C	allele_1_FCGR3A C
ID36	Both	Both	allele_2_FCGR3A A	allele_2_FCGR3A A

6.2 Mutations analysis on FFPE and H&E FFPE

Mutations in the *KRAS* oncogene are frequently found in human cancers and the presence of these mutations correlates with a lack of response to certain EGFR inhibitor cancer therapies for example in metastatic colorectal cancer patients. Therefore KRAS status can predict which patients benefit (KRAS wild-type) or do not benefit (KRAS mutated) from anti-EGFR therapy. For this analysis different tissue type could be used such as fresh frozen samples or FFPE samples. These latter are the most used due to the fact that many samples are already available from tumor biopsy. Another source of genomic DNA evaluated for these assays was Haematoxylin and Eosin staining FFPE (H&E) often used for pathology evaluation and conserved for long period of time.

Since *KRAS* analysis was used for diagnostic analysis, an accurate validation of the method was required.

In the present work two methods based on different technology were used as described in following sections.

6.2.1 Validation of *KRAS* mutation analysis using Pyrosequencing

The analytical performances of *KRAS* mutation analysis was evaluated following the Guidance for Industry on Pharmacogenetic tests and genetic tests for Heritable Markers and on Validation of Analytical Procedures.

In particular, for the validation of qualitative technique such as the mutation detection, the following parameters were considered: accuracy, precision, repeatability (intra-assay precision under the same operating conditions) and ruggedness (reproducibility between two different operators).

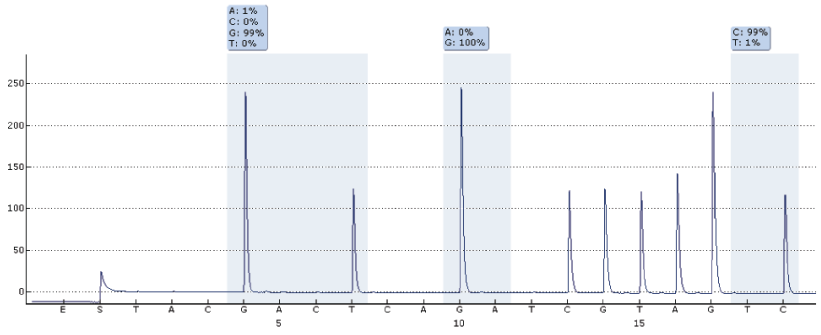
These tests were performed first on good quality genomic DNA obtained from cell line and then on real samples. In particular, the

impact on the analysis of FFPE tissue samples and H&E staining FFPE tissue samples was evaluated.
Additionally, in order to evaluate the assays sensitivity the analysis of different tumor percentage (sensitivity test) and limiting amount of material was performed on known *KRAS* status samples
The analyses were performed on both codon 12/13 and codon 61.

6.2.1.1 Accuracy and Precision on good quality DNA positive control

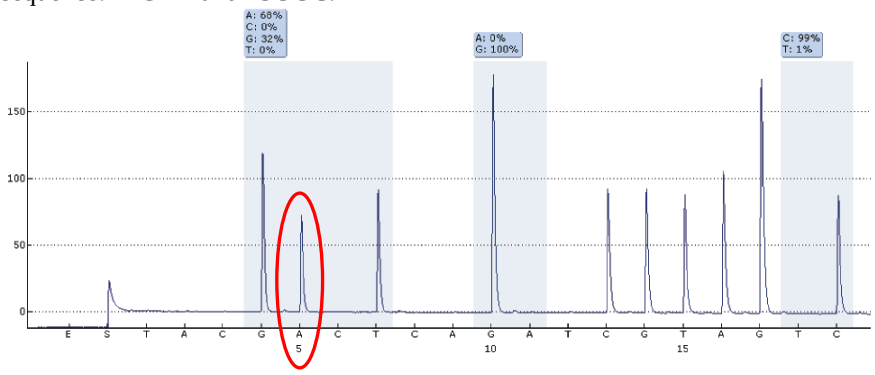
The performance characteristics of the assays were tested using good quality DNA samples of known genotype extracted from cell cultures. In particular, accuracy and precision (intra and inter-run, different operator) were evaluated for both codons.
In the following figures, some examples of pyrograms from cell line samples are reported. In particular, Figure 35 shows the analysis results for a wild-type sample (HT-29) in which the mutation region is highlighted in light blue. The sample resulted wild-type since codon 12 sequence was GGT and codon 13 resulted GGC. Height of peak was proportional to the number of base incorporated and accordingly to sequence. Indeed, peak of nucleotide G was twice as high as peak of nucleotide T in codon 12 and peak of nucleotide G was twice as high as peak of nucleotide C in codon 13.

Figure 35. *KRAS* pyrogram of sample HT-29 for codon 12-13 assay.
Analysis result shows that this DNA from cell line is wild-type with a codon sequence: 12GGT and 13GGC.



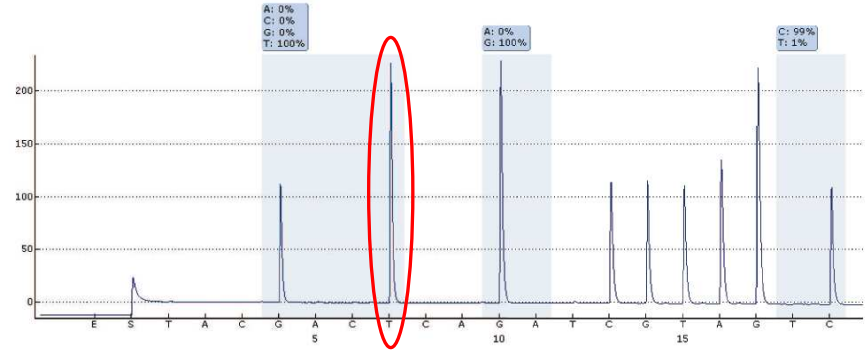
An example of analysis of a mutated cell line is reported in Figure 36. The mutation present in the cell line LS-174 was detected in codon 12. Indeed it was possible to observe the presence of a peak in position 2 of codon 12 (GGT>GAT).

Figure 36. *KRAS* pyrogram of sample LS-174 for codon 12-13 assay.
Analysis result shows that this DNA from cell line is mutated with a codon sequence: 12GAT and 13GGC.



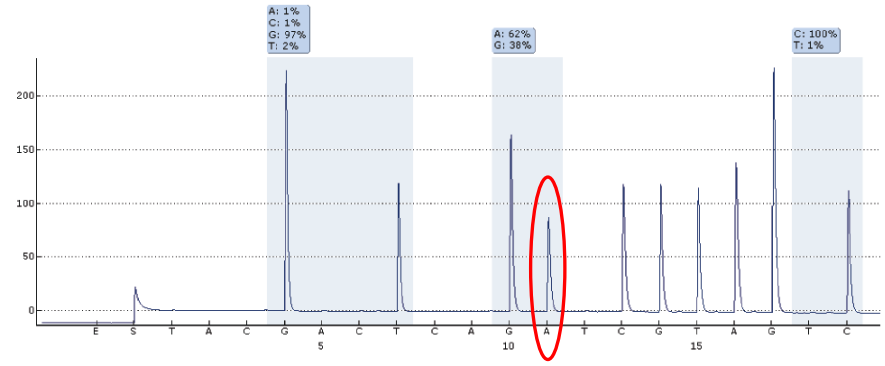
Another example is shown in Figure 37 where the mutation present in the cell line SW480 was detected in position 3 of codon 12 (GGT>GTT). In this case the pyrogram showed a decrease in peak corresponding to nucleotide G and an increase in nucleotide corresponding to nucleotide T.

Figure 37. *KRAS* pyrogram of sample SW480 for codon 12-13 assay.
Analysis result shows that this DNA from cell line is mutated with a codon sequence: 12GTT and 13GGC.



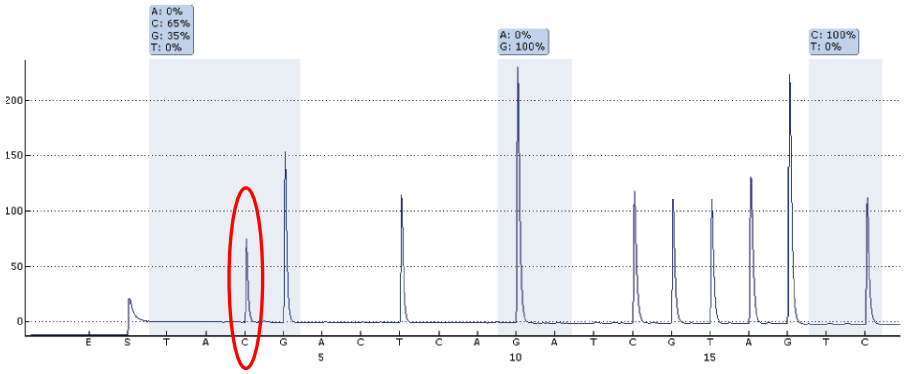
If a mutation was detected in codon 13 it was possible to observe the presence of a peak in position 1 corresponding to nucleotide A and mutation 13ASP as shown in Figure 38 (cell line HCT116).

Figure 38. KRAS pyrogram of sample HCT116 for codon 12-13 assay.
Analysis result shows that this DNA from cell line is mutated in codon 13 with a codon sequence GAC while it is wild-type for codon 12 (sequence GGT).



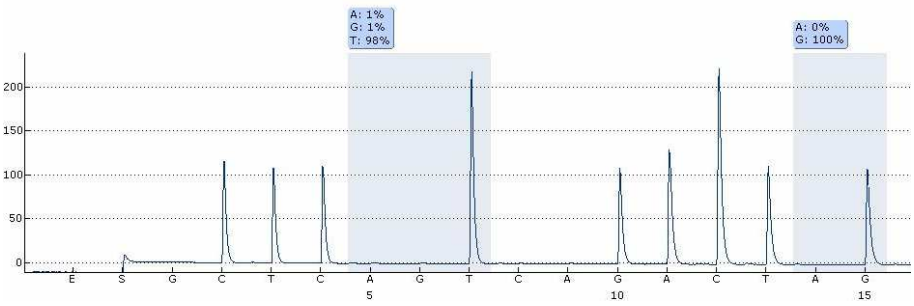
The mutation of cell line HuP-T3 was detected in the first position of codon 12. In this case a peak was visible in the first three positions of the program (GAC codifying for 12ARG) (Figure 39).

Figure 39. *KRAS* pyrogram of sample HuP-T3 for codon 12-13 assay.
Analysis result shows that this DNA from cell line is mutated in codon 12 with a codon sequence CGT while it is wild-type for codon 13 (sequence GGT).



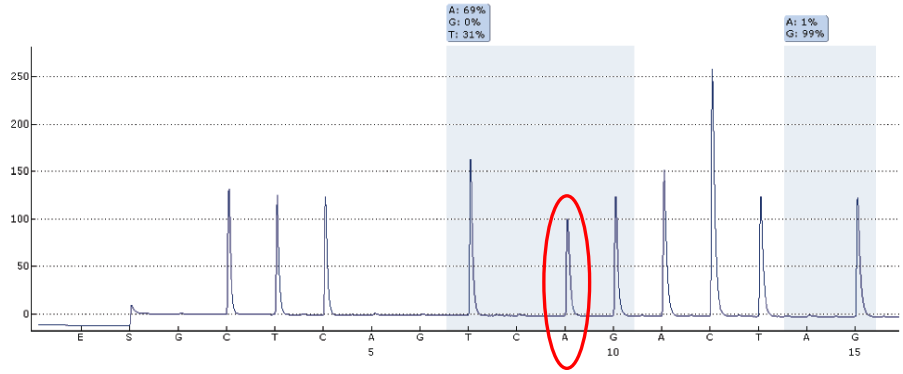
Regarding the analysis of codon 61, Figure 40 showed the result obtained from the wt cell line HT-29.

Figure 40. *KRAS* pyrogram of sample HT-29 for codon 61 assay.
Analysis result shows that this DNA from cell line is wild-type in codon 61 with a codon sequence CAA (reverse sequence in the pyrogram TTG).



In Figure 41 is reported an example of mutation detected in the second based of codon 61 from the cell line SW948 (TTG>TAG).

Figure 41. *KRAS* pyrogram of sample SW948 for codon 61 assay.
Analysis result shows that this DNA from cell line is mutated in with a codon sequence TAG.



The overall results obtained from cell line are reported in Table 18.
Cell lines were used as positive controls since they are well characterized as reported in the Catalogue of Somatic Mutations in Cancer (COSMIC).

Table 18. Summary results for accuracy, repeatability and second operator on cell lines DNA.
The expected *KRAS* status refers to which ones reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (http://www.sanger.ac.uk/genetics/CGP/cosmic/).

Sample ID	Expected *KRAS* status	Obtained *KRAS* status	Accuracy	Concordance of Replicates	Concordance of Operators
HT-29	WT	WT	Passed	Passed	Passed
HuP-T3	12ARG	12ARG	Passed	Passed	Passed
SW480	12VAL	12VAL	Passed	Passed	Passed
SW116	12ALA	12ALA	Passed	Passed	Passed
A549	12SER	12SER	Passed	Passed	Passed
LS-174T	12ASP	12ASP	Passed	Passed	Passed
HCT116	13ASP	13ASP	Passed	Passed	Passed
Calu-1	12CYS	12CYS	Passed	Passed	Passed
SW948	61LEU	61LEU	Passed	Passed	Passed

Results showed concordance with the expected genotype, among replicates and among runs (Table 18).
Additional tests were carried out to evaluate precision under different conditions. In particular, variability between two operators was assessed. As reported in Table 18, it was observed that these conditions had no impact on results and each assay was accurate and reproducible.

6.2.1.2 Impact of fragmented DNA and H&E staining

Since the DNA extracted from biological tissue such as FFPE has low quality (due to degradation, cross-link, limitation of material, methylol derivatives and PCR inhibitors presence) it is important to evaluate the performance of the mutation analysis method on real sample.
For this purpose some FFPE samples of different type of cancer, were commercially purchased. The complete list of samples is reported in Table 19.

Table 19. List of FFPE sample used during the validation.
The summary includes samples according to Proteogenex identification code. Moreover for each sample tumor origin is reported.

Sample ID	Source
21196	Colorectal cancer
21377	Colorectal cancer
66288	Colorectal cancer
21321	Colorectal cancer
66378	Colorectal cancer
21454	Colorectal cancer
66460	Colorectal cancer
21123	Colorectal cancer
21452	Colorectal cancer
13802T2	Cutaneous Malignant Melanoma
13823T2	Cutaneous Malignant Melanoma
13838T2	Cutaneous Malignant Melanoma

As no mutated sample in codon 61 was observed from a preliminary screening (data not shown), some selected FFPE samples (21454, 66460, 21123) which resulted WT for both codon 12-13 and 61 were spiked with SW948 cell line DNA to mimic the 61 LEU mutation.
Results from FFPE sample showed that the assay is robust also on real samples. Indeed, as shown in Figure 42 and Figure 43 mutation are clearly detected when present. Comparing these pyrograms with those obtained from cell line no impact was observed due to DNA degradation.

Figure 42. *KRAS* pyrogram of sample 21452 for codon 12-13 assay.
Analysis result shows that sample 21452 is wild-type with a codon sequence: 12GGT and 13GGC.

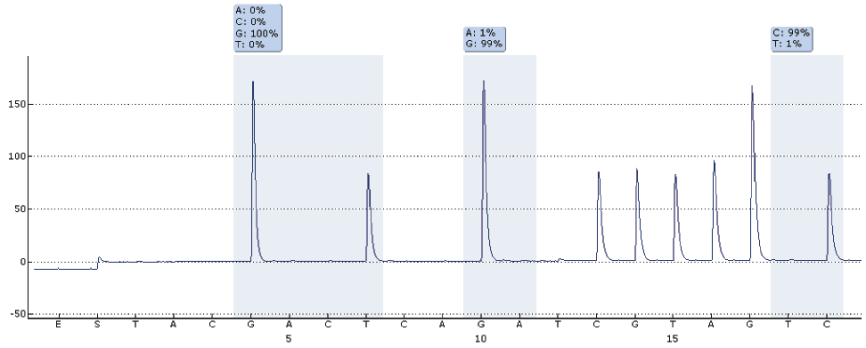
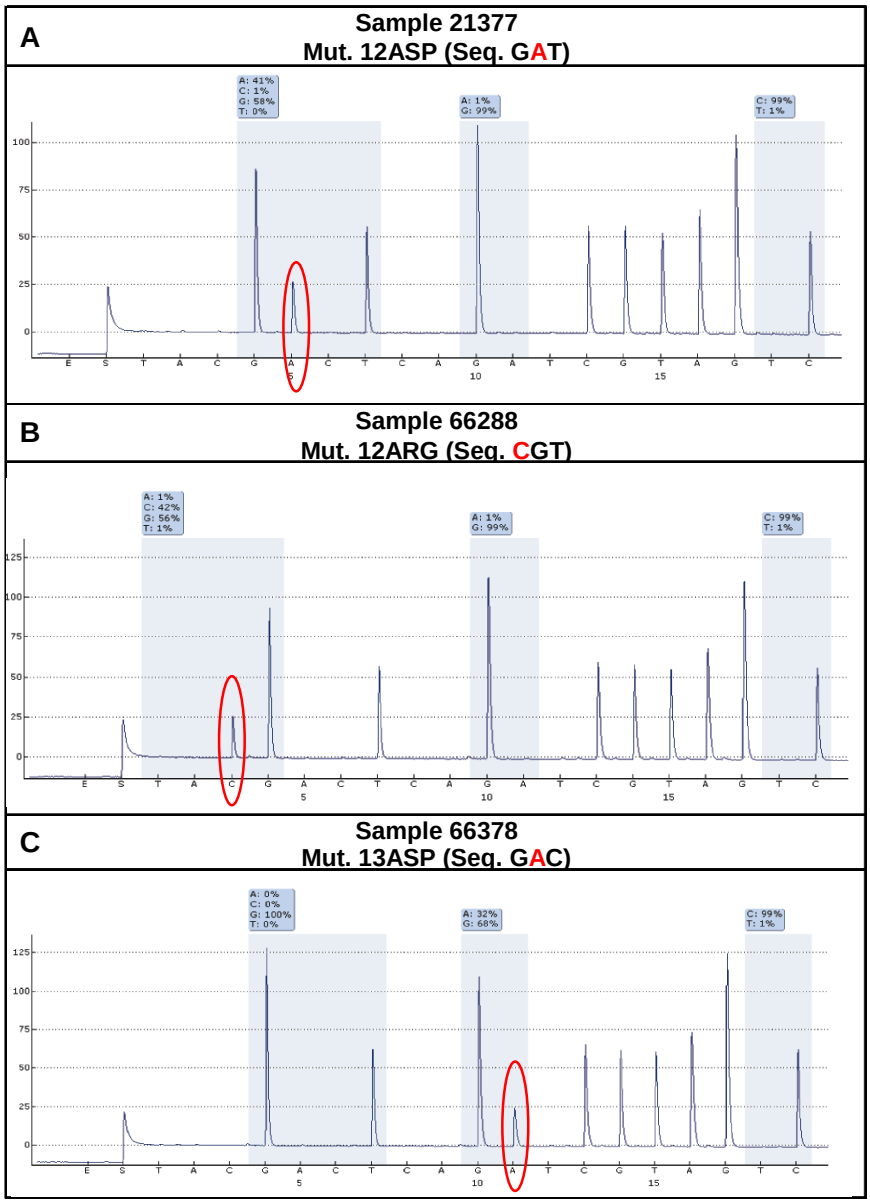


Figure 43. Examples of *KRAS* pyrogram of FFPE samples for assay 12-13.
(A) Analysis results shows that sample 21452 is mutated in codon 12 with sequence GAT corresponding to 12ASP mutation while the sample is wild-type for codon 13.
(B) Analysis results shows that sample 66288 is mutated in codon 12 with sequence CGT corresponding to 12ARG mutation while the sample is wild-type for codon 13.
(C) Analysis results shows that sample 66378 is mutated in codon 13 with sequence GAT corresponding to 13ASP mutation while the sample is wild-type for codon 12.



Concordance among runs and between operators was obtained for both codon 12-13 and codon 61 analysis. Accuracy was evaluated comparing results obtained on the same samples with an independent method (DxS see section 6.2.2).

As shown in Table 20 all analyses met the acceptance criteria.

Table 20. Summary results for accuracy, repeatability, second operator and FFPE impact.

(*) During validation this sample were spiked with SW948 cell line DNA to mimic 61LEU mutation.

(**) Samples were cross-validated using DxS Therascreen KRAS kit as independent analysis method.

Sample ID	Expected *KRAS* status	Obtained *KRAS* status	Accuracy**	Concordance of Replicates	Concordance of operator
21196	12ASP	12ASP	Passed	Passed	Passed
21377	12ASP	12ASP	Passed	Passed	Passed
66288	12ARG	12ARG	Passed	Passed	Passed
21321	12CYS	12CYS	Passed	Passed	Passed
66378	13ASP	13ASP	Passed	Passed	Passed
21454*	61LEU	61LEU	Passed	Passed	Passed
66460*	61LEU	61LEU	Passed	Passed	Passed
21123*	61LEU	61LEU	Passed	Passed	Passed
21452	WT	WT	Passed	Passed	Passed
13802T2	WT	WT	Passed	Passed	Passed
13823T2	WT	WT	Passed	Passed	Passed
13838T2	WT	WT	Passed	Passed	Passed

Finally, an additional test was performed to validate FFPE tissue stained with H&E. Mutational analysis was conducted on DNA extracted from stained slides of the same samples.

The figures below (Figure 44 and Figure 45) show the pyrograms comparison between FFPE and H&E FFPE for codon 12-13 and codon 61 while a summary of genotype obtained from stained FFPE is reported in Table 21. Concordance of genotype between stained and unstained was observed.

Figure 44. Comparison between FFPE and H&E FFPE for codon12-13.
No impact was due to DNA degradation (FFPE samples) or due to H&E staining.

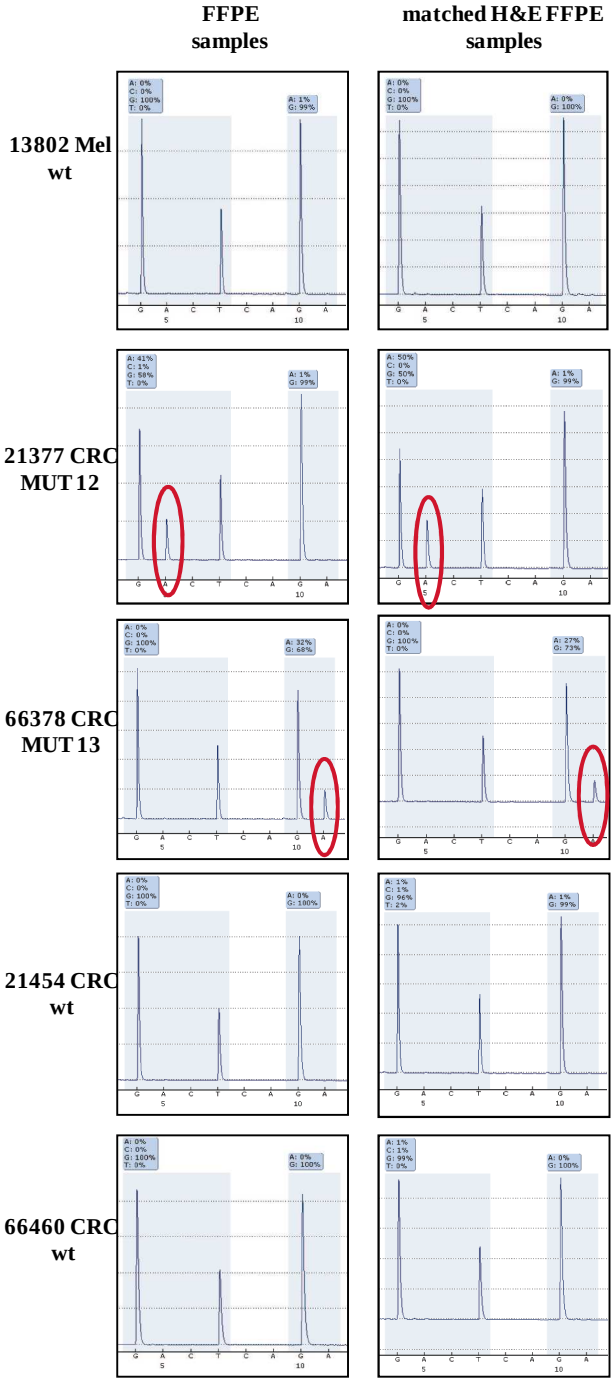


Figure 45. Comparison between FFPE and H&E FFPE for codon 61.
No impact was due to DNA degradation (FFPE samples) or due to H&E staining.

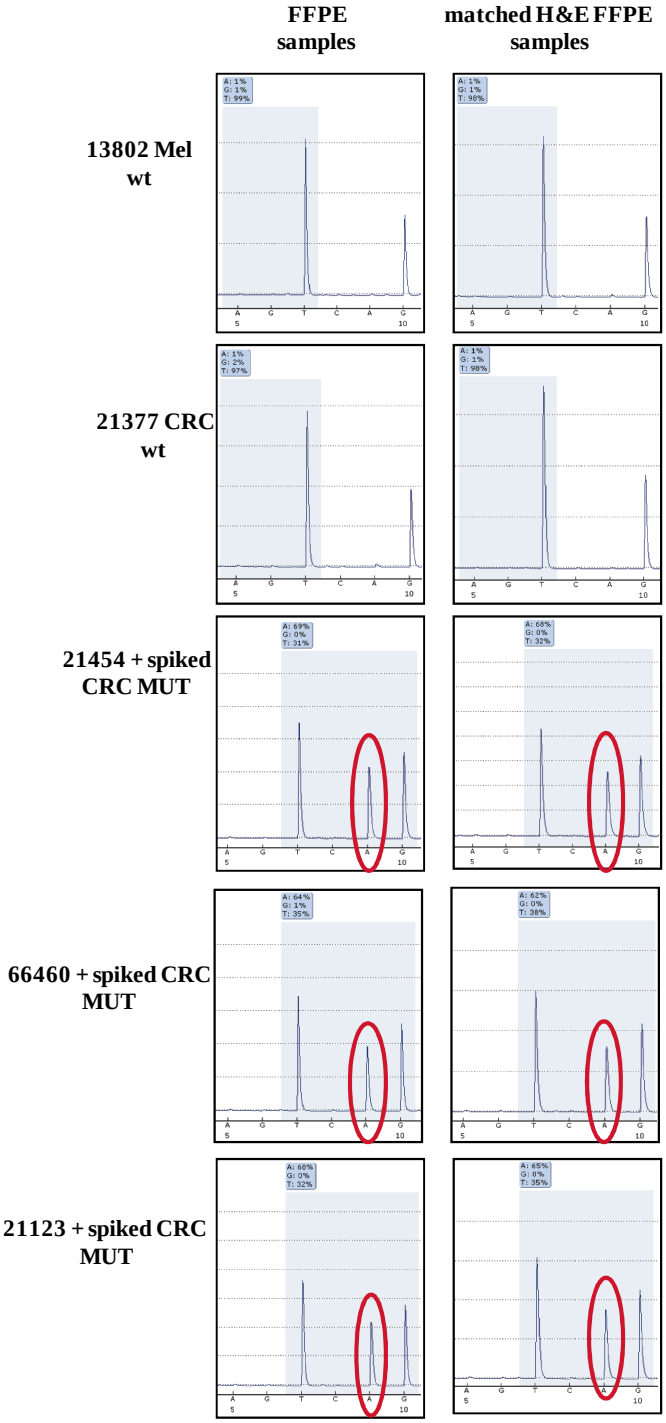


Table 21. Summary of *KRAS* status obtained from H&E FFPE.
As shown in table there was no genotype difference in KRAS status among unstained and stained FFPE samples. (*) During validation this sample were spiked with SW948 cell line DNA to mimic 61LEU mutation.

Sample ID	Expected KRAS status	Concordance with KRAS status obtained from H&E
21196	12ASP	Passed
21377	12ASP	Passed
66288	12ARG	Passed
21321	12CYS	Passed
66378	13ASP	Passed
21454*	61LEU	Passed
66460*	61LEU	Passed
21123*	61LEU	Passed
21452	WT	Passed
13802T2	WT	Passed
13823T2	WT	Passed
13838T2	WT	Passed

6.2.1.3 Impact of limited quantity of material in FFPE samples

In order to evaluate the impact of limited quantity of DNA, a mutated FFPE DNA sample (21196 mutated in codon 12) was serially diluted before analysis. In particular, the analysis was performed at 20 ng, 10 ng and 2 ng of starting DNA material (Figure 46).

The different amount of DNA did not impact the analysis as the profile of the pyrosequencing showed that PCR efficiently amplified the target also at the minimum concentration tested. Moreover the percentage between wild-type and mutated was maintained with the same ratio in all concentrations.
As mentioned before among FFPE samples screened there were no sample mutated in codon 61. To reproduce the same test of codon 12-13 instead of real sample DNA from cell line was used. Cell line DNA SW948, which is mutated in codon 61, was spiked into a wild-type DNA cell line sample (HT-29) with ratio 1:2.

Also this test confirmed that results were not impaired by a decrease in DNA quantity (Figure 47). Moreover the peak percentage of the mutation was comparable among different concentration (mutated base A 18%, 17%, 16%).

Figure 46. Impact of limited quantity of DNA for Codon 12.
Sample obtained by the mutated tumor FFPE DNA sample (21196) loaded at different concentrations: 20, 10 and 2 ng/reaction.

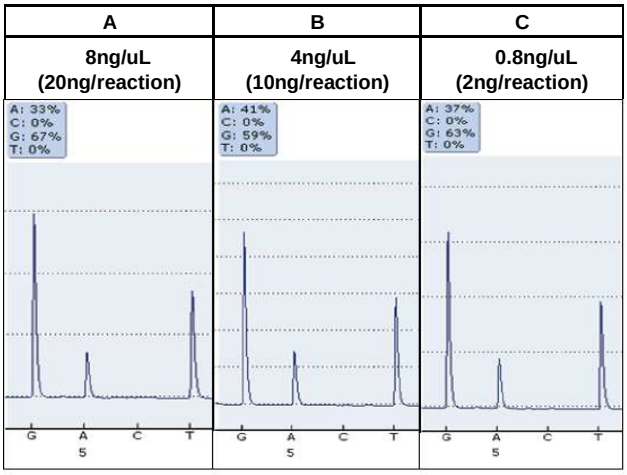
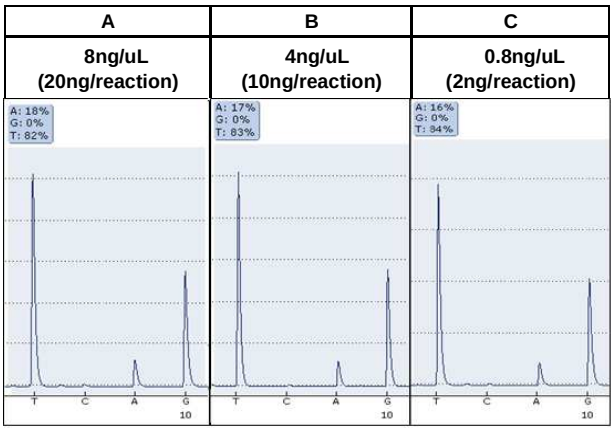


Figure 47. Impact of limited quantity of DNA for Codon 61.
Sample obtained by the mutated cell line DNA sample (SW948) was spiked with the wild-type cell line DNA sample (HT29) at a ratio 1:2 and loaded at different concentrations: 20, 10 and 2 ng/reaction.



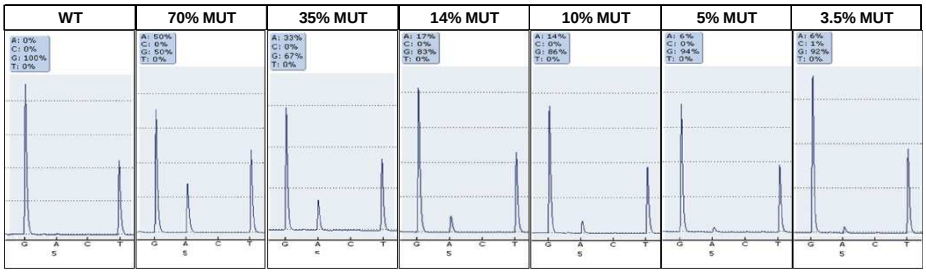
6.2.1.4 Impact on the analysis of different tumor percentage (Sensitivity test)

An additional test has been performed in order to evaluate the PyroMark KRAS mutation kit sensitivity and to determine the minimum percentage of tumor tissue necessary to detect a mutation in order to avoid false negative results.

DNA from a mutated FFPE sample (21196) has been diluted with DNA from wild-type FFPE sample (21123) at different ratio to mimic real samples with different tumor percentages (70, 35, 14, 10, and 5%). This test was performed on codon 12-13 assay and at the concentration of 20 ng/reaction.

As shown in Figure 48 the peak corresponding to mutation (nucleotide A) decreased according to the tumor percentage.

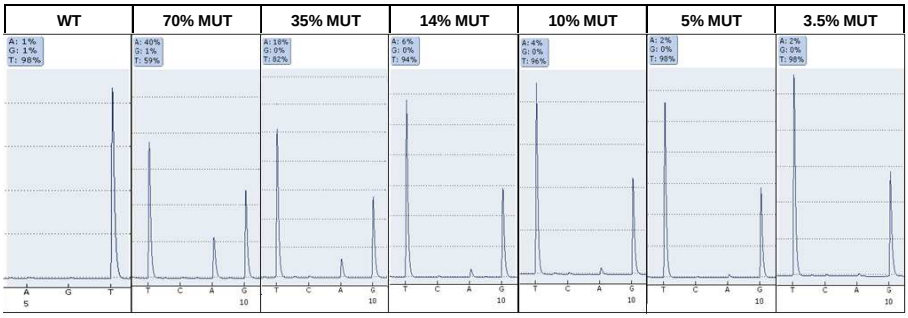
Figure 48. *KRAS* mutation sensitivity test on codon 12-13.
It is represented only the codon 12 area of the mutated sample 21196 (12GGT>GAT) and the wild-type sample (WT, 21123) at different tumor percentage.



The analysis on codon 61 was carried out diluting DNA a mutated cell line DNA (SW948) with a wild-type cell line DNA (HT-29) at different ratio to mimic the different tumor percentages (70, 35, 14, 10, and 5%). For each reaction 20ng of DNA were used.

As expected, the peak corresponding to mutation (nucleotide A) decreased according to tumor percentage and the wild-type nucleotide T peak showed an increase. Pyrograms are shown in Figure 49.

Figure 49. *KRAS* mutation sensitivity test on codon 61.
It is represented only the codon61 area of the mutated sample SW948 (61TTG>TAG) and the wild-type sample (WT, HT29).



Sensitivity test were repeated at different DNA concentration (20, 10, 2 ng/reaction).

Data obtained from the sensitivity experiments were used to establishing the limit of detection of the kit and the threshold mutation peak detection.

Considering the ground noise, a peak threshold for mutated base was established at 7%. In particular, if the peak percentage of all mutated bases was lower than 4% the sample was assigned as wild-type while if the percentage was between or equal to 4% and 7% sample resulted as not-assigned (N.A). If the peak percentage was higher than 7% the sample was assigned as mutated.

Using this threshold, the following results were obtained (Table 22 and Table 23).

Table 22. Pyrosequencing *KRAS* sensitivity test codon 12-13.
Results are accurate till a 10% percentage of tumor and at 10ng of DNA sample. N.A.: not assigned.

Sample concentration	70% MUT	35% MUT	14% MUT	10% MUT	5% MUT	3.5% MUT
20 ng/reaction	Mutated	Mutated	Mutated	Mutated	N.A.	N.A.
10 ng/reaction	Mutated	Mutated	Mutated	Mutated	Mutated	N.A.
2 ng/reaction	Mutated	Mutated	Mutated	Mutated	Wild-Type	N.A.

Table 23. Pyrosequencing KRAS sensitivity test codon 61.
Results are accurate till a 14% percentage of tumor and at 10ng of DNA sample. N.A.: not assigned.

Sample concentration	70% MUT	35% MUT	14% MUT	10% MUT	5% MUT	3.5% MUT
20 ng/reaction	Mutated	Mutated	Mutated	N.A.	Wild-Type	Wild-Type
10 ng/reaction	Mutated	Mutated	Mutated	N.A.	Wild-Type	Wild-Type
2 ng/reaction	Mutated	Mutated	Mutated	N.A.	Wild-Type	Wild-Type

Looking at this data and assay specifications, it could be establishing the minimum percentage of tumor needed to conduct the analysis to reduce the risk of false negative results. In particular the assay for codon12-13 was reliable till 10% of tumor while the assay for codon 61 till 14% of tumor.

6.2.2 Validation of *KRAS* mutation analysis using DxS Therascreen

The analytical performances of *KRAS* mutation analysis was evaluated following the Guidance for Industry on Pharmacogenetic tests and genetic tests for Heritable Markers and on Validation of Analytical Procedures.

For the validation of qualitative technique such as the mutation detection, the same parameters used for *KRAS* pyrosequencing validation were considered. In particular it has been evaluated intra-assay precision under the same operating conditions and reproducibility under different conditions.

Tests were performed first on good quality genomic DNA obtained from cell line and then on real samples. In particular the impact on the analysis of FFPE tissue samples and H&E staining FFPE tissue samples was evaluated. As these samples were not commercially characterized this method allows a cross-validation of results obtained by pyrosequencing. Indeed accuracy could be evaluated as an independent and different technology was used.

Additionally, in order to evaluate the assays sensitivity the analysis of different tumor percentage (sensitivity test) and limiting amount of material was performed on known *KRAS* status samples.

6.2.2.1 Accuracy and Precision on good quality DNA positive control

The first step of this validation was to evaluate the performance characteristics of genetic assays using good quality DNA samples of known genotype extracted from cell cultures. In particular, accuracy and precision (intra and inter-run, different operator, different assay batches) were evaluated.

Genomic DNA from different cell lines with known *KRAS* status has been used as positive controls. The expected *KRAS* status of these cell lines is reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) Database. In addition, a commercial pool of synthetic oligonucleotides (Trimgen) representing all mutations screened by the kit was employed.

As this assay is based on the use of specific primers for each reaction, wild-type samples results only in the amplification of the control curve. On the contrary, if a sample is mutated it was observable another amplification curve corresponding to one of seven mutations.

An example of wild-type and mutated sample is reported in Figure 50.

Concordance with the expected genotype, among replicates and among runs was the acceptance criteria selected for validation. As shown in Table 24, all results generated from cell line samples satisfied the required conditions.

Additional tests were carried out to evaluate precision under different conditions. In particular, variability between two operators and concordance among DxS Therascreen kit batches were assessed. As reported in Table 24, it was observed that these conditions had no impact on results and the assay was accurate and reproducible.

Figure 50. Example of DxS ThreraScreen KRAS mutation results for codon 12/13.
On the left a wild-type sample in which is shown only the control reaction; on the right a mutated sample in which is also shown the mutation amplification curve.

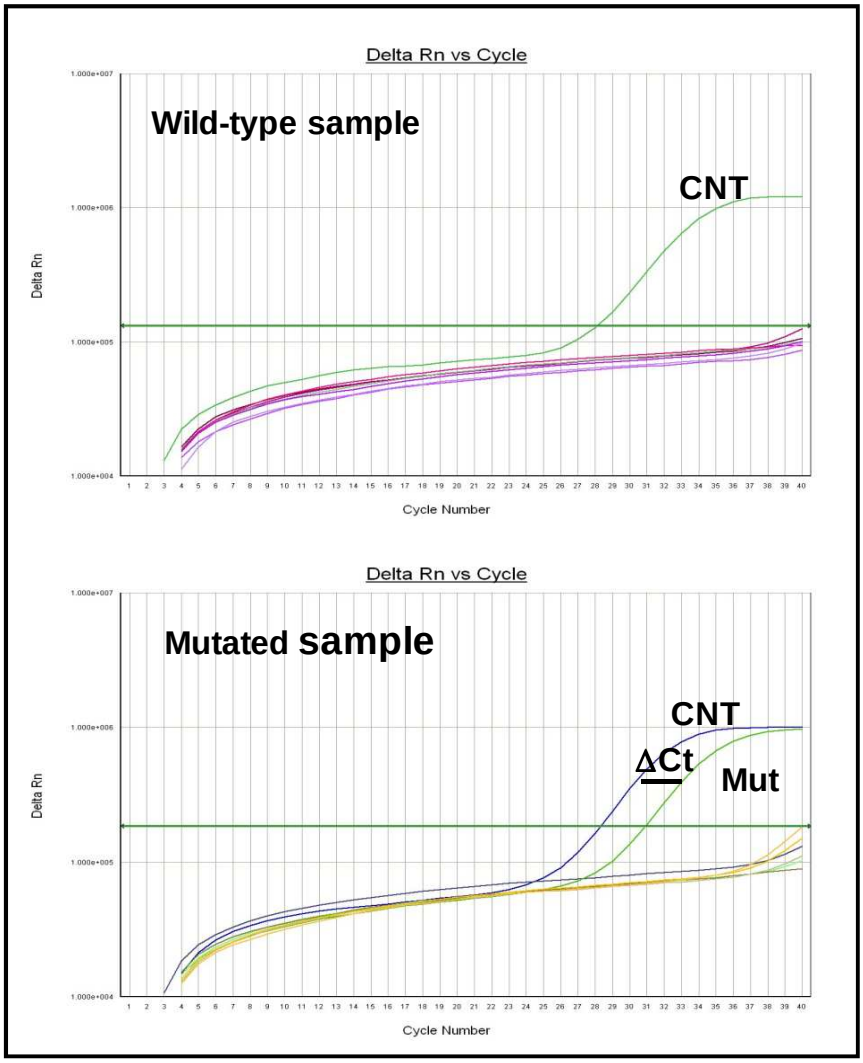


Table 24. Summary results of accuracy and repeatability tests using cell line controls.
The expected *KRAS* status refers to which ones reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (http://www.sanger.ac.uk/genetics/CGP/cosmic/). The synthetic control is used as positive control as it was mutated in all seven mutation detected by the kit.

Controls	Expected classification	Accuracy: Concordance of genotype	Intra-assay precision: Concordance of replicates	Ruggedness: Concordance of operators	Ruggedness: Concordance of batches
HT-29	Wild-type	Passed	Passed	Passed	Passed
Cal-62	12ARG	Passed	Passed	Passed	Passed
SW480	12VAL	Passed	Passed	Passed	Passed
GEO	12ALA	Passed	Passed	Passed	Passed
A549	12SER	Passed	Passed	Passed	Passed
LS-174T	12ASP	Passed	Passed	Passed	Passed
HCT116	13ASP	Passed	Passed	Passed	Passed
Calu-1	12CYS	Passed	Passed	Passed	Passed
Synthetic control	Mutated (all mutations)	Passed	Passed	Passed	Passed
Negative control	Undetermined	Passed	Passed	Passed	Passed

6.2.2.2 Impact of fragmented DNA and H&E staining

The performance of the DxS Therascreen assays was evaluated on matrices that mimic real samples. On this purpose degraded DNA extracted from 9 different FFPE samples was analyzed. Moreover for these samples Hematoxylin and Eosin stained FFPE slides (H&E FFPE) were available and they were analyzed to evaluate the impact of H&E staining on FFPE DNA quality.

The panel included all colorectal samples already characterized and used in the PyroMark Q24 KRAS kit validation.

Accuracy and precision were evaluated. To assess these criteria concordance of genotype under same conditions was evaluated. Additionally, reproducibility of results under a variety of test

conditions such as different operators, different RealTime PCR instrument were tested.
Results are reported in Table 25.

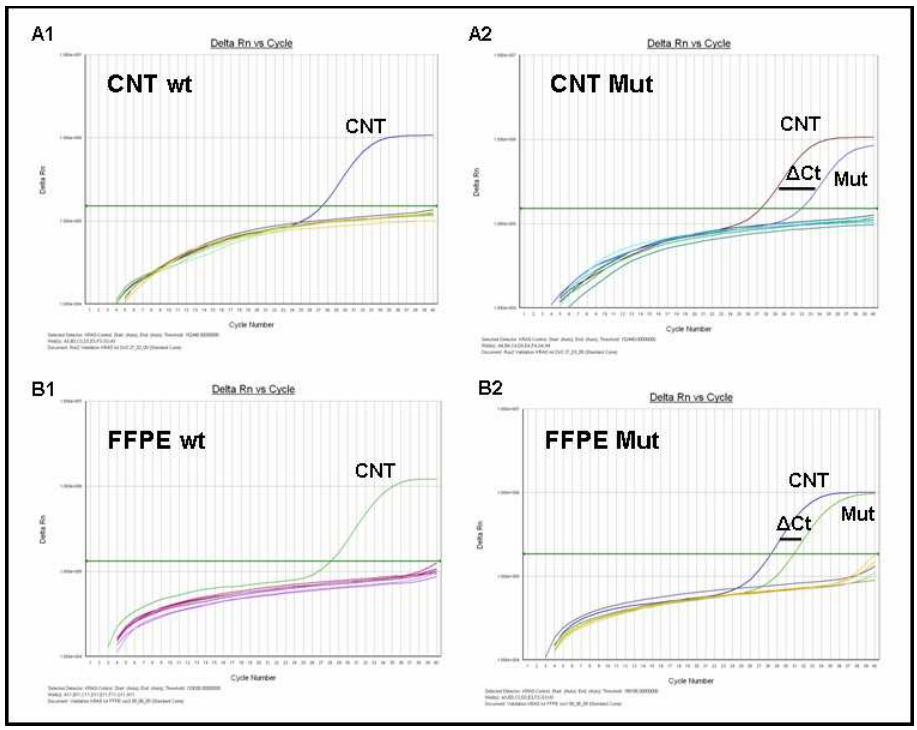
Table 25. Summary results of repeatability and intermediate precision tests using FFPE samples.

Sample ID	*KRAS* status	Real sample precision: Concordance of replicates	Intermediate precision: Concordance of operators	Intermediate precision: Concordance instruments	Concordance with H&E FFPE
21196	12ASP	Passed	Passed	Passed	Passed
21377	12ASP	Passed	Passed	Passed	Passed
66288	12ARG	Passed	Passed	Passed	Passed
21321	12CYS	Passed	Passed	Passed	Passed
66378	13ASP	Passed	Passed	Passed	Passed
21123	Wild-type	Passed	Passed	Passed	Passed
21454	Wild-type	Passed	Passed	Passed	Passed
21452	Wild-type	Passed	Passed	Passed	Passed
66460	Wild-type	Passed	Passed	Passed	Passed
Neg. control	Undeter-mined	Passed	Passed	Passed	Passed

All results met the acceptance criteria. It was noticed, however, a delay in the control amplification curve in degraded DNA obtained from FFPE specimens compared to the amplification of high quality genomic DNA obtained from cell lines as shown in Figure 51.

The amplification delay might result in a decrease of the assay sensitivity. In fact, the sensitivity of 1% is guarantee only if the control assay amplification is equal or lower than 29 cycles. This effect can be balanced increasing the amount of template DNA used in each reaction. However, highly degraded samples showed late amplification regardless of the DNA amount used as template.

Figure 51. Comparison of analysis between cell line DNA and FFPE DNA.
(A1) Wild-type cell line control sample resulted in amplification of only control assay. **(A2)** Analysis of a mutated cell line control sample resulted in amplification of both control assay and specific mutation assay; the DeltaCt value resulted lower than threshold limit. **(B1)** Wild-type FFPE sample resulted in amplification of only control assay. **(B2)** analysis of a mutated FFPE sample resulted in amplification of both control assay and specific mutation assay; the DeltaCt value resulted lower than threshold limit.



Another source of genomic DNA evaluated was Haematoxylin and Eosin staining FFPE (H&E). Unlike unstained FFPE, histological slides are often used for pathology evaluation. To evaluate the impact of the staining on results reliability DNA was extracted from slide obtained from the same donors. Genotype obtained was compared to genotype obtained from unstained slides. In all samples concordance was observed and no impact on genotype analysis was due to H&E staining (Table 25). For these samples accuracy and repeatability were tested and all the acceptance criteria were met.

6.2.2.3 Impact of different tumor percentage (sensitivity test) and limited amount of material

An additional test has been performed in order to evaluate the DxS TheraScreen KRAS mutation kit sensitivity and to determine the minimum percentage of tumor tissue necessary to detect a mutation.
For this purpose, each sample representing different tumor percentage was serially diluted to evaluate the impact of late amplification and to test the assay sensitivity. Indeed, late amplification can be observed in low quality-high fragmented DNA deriving from long stored FFPE blocks.
First DNA from a mutated FFPE sample has been diluted with DNA from wild-type FFPE sample at different ratio to mimic real samples with different tumor percentages (70, 35, 10, and 5%). Then each sample was analyzed as shown if Figure 52.

As expected the control amplification curve resulted stable among samples (with C_t < 29 cycles), while the mutation amplification curve shifted in function of the tumor percentage, determining the increase of the deltaCt value. Despite the increase of deltaC_t value, the kit was able to detect a mutation with tumor percentage down to 5%.
In addition, to evaluate the impact of limited material and late amplification, each sample was then serially diluted and analyzed. An example is reported in Figure 53 where the sample with 35% of tumor was serially diluted down to 1.87 ng for each reaction.

Figure 52. ***KRAS*** **mutation analysis results at different tumor percentage. A** Tumor FFPE DNA sample (~70% of tumor tissue). **B** Tumor FFPE DNA sample spiked in wild-type FFPE DNA sample at a ratio 1:1 (~35% of tumor tissue). **C** Tumor FFPE DNA sample spiked in wild-type FFPE DNA sample at a ratio 1:7 (~10% of tumor tissue). **D** Tumor FFPE DNA sample spiked in wild-type FFPE DNA sample at a ratio 1:14 (~5% of tumor tissue).

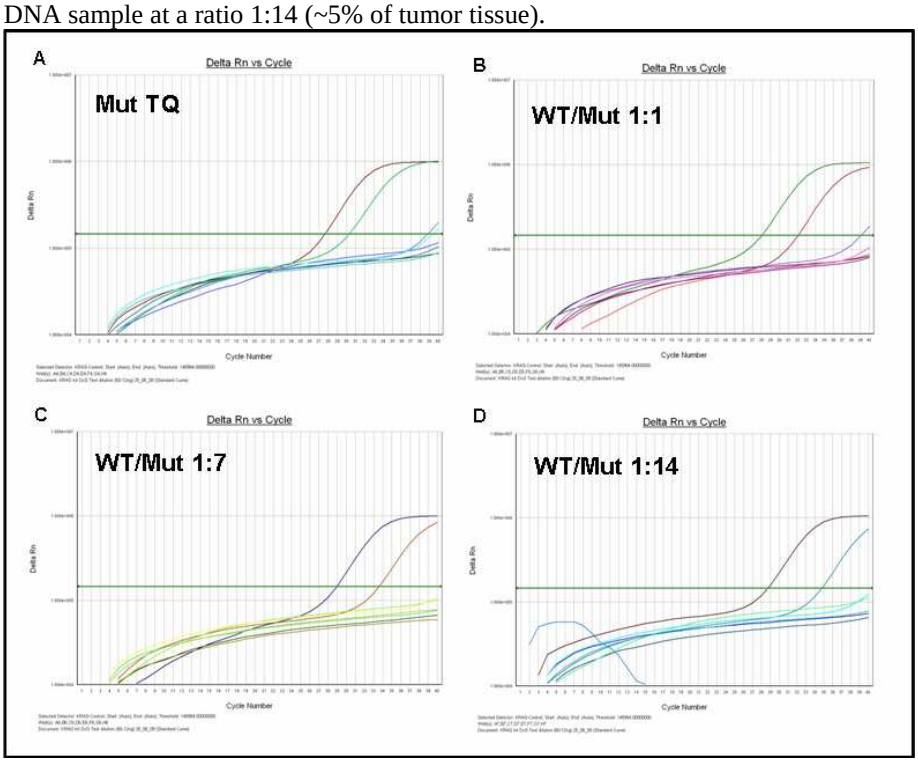
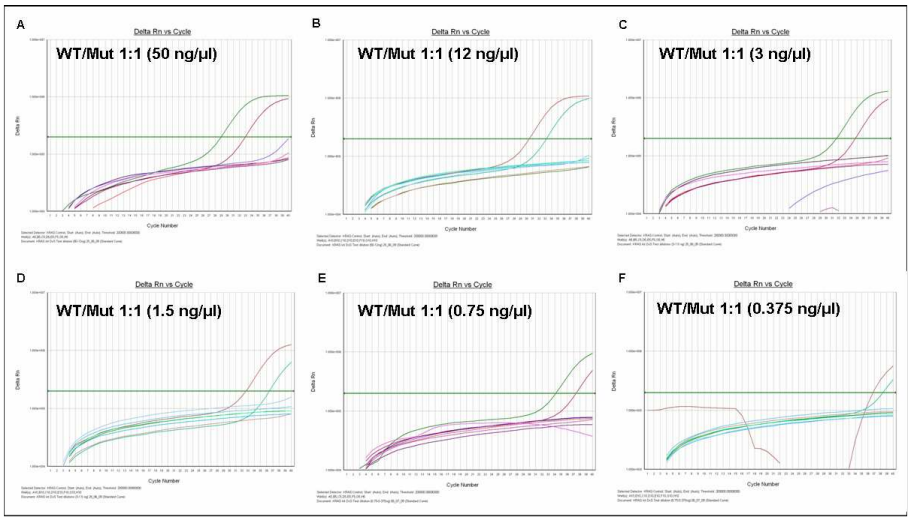


Figure 53. *KRAS* mutation analysis results at different sample dilutions.
Tumor FFPE DNA sample spiked in wild-type FFPE DNA sample at a ratio 1:1 (~35% of tumor tissue) loaded at different concentrations: **A** 50 ng/µl (corresponding to 250ng/reaction). **B** 12 ng/µl (corresponding to 60ng/reaction). **C** 3 ng/µl (corresponding to 15ng/reaction). **D** 1.5 ng/µl (corresponding to 7.5ng/reaction). **E** 0.75 ng/µl (corresponding to 3.5ng/reaction). **F** 0.375 ng/µl (corresponding to 1.87ng/reaction).



As result of the serial dilutions, both the amplification and the mutation curves shifted towards late amplifications, while the deltaC_t value resulted stable among dilutions.

Delay in the control amplification determines a decrease in the assay sensitivity. Results showed that the mutation from 5% tumor percentage was not detected in samples with control C_t higher than 32, and the mutation from 10% tumor percentage was not detected in samples with control C_t higher than 34.

All results from sensitivity tests are reported in Table 26.

Table 26. DxS TheraScreen KRAS sensitivity test codon 12/13.
Results are accurate till a 5% percentage of tumor and at 7.5ng of DNA sample for each reaction.

Sample concentration	**70% MUT**	**35% MUT**	**10% MUT**	**5% MUT**
250 ng/reaction (Ct 27-29)	Mutated	Mutated	Mutated	Mutated
60 ng/reaction (Ct 28-30)	Mutated	Mutated	Mutated	Mutated
15 ng/reaction (Ct 30-31)	Mutated	Mutated	Mutated	Mutated
7.5 ng/reaction (Ct 31-32)	Mutated	Mutated	Mutated	Mutated
3.5 ng/reaction (Ct 32-34)	Mutated	Mutated	Mutated	**Wild-Type**
1.87 ng/reaction (Ct 34-36)	Mutated	Mutated	**Wild-Type**	**Wild-Type**

Overall DNA obtained from unstained FFPE and from H&E stained FFPE slices resulted suitable for analysis. The method resulted sensitive enough to detect a mutation on samples containing at least 5% of tumor.

Reduction in assay sensitivity was evaluated to support the interpretation of results deriving from low quality-highly fragmented DNA. The present study provides a matrix in which tumor percentage and control amplification value were correlated. As results, low quality samples with a late amplification control (C_t range 29-35) can be assigned only if tumor percentage is higher than 35%. Samples with low tumor percentage (< 35%) can be assigned only if the control C_t value is lower than 29 to exclude the risk of false negative results.

6.3 RNA expression analysis on Hair follicle

6.3.1 Collection, Storage and Extraction Method

The first step of the work was to establish an RNA extraction method from hair follicles. Hair follicles are a non-conventional source of nucleic acid that might provide useful information as they can reflect

biological response in epithelial tissue, they are easy to collect (non-invasive), available from most individual.

In the literature there are some examples of hair follicle extraction by TRIAZOL method but in order to improve yield and avoid PCR inhibitor, samples were extracted using binding silica-based membrane. This preliminary extraction was carried on two donors comparing two different methods: Qiagen® RNeasy® Micro Kit and Ambion RNAqueous® Micro Kit.

For each donor four HFs were grasped with forceps and immediately extracted.

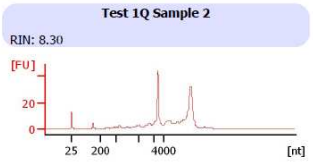
As shown in Figure 54, RNeasy® Micro Kit Qiagen® showed better quality RNA with a yield comparable to the TRIAZOL method reported in the literature (Kim et al., 2006). Regarding the Ambion kit, regardless of the high quantity of RNA quantified by spectrophotometer, electrophoresis analysis showed low extraction yield and quality. Results are shown in Figure 54B.

Figure 54. Comparison of two kits for HFs RNA extraction.
Panel **A** shows results obtained with Micro Kit provided by Qiagen® while panel **B** shows results obtained from RNAqueous® Micro kit provided by Ambion. Quality analysis shows an unexpected peak in Sample 2 extracted by Ambion kit.

A

RNeasy Micro Kit Qiagen - Fresh Sample -

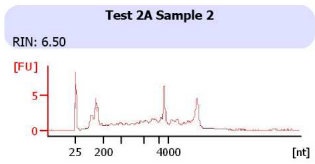
Kit Qiagen	Follicle	Total RNA (ng/µl)	Total yield RNA (ng)	RIN
Sample 1	4	16.3	163.2	8.9
Sample 2	3	6.13	61.3	8.3



B

RNAqueous Micro Kit Ambion - Fresh Sample -

Kit Ambion	Follicle	Total RNA (ng/µl)	Total yield RNA (ng)	RIN
Sample 1	4	87.6	875.9	7.9
Sample 2	4	25.7	256.8	6.5



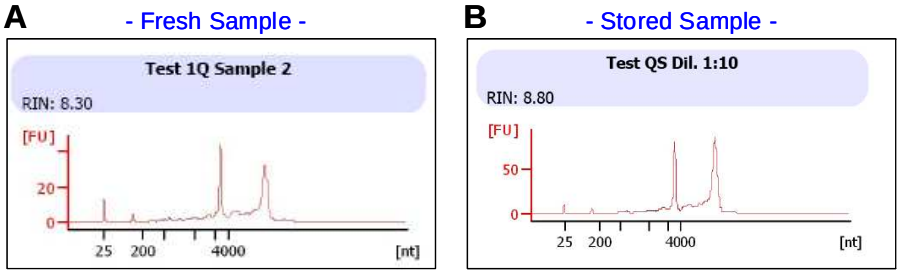
For this reason Qiagen® RNeasy® Micro Kit was chosen for further analysis.

In order to apply gene expression analysis from HFs on clinical studies, it was important to evaluate a suitable sample collection and storage. For this reason, it has been tested a collection method to stabilize biological samples.

The focus was to set up a simple but robust method that allows HF storage and maintenance of RNA integrity. Four hair follicles from the same donors were grasped with forceps and incubated into the RNAlater® Tissue Collection solution which prevents RNA degradation. Samples were maintained at 4°C and then RNA extraction was performed after 5 days.

Qiagen extraction showed high quality RNA on both fresh and frozen tissue as their profile was comparable after Bioanalyzer analysis (Figure 55).

Figure 55. QC comparison between fresh HF and stored HF obtained from extraction by QIAGEN® kit.
(A) Bioanalyzer results of Fresh HF sample. **(B)** Bioanalyzer results of stored HF sample.



RNA quality was not affected by storage procedure and total yield was comparable to the one obtained from fresh follicle as shown in Table 27.

Table 27. Summary of data obtained from methods used for extraction.
The table above refers to yield and quality of both method on fresh hair follicle. The table belove reported data obtained from stored hair follicle.

Fresh Sample	Follicle	Total RNA (ng/µl)	Total yield RNA (ng)	RIN
Sample 1 Qiagen	4	16.3	163.2	8.9
Sample 2 Ambion	4	25.7	256.8	6.5

Stored Sample	Follicle	Total RNA (ng/µl)	Total yield RNA (ng)	RIN
Sample 1 Qiagen	4	23.5	281.5	8.8
Sample 2 Ambion	4	55.9	558.7	8.5

In order to evaluate method reliability, an increased number of samples was tested. In particular, 4 HF were collected from 23 individuals and stored in RNAlater® and maintained at -20°C until extraction. Samples have been stored for a maximum of four weeks.
Yield obtained from samples was very variable: from a minimum of 4.03 ng/µl to a maximum of 99.95 ng/µl .
Inter-individual variability was also observed in RNA quality. Indeed, RNA integrity number (RIN) ranged from 5.3 to 9.2. Electrophoresis profile showed a major degradation for those samples with a low 18S and 28S peaks.

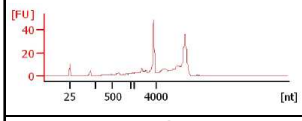
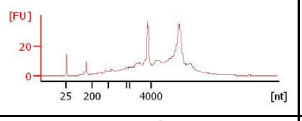
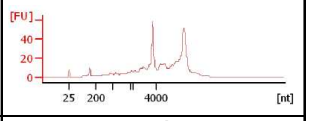
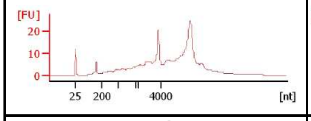
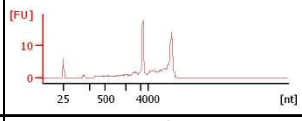
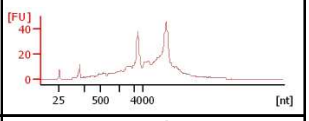
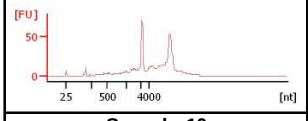
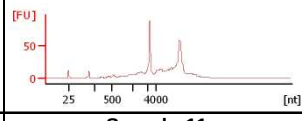
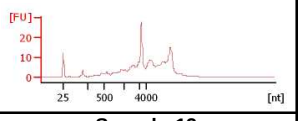
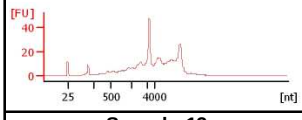
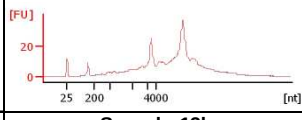
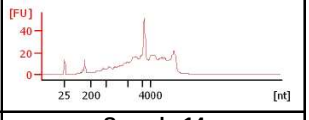
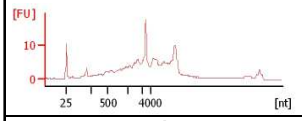
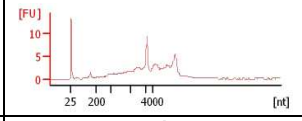
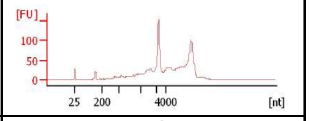
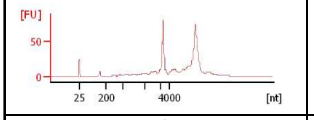
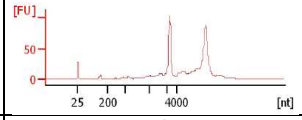
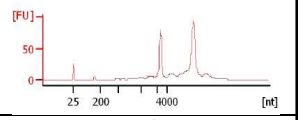
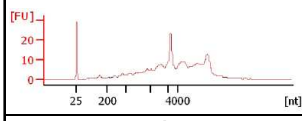
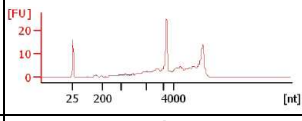
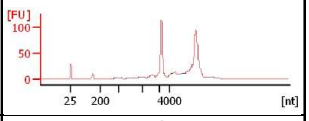
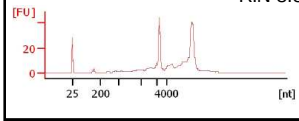
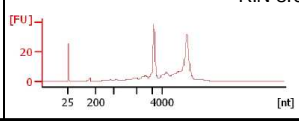
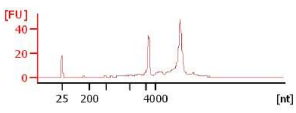
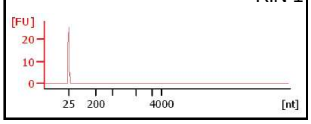
A complete summary of the total yield and the Bioanalyzer data is reported in Figure 56.

Figure 56. Summary of data obtained to evaluate HFs inter-individual variabilità.
(**A**) Summary of RNA extractions for all samples. The table reports the sample list, total yield obtained and the number of HFs used for the extraction.
(**B**) Summary of QC analysis obtained by Bioanalyzer. For each sample RIN value is shown. The negative control (water) is also included.

A

Sample	Total Yield (ng)	HF Number
Sample 1	97	4
Sample 2	775.1	4
Sample 3	1199.4	4
Sample 4	758.3	4
Sample 5	116	4
Sample 6	366.8	4
Sample 7	292.8	4
Sample 8	243.7	4
Sample 9	47.9	4
Sample 10	83.8	4
Sample 11	483.7	4
Sample 12	379.4	4
Sample 13	112	4
Sample 13b	91.9	1
Sample 14	539.5	4
Sample 15	177	4
Sample 16	305	4
Sample 17	230.3	4
Sample 18	71.6	4
Sample 19	63.4	4
Sample 20	241	4
Sample 21	44.3	4
Sample 22	125.5	4
Sample 23	124.2	4

B

Sample 1	Sample 2	Sample 3
RIN 7.6 	RIN 7 	RIN 7.5 
Sample 4	**Sample 5**	**Sample 6**
RIN 7.1 	RIN 8 	RIN 7.2 
Sample 7	**Sample 8**	**Sample 9**
RIN 7.8 	RIN 7.9 	RIN 6.4 
Sample 10	**Sample 11**	**Sample 12**
RIN 6.2 	RIN 7 	RIN 5.3 
Sample 13	**Sample 13b**	**Sample 14**
RIN 5.9 	RIN 6 	RIN 7 
Sample 15	**Sample 16**	**Sample 17**
RIN 8.6 	RIN 8.8 	RIN 9.1 
Sample 18	**Sample 19**	**Sample 20**
RIN 5.9 	RIN 7.3 	RIN 8.9 
Sample 21	**Sample 22**	**Sample 23**
RIN 8.8 	RIN 8.6 	RIN 9.2 
Water		
RIN 1 		

6.3.2 Hair Follicle RNA Characterization

After the establishment of the collection, storage and extraction method, RNA was tested to evaluate feasibility of gene expression analysis.

For this reason, RNA extracted from 8 Hair Follicles donors was reverse transcribed (20 ng for reaction). As preliminary assay, the expression level of housekeeping genes was evaluated using the TaqMan® Endogenous Control Arrays.

As shown in Table 28 all genes resulted to be expressed in this type of tissue. However for most of the genes Ct medium was very high near to limit of detection.

Table 28. Results of TaqMan® Endogenous Control Arrays.

Gene Symbol	Ct range
18S	16-24
PGK1	28-30
ACTB	25-28
POLR2A	31-33
B2M	29-31
PPIA	28-29
GAPDH	28-30
RPLP0	27-29
GUSB	33-35
TBP	33-35
HMBS	33-35
TFRC	30-31
HPRT1	32-34
UBC	28-29
IPO8	32-34
YWHAZ	32-33

As second step, it was investigated the expression of tissue specific genes. To this purpose a literatures research was conducted to select a list of markers (see Table 7) which included:

- ubiquitously expressed genes (*18S rRNA*, interferon recepror *IFNAR1* and *IFNAR2*)
- genes expressed in blood (*FPR1* and *PRF1*)

- genes expressed in epithelial tissues (*EGFR*, *KRT19*, *COL1A1*, *MLANA*)

The expression of all genes has then been tested on three different tissues:

- hair follicles as a source of unconventional sample
- human blood as a common source to conduct clinical analyses
- human reference RNA constituted by a pool of several cell line RNA and used as a positive control.

Based on previous analysis on housekeeping genes, 50 ng of total RNA obtained from 4 HFs was chosen as starting material for analysis.

The Realtime PCR analyses showed that Formyl peptide receptor 1 (*FPR1*, involved in neutrophils activation) and Perforin 1 (*PRF1*, one of the main cytolytic proteins of cytolytic granules) genes were indeed detected in blood; while Epidermal Growth Factor receptor (*EGFR*), Keratin 19 (*KRT19*), Collagen (type I, alpha 1), Melan-A (codes for a melanocyte differentiation antigen) were expressed only in HFs.

Table 29. Summary of tissue specific gene expression.
In the table gene specific for hair follicle are highlighted in yellow while blood specific gene in red.
- = Ct value undetected or > than 36; + = gene detected with a Ct value < than 36.

Gene	Human Reference	Human Blood	Hair Follicles
18S	+	+	+
IFNAR1	+	+	+
IFNAR2	+	+	+
EGFR	+	-	+
KRT 19	+	-	+
COL1A1	+	-	+
MLANA	+	-	+
FPR1	+	+	-
PRF1	+	+	-

These results suggest that HF represents a valuable biological source to study pathways active in epithelial tissue.

In addition, to optimize and to evaluate analytical performance of the RealTime PCR method, 4 kits of reverse transcription have been compared: the High Capacity cDNA Reverse Transcription kit (Applied Biosystems); High Capacity RNA-to-cDNA Master Mix (Applied Biosystems); QuantiTect Reverse Transcription kit (QIAGEN); and the SuperScriptTM III Reverse Transcriptase (Invitrogen).

For this scope, a pool of hair follicle RNA was used (50ng) and *EGFR* and *KTR19* genes were quantified. Results are summarized in Table 30.

Table 30. Summary of data obtained from different methods of reverse transcription.

App. Bio. 1 =High Capacity cDNA Reverse Transcription; App. Bio. 2 =High Capacity RNA-to-cDNA Master Mix; Qiagen=QuantiTect® Reverse Transcription; Invitrogen=SuperScriptTM III Reverse Transcriptase.

Kit	Gene	Ct Average	CV% - 2^DCt
Appl. Bio. 1	18S	10.52	-
Appl. Bio. 2	18S	9.65	-
Qiagen	18S	10.96	-
Invitrogen	18S	11.95	-
Appl. Bio. 1	EGFR	27.70	7.19
Appl. Bio. 2	EGFR	27.53	16.07
Qiagen	EGFR	27.75	50.31
Invitrogen	EGFR	27.27	63.86
Appl. Bio. 1	KRT19	30.56	7.57
Appl. Bio. 2	KRT19	29.78	22.18
Qiagen	KRT19	30.16	25.50
Invitrogen	KRT19	30.59	24.83

In this analysis *18S* was used as housekeeping gene. Results showed that among different protocols, the High Capacity cDNA Reverse Transcription kit resulted more robust and reproducible. This method was chosen for further analysis.

6.3.3 Preliminary in vivo results

To evaluate if HF model can be informative, known PD markers of Interferon treatment were investigated after *in vivo* subministration of Interferon-beta (IFN-β) in Cynomolgus monkey. In particular, known IFN-β responding genes such as *MxA* were investigated in blood and in HF.

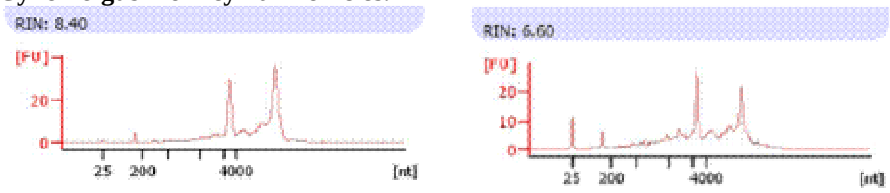
Interferon Stimulated Genes (ISGs) are a diverse group of more than 300 genes that mediate the biological and therapeutic effects of IFN stimulation.

Monkey HFs were grasped with forceps from the same body area at different time points and stored on tubes containing RNAlater®.

Anagen hair follicles (about 10) were selected by optical microscope and the RNA was extracted.

The extracted RNA had a total yield that ranged between 90 and 435 ng RIN numbers that ranged between 3.1 and 7.2. An example of results is shown Figure 57.

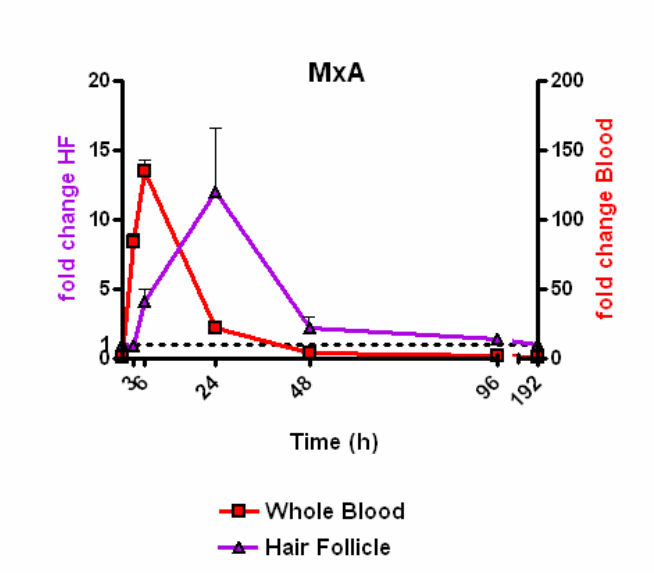
Figure 57. Example of Bioanalyzer analysis of RNA extracted from Cynomolgus monkey hair follicles.



Even if RNA quality was very variable, it was possible to perform gene expression analysis using the interferon responding gene *MxA* (*Mx1* Myxovirus resistance 1). The Housekeeping gene used is *18S* Eukaryotic 18S ribosomal RNA. The analysis was conducted on both blood and HFs. In this experiment an IFN responding gene was chosen since its induction was expected to be present in both tissues.

After reverse transcription and RealTime PCR analysis, *MxA* expression level was normalized to the *18S* housekeeping gene (DCt). Each sample was then normalized to the pre-dose time (DDCt) and expressed as "fold change" (2^-DDCt).
Results are summarized in the following figure (Figure 58).

Figure 58. Expression profile of *MxA* gene after IFN stimulation.
MxA induction in blood and in Hair Follicles is compared.



MxA induction was observed after exposure to IFN-beta both in blood and in HFs as expected.
Profiles from the two tissues differ for the time needed to reach maximum peak probably reflecting the IFN distribution. Moreover, different degree of induction was observed.
Nevertheless, these results suggest that HF might represent a valuable biological source to study gene expression. They can be use as non-invasive collection to study molecular mechanisms in epithelial tissues.

7 CONCLUSION

The present research aimed at developing new genomics-based tools using non-conventional biological samples that might support biomarker investigations in clinical settings.
Methods developed were validated in accordance with the fit-for-purpose approach and Guidance for Industry encouraged by Regulatory Agencies.
Analytical methods developed focused on both DNA and RNA genomic biomarkers analysis; in particular single nucleotide polymorphisms (SNP), mutations and gene expression analyses were evaluated.

SNP analysis was validated on DNA extracted from Formalin-Fixed Paraffin-Embedded (FFPE) tissue, from Hematoxylin and Eosin (H&E) stained FFPE slides, and from serum samples.
First a panel of polymorphism assays was tested on reference samples for accuracy and reproducibility. Moreover, concordance of genotype between high quality DNA, extracted from fresh frozen tissues, and degraded DNA extracted from FFPE samples of the same donors was demonstrated. In addition, it was shown that H&E staining did not impact assay performance.
In conclusion FFPE and H&E FFPE could be a used as DNA unconventional source to investigate polymorphisms in those situations in which blood is not available.

A further effort was carried out to evaluate if serum could be used as alternative source of DNA for genotyping analysis. Results showed that SNP analyses from serum are reliable but a high failure rate was observed probably due to low quantity DNA extracted from samples. To overcome the problem whole genome amplification was applied to serum DNA. However the method did not result accurate as allelic drop out effect and change of allele were observed.

Overall, the validation conducted showed that serum could be used as source of biological material to conduct genetic analyses. However limitation of DNA does not consent to perform a large panel of analysis. It has been considered that analysis performed in this thesis were conduct on healthy volunteers. SNP analysis on serum could be further explored in samples deriving from patients since circulating DNA is present at higher levels in several diseases.

An additional study was conducted to validate two different methods for *KRAS* mutation detection on FFPE tumor specimens and on H&E stained FFPE which represent an unconventional source of samples for this type of analysis.
In particular, the DxS ThreraScreen KRAS mutation kit based on Real-Time PCR assay and the PyroMark KRAS Kit based on pyrosequencing technology were compared and validated.
Results from validation showed that both Pyrosequencing assay both DxS ThreraScreen assay are accurate and reproducible on good quality DNA and on unconventional samples. For instance, no impact of degraded DNA deriving from FFPE tissues or influence due to the H&E staining was observed in both methods.
In conclusion PyroMark KRAS Kit showed advantages such as lower amount of DNA, detection of additional mutations in cod.12/13 and codon 61than DxS TheraScreen kit; on the other hand PyroMark KRAS showed lower sensitivity and it is more time consuming.
The experience gained through the establishment of the *KRAS* analysis, put solid bases for the evaluation of a larger panel of mutations of interest in clinical trials.
The application of genomic markers in diagnosis, prognosis and treatment could be useful to move towards a personalized medicine which could improve patient outcome.

Concerning RNA biomarkers, gene expression investigation on hair follicles (HF) was evaluated to assess method applicability in clinical trials. Hair follicles may offer a viable alternative to study epithelial tissue cancers since they can reflect their biological response, they are easy to collect (non-invasive) and available from most individuals.
First a method for sample collection and RNA extraction was established. Then, gene expression analysis was conducted on a panel of 16 housekeeping genes to assess the feasibility of the analysis. It was shown that in HF it was possible to analyze the expression of epithelial specific genes not present in blood suggesting that HF represents a valuable biological source to study pathways active in epithelial tissues.
Finally, gene expression analysis was conducted on an *in vivo* experiment to evaluate if a response to treatment could be observed in anagen HFs. In particular, PD markers of Interferon treatment were investigated after *in vivo* subministration of Interferon-beta (IFN-β) in Cynomolgus monkey. The expression of the known IFN-β responding genes MxA resulted stimulated both in blood and in HF. Gene induction in blood was observed at 6 hours after subministration while in HF at 24hours probably due to a different IFN-β distribution.
These data suggest that gene expression analysis can be carried out in HF samples. However, it is important to highlight that in HF the response had a lower degree of induction and higher variability than in blood.
These preliminary results suggest that HF might represent a non-invasive valuable biological source to study molecular mechanisms in epithelial tissues. However this preliminary observations need to be further explored in pilot clinical studies to evaluate its applicability.

Overall the validation of different genomic analysis on unconventional samples opened the possibility to conduct biomarker investigations on several clinical trials conducted in the past or to plan new investigations with non invasive methods.

In addition, from the deep evaluation of the current guidelines from the Regulatory Agencies (and from the open debate in the field) a proper strategy to validate genomic analytical assays was designed according to fit-for-purpose criteria.

8 REFERENCES

Abd El-Rehim DM, Pinder SE, Paish CE, Bell JA, Rampaul RS, Blamey RW, Robertson JF, Nicholson RI, Ellis IO.
Expression and co-expression of the members of the epidermal growth factor receptor (EGFR) family in invasive breast carcinoma. Br J Cancer. 2004; 91(8):1532-42.

Alizadeh BZ, Valdigem G, Coenen MJ, Zhernakova A, Franke B, Monsuur A, van Riel PL, Barrera P, Radstake TR, Roep BO, Wijmenga C, Koeleman BP. Association analysis of functional variants of the FcγRIIa and FcγRIIIa genes with type 1 diabetes, celiac disease and rheumatoid arthritis. Hum. Mol. Genet. 2007; 16:2552-9.

Allegra CJ, Jessup JM, Somerfield MR, Hamilton SR, Hammond EH, Hayes DF, McAllister PK, Morton RF, Schilsky RL. American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. J Clin Oncol. 2009; 27(12):2091-6.

Alonso L, Fuchs E. The hair cycle. J Cell Sci. 2006; 119(Pt 3):391-3.

Anderson JE, Hansen LL, Mooren FC, Post M, Hug H, Zuse A, Los M. Methods and biomarkers for the diagnosis and prognosis of cancer and other diseases: towards personalized medicine. Drug Resist Updat. 2006; 9(4-5):198-210.

Andreassen CN, Sørensen FB, Overgaard J, Alsner J. Optimisation and validation of methods to assess single nucleotide polymorphisms (SNPs) in archival histological material. Radiother Oncol. 2004; 72(3):351-6.

Banaschak S, Rolf B, Brinkmann B.
Influence of different staining techniques on the DNA analysis of histological sections. Int J Legal Med. 2000;113(2):114-6.

Bardelli A and Siena S. Molecular mechanisms of resistance to cetuximab and panitumumab in colorectal cancer. J Clin Oncol. 2010; 28(7):1254-61.

Barker DL, Hansen MS, Faruqi AF, Giannola D, Irsula OR, Lasken RS, Latterich M, Makarov V, Oliphant A, Pinter JH, Shen R, Sleptsova I, Ziehler W, Lai E.
Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. Genome Res. 2004; 14(5):901-7.

Bernard BA. The life of human hair follicle revealed.
Med Sci (Paris). 2006; 22(2):138-43.

Bibeau F, Lopez-Crapez E, Di Fiore F Thezenas S, Ychou M, Blanchard F, Lamy A, Penault-Llorca F, Frébourg T, Michel P, Sabourin JC, Boissière-Michot F. Impact of FcγRIIa-FcγRIIIa polymorphisms and KRAS mutations on the clinical outcome of patients with metastatic colorectal cancer treated with cetuximab plus irinotecan. Clin. Oncol. 2009; 27:1122-9.

Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework.
Clin Pharmacol Ther. 2001; 69(3):89-95.

Board RE, Williams VS, Knight L,Shaw J, Greystoke A, Ranson M, Dive C, Blackhall FH, Hughes A. Isolation and Extraction of Circulating Tumor DNA from Patients with Small Cell Lung Cancer. Ann. N.Y. Acad. Sci. 2008; 1137: 98–107.

Booth DR, Arthur AT, Teutsch SM, Bye C, Rubio J, Armati PJ, Pollard JD, Heard RN, Stewart GJ; Southern MS Genetics Consortium.
Gene expression and genotyping studies implicate the interleukin 7 receptor in the pathogenesis of primary progressive multiple sclerosis.
J Mol Med. 2005; 83(10):822-30.

Borden EC, Sen GC, Uze G, Silverman RH, Ransohoff RM, Foster GR, Stark GR. Interferons at age 50: past, current and future impact on biomedicine.
Nat Rev Drug Discov. 2007; 6(12):975-90.

Brady G, Hurley PT, Grimes E, Herath A, Wilson C, Moore JG, Potten CS, Graham A, Haggerty C, Vinsun D, Hodgson D. Comprehensive expression analysis of single plucked human hairs as means of assisting clinical trials. In American Association for Cancer Research conference, San Francisco, October 2007.

Brash AR, Boeglin WE, Chang MS.
Discovery of a second 15S-lipoxygenase in humans.
Proc Natl Acad Sci U S A. 1997; 94(12):6148-52.

Bredel M, Bredel C, Juric D, Kim Y, Vogel H, Harsh GR, Recht LD, Pollack JR, Sikic BI. Amplification of whole tumor genomes and gene-by-gene mapping of genomic aberrations from limited sources of fresh-frozen and paraffin-embedded DNA. J Mol Diagn. 2005; 7(2):171-82.

Bruhns P, Iannascoli B, England P et al.: Specificity and affinity of human Fcγ Receptors and their polymorphic variants for human IgG subclasses. Blood 2009; 113(16):3716-25.

Brynedal B, Khademi M, Wallström E, Hillert J, Olsson T, Duvefelt K.
Gene expression profiling in multiple sclerosis: a disease of the central nervous system, but with relapses triggered in the periphery?
Neurobiol Dis. 2010; 37(3):613-21.

Cañete JD, Suárez B, Hernández MV, Sanmartí R, Rego I, Celis R, Moll C, Pinto JA, Blanco FJ, Lozano F. Influence of variants of Fc gamma receptors IIA and IIIA on the American College of Rheumatology and European League Against Rheumatism responses to anti-tumour necrosis factor alpha therapy in rheumatoid arthritis. Ann. Rheum. Dis. 2009; 68:1547-52.

Chau CH, Rixe O, McLeod H, Figg WD.
Validation of analytic methods for biomarkers used in drug development.
Clin Cancer Res. 2008; 14(19):5967-76.

Chen J, Guo L, Peiffer DA, Zhou L, Chan OT, Bibikova M, Wickham-Garcia E, Lu SH, Zhan Q, Wang-Rodriguez J, Jiang W, Fan JB. Genomic profiling of 766 cancer-related genes in archived esophageal normal and carcinoma tissues. Int J Cancer. 2008; 122(10):2249-54.

Colburn WA, Lee JW. Biomarkers, validation and pharmacokinetic-pharmacodynamic modelling. Clin Pharmacokinet. 2003;42(12):997-1022.

Curry SH. Translational science: past, present, and future. Biotechniques. 2008; 44(2): ii-viii.

De Hertogh G and Geboes KP. Practical and Molecular Evaluation of Colorectal Cancer. Arch Pathol Lab Med—Vol 134, 853-63. 2010

De Roock W, De Schutter J, De Hertogh G, Janssens M, Biesmans B, Personeni N, Geboes K, Verslype C, Van Cutsem E, Tejpar S. KRAS mutations preclude tumor shrinkage of colorectal cancers treated with cetuximab. J Clin Oncol 2007; 25 (18S): 4132.

Der SD, Zhou A, Williams BR, Silverman RH. Identification of genes differentially regulated by interferon alpha, beta and gamma using oligonucleotide arrays. Proc. Natl Acad. Sci. USA 1998; 95:15623–15628.

Doleshal M, Magotra AA, Choudhury B, Cannon BD, Labourier E, Szafranska AE.
Evaluation and validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues.
J Mol Diagn. 2008; 10(3):203-11.

Dutt A and Beroukhim R. Single nucleotide polymorphism array analysis of cancer. Curr Opin Oncol 2007; 19:43-9.

EMA, CHMP. Reflection paper on co-development of pharmacogenomic biomarkers and Assays in the context of drug development. 2010.

EMA. ICH Topic E15 - Definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories. 2007.

EMA. ICH-Q2A - Guideline for Industry - Text on Validation of Analytical Procedures. 1995.

Farragher SM, Tanney A, Kennedy RD, Paul Harkin D. RNA expression analysis from formalin fixed paraffin embedded tissues. Histochem Cell Biol. 2008; 130(3):435-45.

FDA, CDRH. Guidance for Industry - Pharmacogenomic Data Submissions - Companion Guidance. 2007.

FDA, CDRH. Guidance for Industry and FDA staff - Pharmacogenetic Tests and Genetic Tests for Heritable Markers. 2007.

Gahana PB and Swaminathanb R. Circulating Nucleic Acids in Plasma and Serum. Ann. N.Y. Acad. Sci. 2008; 1137: 1–6.

Garm Spindler KL, Pallisgaard N, Rasmussen AA, Lindebjerg J, Andersen RF, Crüger D, Jakobsen A. The importance of KRAS mutations and EGF61A>G polymorphism to the effect of cetuximab and irinotecan in metastatic colorectal cancer. Ann Oncol. 2009; 20:879-84.

Gormally E, Caboux E, Vineis P, Hainaut P. Circulating free DNA in plasma or serum as biomarker of carcinogenesis: Practical aspects and biological significance. Mutation Research. 2007; 635: 105–117.

Goswami RS, Waldron L, Machado J, Cervigne NK, Xu W, Reis PP, Bailey DJ, Jurisica I, Crump MR, Kamel-Reid S. Optimization and analysis of a quantitative real-time PCR-based technique to determine microRNA expression in formalin-fixed paraffin-embedded samples. BMC Biotechnol. 2010; 10:47.

Habek M, Borovecki F, Brinar VV. Genomics in multiple sclerosis.
Clin Neurol Neurosurg. 2010; 112(7):621-4.

Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast RC Jr; American Society of Clinical Oncology. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer.
J Clin Oncol. 2007 Nov 20;25(33):5287-312.

Hong Y, Eu KW, Seow-Choen F, Fook-Chong S, Cheah PY. GG genotype of cyclin D1 G870A polymorphism is associated with increased risk and advanced colorectal cancer in patients in Singapore.
Eur. J. Cancer 2005; 41:1037-44.

Hu YF, Kaplow J, He Y. From Traditional Biomarkers to Transcriptome Analysis in Drug Development. Curr Mol Med. 2005; 5(1):29-38.

Huang YT, Heist RS, Chirieac LR, Lin X, Skaug V, Zienolddiny S, Haugen A, Wu MC, Wang Z, Su L, Asomaning K, Christiani DC. Genome-wide analysis of survival in early-stage non-small-cell lung cancer.
J Clin Oncol 2009; 27:2660-7.

Ito T, Ito N, Saathoff M, Bettermann A, Takigawa M, Paus R. Interferon-gamma is a potent inducer of catagen-like changes in cultured human anagen hair follicles. Br J Dermatol. 2005 Apr;152(4):623-31.

Karoline Krause, MD, and Kerstin Foitzik, MD. Biology of the Hair Follicle: The Basics. Semin Cutan Med Surg. 2006; 25:2-10 .

Kim SJ, Dix DJ, Thompson KE, Murrell RN, Schmid JE, Gallagher JE, Rockett JC.
Gene Expression in Head Hair Follicles Plucked from Men and Women.
Ann Clin Lab Sci. 2006; 36(2):115-26.

Kirchheiner J and Brockmöller J. Clinical consequences of cytochrome P450 2C9 polymorphisms. Clin Pharmacol & Therap 2005; 77:1-16

Kirk BW, Feinsod M, Favis R, Kliman RM, Barany F. Single nucleotide polymorphism seeking long term association with complex disease.
Nucleic Acids Res 2002; 30:3295-311.

LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances.
Nucleic Acids Res 2009; 37:4181–4193.

Lathia CD, Amakye D, Dai W, Girman C, Madani S, Mayne J, MacCarthy P, Pertel P, Seman L, Stoch A, Tarantino P, Webster C, Williams S, Wagner JA. The value, qualification, and regulatory use of surrogate end points in drug development. Clin Pharmacol Ther. 2009; 86(1):32-43.

Lee CI, Leong SH, Png AE, Choo KW, Syn C, Lim DT, Law HY, Kon OL.
An isothermal method for whole genome amplification of fresh and degraded DNA for comparative genomic hybridization, genotyping and mutation detection. DNA Res. 2006; 13(2):77-88.

Lee JW, Devanarayan V, Barrett YC, Weiner R, Allinson J, Fountain S, Keller S, Weinryb I, Green M, Duan L, Rogers JA, Millham R, O'Brien PJ, Sailstad J, Khan M, Ray C, Wagner JA.
Fit-for-purpose method development and validation for successful biomarker measurement. Pharm Res. 2006; 23(2):312-28.

Lee JW, Weiner RS, Sailstad JM, Bowsher RR, Knuth DW, O'Brien PJ, Fourcroy JL, Dixit R, Pandite L, Pietrusko RG, Soares HD, Quarmby V, Vesterqvist OL, Potter DM, Witliff JL, Fritche HA, O'Leary T, Perlee L, Kadam S, Wagner JA.
Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report.
Pharm Res. 2005; 22(4):499-511.

Lehmann U. MicroRNA-profiling in formalin-fixed paraffin-embedded specimens. Methods Mol Biol. 2010; 667:113-25.

Lesko LJ, Atkinson AJ Jr.
Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies.
Annu Rev Pharmacol Toxicol. 2001; 41:347-66.

Lesko LJ, Salerno RA, Spear BB, Anderson DC, Anderson T, Brazell C, Collins J, Dorner A, Essayan D, Gomez-Mancilla B, Hackett J, Huang SM, Ide S, Killinger J, Leighton J, Mansfield E, Meyer R, Ryan SG, Schmith V, Shaw P, Sistare F, Watson M, Worobec A.

Pharmacogenetics and pharmacogenomics in drug development and regulatory decision making: report of the first FDA-PWG-PhRMA-DruSafe Workshop. J Clin Pharmacol. 2003; 43(4):342-58.

Li J, Harris L, Mamon H, Kulke MH, Liu WH, Zhu P, Mike Makrigiorgos G. Whole genome amplification of plasma-circulating DNA enables expanded screening for allelic imbalance in plasma. J Mol Diagn. 2006; 8(1):22-30.

Lièvre A, Bachet JB, Boige V, Cayre A, Le Corre D, Buc E, Ychou M, Bouché O, Landi B, Louvet C, André T, Bibeau F, Diebold MD, Rougier P, Ducreux M, Tomasic G, Emile JF, Penault-Llorca F, Laurent-Puig P. KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. J Clin Oncol. 2008; 26(3):374-9.

Lips EH, Dierssen JW, van Eijk R, Oosting J, Eilers PH, Tollenaar RA, de Graaf EJ, van't Slot R, Wijmenga C, Morreau H, van Wezel T.
Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed, paraffin-embedded tumors using single nucleotide polymorphism arrays. Cancer Res. 2005; 65(22):10188-91.

Liu A, Tetzlaff MT, Vanbelle P, Elder D, Feldman M, Tobias JW, Sepulveda AR, Xu X. MicroRNA expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-embedded tissues.
Int J Clin Exp Pathol. 2009; 2(6):519-27.

Lu Y, Gioia-Patricola L, Gomez JV, Plummer M, Franceschi S, Kato I, Canzian F.
Use of whole genome amplification to rescue DNA from plasma samples.
Biotechniques. 2005; 39(4):511-5.

Marrer E, Dieterle F.
Impact of biomarker development on drug safety assessment.
Toxicol Appl Pharmacol. 2010; 243(2):167-79.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN**.** Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 2008; 9:356-69.

McWhinnie AJ, Holman R, Fowels F, Casanova AA, Shah A, Wallis-Jones S, Stevens S, Madrigal JA, Marsh SGE. Saliva specimens collected with Oragene ONE are a reliable alternative to blood and buccal swabs large scale DNA extraction and HLA typing of recruits for Hematopoietic stem cell donor registries. In European Federation of Immunogenetics, Florence, May 2010.

O'Doherty C, Villoslada P, Vandenbroeck K.
Pharmacogenomics of Type I interferon therapy: a survey of response-modifying genes. Cytokine Growth Factor Rev. 2007; 8(3-4):211-22. .

Oliveira IO, Lhullier C, Brum IS, Spritzer PM.
The 5alpha-reductase type 1, but not type 2, gene is expressed in anagen hairs plucked from the vertex area of the scalp of hirsute women and normal individuals. Braz J Med Biol Res. 2003; 36(10):1447-54.

Oliveira IO, Lhullier C, Brum IS, Spritzer PM.
Gene expression of type 2 17 beta hydroxysteroid dehydrogenase in scalp hairs of hirsute women. Steroids. 2003; 68(7-8):641-9.

Pachner AR, Bertolotto A, Deisenhammer F.
Measurement of MxA mRNA or protein as a biomarker of IFNbeta bioactivity: detection of antibody-mediated decreased bioactivity (ADB).
Neurology. 2003; 61(9 Suppl 5):S24-6.

Penland SK, Keku TO, Torrice C, He X, Krishnamurthy J, Hoadley KA, Woosley JT, Thomas NE, Perou CM, Sandler RS, Sharpless NE. RNA expression analysis of formalin-fixed paraffin-embedded tumors.
Lab Invest. 2007; 87(4):383-91.

Ralf Paus, M.D., and George Cotsarelis M.D. The biology of hair follicle. New England Journal of Medicine. 1999; 341(7):491-97

Rani MR, Shrock J, Appachi S, Rudick RA, Williams BR, Ransohoff RM.
Novel interferon-beta-induced gene expression in peripheral blood cells.
J of Leukocyte Biology 2007; 82:1353-1360.

Rendl M, Polak L, Fuchs E.
BMP signaling in dermal papilla cells is required for their hair follicle-inductive properties. Genes Dev. 2008; 22(4):543-57.

Riveros C, Mellor D, Gandhi KS, McKay FC, Cox MB, Berretta R, Vaezpour SY, Inostroza-Ponta M, Broadley SA, Heard RN, Vucic S, Stewart GJ, Williams DW, Scott RJ, Lechner-Scott J, Booth DR, Moscato P; ANZgene Multiple Sclerosis Genetics Consortium.
A transcription factor map as revealed by a genome-wide gene expression analysis of whole-blood mRNA transcriptome in multiple sclerosis.
PLoS One. 2010; 5(12):e14176.

Roberts PJ, Stinchcombe TE, Der CJ, Socinski MA.
Personalized medicine in non-small-cell lung cancer: is KRAS a useful marker in selecting patients for epidermal growth factor receptor-targeted therapy?. J Clin Oncol. 2010; 28(31):4769-77.

Saridaki Z, Georgoulias V, Souglakos J. Mechanisms of resistance to anti-EGFR monoclonal antibody treatment in metastatic colorectal cancer.
World J Gastroenterol. 2010; 16:1177-87.

Scagnolari C, Bellomi F, Trombetti S, Casato M, Carlesimo M, Bagnato F, Lavolpe V, Bruno R, Millefiorini E, Antonelli L, Girardi E, Turriziani O, Antonelli G. Expression of biomarkers of interferon type I in patients suffering from chronic diseases. Clin. Exp. Immunol. 2007; 147(2):270-276.

Schneider MR, Schmidt-Ullrich R, Paus R. The hair follicle as a dynamic miniorgan. Curr Biol. 2009 Feb 10;19(3):R132-42.

Serana F, Sottini A, Ghidini C, Zanotti C, Capra R, Cordioli C, Caimi L, Imberti L. Modulation of IFNAR1 mRNA expression in multiple sclerosis patients. J Neuroimmunol. 2008; 197(1):54-62.

Shastry BS. SNPs in disease gene mapping, medicinal drug development and evolution. J Hum Genet. 2007; 52:871–880.

Shastry BS. SNPs: impact on gene function and phenotype.
Methods Mol Biol. 2009; 578:3-22

Silvestris N, Tommasi S, Santini D, Russo A, Simone G, Petriella D, Maiello E, Tonini G, Colucci G. KRAS mutations and sensitivity to anti-EGFR monoclonal antibodies in metastatic colorectal carcinoma: an open issue.
Expert Opin Biol Ther. 2009; 9(5):565-77.

Smolec J, DeSilva B, Smith W, Weiner R, Kelly M, Lee B, Khan M, Tacey R, Hill H, Celniker A, Shah V, Bowsher R, Mire-Sluis A, Findlay JW, Saltarelli M, Quarmby V, Lansky D, Dillard R, Ullmann M, Keller S, Karnes HT.
Bioanalytical method validation for macromolecules in support of pharmacokinetic studies. Pharm Res. 2005; 22(9):1425-31.

Socinski MA.
The emerging role of biomarkers in advanced non-small-cell lung cancer.
Clin Lung Cancer. 2010; 11(3):149-59.

Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids.
Am. J. Pathol. 2002; 161(6):1961-71.

Stenn KS, Paus R. Controls of hair follicle cycling.
Physiol Rev. 2001; 81(1):449-494.

Taylor JM, Ankerst DP, Andridge RR.
Validation of biomarker-based risk prediction models.
Clin Cancer Res. 2008; 14(19):5977-83.

van Krieken JH, Jung A, Kirchner T, Carneiro F, Seruca R, Bosman FT, Quirke P, Fléjou JF, Plato Hansen T, de Hertogh G, Jares P, Langner C, Hoefler G, Ligtenberg M, Tiniakos D, Tejpar S, Bevilacqua G, Ensari A. KRAS mutation testing for predicting response to anti-EGFR therapy for colorectal carcinoma: proposal for an European quality assurance program. Virchows Arch. 2008;45:417-31.

VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques. 2008; 44(5):619-626.

Weinstock-Guttman B., Ramanathan M., and Zivadinov R. 2008 Interferon-beta treatment for relapsing multiple sclerosis.
Expert Opin. Biol. Ther 8(9):1435-1447.

Williams BR. Transcriptional regulation of interferon-stimulated genes.
Eur J Biochem. 1991; 200(1):1-11.

Wu GY, Hasenberg T, Magdeburg R, Bönninghoff R, Sturm JW, Keese M. Association Between EGF, TGF-b1, VEGF Gene Polymorphism and Colorectal Cancer. World J. Surg. 2009; 33:124–129.

Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases.
Nat Rev Genet. 2009;10(1):43-55.