Università degli
Stuidi di Pavia

DIPARTIMENTO DI
INGEGNERIA INDUSTRIALE E
DELL'INFORMAZIONE

Fondazione IRCCS
Policlinico San Matteo

DOTTORATO DI RICERCA
IN BIOINGEGNERIA E BIOINFORMATICA

XXV CICLO

New technologies for novel diagnostic strategies in inherited cardiovascular diseases

DOCTORAL DISSERTATION OF:
Ing. Valentina Favalli

Advisor: Prof. Eloisa Arbustini

Tutor: Ch.mo Prof. Riccardo Bellazzi

SUPERVISOR OF THE PHD PROGRAM: Ch.mo Prof. Riccardo Bellazzi

2009 - 2012

to my mom,
"...because a Bioengineer
is always useful..."

Abstract

Il **Centro Malattie Genetiche Cardiovascolari (CMGCV)** dell'IRCCS Fondazione Policlinico San Matteo di Pavia, sede della mia attivitá di ricerca, si occupa della diagnosi e della cura di patologie rare su base ereditaria. Nel 2008, alla fine del mio corso di laurea, mi fu chiesto se ero interessata a lavorare in ambito di ricerca traslazionale presso il CMGCV, dove avevo completato il mio lavoro di tesi.
La sfida che mi venne proposta fu quella di integrare competenze bioingegneristiche all'interno di un team di ricerca multidisciplinare in cui operano biologi, biotecnologi, cardiologi, genetisti, anatomopatologi, infermieri e personale amministrativo. Il CMGCV, riconosciuto unitá complessa di eccellenza nel 2007 (delibera n.0080/C.d.A del 24.04.07), comprende un centro clinico, un laboratorio di genetica molecolare ed un laboratorio di patologia cellulare. Il personale del centro é formato da 27 professionisti con competenze diverse, di cui 8 strutturati all'interno del Policlinico ed i restanti sostenuti da contratti di ricerca e borse di studio: da questo si evince come in un IRCCS, assistenza e ricerca siano due facce della stessa medaglia, previlegiando nettamente la ricerca di tipo traslazionale, direttamente applicata all'assistenza, che genera nuove necessitá di ricerca, in un circolo virtuoso che ha portato il CMGCV ad una crescita esponenziale, misurabile sia come attivitá clinica che scientifica.
Data l'elevata complessitá delle patologie di cui si occupa, il CMGCV ha acquisito la strumentazione piú all'avanguardia presente sul mercato per lo studio clinico, molecolare ed ultrastruttrale delle malattie genetiche cardiovascolari.
Il CMGCV é riconosciuto centro di riferimento per le patologie rare di cui si occupa, che comprendono:

- Le malattie ereditarie del tessuto connettivo.
- Le malattie primitive ereditarie del muscolo cardiaco.
- Altre sindromi rare con coinvolgimento cardiovascolare.

In questa complessa realtá, nella quale la ricerca si alimenta real-time di nuove informazioni tratte sia da osservazioni cliniche che da strumentazioni ad alta tecnolgia anche molto diverse tra loro, il mio compito é stato quello di

integrare tecnologie, modelli e dati per gestire tutte queste informazioni al fine di trarne la massima resa e generare piani di implementazione.

Tratteró in questo lavoro uno dei problemi chiave che mi sono trovata ad affrontare durante il mio dottorato di ricerca, ovvero la diagnostica delle malattie genetiche cardiovascolari. In particolare la necessitá cui mi sono trovata di fronte é stata quella di stabilire il migliore utilizzo di tecnologie e metodi all'avanguardia per fare diagnosi appropriate nel minor tempo possibile, al minor costo e con le maggiori efficienza ed appropriatezza. La diagnosi genetica infatti si sta progressivamente rivelando un fattore chiave per impostare piani terapeutici farmacologici o inteventistici, per la diagnosi precoce di malattia o per la diagnosi prenatale.

Un esempio chiave dei miei ruoli riguarda i test genetici.
La tecnica attualmente considerata il gold standard per formulare una diagnosi genetica di malattia é il sequenziamento tramite metodo Sanger. In questa tecnica il gene viene suddiviso in regioni comprendenti le sequenze codificanti e le regioni fiancheggianti, quindi completamente sequenziato .
Questo tipo di approccio, seppur efficace nel caso di patologie con diagnosi clinica molto chiara a indirizzare l'indagine genetica (es: criteri di Ghent per la Sindrome di Marfan positivi per lo studio del gene *FBN1*), é stato essenziale per i percorsi diagnostici delle malattie monogeniche e non geneticamente eterogenee, ma é oggi difficile da sostenere per malattie monogeniche con eterogeneitá genetica come ad es. le cardiomiopatie dilatative, con altri 40 geni-malattia noti. In questi casi, infatti, individuato un gruppo di possibili geni-malattia, questi vengono tutti testati tramite sequenziamento Sanger sulla base di priorità definite dall'esperienza clinica del genetista, dai piú frequenti fino ai piú rari.

In questo contesto lo scopo che il mio lavoro di dottorato é stato quello di elaborare un un sistema che sistematizzasse le osservazioni del genetista basate su procedure standardizzate di diagnosi di patologie con base genetica, integrando dati derivati da indagine clinica, studio di biomarcatori specifici e quindi test genetici. Il progetto intendeva rispondere non solo ad un problema clinico di inquadramento del singolo paziente, ma anche ad un problema di indirizzo appropriato dell'indagine molecolare(test genetici guidati clinicamente) al fine di aumentare l'efficienza e l'efficacia dell'attivitá genetico-molecolare, diminuendo i costi tramite eliminazione di test superflui eseguiti su errate interpretazioni cliniche, diminuendo i tempi di indagine ed aumentando il numero di test appropriati.

Inizialmente lo studio prevedeva 3 step:

- Step 1 : identificazione di marcatori clinici (biochimici, imaging) malattia-specifici.

- Step 2 : identificazione di biomarcatori molecolari malattia-specifici.
- Step 3 : comparazione dei risultati e sistematizzazione del percorso diagnostico.

Successivamente, pur avendo acquisito sempre maggiore conoscenza delle malattie, ci si é resi conto che proprio a causa dell'eterogeneitá genetica sopracitata, restava comunque una percentuale di casi in cui il sequenziamento Sanger di tutti i geni candidati era necessario, ma diventava difficilmente sostenibile in un contesto diagnostico.
Per questo motivo si é deciso di acquisire un nuovo strumento che introducesse una nuova tecnologia di sequenziamento (massive parallel sequencing), che potesse rendere sostenibile la diagnosi genetica in questi casi.

IMAGING

I dati, collezionati dal 2000 a oggi, comprendono misure elettrocardiografiche, ecocardiografiche transtoraciche e trans-esofagee, angioCT scan, RMN cuore, esami obiettivi, esami di laboratorio, collezioni di eventi di diversa natura, alcune ripetute nel tempo e prese secondo linee guida standardizzate, altre piú descrittive. Dopo il primo filtraggio dei dati sopracitati, ho separato i dati ricavati da quelli piú raramente presenti o derivanti da procedure non routinariamente eseguite. Questo al fine da un lato di salvaguardare la significativitá dell'informazione, dall'altro non perdere dati anche poco rappresentati ma potenzialmente utili in un contesto di malattie rare. Inoltre mi sono confrontata con i clinici in modo da ottenere un dataset snello da utilizzare nelle analisi di sottopopolazioni, cercando di tenere conto anche dell'evoluzione temporale della patologia.
Quindi ho cercato di identificare marcatori aggiuntivi, operando sulle indagini strumentali maggiormente accessibili nel CMGCV.

Tra questi marcatori ho incluso lo studio di e-tracking per l'analisi della rigiditá (perdita di elasticitá) dell'albero arterioso. Questo parametro é candidato a svolgere un ruolo chiave nel monitoraggio di patologie quali quelle del tessuto connettivo, per valutare la risposta al trattamento farmacologico. Questo obiettivo ha richiesto di generare valori normali di riferimento. I dati ottenuti nella popolazione normale hanno consentito di studiare gruppi di pazienti che derivano da un trial clinico in doppio cieco in corso presso il CMGCV. Il follow-up si concluderá nel 2016. Per questo motivo le uniche analisi statistiche sono di carattere generale, non potendo ancora analizzare i dati per braccio di trattamento.

I BIOMARCATORI

Espressione genica
Lo scopo di questa parte dello studio é stato quello di individuare biomarcatori

da prelievi di sangue periferico confrontabili con i campioni ottenuti dal tessuto cardiaco (biopsie, espianti di cuore) generare test malattia-specifici, predittivi di mutazione. L'ipotesi di lavoro é che l'mRNA da gene mutato possa essere presente in quantitá diversa rispetto al normale. Pur essendo significative le differenze di espressione del gene mutato rispetto alla popolazione normale, con elevata sensibilitá, la specificitá é risultata bassa, con 10% di falsi positivi.

Marcatori biochimici
Lo scopo di questa parte dello studio, parallelamente alla gene expression, era quello di individuare marcatori malattia-specifici dal sangue. I primi risultati dello studio sono promettenti, indicando una differenza significativa della molecola target tra mutati e normali. Limitazione dello studio é la determinazione di valori normali di riferimento, che presentano una variabilitá molto elevata. Sviluppo futuro, in attesa anche in questo caso di poter usufruire dei dati derivati dal trial clinico, sará quello di selezionare una popolazione normale controllata, in modo da verificarne la variabilitá.

MASSIVE PARALLEL SEQUENCING

Nel 2011 il CMGCV ha acquisito lo strumento ROCHE GS Junior, sequenziatore di seconda generazione, tramite il quale si stanno mettendo a punto pannelli di screening genetico malattia-specifici, che permettano di sequenziare in una sola corsa dello strumento tutti i geni candidati a causare fenotipi simili (per esempio, tutti i 40 geni causali di cardiomiopatia dilatativa). Abbiamo svolto uno studio pilota utilizzando una piastra commerciale per l'identificazione di mutazioni sui geni *BRCA1* e *BRCA2*, che hanno riportato risultati incoraggianti dal punto di vista dell'individuazione di mutazioni effettivamente presenti (validate col metodo Sanger) e di tempi di preparazione. Ció che sié potuto fino'ora concludere é che il massive parallel sequencing puó rivelarsi uno strumento estremamente prezioso laddove vi sia la necessitá di studiare pannelli di geni e non si possa indirizzare su base clinica l'indagine genetica, anche se ancora nessuna considerazione puó essere fatta in rapporto ai costi/beneficio clinico effettivo di questa scelta.

CONCLUSIONI

Dai risultati fin'ora conseguiti emerge come un approccio di tipo bioingegneristico abbia permesso di contribuire in modo significativo e continuativo allo sviluppo gestione, e controllo sia delle attivitá di ricerca clinica sia delle attivitá quotidiane di un centro avanzato di diagnostica molecolare.

List of Abbreviations

Abbreviations	Description
BAV	Bicuspid Aortic Valve
DCM	Dilated Cardiomyopathy
EDS	Elhers-Danlos Syndrome
HCM	Hypertrphic Cardiomyopathy
LDS	Loeys-Dietz Syndrome
MFS	Marfan Syndrome
NGS	Next generation sequencing
RCM	Restrictive Cardiomyopathy
TAAD	Thoracic Aortic Aneurysm Disease

Contents

The role of a bioengineer in the real-life of a Public Research and Care Hospital can reflect existing models of bioengineering and clinical engineering services of the hospitals and contribute to exploit engineering knowledge and expertise in a biomedical environment.

After my graduation in Bioengineering I was challenged by the request of entering as bioengineer in the CMGCV with the aim of contributing to translational research of the Center focusing on the following issues:

A. Data: best collection, maintainance, management, use and control.

B. Technology: best use of the novel technologies entering the center.

C. Large numbers of families with different "rare" diseases are cared for at the CMGCV, requiring innovation in the models of care.

D. International research projects and evolving clinical protocols at the Center calling for the need of translational networks and collaborations with exchange of high volume data.

Based on the above issues, I was challenged to give my best contribution in the translational biomedical research of the Center.

In this PhD thesis I'll describe the projects in which I had a major role, as bioengineer, in the activities of a clinical and research Unit of a Research and Care Hospital of Public Right.

Chapter 1

The diagnostic work-up in inherited diseases

In this chapter, I will introduce the setting in which I worked during my PhD and the main topics of my activity.

CONCEPTS

- The research activity of the Center for Inherited Cardiovascular Diseases (CMGCV) deals with genetic cardiovascular diseases.
- Inherited cardiovascular diseases are complex diseases often demonstrating multiorgan/tissue involvement (clinical heterogeneity).
- Their care implies the clinical and genetic characterization of the disease in the proband and cascade family screening.
- The CMGCV developed a patient-centered model for the clinical work-up of inherited diseases, both diagnosis and treatments.
- Genetic heterogeneity is at the basis of several inherited diseases.
- In 2009 the gold standard for genetic testing was Sanger method.
- Combining clinical markers with Sanger sequencing, we achieved major results. As far as the number of novel disease genes increased, Sanger sequencing was too expensive to allow the screening of all known genes (up to 40 genes for one disease) in all probands.
- The working hypothesis was to explore disease-specific clinical-molecular-biochemical features (red flags) able to guide the genetic tests.
- My PhD work was based on the exploration of any potential contributor, clinical and technological, in order to support the exploitation of all research activities of the Center.

1.1 Background

The **Center for Inherited Cardiovascular Diseases (CMGCV)** of the Fondazione IRCCS Policlinico San Matteo, Pavia, in which I conducted my research activity, is highly specialized in the study and treatment of complex diseases. The CMGCV integrates three specialties: medical and molecular genetics, clinical cardiology and cardiovascular pathology.

The mission of the CMGCV can be summarized as follows:

- care for families with genetic cardiovascular diseases characterized by high complexity and depending on advanced biotechnology for diagnosis and treatment;
- translational research of the causes of diseases (from bench to bedside and viceversa) and validation of the most appropriate disease-specific treatments.

Patients and families with genetic cardiovascular diseases should meet a model that can cover all their needs without fragmentation of care.
In the vision of CMGCV patients and families are the center of multidisciplinary medical work-up; specialists converge their expertise in a unique clinical appointment that usually takes one single day in an outpatient setting. The patient/family-centered model replaces the discipline-centered model.

CMGCV takes care of patients and families with inherited cardiovascular diseases. These diseases are often syndromes with involvement of multiple organs and systems ("*a patient with a cardiovascular genetic disease is not only his/her heart or his/her vessels*"), in which the cardiovascular involvement is critical for the survival and quality of life. Major groups of diseases include: the primary myocardial tissue diseases (cardiomyopathies) commonly requiring heart transplantation, and connective tissue disease with aortic/arterial aneurysms and dissections, a major cause of morbidity and mortality in young-adult patients. Patients and families followed-up in the CMGCV come from the region Lombardia (40%) and from extra-region national ground (60%).

Facilities and CMGCV structure

The CMGCV consists of three units:

1. The Center of clinical cardiology and genetics for outpatients and families;
2. The Molecular genetic laboratory.
3. The Cardiovascular pathology and cell culture laboratory.

It harbors 27 between cardiologists, geneticists, researchers, biomedical engineers, technicians, nurses and administrative personnel.

The center is fully equipped for state of the art cardiologic, genetic, molecular and pathologic research, including tissue cell culture, and bioinformatics. Sponsored by the OSM Foundation, Ministry of Health, Telethon, INSERM, Region Lombardia, EU and local Foundations, the Centre runs research projects on clinical and molecular genetics of inherited cardiovascular diseases; family screening studies; genomic and transcriptomic research; cardiovascular pathology and tissue study of protein expression in cardiovascular diseases. The combination of patients and families from outpatient clinic, the molecular genetics and cardiovascular pathology provides unique facilities for **translational research** on inherited cardiovascular diseases.

1.1.1 Inherited rare diseases

Speaking about *INHERITED RARE* diseases, it's necessary to underline some key concepts:

1. The diseases are "chronic ", starting in pediatric age and progressing throughout the life span;

2. The diseases are familial, which means that each single member of the family has to be screened in order to identify the disease even in the pre-clinical phase (genetic testing or silent clinical markers, i.e. biochemical markers).

3. The presence of a genetic diagnosis can guide clinicians to use the best therapeutic strategy to prevent hard events. In the absence of a genetic diagnosis, the clinicians commonly rely on clinical manifestations, with the risk of missing the real problem when symptoms are not specific (i.e.: arrythmogenic risk and sudden death).

4. The management of these diseases is based on a multidisciplinary strategy, starting from the cardiac traits, which are the major causes of hard events, but also considering not-cardiac traits and markers.

As anticipated, the two major groups of diseases are:

- Hereditary connective tissues diseases.
- Hereditary diseases of the heart muscle.

In the above descriptive nosology there are two major genetic scenarios:

- Monogenic disorders without genetic heterogeneity (one gene $\rightarrow$ one disease) but with partially overlapping phenotypes. These diseases are caused by a single disease gene such as Marfan syndrome (MFS) or Ehler-Danlos syndrome (EDS). The overlapping phenotypes (i.e. MFS and LDS syndromes) were only appearently misleading when the clinical spectrum of the two diseases was not fully elucidated. Experienced teams have now the capacity of making clinical diagnosis in about 90% of cases, thus addressing genetic tests to the right genes.

- Monogenic disorders with genetic heterogeneity (one disease $\rightarrow$ more genes), namely diseases that are caused by different genes when the major phenotype is considered on the basis of old descriptive nosologies. When the major phenotype is flanked by minor disease-specific markers, different in the diseases caused by different genes, the specificity of diagnostic hypothesis increases. The main aim in this setting is to reach the final ideal scenario: one gene $\rightarrow$ one disease, as for genetically homogeneous diseases.

The strategy developed at the CMGCV is to take care of the patient and his/her family and identify the genetic cause of his/her disease; to find the best known therapy for the specific patient/disease; to monitor the patient during all his/her life to prevent hard events; to perform all the examinations in the same day in order to permit to the patient a regular cohabitation with his/her disease with minimal restrictions of his/her life style. For these reasons all clinical and genetic extracardiac evaluations that do not require the use of specialized equipments are carried out at the center. All the assessments requiring investigations that depend on instruments placed in different clinics imply the transfer of patients within the hospital (i.e.: imaging tests such as CT angiography and magnetic resonance imaging are performed in the Radiology Unit).

1.1.2 The CMGCV diagnostic model

The CMGCV was formally established in 2006 to cover a need not met by the existing clinical care models, namely to place the patient and his/her family at the center of the health-care model. Today the CMGCV follows-up about 4000 families.

Because of their complexity, these diseases [involving multiple organs and systems and with many of these with a recent nosological definition] often encounter a mismatch between clinical needs and diagnostic-therapeutic offer. The clinical signs at onset can be variable (i.e.: ocular, skeletal, cardiac, vascular, neurological, gastrointestinal, cutaneous). The individual specialist evaluations rarely integrate markers or symptoms involving other organs or systems.

These diseases demonstrate an age-dependent expression, from pediatric to adult age. Further, the clinically overt onset of these disease may occur in previously "healthy" individuals in a normal familal, social and economical context. Often catastrophic events occur suddenly, with impact not only on the patient but also on the whole family. Inheritance is a major determinant of the family fate: part of the family members are often affected but unaware of their disease or will develop the disease during their lifetime. In particular, since many of these diseases are transmitted as autosomal dominant Mendelian characters (50% risk of transmission for each new pregnancy) and since fertility is not affected and the diseases themselves can be asymptomatic for long periods of time, preventive actions may appear impossible. Vice versa, preclinical (genetic) and early (presymptomatic) diagnoses in families are possible when the disease is recognized in the proband. The preclinical diagnosis involves monitoring and early treatments, aimed at delaying the onset and the progression of the disease and preventing potentially fatal complications (rupture of aorta for aneurysmal disease of the aorta and sudden arrhythmic death in cardiomyopathies) or severely debilitating events (stroke, chronic heart failure, chronic renal failure).
More organs and systems can be involved, requiring a novel view of patient and family healthcare management. For this reason the CMGCV trained administrative personnel specialized in the organization of disease specific PDTA (Percorsi Diagnostico-Terapeutici Assistenziali = clinical work-up) to avoid fragmentation of care, and to introduce each patient in tailored programs of global care, with a balanced combination of quality of life and protection from catastrophic preventable events.

In this context, research may explore any novel frontier to generate knowledge and translation. New knowledge can be achieved using old methods when clear disease-specific mindset is the guide. This means that even old tools can generate innovation. New knowledge requires on one hand, substantial changes of the structure of the diagnostic work-flow and, on the other hand, the ability to respond to new queries. The **translation** must be done in real time to

be clinically effective: today specific diagnoses based not on the descriptive phenotype (enlarged heart, dilated aorta) but on genetic tests, save lives, and often allow patients to continue a normal family, social and economic life.

The keys of sustainability of the model are knowledge and investigation: observations and recording of data, listening to the patient and relatives, study of documental material, history of families, driving the summary that every patient makes of his history, with specific questions that are finalized to capture the most useful information to guide clinical suspicion to specific genes. A patient who goes to a cardiologist, rarely tells about his kidney stones or its visual disturbances. When asked about the family the narrative concentrates on heart attacks, sudden deaths but not on stroke or kidney transplantation in parents or siblings or uncles or cousins. Likewise, if one patients goes to the nephrologist, he will not mention a defect of vision, or skin lesions. The examples are numerous. Therefore each specialist handles the problem of his pertinence and the patient lives booking visits and never reaching a specific diagnosis. In the setting of chronic diseases, the absence of specific diagnosis, often makes treatment repeatedly unsatisfactory. This leads to search for second opinions, further specialists, and evaluations.
The costs of clinical knowledge, observation, listening, the study of signs and symptoms and information assets are very low, certainly less expensive than other instrumental procedures often asked to substitute the clinical skills. The low-cost investigative activity is the basis of medical clinical governance ensuring that resources are well targeted and used for most appropriate aims, for the patient and the family.

1.1.3 The genetic strategy

The CMGCV strategy in genetic testing, in particular for diseases with genetic heterogeneity (TAAD, cardiomyopathies, different types of cancer, melanoma, etc.), is to take as much clinical information as possible to guide focused genetic testing (the right gene for the right disease, Figure 1.1). This procedure is obviously easy to state but difficult to realize. However, we have now evidence that similar phenotypes associated with defects of different genes are different diseases and that deepening the clinical evaluation of both cardiac and non cardiac traits in both probands and relatives may highlight clinical traits recurring more frequently in association with defects of one or few genes. Key clinical traits represent **red flags** useful to guide genetic testing. Once clinical genetic evaluation has been completed [and cascade family screening may help to identify markers that are perhaps present in early phases of the disease and no longer recognizable in later phases], probands are counseled and asked to sign the informed consent form for the genetic testing.

The mutation analysis of the disease genes is performed according to three levels:

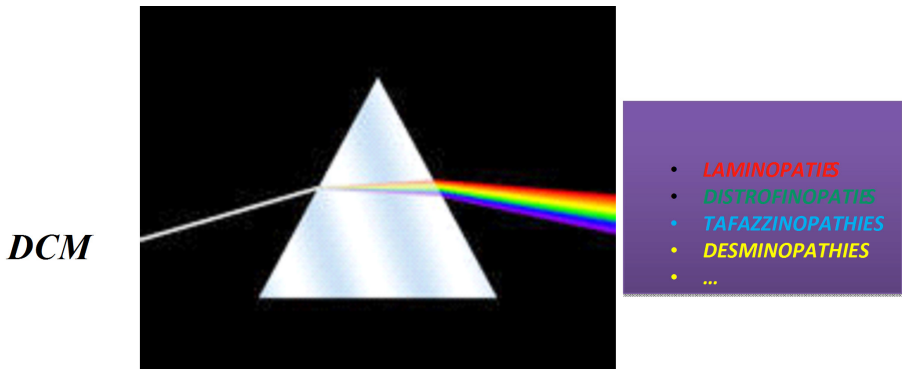


Figure 1.1: The right gene for the right disease.

1. I level: DHPLC screening of amplicons including exons and flanking regions of each tested gene. Exceptions for starting from II level testing are: urgent request (pregnancies, surgeries with high risk, fast progression of a risky phenotype, special medical request for treatments)

2. II level: Direct sequencing of all coding and flanking regions when the first level test is negative and the clinical suspicion remains strong.

3. III level test: MLPA, when the former two steps tested negative.

The requirements for interpretation of the mutations include: a) absence of the mutation in large control series; b) segregation of the mutation in the family by age, if the disease is familial; c) prediction of a severe effect in the protein (i.e. splice site, stop, frame shift ins/del); d) for missense mutations, evolutionarily conservation of mutated residues and potential damaging effect of the mutation predicted by *in sylico* analysis (i.e.Polyphen, SIFT, etc.).

With this approach the probability of mutation identification approaches values up to 95% patients with clear phenotypical manifestations (ex: MFS and ectopia lentis associated with *FBN1* gene mutations).

1.2 Study aims and study design

In 2009, when I started my doctoral course, it was clear that at the basis of phenotypically similar diseases there was a genetic heterogeneity with i.e. more than 40 diseases genes. Molecular diagnostics, with the identification of the genetic defect causing the specific disease, is the only way to create a novel nosology based on the genetic cause, suitable for a disease-specific clinical management and search for new specific-diseases therapies.

One of the end points of my PhD project was to integrate all data collected by molecular and clinical work-up, and to find phenotype-genotype correlation in term of red flags in order to generate disease-specific check lists useful for addressing genetic tests. The project aimed at responding not only to a diagnostic problem of classification of the individual patient, but also to the need of appropriate assignment of resources. As a result, the research of the causative mutation increases its efficiency and effectiveness, reducing costs through the elimination of unnecessary testing due to clinical misinterpretation, lowering the time of the investigation and increasing the number of final tests per year.

Different types of genetic diseases were studied with the similar aims, and the study was supported by:

- The European project INHERITANCE on dilated cardiomyopathy and the project from Ministry of Health 2008 on hypertrophic cardiomyopathy;
- The project from Ministry of Health for Young Researchers 2009 on aneurysmal inherited syndromes and the Telethon project GGP08238 on Marfan syndrome.

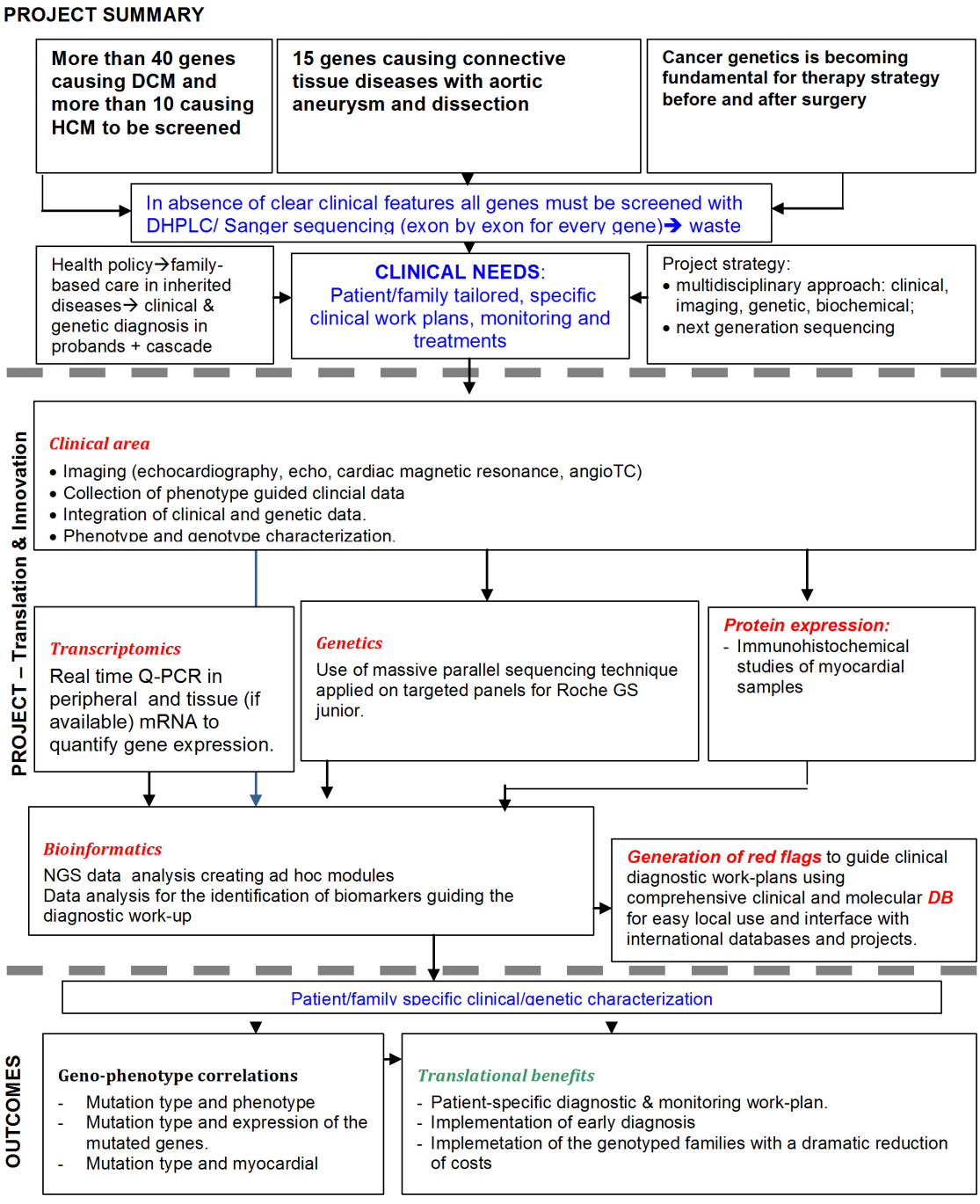


Figure 1.2: Study design

Chapter 2

Inherited Connective Tissue Diseases

In this chapter I will analyze inherited diseases of the connective tissue. In particular I will focus on MFS, which was one of the main topic of my analysis. This project was supported by Telethon grant GGP08238. I will detail the project aims and the connections with the aims of my PhD, and I will present results.

CONCEPTS

- Heritable connective diseases are life-threatening disorders. Main complications are aortic aneurysms (AA), dissection (AD) and rupture.
- Dissection and rupture of the aorta are potentially fatal events.
- Several similar disorders are termed MFS-like syndromes because of the overlapping phenotypes with MFS.
- MFS is the most known disease, but key concepts on the mechanisms of this disease are only partially elucidated.
- MFS is a monogenic disorder. In most cases (about 90%), clinical signs guide to a certain genetic diagnosis.
- About 30% of patients addressed to our attention present confounding MFS-like features.
- MFS and MFS-like populations regularly monitored by the CMGCV includes more than 50% of the overall families cared for the Center.
- For above reasons I decided to focus on MFS for the initial evaluation of my project plans, to establish whether the MFS model of care generated in the CMGCV can be exported and etended to other complex multi-organ syndromes.

2.1 Heritable connective tissue diseases

Heritable disorders of connective tissue are multi-organ/tissue rare diseases caused by defects of molecules involved in the collagen and microfibrillar structure, synthesis, degradation or signalling pathways. These disorders include several clinically defined conditions and as many as 2 million people in Europe may be affected. This proposal focuses on a few more prevalent and malignant diseases, namely Marfan Syndrome (MFS), Loeys-Dietz Syndromes (LDS1, LDS2, LDS3 and LDS4),familial Thoracic Aortic Aneurysm and Dissection (TAAD) and familial Biscupid Aortic Valve (BAV) with and without aortic aneurysm.

Short description of the syndromes

Marfan Syndrome (prevalence 3:10,000) is one of the best clinically and genetically characterized disease [1, 2]. Although it affects cardiovascular, skeletal, ocular lung, skin and nervous systems, the major clinical problem lies in the aortic root dilatation and dissection with impact on morbidity and mortality [3]. MFS is caused by defects in the gene encoding *fibrillin-1* (*FBN1*). Other than being a protein that constitutes the structural support for connective tissue, *fibrillin-1* binds Transforming Growth Factor β (TGFβ) whose effect on lungs, heart valves and aorta weakens the tissues and causes the cardiovascular and lung features of MFS. Existing angiotensin II receptor blockers (ARBs) reduce the levels of TGFβ [4].

The Loeys-Dietz syndrome types 1 and 2, first described in 2005 [5, 6], are caused by defects of *TGFβR1* and *TGFβR2* genes. Controversial nosology is a matter of current scientific attention because mutations of these genes have been associated with phenotypes that may partly overlap with TAAD [7], LDSs were originally reported as not including major ocular traits [8, 9] but recently reported as rarely presenting with ectopia lentis [10]. Whatever will be the final nosological position of diseases associated with mutations of these two genes, the vascular phenotype of LDS patients seems to be extremely severe with higher risk of aortic dissection and rupture at diameters lower than those observed for example in MFS [11]. Recently LDS type 3 and 4, caused respectively by *SMAD3* and *TGFβ2* genes were described [12, 13]. They present overlapping manifestations with LDSs 1 and 2 characterized by early-onset osteoarthritis (LDS3) and multiple combination of non-vascular traits (LDS4).

TAAD has been further and recently associated with mutations of the *ACTA2* [14] and *MYH11* [15] genes (encoding smooth muscle cell actin and myosin, respectively) in patients who show different extra vascular phenotypes but the number of described families/cases is still very low, and a major effort to characterize geno-phenotype correlations should be made before generating

a precise and complete profile of the disease.

An additional condition that was simply defined as congenital in the past has been recently demonstrated to have a genetic origin: BAV (1-2% of the general population) [16]. The calcified type can be associated with mutations of the *NOTCH1* gene. Non-syndromic BAV can segregate in families according to an autosomal dominant mode of inheritance. Since BAV is a common condition and 25% of the cases develop aneurysms of the ascending aorta, the potential role of the defective NOTCH1 protein on the aortic wall needs to be explored in order to establish whether the aortic aneurysm is exclusively due to valve dysfunction or if there is a role of the gene defect in the aortic wall weakness. However, since isolated BAV can function normally until a very late age, the open key question is how to detect BAV cases/families in which the natural history of the disease will lead to valve dysfunction and aortic aneurysm.

Ehlers-Danlos Syndrome type IV (caused by mutation in *COL3A1* gene) is a rare monogenic vascular disease that often presents with arterial dissection and rupture as well as spontaneous organ rupture. Aortic aneurysm and dissection may also occur [17].

Filling the gap between existing knowledge and translational benefits

European specialized centers are focusing on both research and clinical programs dedicated to these patients. However, translation programs are still limited and only a few thousand patients benefit clinical and molecular diagnostic tools. For example, the estimate of EU individuals affected by MFS (about 1,500,000) vs. existing genotyped series which roughly coincide with best clinically-cared patients (a few thousands) indicates a mismatch between the potential offer and population benefits. When I checked the appropriatness of patients afference to the center I found that they come to the clinical attention:
a) after a first aortic dissection (about 40%);
b) after a second event in the family (about 10%);
c) for isolated cardio-aortic problems (10%);
d) for phenotype traits such as tall stature, skeletal abnormalities, ocular problems, cardiovascular phenotypes without acute events (about 30%);
e) for major isolated ocular problems, ectopia lentis in particular (2-5%);
f) for a combination of skeletal and other organ/tissue phenotypes and mental retardation (1-2%);
g) for inappropriate request of clinical and genetic evaluation in tall individuals with associated minor signs such as myopia or mitral valve prolapse (<10%) after several specialist evaluations in patients whose phenotypes are still clinically uncharacterized and have no conclusive diagnosis nor likely diagnostic

hypotheses.
Overall, the rate of non appropriate access (affected by diseases that do not meet the expertise of the Center) to our Center was about 30% in 2007.

Past actions: We established a pre-admission evaluation of all clinical reports existing foe each patients.

In 3 years, we were able to decrease the number of non-appropriate access to the Center by 20%. In 2012, less than 10% of patients were admitted to the Center with non-appropriate indication.

Future actions: by increasing the number of observations (and recording disease-specific markers) we plan to further decrease the number of non-appropriate access to the center and to improve the timing priority selection.

The present health care for these patients at the national level appears fragmented, with only a few thousands of patients and families best cared for, with respect to the existing possibilities. The above data indicate that:

1. There is a gap in translation from bench to bed-side;
2. There are disorders that are still genetically and phenotypically orphan;
3. Only clinically overt phenotypes capture the clinical attention and patients are alerted to the referral centers. Prior efforts made to pool data and generate large genotype-phenotype correlation studies highlighted the role of family screening and genetic testing which are now routinely performed in all referral centers [18, 19, 20, 21].

The aim of my PhD project in the study of these diseases was to find new **clinical**, **molecular** and **biochemical** non-invasive disease-specific markers.

2.2 Specific project on novel biomarkers in MFS

Telethon project GGP08238 started in Oct 2008 and closed in Oct 2012. The overall objective of the project was to characterize the differential serological levels of TGFβ and transcriptomic signature of the TGFβ pathway in MFS patients carrying *FBN1* mutations, treated with the Angiotensin II receptor 1 blocker (ARB) Losartan vs. β selective blocker (BB) (Nebivolol) vs. the association of both.

Specific aims:

1. to comparatively evaluate the quantitative expression of the mutated *FBN1* and the trascriptomic profile of the TGFβ pathway genes.
2. to quantify serological levels of the active TGFβ.
3. to correlate transcriptomic and biochemical data with quantitative aortic root data.

Background/rationale

More than 18.000 patients are expected to be affected by MFS in Italy (3:10000). The major clinical problem is the potentially fatal aortic aneurysm/ dissection. There is no cure available, but recently drugs that were routinely used to treat hypertension, AT1R-blockers (ARB), were shown to prevent the aortic damage that underlies the aneurysm and dissection, opening new hopes for preventing life-threatening events. The benefits are being comparatively evaluated measuring, with non-invasive tools, the aortic root diameter and the indexes calculated on the body surface area by age.

The potential key molecule that negatively influences cell growth, differentiation, survival and death in MFS is TGFβ, which is antagonized by ARBs. In experimental setting, ARBs proved to be superior to the previously used β-blockers (BB) in the prevention and limitation of aortic wall damage and ACE-Inhibitors, active at an upstream level of ARB, and were proven to be clinically superior to BB in recent controlled small clinical trial. Both ARBs and ACE-I antagonize Angiotensin II at different levels.

Description of the project

We systematically measured the quantitative expression of the causative mutated gene and downstream key genes of the TGFβ pathway as well as the serological levels of the active/total TGFβ in 291 genotyped non-operated patients with MFS and aortic root dilation treated for 48 months with Nebivolol (n=97), vs Losartan (n=97) vs the association of both (n=97). The results will be correlated with quantitative imaging data of the aortic root.

2.3 The clinical trial

We designed and performed an open-label phase III study with a sample size of 291 patients including 20% dropout [22].
The criteria of inclusion were the diagnosis of MFS, the identification of mutations in Fibrillin1 (*FBN1*) gene, the aortic root aneurysm.
The primary end point was the comparative evaluation of the effects of losartan, nebivolol and the association of both on the progression of aortic root growth rate.
The official start-up date was October 15, 2008 and the last enrollment was July 20, 2012, acknowledged by the Ethical Committee in Sep. 15, 2012. We closed the enrollment with 262 patients because the drop out was much lower than planned (8.5% vs. 20%). This was formally communicated to the local Ethical Committee on September 10, 2012.
Based on the need of maintaining the blindness until the end of the study, patients who completed the follow-up are still blind and treated according to

their arm of randomization.

The primary outcome was the decrease of rate of aortic root growth measured by echocardiography at the level of sinuses of Valsalva. The measure will be expressed in mm of growth per year, ARR and Z-score.
Secondary endpoints were comparative arterial stiffness, levels of total and active TGFβ1, *FBN1* mRNA in the groups before, during and at the end of the study, in order to find MFS-specific biomarkers as previously anticipated.

My role in this study was to contribute to the design of the clinical trial and to analyze data collected during the project.

2.3.1 Trial methods and genotyping

The series is constituted of 291 patients, aged 1-55 years, who did not undergo prior aortic surgery, diagnosed with MFS according to Ghent criteria and carriers of FBN1 gene mutations. The trial is designed as a randomized open phase III trial. The patients have been randomized to Losartan (n=97), Nebivolol (n=97) and both drugs (n=97). They have been treated for 48 months. We tested the hypothesis that the combination of Angiotensin-II receptor blockers ARB (anti-TGFβ effects) (losartan) and selective b-receptor blocker (nebivolol) (BB) (with anti-shear stress, anti-stiffness and heart rate lowering effects) were superior to the single ARB or BB in reducing the progression of the aortic root dilation (primary end-point). We enrolled only fully informed and consenting individuals (legal guardians for minors). The clinical series included both adult and children, males and females.

Power of the study and sample size
A total sample size of 291 patients (97 per group) was calculated to detect a difference in the progression of the ARD between any 2 groups of 0.75 mm per year, with a power of 90% and an overall type I error of 5%, while allowing for multiple comparisons between groups, 20% attrition and one interim analysis planned at 24 months.
Based on the literature, we retrieved a potential change in ARD of 1.5 mm/year in the Nebivolol group; we expected an ARD change of 0.75 mm/year and 0 mm/year in the Losartan and Nebivolol+Losartan groups, respectively. We assumed a common standard deviation 1.3 mm/year. The values will be provided at the end of follow-up both as numbers and z-score.

Randomization and stratification.
Eligible subjects had been randomly assigned in a 1:1:1 ratio to receive Nebivolol or Losartan or Nebivolol+Losartan. Random allocation of treatments balanced in blocks of different size was performed, while stratifying on age (child,<16 and adult >16 years), with the use of the Stata 10 software.

The strata were defined by the BSA-adjusted aortic root z score at baseline within the interval <4.5/4.5 SD. The study was conducted entirely in the single referral center. The patients had been monitored (and part of them are still in follow-up) with 2D echocardiography using two Aloka Prosound Alpha10 instruments equipped with e-tracking software every 12 months ± 3 months.

Patient Enrollment

Before starting the trial, we had identified about 600 carriers of *FBN1* gene mutations.

At the starting date, genotyped patients eligible for the clinical trial were contacted and informed by geneticists about the possibility to participate. A three week washout period was indicated to previously treated patients. At baseline examination they signed the informed consent.

The protocol study was registered at Clinical Trials.gov with the number NCT00683124 and it is recognized by the Italian Drug Agency (EudraCT number 2008-001462-81). Telethon Grant GGP08238 funded the molecular and biochemical studies of the trial.

2.4 Functional Imaging and aortic stiffness: e-tracking method

This section deals with a secondary end-point of the trial: the evaluation of the arterial stiffness in patients treated with different drugs.
Arterial stiffening is a hallmark of aging and a correlate of major atherosclerosis risk factors. Progressive spatial disorganization and fragmentation of elastin fibers, together with an increase in collagen, stiffen arterial walls. Indices of aortic stiffness which predict cardiovascular mortality are being used to detect preclinical vascular disease and to monitor responses to treatment.
Arterial stiffness has been reported to be increased in MFS, thus supporting the rationale for using drugs with anti-stiffness effects. Serial measurements of arterial stiffness could be useful for monitoring the effectiveness of medical treatments. We aimed at comparing stiffness in patients with MFS vs. control population [23].
Conventional two-dimensional ultrasound imaging has been used to measure changes in vessel diameter, but it is not accurate enough for the assessment of stiffness because of the low sampling rate of B-mode images. Ultrasound wall tracking produces precise waveforms of changes in diameter and allows the calculation of various indices of arterial stiffness in relation to changes in blood pressure (BP).

The hypothesis of this part of the study was that values of arterial stiffness of the carotid artery can mimic the stiffness of the aorta.

Arterial stiffness had been measured at the carotid level using the ALOKA Prosound Alpha 10 echocardiographic machine, equipped with e-tracking software, using the 10MHz array probe.
The carotid stiffness had been recorded using the long-axis view (Figure 2.1). The recorded track of at least 5 consecutive beats had been processed. Subjects were studied after resting supine for more than 10 minutes. The change in diameter of the vessel was measured as the difference between the displacement waveforms of the anterior and posterior walls, using the e-tracking technique, with the cursors set manually to track the media-adventitia boundaries in the arterial wall approximately 1 cm proximal to the carotid sinus. Brachial systolic and diastolic pressures measured by sphygmomanometry were used as surrogates for the carotid systolic and diastolic pressures.

Four parameters of local arterial stiffness were calculated: pressure-strain elastic modulus (epsilon, Ep), stiffness parameter (beta, β), arterial compliance (AC), and local PWV.

Ep = (Ps Pd) / [(Ds Dd) / Dd] (in kPa)

β = ln (Ps / Pd) / [(Ds Dd) / Dd] (units)

AC = (Ds Dd) / (Ps Pd) (mm^2/kPa)

PWV = $\sqrt{(\beta Ps/2\rho)}(m/s)$

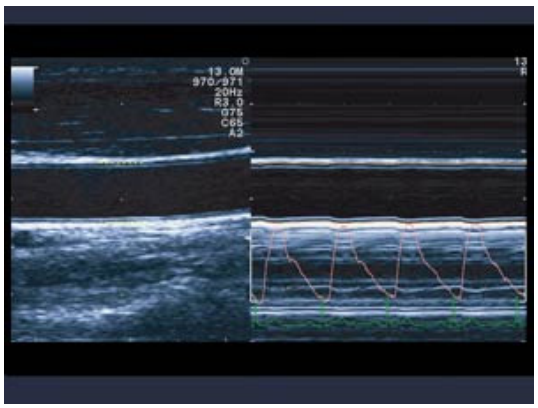


Figure 2.1: 2D ecocardiographic long axis carotid image for wall tracking.

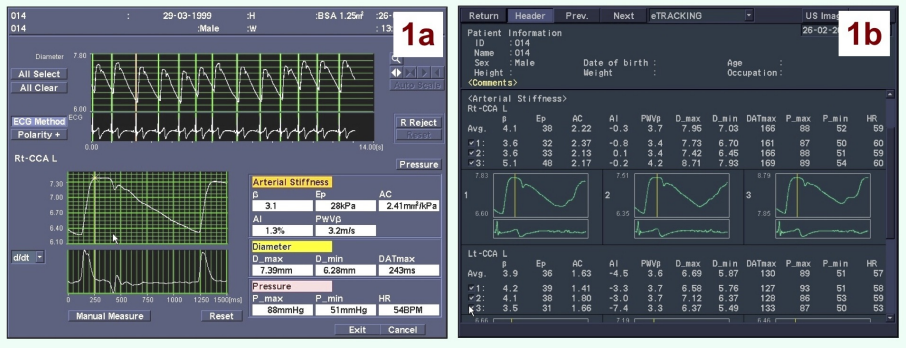


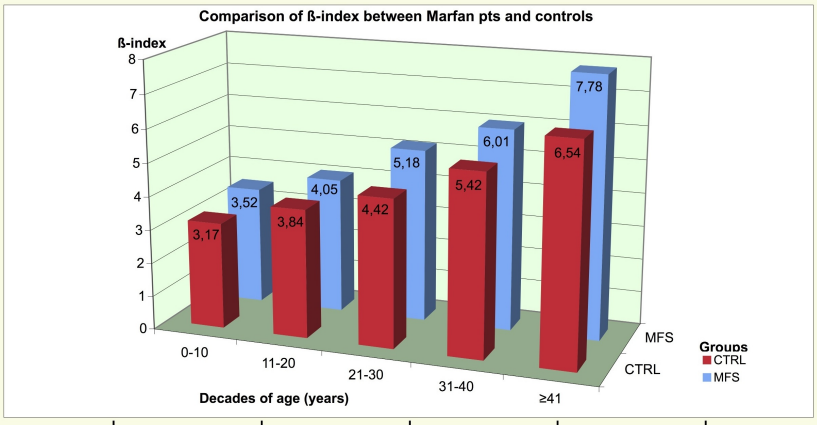
Figure 2.2: a) one of the 3 consecutive measures taken in each patient. Five consecutive beats were chosen; b) in each patient the software calculates the average between three measures.

Where Ps and Pd are systolic and diastolic BP; Ds and Dd are maximal and minimal diameters of the artery; and ρ is the density of blood (1050 Kg/m^3). These parameters were analyzed using a dedicated software package (DAS-RS1, Hitachi-Aloka, Tokyo, Japan).

2.4.1 MFS preliminary results

We measured carotid stiffness in 182 patients enrolled in the clinical trial at baseline evaluation and age- and gender-matched normal individuals (CTRL) using Prosound Alpha 10echo machines (ALOKA, Tokyo, Japan). Mean age was similar in pts and controls (MFS: mean $\pm$ DS 22.47 $\pm$ 13.26 years; CTRL: 27.51 $\pm$ 15.28 years). Arterial stiffness was evaluated as β index at the level of the right carotid common artery using linear-array ultrasound transducers, 1 cm upstream the carotid body. The value of β index for each patient was the average of 3 measurements, recording 5 consecutive beats each, with concurrent arterial pressure measurement (Figure 2.2).

Three independent operators performed the study with interobserver variability of 12% calculated by the Bland-Altmann test. In the 2 groups β-



Decades of age:	*0-10 (p<0.05)*	*11-20 (p=0.05)*	*21-30 (p<0.05)*	*31-40 (p<0.05)*	*≥41 (p<0.05)*
Marfan Group:	3.52±0.89 (N=50)	4.05±0,77 (N=50)	5.18±1.51 (N=33)	6.01±1.76 (N=34)	7.78±3.01 (N=15)
Controls:	3.17±0,91 (N=115)	3.84±0.99 (N=195)	4.42±1.18 (N=103)	5.42±1.46 (N=116)	6.54±1.94 (N=124)

Figure 2.3: In each class of age, β-index is significantly higher in MFS patients than in controls: p<0,05 in each class of age.

indexes were compared using a two-tailed Student t-test for unpaired data (alpha=0.05). Statistical analysis was performed with STATA 10. The β-index value increases with age both in controls and Marfan patients (Figure 2.3) and remains significantly higher in Marfan patients than in controls in each decade of age.

This new marker could contribute to monitor the effects of treatments in connective tissue diseases. However reference values in the normal population were limited. Therefore specific reference values in large samples of normal individuals were needed.

2.4.2 Normality values for e-tracking measures

We participated to the E-tracking International Collaboration (ETIC) in order to create a large database of measurements of carotid arterial stiffness in healthy subjects from centers across Europe, by pooling existing data obtained using the echo-tracking (e-tracking) method (Hitachi-Aloka, Tokyo, Japan). The objectives of the present study were to establish age-specific reference values for arterial stiffness indices, and to evaluate their relationships with blood pressure.

ETIC study methods

The ETIC database included healthy subjects aged 3-85 years who had ultrasound scans of the right or left common carotid artery for the assessment

of arterial stiffness and wave travel, using the e-tracking technique, and for whom background characteristics and risk factors were available. Data were collected in 14 centers across Europe.
Data recorded at the time of study included age, gender, body height and weight, body mass index (BMI), heart rate, and brachial systolic and diastolic blood pressure. The data were inspected for aberrant values, and summary statistics for the measures of stiffness, blood pressure, heart rate and body mass index (BMI) were calculated and presented as mean and standard deviation. Inter-center comparisons were made to ensure there was no systematic variation between centers.
Altmans method was used to derive age-related centiles for the stiffness parameters.
The data were transformed, using the Box-Cox procedure to select an appropriate transformation, to produce approximately normally distributed (Gaussian) data. The dependence of the mean on age was modeled by a cubic function, fitted by linear regression, with a linear function of age for the standard deviation based on regression modeling of the absolute residuals.
For some of the parameters there was evidence of different relationships with age in subjects aged <18 years, compared to adults, and so segmented regression was used to fit different models in these two broad age groups; this method ensured continuity in predicted values for ages below and above 18. Age-dependent centiles were calculated based on these Gaussian distributions and the transformations were inverted to produce centiles for the original variables.
Z-scores of $\pm$ 2 , $\pm$ 1 and 0 were calculated and plotted; these correspond to the 2.5^{th}, 16^{th}, 50^{th}, 84^{th} and 97.5^{th} centiles (encompassing mean values $\pm$ 1 or $\pm$ 2 standard deviations).
Interactions between age and gender were tested in all the subjects by general linear modeling on the transformed data, with age treated as a categorical variable.
The total number of healthy subjects was 2,042 (1,127 male and 915 female), and they were aged from 3-85 years. We contributed with 944 controls.
Systolic BP and pulse pressure were both higher in men ($p<0.001$); they increased with age more in women so that the difference between genders narrowed (interaction, $p<0.001$).
Diastolic BP was higher in men ($p<0.001$) and showed no significant interaction with gender and age ($p=0.094$).
BMI was higher in men ($p<0.001$); the difference was reasonably constant across ages but there was some evidence for an interaction by gender ($p=0.034$) with a steeper increase in BMI in men.
Figure 2.8 show the 2.5^{th}, 16^{th}, 50^{th}, 84^{th} and 97.5^{th} centiles for all the parameters, presented as median values and z-scores; these relate to $\pm$ 1 and $\pm$ 2 standard deviations from the mean, of the transformed, approximately normally distributed, variables.

Epsilon: Ep increased from the third to the seventh decade (ages 18-29 to 60-74 years) by a factor of 2.4 in men compared with 2.7 in women. There was no significant overall difference between the genders (p=0.88) and no significant interaction with age (p=0.12).

β index: β increased over the age range by a factor of 2.1 in men and 2.4 in women. It was slightly higher in women (p=0.034). There was a marginally non-significant interaction with age and gender: the values were comparable in young adults but higher in women at older ages (p=0.07).

Arterial compliance: The relative reductions in AC (the inverse of stiffness) between ages 18-29 and 60-75 years were 41% for men and 49% for women. AC was higher in men (p<0.001); the gap between males and females gradually widened with increasing age (interaction, p=0.004).

Pulse wave velocity: The relative increases in local PWV over the observed age span were 1.5 in men and 1.6 in women; there was no significant difference between genders (p=0.62) and no significant interaction with age (p=0.36).

Ep and β increased with age, with slightly accelerated changes in later decades. Reference ranges gradually widened with age, as also reported for carotid-femoral PWV [24].

Limitations

Brachial BP was used as a substitute for carotid BP. In younger subjects, the amplitude of the pressure wave increases as it travels towards the periphery, so systolic BP and pulse pressure are generally higher in the brachial artery than the central aorta (pulse pressure amplification). In older subjects, there is less amplification and brachial and central BPs are more similar. Therefore, carotid Ep, β and PWV may have been overestimated especially in younger subjects by using brachial rather than carotid pressures. Nevertheless, several epidemiological studies have used similar substitutions and brachial BP remains the standard measure in clinical practice.

The results of the present study have been used to produce a simple calculator that can convert measurements obtained in an individual patient into z scores, in order to help clinicians to assess the significance of tests of carotid arterial function and to provide a tool that could be used to track and compare results over time.

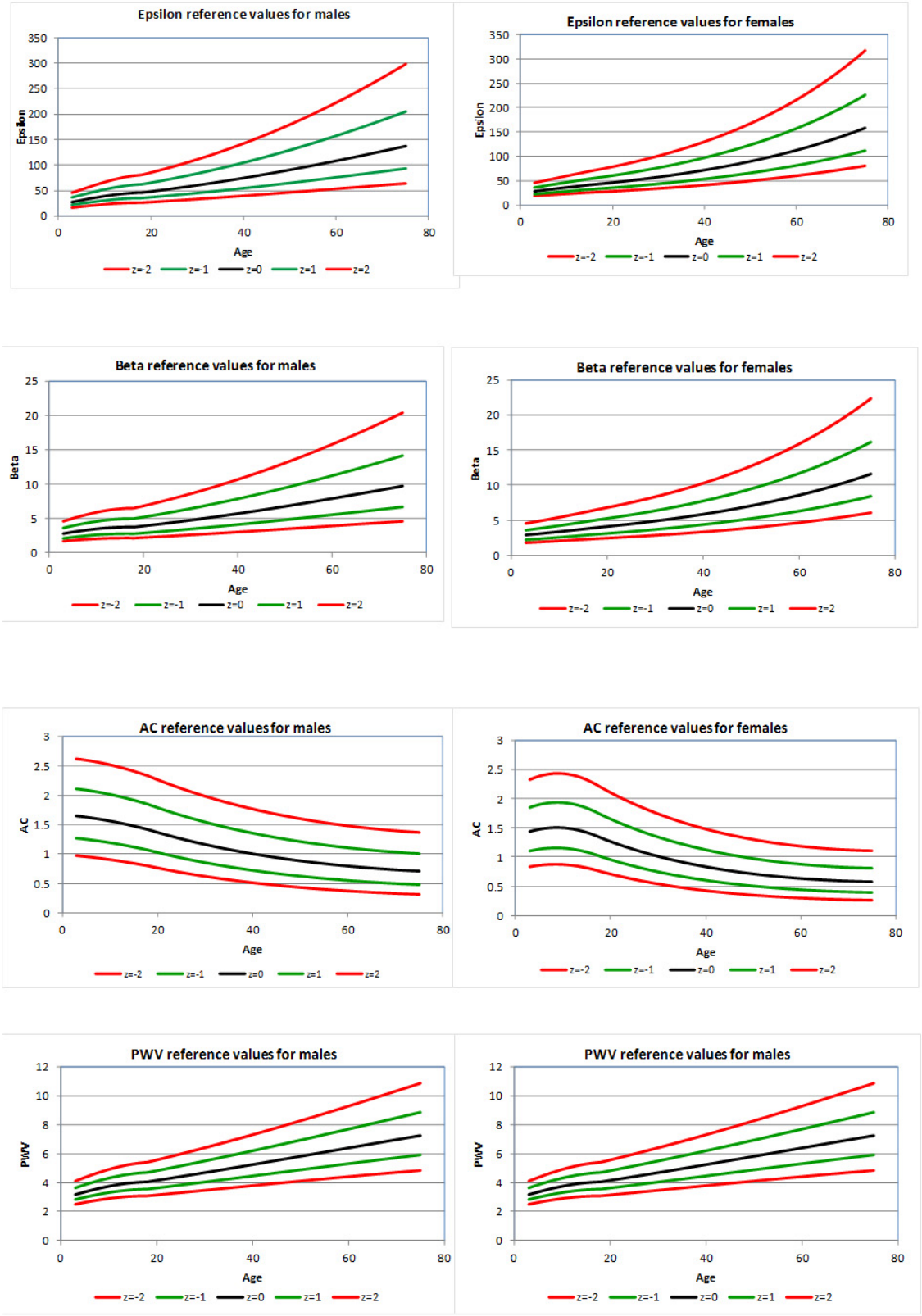


Figure 2.4: Reference chart showing normal values for EP,β, AC, PWV

2.5 Molecular biomarkers: Gene expression studies

Gene expression analysis is a contributory technique that may help understanding complex mechanism of diseases and identifying new biological processes. Two main methods were developed to measure the quantity of transcript: Real-time PCR (RT-PCR), and Microarrays. RT-PCR provides the simultaneous measurement of gene expression in many different samples for a limited number of genes. Microarrays allow parallel analysis of thousands of genes in two different labeled populations.

The gene expression profile of the mutated gene *FBN1* in Marfan patients VS. controls, and of the TGFβ1 pathway, before and during treatment were analyzed using RT-PCR.

2.5.1 Real time PCR for gene expression quantization

The advantages of using RT-PCR for gene expression quantization were very short time preparation and data analysis, and the possibility to study one or a little group of genes in a great number of samples.

The amount of DNA theoretically doubles with every cycle of PCR. After each cycle , the amount of DNA is twice what it was before, so after two cycles we have 2 X 2 times as much, after 3 cycles - 2 X 2 X 2 times as much or 8 (2^3) times as much, after 4 cycles 2 X 2 X 2 X 2 times as much or 16 times (2^4) as much, etc. Thus, after N cycles we shall have 2^N times as much. Since the reaction cannot go on forever, it tails off and reaches a plateau phase. Amplification in the earlier cycles cannot be detected because the changes do not show up on the PCR linear scale. Eventually, the last few cycles of the linear phase can rise above the baseline to a non-linear or plateau phase. If these values are plotted on a logarithmic scale, small differences can be seen at earlier cycles. In real time PCR both types of graphs are used to examine the data. When using a logarithmic scale, there is a straight line relationship between the amount of DNA and cycle number. This is because PCR amplification is a exponential reaction. Real-time PCR analyzes the relative abundance of PCR products during the exponential phase, in which reagents are not limited (Figure 2.5).

Since real-time PCR system is based on the detection and quantitative evaluation of a fluorescent reporter, and since this signal directly increases in proportion to the amount of PCR product in a reaction, by recording the amount of fluorescence emission at each cycle, it is possible to monitor the PCR reaction during the exponential phases where the first significant increase in the amount of PCR product correlates with the initial amount of target template. The higher the starting copy number of the nucleic acid target, the

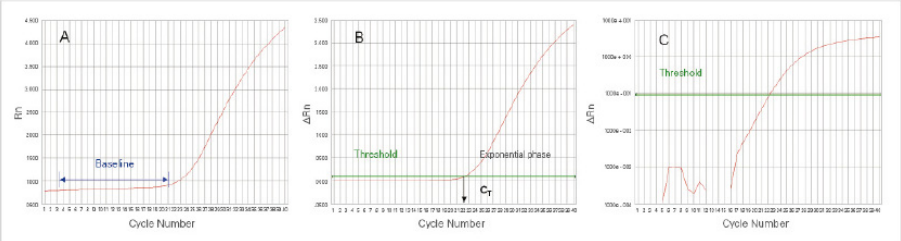


Figure 2.5: PCR reaction phases

sooner a significant increase in fluorescence is observed. A fixed fluorescence threshold is set significantly above the baseline that can be eventually altered by the operator. The parameter CT (threshold cycle) is defined as the cycle number at which the fluorescence emission exceeds the fixed threshold.

House Keeping Gene (HKG)

HKG (loading control or internal standard) is essential to normalize results from quantitative real-time PCR experiments to a fixed reference not affected by experimental conditions. Therefore, we first developed assays aimed at identifying the ideal/best reference gene (HKG) for the study. HKG to be tested were selected on the basis of the following criteria:

- same copy number in all cells
- expressed in all cells
- medium copy number advantageous since the correction should be more accurate

We tested the following 7 genes:

- Glyceraldehyde-3-phosphate dehydrogenase mRNA (*GAPDH*)
- beta actin mRNA (*ACTB*)
- 28S or 18S rRNAs (ribosomal RNAs) (*28SrRNA* and *18SrRNA*)
- Hypoxanthine phosphoribosyltransferse (*HPRT*)
- Glucorodinase (*GUS*)
- TATA-box binding protein (*TBP*)
- GATA-binding factor 1 (*GATA1*)

Sample preparation and performance of the assay.

Total RNA was extracted from peripheral blood and from myocardial tissue specimens with the Maxwell 16 technology (Promega). After checking quality

and quantity of the extracted RNA by means of Nanodrop 1000 (Thermo Scientific), a total 500 ng were reverse transcribed (High Capacity cDNA Archive Kit, Applied Biosystems). First-strand cDNA was synthesized from RNA according to the high capacity cDNA Reverse Transcription Kit protocol (Applied Biosystems, Foster City, CA, USA). Analysis was performed using Applied Biosystems 7900HT Real-Time PCR System following manufacturers protocol for samples preparation and data analysis.

Each sample run in 3 technical replicates: mean Ct value was used for further calculation. Ct is the point at which the fluorescence crosses the threshold and is automatically calculated by the Q-PCR software based on the standard threshold. Since the quantitative evaluation of the amount of cDNA in the original sample must be done where the amplification is exponential and this occurs at the very beginning of the upturn of the curve, we measured the cycle number at which the increase in fluorescence (and therefore cDNA) is exponential (threshold).
Relative quantification of the abundance of the gene at each time point was performed using the comparative $\Delta\Delta$ CT method for population studies. The method is based on the comparative expression of the gene of interest in mutated samples vs. controls. HKG Ct values were subtracted to the Ct values of all samples in both samples and controls (equation 1), in order to delete noise differences generated by the reactions.

Equation 1 :

$$\Delta Ct_i = Ct_{i,gene} Ct_{i,HKG}$$

Then, $2^{-\Delta Ct}$ was calculated in both mutated and controls, and mean $2^{-\Delta Ct} \pm$SD ratio (equation 2) was used for comparison.

Equation 2 :

$$Fold - change = mean 2^{-\Delta Ct} \pm SD_{genemutated} / mean 2^{-\Delta Ct} \pm SD_{Controls}$$

Student's t test with Welch correction for unequal variances (F test p value <0.001) was used to determine statistical significance of the differences of expression between the average values of relative amount of wild-type vs. mutant samples. ANOVA test for paired data was used for follow-up comparisons. Stata 10.1 (StataCorp, College Station, TX) was used for computation. A p-value < 0.05 was considered significant.

2.5.2 Real time PCR of TGFβ Pathway

We analyzed the quantitative gene expression (QGE) in the RNA from peripheral blood of the genes playing in the TGFβ1 pathway; the tested genes

are listed in Table 2.1.

Table 2.1: Genes analyzed from TGFβ1 Pathway

ACVR1	*GATA1*	*PIK3CG*	*ZFYVE9*
AMH	*GDF10*	*RRAS*	
BAD	*GDF6*	*SMAD1*	
BAMBI	*GSK3B*	*SMAD2*	
BMP1	*HRAS*	*SMAD3*	
BMPR2	*INHBA*	*SMAD5*	
FAS	*JUNB*	*SMAD6*	
FBN1	*JUND*	*SMAD9*	
FOSL1	*KRAS*	*SMURF1*	
FOXA1	*MAPK1*	*TERT*	
FOXE1	*MAPK7*	*TGFB1*	
FOXN3	*MMP2*	*TGFB12*	
FOXO1	*MMP3*	*TGFB1R1*	
FOXO3	*MMP9*	*TGFB1R2*	
GADD45A	*MYST3-CREBBP*	*TGFB1R3*	

QGE tests were performed in RNA from peripheral blood of patients in which we identified mutations of *FBN1* gene. The experiments run in 384 optical reaction plates. Both Target Gene and HKG were amplified in separate wells (singleplex reaction) using Pre-Designed Assay Taqman probe (Applied Biosystems, Foster City, CA). Probes used for pathways genes are described in Table 2.2.

GAPDH, *GATA1* and *18S*, three of the most popular housekeeping genes, were tested and *18S* was selected. *18S* Ct values were subtracted to the Ct values of all samples in both pathway genes in mutated samples and controls, in order to delete noise differences generated by the reactions.

Results

The Figure 2.8 shows the graphical view of the $2^{-\Delta\Delta Ct}$ values shown in the above table with fold changes between the baseline and after 1 year of follow-up.

The results achieved to date show differential expression (baseline, pre-treatment vs. treated, any arm, 1 years later) of the majority of the genes in the baseline and in the 1 year follow-up assay. The treatment seems not significantly change the gene expression of *FBN1*, but to influence genes downstream the activation of TGFβ1 (underexpressed). This effect could be caused by the

Table 2.2: Genes analyzed from TGFβ1 Pathway: probes

ACVR1-Hs00153836_m1	*GATA1-Hs01085824_gH*	*PIK3CG-Hs00176916_m1*
AMH-Hs00174915_m1	*GDF10-Hs01086867_m1*	*RRAS-Hs00196699_m1*
BAD-Hs00188930_m1	*GDF6-Hs01377663_m1*	*SMAD1-Hs00195432_m1*
BAMBI-Hs03044164_m1	*GSK3B-Hs00275656_m1*	*SMAD2-Hs00183425_m1*
BMP1-Hs00986789_m1	*HRAS-Hs00610483_m1*	*SMAD3-Hs00706299_s1*
BMPR2-Hs00176148_m1	*INHBA-Hs00170103_m1*	*SMAD5-Hs00195437_m1*
FAS-Hs00910107_m1	*JUNB-Hs01564117_s1*	*SMAD6-Hs00178579_m1*
FBN1-Hs00171191_m1	*JUND-Hs02330233_u1*	*SMAD9-Hs00195441_m1*
FOSL1-Hs00759776_s1	*KRAS-Hs00932330_m1*	*SMURF1-Hs00410929_m1*
FOXA1-Hs00270129_m1	*MAPK1-Hs01052196_m1*	*TGFB1-Hs00998133_m1*
FOXE1-Hs00916085_s1	*MAPK7-Hs00177079_m1*	*TGFB12-Hs01548875_m1*
FOXN3-Hs00231993_m1	*MMP2-Hs01548724_m1*	*TGFB1R1-Hs00610318_m1*
FOXO1-Hs00231106_m1	*MMP3-Hs00968308_m1*	*TGFB1R2-Hs00559660_m1*
FOXO3-Hs00818121_m1	*MMP9-Hs00957562_m1*	*TGFB1R3-Hs01114253_m1*
GADD45A-Hs00169255_m1	*MYST3-CREBBP fusion-Hs03024544_ft*	*ZFYVE9-Hs00245110_m1*
GAPDH-Hs00266705_g1	*18S-Hs99999901_s1*	

interaction between the treatment and the TGFβ1 circulating levels. We will complete the assay at the completion of the follow-up (expected in July 2016).

2.5.3 Real time PCR of *FBN1*

Then we tested the QGE^{FBN1} for further follow-up, thus giving priority to the target gene of the study (*FBN1*).
We measured the quantitative gene expression (QGE) of the *FBN1* in mutated patients (QGE$^{FBN1\text{mut}}$) with MFS vs. wild type controls (QGE$^{FBN1\text{wt}}$) at the baseline and the QGE$^{FBN1\text{mut}}$ after 12 and 24 months of treatment, any arm. The main idea was to determine if QGE^{FBN1} could be a easy,cheap, quick, non invasive pre-genetic screening method for MFS, and /or could help in monitoring the effectiveness of treatment.

Both ^{FBN1} Gene and HKG were amplified in separate wells (singleplex reaction) using Hs00758098_m1 probe for the FBN1 gene, and Hs99999909_m1 probe for *ACTB* (Applied Biosystems, Foster City, CA). *ACTB* Ct values were subtracted to the Ct values of all samples in both *FBN1* mutated samples and controls, in order to delete noise differences generated by the reactions. *ACTB* was compared with *HPTR* and *GAPDH* and was considered the more stable gene, independently from disease condition or treatment. Patients carriers of mutations of the *FBN1* gene demonstrated 33% underexpression of the gene vs. normal controls with wild type gene (QGE$^{FBN1\text{wt}}$).

Although this result confirmed our preliminary hypothesis of loss of *FBN1*,

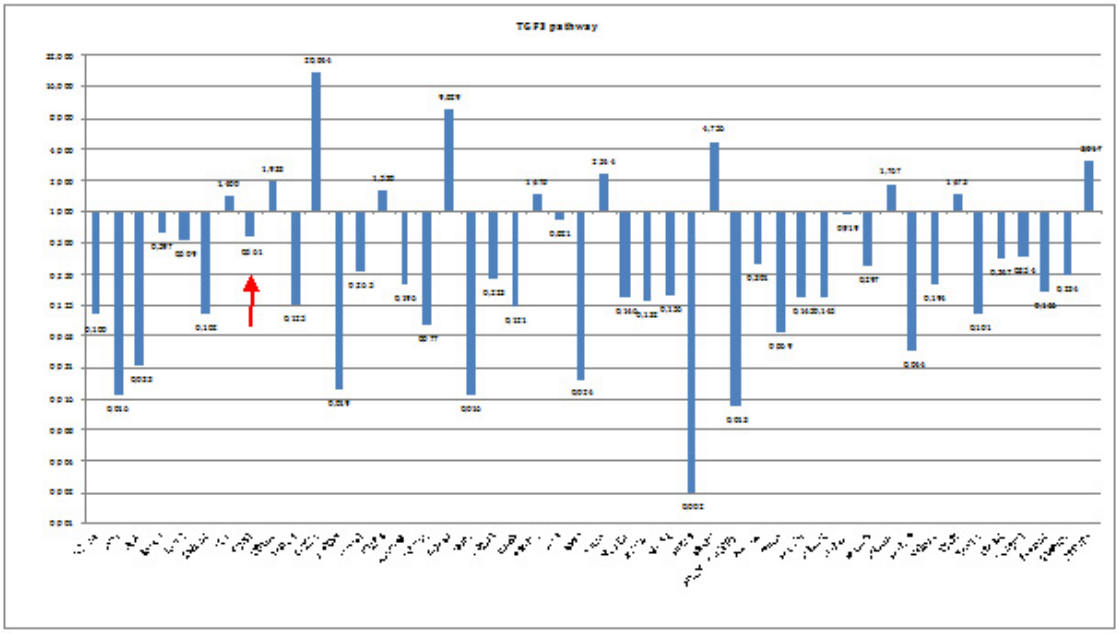


Figure 2.6: Gene expression in MFS mutated patients after 12 months treatment.

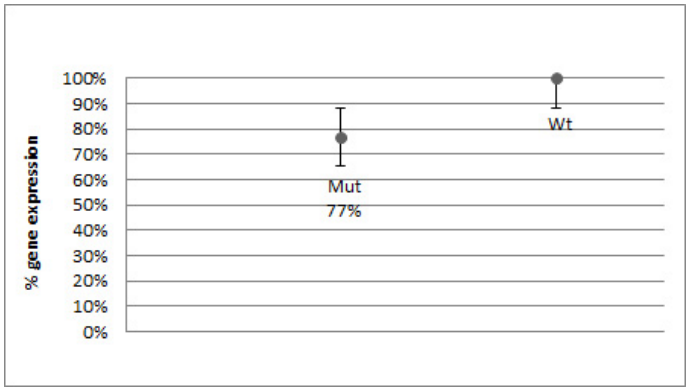


Figure 2.7: Gene expression in MFS mutated patients vs. controls

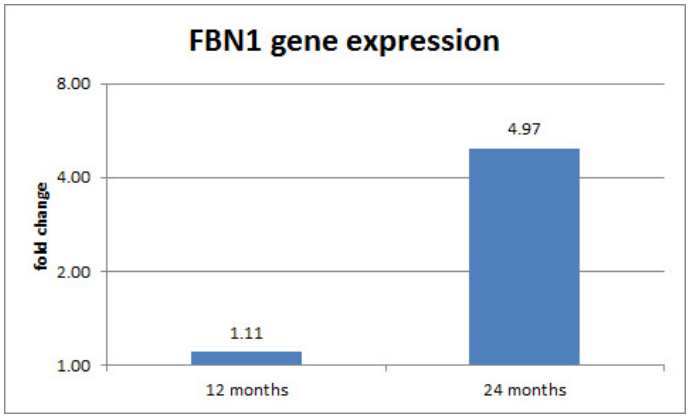


Figure 2.8: Gene expression in MFS mutated patients after 12 and 24 months treatment.

which can cause an excessive activation of TGFβ1 circulating leading to the pathologic state, the underexpression was too small to justify the generation of a pre-diagnostic reliable test.

QGE$^{FBN1\,\mathrm{mut}}$ modified during treatment, any arm, demonstrating a progressively increase of the expression in the same patients not significant from baseline to 1st year, but highly prominent in second year with respect to the 1$^{\mathrm{st}}$ year of treatment.

Given that these data are in the overall population of enrolled and treated patients, we need the completion of the follow up to evaluate the potential different effect of the treatments in the three groups of patients.
The results obtained in the overall groups, constitute the content of a paper that is going to be submitted to PLOS ONE (by Favalli et al.)

2.6 Biochemical biomarkers: TGFβ circulating dosage

Circulating active TGFβ1 in MFS is hypothesized to make extracellular matrix structure weaker and to increase vessels wall stiffness, thus leading to the complex pathological phenotype of MFS. The scientific data available to date indicate that TGFβ1 is the major player in MFS: on this concept we decided to study the circulating level of TGFβ in mutated patients at baseline, compare baseline values of not mutated controls, and monitor the levels during the duration of the clinical trial.
The hypothesis to be tested was whether TGFβ1 level could be a non invasive marker for monitoring the effectiveness of treatment.

We developed an ELISA assay to measure circulating TGFβ1 levels at baseline, during and after treatment in patients diagnosed with MFS and enrolled in the study Telethon GGP08238. We collected and analyzed serum obtained from 267 patients, ages 1 to 55, diagnosed with MFS and carrying an *FBN1* mutation, and those retrieved from 98 age- and gender-matched healthy, voluntary controls. In addition, we measured TGFβ1 levels at 1 st, 2 nd, 3 rd and 4 th control, according to the follow-up schedule.

TGFβ1 levels were measured by the DuoSet ELISA kit (RD System) TGFβ1 immunoassay (from DY240). In this type of assay, the minimal levels of detectable TGFβ1 range from 0.047 ng/ml to 24 ng/ml. The assay displays cross-reactivity with the latent TGFβ1 complex by less than 1%. The TGFβ1 total levels were measured after acid activation (with 2.5N acetic acid/10M Urea) and subsequent neutralization (with 2.7N NaOH/1M HEPES) to yield pH 7.2. The ELISA immunoassay was performed according to the manufacturer‘s protocol. All samples were run in duplicate.
One-tailed Student t-test for unpaired data (alpha$<$0.05) with unequal variance (tested with F-test, p$<$0.001) was used to compare MFS and normal population. ANOVA test for paired data (alpha$<$0.05) was used for follow-up comparisons. Statistical analysis was performed with STATA 10.

In MFS patients, TGFβ1 levels were significantly lower than those measured in the control population, (p$<$0.001, Figure 2.9 A). We grouped controls and MFS patients by age (MFS series: 1-16 yrs, n = 136; 17-55 yrs, n = 131; CTRL series: 80 adults and 18 children respectively). In the control group, TGFβ1 levels did not significantly increase with age (p=0.14). In MFS patients, TGFβ1 levels decreased significantly by age (p=0.004, Figure 2.9 B). The difference was not significant between children mutated vs. control, while it was significant in adults mutated vs controls (p=0.61 and p$<$0.001 respectively). We did not observe gender-related differences (p=0.78).

However we found a great variability in normal population (mean TGFβ1 level 16 $\pm$ 10.2). These normal population was randomly chosen from AVIS

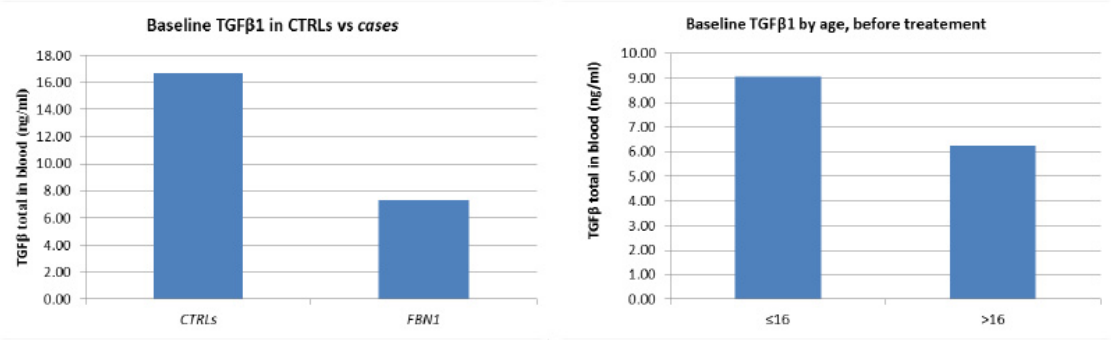


Figure 2.9: TGFβ1 level in MFS vs. controls

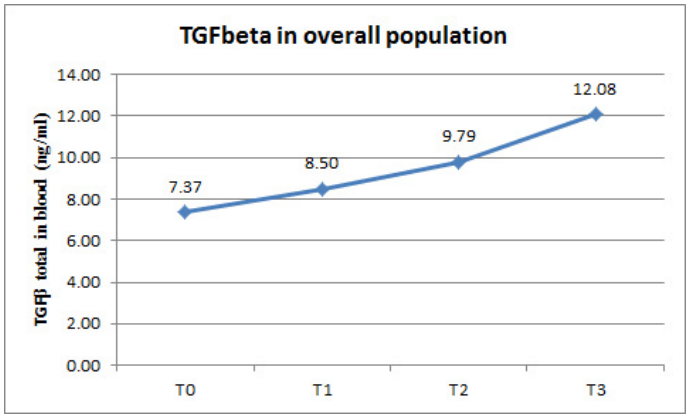


Figure 2.10: TGFβ1 levels after 12-24-36 months treatment

donors population: the only information we have about them is about their normal phenotype, ECG and biochemical evaluations, but nothing about possible treatment (eg: for flu) they assumes before the blood sampling. For these reason we will need to study a controlled normal population in order to develop reference value for further evaluation about TGFβ1 as biomarker.

Considering the overall population of MFS patients, the difference between values measured at 12-24-36 months of follow-up was significantly different ($p<0.001$), underling an increase of the average TGFβ1 total value, blind to the arm of treatment (Figure 2.10).

2.7 Conclusions

All data presented in above sections are relative to baseline measurements or to general population, without any comparative evaluation in the three arms of the trial. The clinical trial will close in July 2016 and after follow-up completion of all patients.

Bibliography

[1] Dean JC. Marfan syndrome: clinical diagnosis and management. Eur J Hum Genet 2007;15:724-33

[2] De Paepe et al. Revised diagnostic criteria for the Marfan syndrome. Am J Med Genet. 1996;62:417-26.

[3] Groenink et al. Aneurysm and dissection. Survival and complication free survival in Marfan's syndrome: implications of current guidelines. Heart. 1999;82:499-504

[4] Habashi et al. Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome. Science. 2006;312:117-21

[5] Loeys et al. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in *TGFβR1* or *TGFβR2*. Nat Genet 2005;37:275-8

[6] Loeys et al. Aneurysm syndromes caused by mutations in the TGF-β receptor. N Engl J Med.2006;355:788-98.

[7] Pannu et al. Mutations in transforming growth factor-β receptor type II cause familial thoracic aortic aneurysms and dissections. Circulation. 2005;112:513-20.

[8] Mizuguchi et al. Heterozygous *TGFbetaR2* mutations in Marfan syndrome. Nat Genet. 2004;36:855-60.

[9] Disabella et al. Two novel and one known mutation of the *TGFβR2* gene in Marfan syndrome not associated with *FBN1* gene defects. Eur J Hum Genet. 2006;14:34-8

[10] Stheneur et al. Identification of 23 *TGFbetaR2* and 6 *TGFbetaR1* gene mutations and genotype-phenotype investigations in 457 patients with Marfan syndrome type I and II, Loeys-Dietz syndrome and related disorders. Hum Mutat. 2008;29:E284-95.

[11] Williams et al. Early surgical experience with Loeys-Dietz: a new syndrome of aggressive thoracic aortic aneurysm disease. Ann Thorac Surg. 2007;83:S757-6

[12] Regalado et al. Exome sequencing identifies *SMAD3* mutations as a cause of familial thoracic aortic aneurysm and dissection with intracranial and other arterial aneurysms. Circ. Res. 109: 680-686, 2011

[13] Lindsay, M. E.et al Loss-of-function mutations in *TGFβ2* cause a syndromic presentation of thoracic aortic aneurysm. Nature Genet. 44: 922-927, 2012

[14] Pannu et al. Mutations in smooth muscle alpha-actin (*ACTA2*) lead to thoracic aortic aneurysms and dissections. Nature Genet. 39: 1488-1493, 2007.

[15] Zhu et al. Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. Nature Genet. 38: 343-349, 2006.

[16] McKellar et al. Novel *NOTCH1* mutations in patients with bicuspid aortic valve disease and thoracic aortic aneurysms. J Thorac Cardiovasc Surg. 2007 134: 290-296

[17] Germain DP. Ehlers-Danlos syndrome type IV. Orphanet J Rare Dis. 2007;2:32

[18] Faivre et al. Clinical and molecular study of 320 children with Marfan syndrome and related type I fibrillinopathies in a series of 1009 probands with pathogenic *FBN1* mutations. Pediatrics. 2009 ;123:391-8.

[19] Faivre et al. Clinical and mutation-type analysis from an international series of 198 probands with a pathogenic *FBN1* exons 24-32 mutation. Eur J Hum Genet. 2008 12

[20] Faivre et al. Contribution of molecular analyses in diagnosing Marfan syndrome and type I fibrillinopathies: an international study of 1009 probands. J Med Genet. 2008a ;45:384-90.

[21] Faivre et al. Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. Am J Hum Genet. 2007;81:454-66

[22] Rationale and design of a trial evaluating the effects of losartan vs. nebivolol vs. the association of both on the progression of aortic root dilation in Marfan syndrome with FBN1 gene mutations. Gambarin FI, **Favalli V**, Serio A, Regazzi M, Pasotti M, Klersy C, Dore R, Mannarino S, Viganó M, Odero A, Amato S, Tavazzi L, Arbustini E. J Cardiovasc Med (Hagerstown). 2009 Apr;10:354-62.

[23] Aortic stiffness and diameter predict progressive aortic dilatation in patients with Marfan syndrome. Nollen GJ, Groenink M, Tijssen JG, Van Der Wall EE, Mulder BJ. Eur Heart J. 2004 Jul;25(13):1146-52.

[24] Reference Values for Arterial Stiffness Collaboration. Determinants of pulse wave velocity in healthy people and in the presence of cardiovascular risk factors: 'establishing normal and reference values. Eur Heart J 2010;31:2338-50.

Chapter 3

The Cardiomyopthies

In this chapter I will deal with inherited diseases primarily involving heart muscle. In particular I will focus on DCM, which was the main subject of my analysis, with the support of EU project INHERITANCE (n° 241924). I will detail the project aims and the connections with the overall aim of my doctoral project, and I will present results.

CONCEPTS

- Inherited cardiomyopathies (CMPs) affect about 180,000 Italian citizens and about 2,250,000 European citizen.
- Inherited CMP show clinical and genetic heterogeneity. Genes causing HCM may also cause DCM or RCM, while DCM genes may also cause ARVC and vice versa. Phenotypes may also overlap in different members of the same family.
- >10% of CMPs show double or triple mutations in different genes, with different combinations in family members.
- The clinically guided genetic testing may not be sufficient to address genetic screening with about 50% DCM, 30% HCM/RCM and 50% ARVC that remain genetically orphan after completion of the screening of known candidate genes.
- A novel diagnostic viewpoint of the familial phenotype and genetic makeup, including carriers of single or double mutations, may contribute to clarify geno-phenotype correlations and stratify gene-specific prognosis.

3.1 Cardiomyopathies

Cardiomyopathies are defined as primary myocardial disorders of unknown cause and are classified into four main subtypes, based on ventricular morphology and physiology: hypertrophic (HCM); dilated (DCM); restrictive (RCM); and arrythmogenic right ventricular cardiomyopathy (ARVC) [1]. Overlapping phenotypes (for example, dilatation of a hypertrophic ventricle) can occur in advanced phases of the disease.

DCM is defined as a myocardial disorder characterized by the presence of left ventricular dilatation and systolic impairment, in the absence of abnormal loading conditions (e.g. hypertension, valve disease) or coronary artery disease sufficient to cause global systolic dysfunction. Right ventricular dilatation and dysfunction may also be present.
The prevalence of DCM is approximately 1 in 2,500 adults, with an annual incidence of between 5 and 8 per 100,000 [2]. In children, the incidence is much lower (0.5 to 0.8 per 100,000 per year), but DCM is the commonest cardiomyopathy in the pediatric population [3, 4]. DCM is responsible for 10,000 deaths per year in Europe and is the commonest indication for cardiac transplantation in adolescents and adults. Up to 35% of individuals with DCM have at least one other first-degree relative that is affected [5, 6, 7, 8, 9]; a further 20% of family members have isolated left ventricular enlargement with preserved systolic function, many of whom subsequently develop overt DCM [8, 9]. In spite of this knowledge, the majority of patients are neither made aware of the familial risk nor offered genetic testing and less than 1% of patients in Europe benefit of genetic testing.

Family screening and genetic studies have identified more than 40 disease-causing genes to date [10]. The commonest genes that cause DCM are those encoding proteins of the sarcolemma [11, 12, 13], cytoskeleton [14], sarcomere [15, 16, 17, 18, 19], nuclear envelope [20, 21, 22] and energy generation [23]. Recent evidence suggests that mutations in genes encoding desmososmal genes, more typically associated with the development of ARVC [10], also cause disease in some patients. Most DCMs (up to 90%) are inherited as autosomal dominant traits [10], while a minority are X-linked recessive [1, 24], autosomal recessive [25] and matrilineal [23, 26].

Phenotype characterization

Clinical assessment of the different genetic disorders that cause DCM is relatively crude, with little emphasis given to the search for potential diagnostic cardiac and systemic markers. As a result, specific genetic diagnoses are often missed in everyday clinical practice and the implications for other family members overlooked. The symptoms and signs of DCM are highly variable and depend partly on the degree of left ventricular dysfunction and on the

underlying aetiology. The majority of patients present with symptoms of high pulmonary venous pressure and a low cardiac output. Presentation can be acute, often precipitated by intercurrent illness or arrhythmia [27], or chronic, preceding the diagnosis by many months or years. Increasingly, DCM is diagnosed as an incidental finding in asymptomatic individuals during routine examination or family screening [6, 7].
Associated cardiac traits (such as left ventricular non-compaction, conduction tissue disease, WPW, short PR) and non-cardiac traits such as hearing loss, retinitis pigmentosa, myopathy, palpebral ptosis, granulcytopenia, methylglutaconic aciduria, premature cataract, learning difficulties, can be present and are important in guiding physicians towards a possible aetiology [1, 28].

The prognosis of DCM is highly variable and may also depend on aetiology. Early survival studies suggested mortality in symptomatic adults with idiopathic DCM approaching 25% at one year and 50% at 5 years [29]. More recent reports have shown better outcomes, with five-year survival rates of approximately 20%, perhaps reflecting earlier disease recognition and treatment, and advances in medical therapy. However, recent studies have shown that the arrhythmogenic risk is extremely high in certain subtypes of DCM, such as cardio-laminopathies.

Role of biochemical markers in diagnosis and management

The role of biochemical markers in the diagnosis and monitoring of specific sub-types of familial DCM, is largely unexplored. Blood biomarkers such as B-type natriuretic peptide levels are elevated in children and adults with chronic heart failure and predict survival, hospitalization rates and listing for cardiac transplantation [29], but are not specific to any subtype of DCM. Elevated levels of serum creatine kinase (sCPK) can be raised in dystrophin-related DCM in X-linked forms, or a *LMNA* gene defect in autosomal dominant DCM, but more disease-specific biomarkers are unknown.
INHERITANCE searched for novel peripheral markers using proteomic and metabolomic analysis in paired myocardial samples and biological fluids (blood and urine) from genotyped patients and healthy carriers of causative mutations. Any metabolite potentially measurable in DCM was tested in larger series of un-genotyped patients to test its predictive value and usefulness in monitoring disease progression as part of a rigorous train and test routine within the multivariate analysis used to identify new potential metabolic biomarkers of disease.

Influence of a genetic diagnosis on treatment

Currently patients with DCM are treated in accordance with international guidelines for the management of heart failure with little consideration of the possible influence of the underlying aetiology on the response to treatment.

Recent studies suggest that this might result in sub-optimal or inappropriate therapy in some patients [30]. For example, patients with mutations in the nuclear envelope protein Lamin AC (*LMNA*) are at high risk of conduction disease relatively early in the course of their disease, but several studies have shown that implantation of a pacemaker alone does not prevent sudden death. Therefore, knowledge that a DCM patient is carrier of a *LMNA* gene mutation might be of major importance when deciding on device therapy. The influence of genetic factors in determining the response (and timing) of drug therapy is largely unstudied in DCM. We also speculate, that greater understanding of the structural consequences of specific mutations on cardiac proteins can generate new disease specific therapeutic targets.

DCM is the best example of a clinically and genetically heterogeneous disease [10], with more than 40 known disease genes. The strategy of diagnosing DCM is still based on clinical screening and identifying the disease-causing mutation in each family, which can be clinically guided or based on screening of unselected consecutive series of probands. The biotechnology tools supporting genetic screening are now offering more than Sanger-based technologies, including the possibility of testing hundreds of genes at a time in large clinical series. Clinical genetics and related rules as well as clinical cardiology should govern molecular genetic information to maximize the benefits of clinical translation for patients and families.
Currently patients with DCM are treated in accordance with international guidelines for the management of heart failure with little consideration of the possible influence of the underlying aetiology on the response to treatment. Recent studies suggest that this might result in sub-optimal or inappropriate therapy in some patients [31]. For example, patients with mutations in the nuclear envelope protein Lamin AC (*LMNA*) are at high risk of conduction disease relatively early in the course of their disease, but several studies have shown that implantation of a pacemaker alone does not prevent sudden death. Therefore, knowledge that a DCM patient is carrier of a *LMNA* gene mutation might be of major importance when deciding on device therapy.

The INHERITANCE project (Integrated Heart Research In Translational Genetics of Cardiomyopathies in Europe), a multidisciplinary, multi-center research project funded by the European Commission, seeks to study the genetic of inherited DCM and to understand the impact and management of the condition within families that suffer from DCMs. About 40,000 DCM cases are expected in the INHERITANCE European countries. DCM is responsible for 10,000 deaths per year in Europe and is the commonest indication for cardiac transplantation in adolescents and adults. The INHERITANCE translational strategy was based on a clinical algorithm that seeks to determine disease-specific features (red-flags) that associate with different types of DCM or suggest specific genetic or metabolic pathways of disease. A reverse translational strategy was run in parallel to establish or confirm the association of DCM

phenotypes with clinical markers that occur in patients with DCM caused by the different genes.

The aim of INHERITANCE was to develop a more sophisticated diagnostic algorithm using existing clinical methods and tools supplemented by genetic information, novel imaging and biochemical markers validated during the project in order to define specific phenotypes that can be used to guide genetic testing.

The INHERITANCE project is structured into 6 research areas that study different facets of the DCM condition, including clinical cardiogenetics, -omics, i.e. genetic testing, transcriptomics, proteomics and metabolomics, animal studies, structural studies, treatments, and biomedical informatics, which aims to implement information technology solutions to support the project team in managing the huge quantity of scientific, clinical and patient data generated by the project.

I focused on the research of a non-invasive, cheap, quick and prediagnosic biomarker which can pre-test the presence of absence of mutation in affected subjects. The test was based on quantization of peripheral blood mRNA of the mutated gene using RT-PCR.

We performed a pilot study in 47 consecutive *LMNA*-mutated and wild type controls. In peripheral blood mRNA we observed a down regulation in patients with mutation with respect to controls. The presence of a mutation involving splice-sites leads to about 40% reduction of the mRNA levels. This can be explained by invoking the presence of NMD mechanisms causing potential haploinsufficiency (partially explaining the variable presentation of the disease within the same family) whilst insufficient NMD can account to the presence of the disease, allowing the truncated proteins exerting their dominant negative effect on the nuclear lamina.

The first aim of my research was to validate results of the pilot study on the population enrolled in INHERITANCE project and evaluate the possibility to create a pre-diagnostic-mutation test for *LMNA* gene. In case of positive results in the first phase, the second aim was to extend this type of tests to subjects with suspected Dystrophinopathies and Zaspopathies.

3.2 *LMNA* and the new diagnostic test

LMNA is probably the gene is most commonly tested in DCM patients. Its small size makes clinical screening easy to perform, and the malignant cardiac phenotype justifies a genetic test. During INHERITANCE we enrolled a series of 1009 unselected consecutive probands in which the *LMNA* gene was systematically screened (unpublished data from the INHERITANCE Project EU291924; August 2012). We identified 81 probands as carriers of *LMNA* mutations (8%). Family screening identified additional 253 mutated relatives, with 130 affected members and 123 healthy carriers who undergo regular clinical surveillance. In these families, the penetrance by age of 60 is nearly 100%, conduction disease characteristically occurs in up to 80% of cases, and the arrhythmogenic risk is high, even with mild LV dilatation and dysfunction.

3.2.1 Real Time PCR expression of *LMNA*

Phase 1: Validation of the pilot study and comparison with tissue gene expression

Cases and controls: blood samples The case series includes three major sample series collected between January and February 2011. The three subgroups are constituted of:

1. Group A: 67 patients with dilated cardiolaminopathy in which a pathologic *LMNA* gene mutation has been identified as the cause of the disease in the proband and in the family.

   - median age 39.5 years, IQR 30-52
   - M vs. F = 34 (median age 36.5 years, IQR 25-45), 33 (median age 44 years, IQR 32-52), p-value 0.207

2. Group B: 95 patients with DCM in which the sequencing analysis identified *LMNA* SNPs of uncertain significance without mutations

   - median age 38 years, IQR 27-49
   - M vs. F = 57 (median age 40 years, IQR 25-49) , 39 (median age 43 years, IQR 31-50), p-value 0.617

   These 95 patients were selected from the overall series of 936 DCM who underwent *LMNA* test.

3. Group C: 115 healthy individuals, with normal ECG and echocardiographic data, in which *LMNA* genetic testing excluded both mutations and the polymorphisms identified in the group B.

   - median age 41 years , IQR 28.5-49.5.

Table 3.1: Case and controls population in blood samples study

Group	**N.**	**Median age (yrs) IQR**	**P-value**
A	68	39.5 (30-52)	Ns
B	115	38 (27-49)	Ns
C	95	41 (28.5-49.5)	Ns

Table 3.2: Case and controls population in tissue samples study

Group	**N.**	**Median age (yrs) IQR**	**P-value**
A	20	40 (28.5-52)	Ns
B	20	39 (26-53.5)	Ns

- M vs. F = 74 (median age 37.5 years, IQR 26-49) , 41 (median age 40 years, IQR 31-49), p-value 0.487

Cases and Controls: tissue samples

The case series includes three major subgroups:

1. Group A: 20 patients who underwent biopsy or heart transplantation:
   - median age 37 years, IQR 24.5-47.5
   - M vs. F = 9 (median age 27 years, IQR 24-41) /11 (median age 45 years, IQR 33-51.5), p-value 0.09
2. Group B: 20 patients diagnosed with DCM and with wild type *LMNA* at genetic testing, who underwent endomyocardial biopsy or heart transplantation:
   - median age 39 years , IQR 26-53.5.
   - M vs. F = 8 (median age 31 years, IQR 25-48) / 12 (median age 42 years, IQR 32.5-47.5), p-value 0.371.

House-Keeping Gene selection

The expression stability of each HKG was calculated through normalization by ranking using geNorm VBA applet for Microsoft Excel. The average

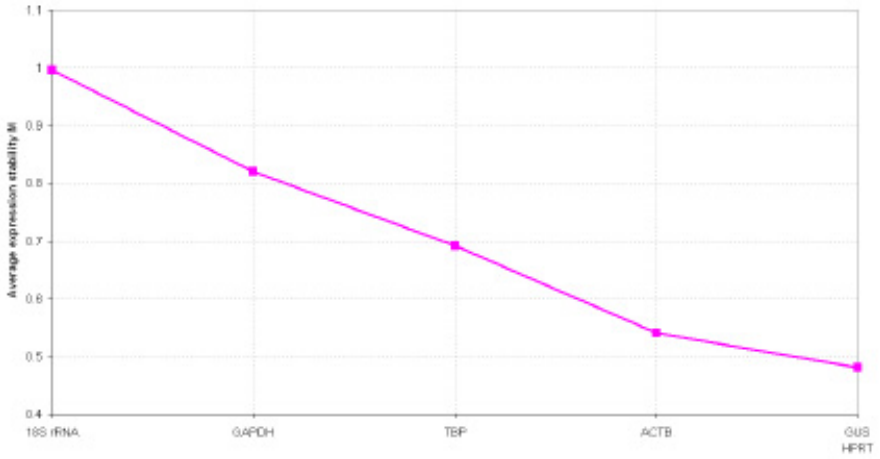


Figure 3.1: Average expression stability in blood samples of tested HKGs

expression stability value M of reference genes was measured at each step during stepwise exclusion of the less stable one . The gene showing the lowest expression stability was *HPRT* (Figure 3.1).

Sample preparation and performance of the assay.

Total RNA was extracted from peripheral blood and from myocardial tissue specimens with the Maxwell 16 technology (Promega). After checking quality and quantity of the extracted RNA by means of Nanodrop 1000 (Thermo Scientific), a total 500 ng were reverse transcribed (High Capacity cDNA Archive Kit, Applied Biosystems). First-strand cDNA was synthesized from RNA according to the high capacity cDNA Reverse Transcription Kit protocol (Applied Biosystems, Foster City, CA, USA). Analysis was performed using Applied Biosystems 7900HT Real-Time PCR System following manufacturers protocol for samples preparation and data analysis. The experiment run in 384 optical reaction plate. Both Target Gene and HKG were amplified in separate wells (singleplex reaction) using Pre-Designed Assay Taqman probe (Applied Biosystems, Foster City, CA) respectively Hs00153462_m1* for *LMNA* gene and Hs99999909_m1 for *HPRT*. A standard/default thermal profile with an initial denaturation at 95°C for 10 min was used , and then cycling for 40 cycles with denaturation at 95°C for 15s and an extension at 60°C for 1 min. Data collection was done at the extension stages during the thermal cycling. Each reaction was conducted in a total volume of 12ul.

Each sample run in 3 technical replicates: mean Ct value was used for further calculation.
Student's t test with Welch correction for unequal variances (F test p value <0.001) was used to determine statistical significance of the differences of expression between the average values of relative amount of wild-type vs. mutant

samples. Stata 10.1 (StataCorp, College Station, TX) was used for computation. A p-value < 0.05 was considered significant.

Myocardial samples

A difference in gene expression of 71% was found in mutated group vs. wt ($p < 0.001$) showing a marked under-expression of *LMNA* in mutated samples with respect to control samples. in cardiac tissue. (Figure 3.2)

$41.24 \pm 11.6_{LMNAmutated}/143.34 \pm 53.96_{Controls}$

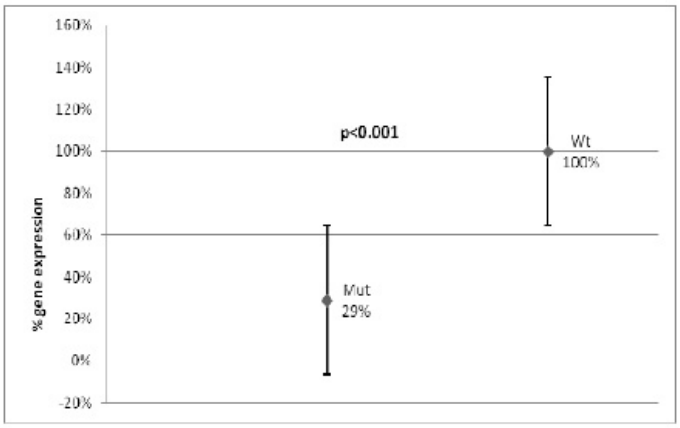


Figure 3.2: Gene expression in *LMNA* mutated vs. controls; tissue samples

Blood samples

A difference in gene expression of 30% was found in mutated group vs wt ($p=0.0075$), being *LMNA* underexpressed in mutated vs. control samples. (Figure 3.3)

$$1.37 \pm 0.92_{LMNAmutated}/1.92 \pm 2.02_{Controls}$$

As expected in blood samples, the difference of expression was less marked in blood samples.

Development of a predictive *LMNA* expression assay: first evaluation

I generated an independent set of 100 randomized RNA samples (testing set) by including 80 samples from patients with wild type *LMNA* and 20 from *LMNA* mutation carriers. The samples were made blind and then subjected to the assay for quantitative expression. The aim was to preliminarily test specificity, sensitivity and predictive value of *LMNA* expression with the normal median Δ Ct value.
I pre-established the cut-off limit of the ROC curve at values $> 50\%$ under which we consider further efforts at risk of no significant results. The ROC

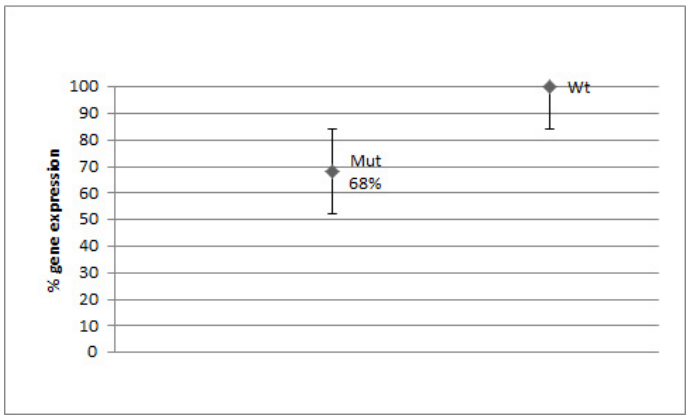


Figure 3.3: Gene expression in *LMNA* mutated vs. controls; blood samples

analysis (Figure 3.4) showed a significant AUC value of 0.61 (p<0.001). $2^{-\Delta Ct}$ values obtained for each sample in the testing set were divided for 1.92 value (mean $2^{-\Delta Ct}$ obtained in controls population). Considering a cutoff of 0.71 (ratio value previous obtained comparing our mutated population vs controls) the test presented a specificity of 50% and a sensitivity of 62%. Given the value over 50%, the progression of the experiments by adding cases and grouping by age is feasible and based on promising results.

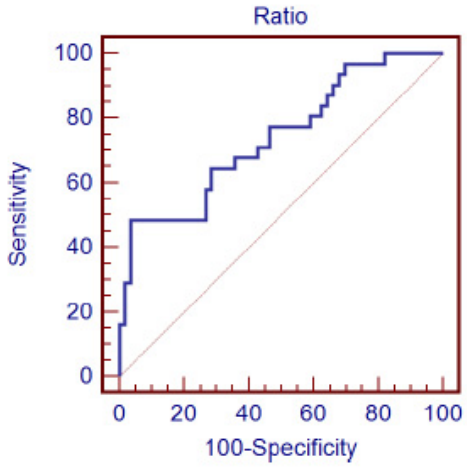


Figure 3.4: ROC curve, AUC 0.61

Results document that *LMNA* gene is significantly underexpressed in the RNA from peripheral blood of patients carriers of *LMNA* mutations. Results obtained in the peripheral blood samples confirm the hypothesis that the mutated *LMNA* is transcribed less efficiently than the wild type gene and provide the basis for inferring about pathogenetic mechanisms of the diseases associated with *LMNA* mutations.

Immunohistochemical tests may show the expression of the Lamin AC protein at the nuclear membrane level in endomyocardial biopsy samples and in

samples of myocardium obtained at cardiac transplantation. This evaluation has been party done in the past, documenting the loss of expression of Lamin AC in the affected myocardium and further progressed during the first year of activity in this project to be continued throughout the project as far as diagnostic immunohistochemical test for cardiolaminopathies on endomyocardial biopsies entered the enrolled series. The activity done to assess the potential predictive value of the transcriptomic assays provided encouraging results, although not sufficient, with the present number of test and representative mutations, to predict *LMNA* mutations. Based on the present results we could spare 40% of sequencing of the gene and therefore lowering the costs and time of analysis, but not enough for considering the transcriptomic assay sufficient, by itself, to fully predict *LMNA* mutations. On the basis of these results, we decided to go on with the confirmation and evaluation of QGE^{LMNA} as a predictive mutation test.

Phase 2: Development of a pre-diagnostic mutation test for *LMNA* gene

The Quantitative Gene Expression (QGE^{LMNA}) series for the development of a pre-diagnostic *LMNA* mutation test, included 311 samples consecutively collected during the semester November 2010-April 2011, independently on the phenotypes, age and gender, and including probands and family members addressed to our center for scheduled monitoring, or first clinical evaluation. The QGE experiments were conducted in parallel but blind to the results of *LMNA* sequencing (Figure 3.5).

Sample series

The series includes 311 samples consecutively collected during the semester November 2010-April 2011, independently on the phenotypes, age and gender, and including probands and family members addressed to our center for scheduled monitoring, or first clinical evaluation. The phenotypes of the 311 patients are summarized in Table 3.3.

The rationale for designing this second part of the study on unselected, consecutive and clinically heterogeneous series was related to our scientific aim that was the assessment of the predictive value of QGE^{LMNA} assay for *LMNA* mutations. The optimization of the assay obtained in the first period of research activity also allowed to abate costs making sustainable the experiment in such a large series.

Quantitative Real time PCR assay for *LMNA*

Real-time PCR methods used to perform the GEP experiments have been reported in the precedent section. The $2^{-\Delta Ct}$ was calculated as $\Delta Ct_{LMNA} -$

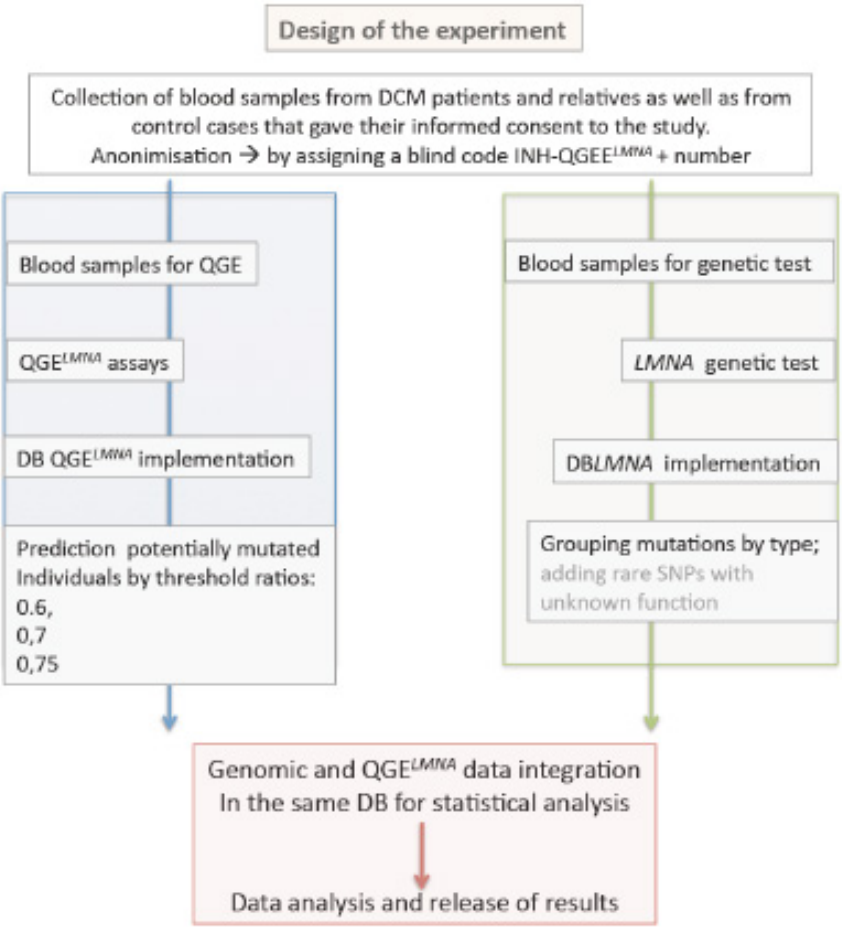


Figure 3.5: QGE experimental design

Table 3.3: Series phenotypical features.
Legend: DCM= Dilated Cardiomyopathy; AVB= AtrioVentricular Block; sCPK= serumCPK; Vn=Ventricle; AF= Atrial Fibrillation; WPW=; LVH; ECG= electrocardiogram; LVNC= Left Ventricular Non Compaction; RV= Right Ventricle; SD= Sudden Death; LV=Left Ventricle; HCM= Hypertrophic cardiomyopathy.

Phenotypes	**N**
DCM	66
DCM+AVB	41
DCM+AVB+<sCPK	10
DCM+ >sCPK	9
DCM+Vn arrhythmias	6
DCM+AF	5
AF in DCM family screening	1
DCM+WPW	1
DCM+LVH ECG	2
DCM+LVNC	7
DCM+RV involvement + vn arrhythmias	3
Vn arrhythmias in familial DCM with Vn arrhythmias	1
Healthy relatives later shown to be *LMNA* mutation carriers	16
Clinically healthy in DCM family screening	96
Early; ECG and/or Echo abnormalities	24
Relatives of pts with SD	3
LVNC without LV dysfunction	6
AVB+LVNC	1
Isolated AVB	4
Syncope	3
HCM $\rightarrow$ DCM	6
	311

ΔCt_{HPRT} for each sample. Then each individual value was divided by the CTRL value previously obtained in normal controls, thus generating the

$$ratio = 2^{-\Delta Ct}{}_{eachsample}/2^{-\Delta Ct}{}_{CTRLvalue}.$$

The CTRL value corresponded to 1.92, which was the mean $2^{-\Delta Ct}$ value in our normal series.

Calculation of the threshold

In the previously described experiment, we established that the ratio between *LMNA* gene expression in overall cases / overall controls was .71. According to the above results, we decided to assess the sensitivity and specificity of QGE^{LMNA} considering three threshold values below which we could consider patients as candidates for possible *LMNA* gene mutations.

- 0.60, which is slightly lower that the lowes ratio found in mutated pantients
- 0.70, which is the closest value to the mean ration of all mutated cases
- 0.75, which is slightly lower to the highest ratio found in mutated patients.

Therefore we calculated the 3 test of data in the entire population of 311 individual blindly to sequencing data:

- Specificity,
- Sensitivity,
- ROC curve.

LMNA sequencing

The *LMNA* gene was analyzed by direct automated sequencing in the 311 patients and controls, according to established and validated methods. Results of sequence analysis were then integrated with QGE^{LMNA} data before progressing with the calculation of sensitivity, specificity and ROC curves.

Specificity and Sensitivity

The specificy and sensitivity of the QGE^{LMNA} were calculated at each of the 3 threshold ratio values

- Threshold value 0.75
- Threshold value 0.7

Table 3.4: Threshold 0.75.

	Mut +	**Mut -**	**tot**
<0.75	72	30	102
>0.75	0	209	209
tot	72	239	311

Table 3.5: Threshold 0.7

	Mut +	**Mut -**	**tot**
<0.70	68	24	92
>0.70	4	215	219
tot	72	239 311	

- Threshold value 0.6

The sensitivity and specificity for each threshold value are given in the Table 3.7.

Then we performed the ROC analysis (Figure 3.6), in which the best threshold value calculated was 0.75 with an AUC of 0.957 ($p<0.001$), a specificity of 87% and a sensitivity of 100%.

The results of the assay are rather encouraging in terms of specificity and sensitivity.

The results reported confirmed prior data released and document a very high sensitivity and specificity of the QGE^{LMNA} assay for the threshold ratio value of 0.75. The assay recognizes amongst all samples with a ratio ≤ 0.75, all mutated patients but also 30 patients/individuals with wild type *LMNA* gene thus preventing immediate diagnostic applications. By lowering the threshold ratio to 0.7, 24 patients with wild type *LMNA* show underexpression of the gene and using the threshold of 0.6 the number of false positive cases is 17.

Overall, the best threshold ratio is 0.75, confirmed also by the ROC analysis. Therefore the cost/benefit ratio against the routine use of the assay as pre-genetic test is that 30 of 100 cases would lead to unclear sequencing in wild type individuals. Viceversa, 70 of 100 individuals would be appropriately sequenced: in terms of costs for consumables and personnel, the QGE^{LMNA} costs are much lower than sequencing the gene in the same sample-sized population. The generation of a fast, low cost kit which is the potential exploitation of the

Table 3.6: Threshold 0.6

	Mut +	**Mut -**	**tot**
<0.60	56	17	73
>0.60	16	222	238
tot	72	239	311

Table 3.7: Specificity and sensitivity of threshold values.

Threshold	0.75	0.70	0.60
Sensitivity	100%	94%	78%
Specificity	87%	90%	93%

ROC curve

Variable	ratio
Classification variable	LMNA_mutated LMNA mutated

Positive group	
LMNA mutated	= 1
Sample size	72
Negative group	
LMNA mutated	= 0
Sample size	239

Disease prevalence (%)	unknown

Area under the ROC curve (AUC)	0.957
Standard error	0.0105
95% Confidence interval	0.928 to 0.977
z statistic	43.695
Significance level P (Area=0.5)	0.0001

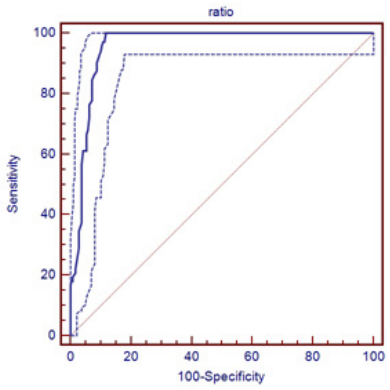


Figure 3.6: ROC analysis

results of this part of the research, definitely requires benchmarking analysis amongst techniques, in particular Sanger-based direct sequencing and NGS. We are in progress with this evaluation.

These results constitute the content of a manuscript published on JACC in 2012 (Narula N. et al).

3.3 New pre-genetic tests in *DYS* and *LDB3* mutations

I investigated patients with familial DCM and relatives who carry pathologic mutations of the *DYS* and *LDB3* genes. We performed QGE^{DYS} and QGE^{LDB3} according to previously optimized methods reported for *LMNA* gene. The assay aimed at assessing the effects on QGE^{DYS} and QGE^{LDB3} of mutations in *Dystrophin* (*DYS*) and in *LDB3* (CYPHER-ZASP) genes. The results showed a significant under expression of the mutated *DYS* gene in the myocardium. Accordingly, QGE^{DYS} correlates with the presence of gene mutations. Vice versa, we fail to demonstrate significant changes in QGE^{LDB3} in patients carriers of *LDB3* mutations.

Critical considerations While results achieved in the group of *DYS* mutation carriers encourage future implementation of the QGE^{DYS} test in the myocardial samples, those obtained in *LDB3* mutation carriers are non significant and do not support further evolution of the QGE^{LDB3}. Due to this disappointing result, we critically revised the list of mutations found in *LDB3* to understand whether this gene is a true candidate gene for DCM as originally reported. We further considered the overall genetic make-up of carriers of mutations of *LDB3* and found that 11 of 26 probands were also carriers of a second mutation in another disease gene and that in families only carrier of a single *LDB3* mutation this latter did not segregate with the phenotype.

DYS mutated cases vs. controls

A difference in gene expression of 50% was found in the RNA from the myocardium in patients with DYS mutations vs. control myocardium with wild type DYS ($p < 0.001$) showing a marked underexpression of DYS in mutated samples with respect to control samples Figure 3.7.

Figure 3.7: QGE^{DYS}

LDB3 mutated cases vs. controls

1. Blood samples: *LDB3* was not expressed in the RNA of the peripheral blood.

2. Myocardial samples

A difference in gene expression of 11% was found in the RNA from the myocardium vs. wild type samples (p = ns).

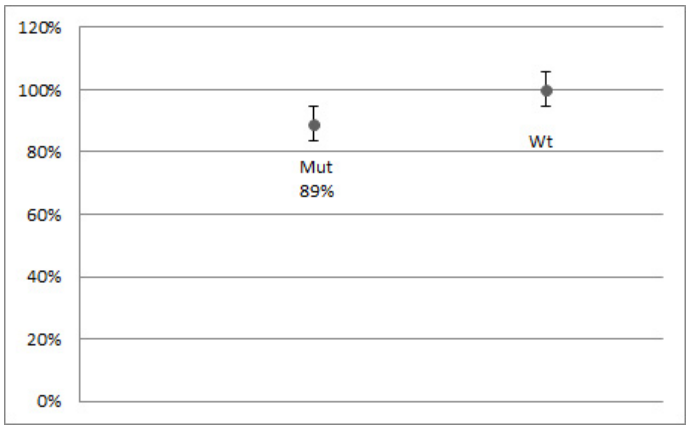


Figure 3.8: QGE^{LDB3}

Overall the scientific and technical contributions of this activity are:

1. Confirmation that mutated *DYS* gene is underexpressed in the myocardial tissue, independently on the mutation type. This results matches the decreased expression of the protein observed with both immunohistochemical and Western blot studies.

2. The mutated *LDB3* gene is not associated with decreased expression of the gene in the affected myocardium. Furthermore, the expression of the *LDB3* gene is not detectable in the peripheral blood RNA, thus making impossible to implement QGE^{LDB3} from peripheral blood. Further, in the myocardial tissue, QGE^{LDB3} does not significantly differ in mutated and wild type subjects. This observation, added to the clinical evidence of little segregation of mutations with the phenotypes, raises serious doubts about the real causative role of *LDB3* mutations in familial DCM.

3.3.1 Conclusions

The major result of INHERITANCE project was the understanding that the possibility to directly guide the genotyping activity from clinical data was

effective in a minority of cases presenting clear clinical manifestations (eg: DCM + AVB + increased CPK level ⇒ Laminopathy). In the majority of cases the descriptive nosology of the four major types of cardiomyopathies, that has been, and still is, extraordinarily useful for current clinical programs, does not yet match the genetic scenario. In fact in an increasing number of families genotyped, different members presented different phenotypical manifestations (eg: father DCM and son HCM), with mutations in the same gene/genes or different combination of mutations. If known disease genes cause more subtypes of cardiomyopathies, testing groups of genes per major phenotype (sarcomere gene in HCM, or cytoskeletal genes in DCM or desmosome genes in ARVC) may negatively impact the possibility of identifying disease-causing mutations.

A solution to this heterogeneity problem can be the design of disease-based experiments using NGS. This implies both the risk of over-sizing genetic testing in hundreds of patients and of detectioning of several genetic variants with unknown functional significance. Possible support for not getting confounded by the expected mass of genetic variations in genetically heterogeneous diseases such as DCM is the respect of clinical genetic rules, genotype-phenotype correlation studies, the observation of families (both affected and healthy mutation carriers) in long term follow-up and, when possible and feasible, functional and pathologic studies of the affected myocardium: this will be the object of my further research work.

Bibliography

[1] Elliott P, Andersson B, Arbustini E, Bilinska Z, Cecchi F, Charron P, Dubourg O, Kühl U, Maisch B, McKenna WJ, Monserrat L, Pankuweit S, Rapezzi C, Seferovic P, Tavazzi L, Keren A. Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases. Eur Heart J. 2008;29:270-6.

[2] Codd MB, Sugrue DD, Gersh BJ, Melton LJ 3rd. Epidemiology of idiopathic dilated and hypertrophic cardiomyopathy. A population-based study in Olmsted County, Minnesota, 1975-1984. Circulation. 1989;80:564-72.

[3] Nugent AW, Daubeney PE, Chondros P, Carlin JB, Cheung M, Wilkinson LC, Davis AM, Kahler SG, Chow CW, Wilkinson JL, Weintraub RG; National Australian Childhood Cardiomyopathy Study. The epidemiology of childhood cardiomyopathy in Australia. N Engl J Med. 2003;348:1639-46.

[4] Grünig E, Tasman JA, Kücherer H, Franz W, Kübler W, Katus HA.Frequency and phenotypes of familial dilated cardiomyopathy. J Am Coll Cardiol. 1998;31:186-94.

[5] Gavazzi A, Repetto A, Scelsi L, Inserra C, Laudisa ML, Campana C, Specchia C, Dal Bello B, Diegoli M, Tavazzi L, Arbustini E. Evidence-based diagnosis of familial non-X-linked dilated cardiomyopathy. Prevalence, inheritance and characteristics. Eur Heart J. 2001;22:73-81

[6] Baig MK, Goldman JH, Caforio AL, Coonar AS, Keeling PJ, McKenna WJ. Familial dilated cardiomyopathy: cardiac abnormalities are common in asymptomatic relatives and may represent early disease. J Am Coll Cardiol. 1998;31:195-201.

[7] Mahon NG, Murphy RT, MacRae CA, Caforio AL, Elliott PM, McKenna WJ. Echocardiographic evaluation in asymptomatic relatives of patients

with dilated cardiomyopathy reveals preclinical disease. Ann Intern Med. 2005;143:108-15.

[8] Repetto A, Serio A, Pasotti M, Fontana, A, Bertoletti, A, Scelsi L, Magrini G, Monti L, Campana C, Ghio S, Tavazzi L, Arbustini E. Re-screening of healthy relatives of patients with dilated cardiomyopathy identifies subgroups at risk of developing the disease. Eur Heart J, 2004, S6: F54-F60.

[9] Ahamad F, Seidman JG, Seidman CE. The genetic basis of cardioac remodelling. Annu Rev Genomics. Hum Genet 2005;6:185-216-

[10] Arbustini E, Diegoli M, Morbini P, Dal Bello B, Banchieri N, Pilotto A, Magani F, Grasso M, Narula J, Gavazzi A, Viganó M, Tavazzi L. Prevalence and characteristics of dystrophin defects in adult male patients with dilated cardiomyopathy. J Am Coll Cardiol. 2000;35:1760-8.

[11] Tsubata S, Bowles KR, Vatta M, Zintz C, Titus J, Muhonen L, Bowles NE, Towbin JA.Mutations in the human delta-sarcoglycan gene in familial and sporadic dilated cardiomyopathy.J Clin Invest. 2000 Sep;106(5):655-62.

[12] Barresi R, Di Blasi C, Negri T, Brugnoni R, Vitali A, Felisari G, Salandi A, Daniel S, Cornelio F, Morandi L, Mora M. Disruption of heart sarcoglycan complex and severe cardiomyopathy caused by beta sarcoglycan mutations. J Med Genet. 2000 Feb;37(2):102-7.Click here to read.

[13] Taylor MR, Slavov D, Ku L, Di Lenarda A, Sinagra G, Carniel E, Haubold K, Boucek MM, Ferguson D, Graw SL, Zhu X, Cavanaugh J, Sucharov CC, Long CS, Bristow MR, Lavori P, Mestroni L; Familial Cardiomyopathy Registry; BEST (Beta-Blocker Evaluation of Survival Trial) DNA Bank.Prevalence of desmin mutations in dilated cardiomyopathy. Circulation. 2007;115:1244-51.

[14] Olson TM, Michels VV, Thibodeau SN, Tai YS, Keating MT. Actin mutations in dilated cardiomyopathy, a heritable form of heart failure. Science. 1998;280:750-2.

[15] Kamisago M, Sharma SD, DePalma SR, Solomon S, Sharma P, McDonough B, Smoot L, Mullen MP, Woolf PK, Wigle ED, Seidman JG, Seidman CE. Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. N Engl J Med. 2000;343:1688-96.

[16] Kaski JP, Burch M, Elliott PM. Mutations in the cardiac Troponin C gene are a cause of idiopathic dilated cardiomyopathy in childhood.Cardiol Young. 2007 Dec;17(6):675-7.

[17] Mogensen J, Murphy RT, Shaw T, Bahl A, Redwood C, Watkins H, Burke M, Elliott PM, McKenna WJ. Severe disease expression of cardiac troponin C and T mutations in patients with idiopathic dilated cardiomyopathy. J Am Coll Cardiol. 2004;44:2033-40.

[18] Olson TM, Kishimoto NY, Whitby FG, Michels VV. Mutations that alter the surface charge of alpha-tropomyosin are associated with dilated cardiomyopathy. J Mol Cell Cardiol. 2001 Apr;33(4):723-32.

[19] Fatkin D, MacRae C, Sasaki T, Wolff MR, Porcu M, Frenneaux M, Atherton J, Vidaillet HJ Jr, Spudich S, De Girolami U, Seidman JG, Seidman C, Muntoni F, Müehle G, Johnson W, McDonough B.Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction-system disease.N Engl J Med. 1999;341:1715-24.

[20] Arbustini E, Pilotto A, Repetto A, Grasso M, Negri A, Diegoli M, Campana C, Scelsi L, Baldini E, Gavazzi A, Tavazzi L. Autosomal dominant dilated cardiomyopathy with atrioventricular block: a lamin A/C defect-related disease. J Am Coll Cardiol. 2002;39:981-90.

[21] Bonne G, Di Barletta MR, Varnous S, Bécane HM, Hammouda EH, Merlini L, Muntoni F, Greenberg CR, Gary F, Urtizberea JA, Duboc D, Fardeau M, Toniolo D, Schwartz K. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. Nat Genet. 1999;21:285-8.

[22] Arbustini E, Diegoli M, Fasani R, Grasso M, Morbini P, Banchieri N, Bellini O, Dal Bello B, Pilotto A, Magrini G, Campana C, Fortina P, Gavazzi A, Narula J, Viganó M. Mitochondrial DNA mutations and mitochondrial abnormalities in dilated cardiomyopathy. Am J Pathol. 1998;153:1501-10.

[23] Marshall JD, Beck S, Maffei P, Naggert JK. Alström syndrome. Eur J Hum Genet. 2007;15:1193-202

[24] Cohen N, Muntoni F. Multiple pathogenetic mechanisms in X linked dilated cardiomyopathy. Heart. 2004;90:835-41.

[25] Brega A, Narula J, Arbustini E. Functional, structural, and genetic mitochondrial abnormalities in myocardial diseases. J Nucl Cardiol. 2001;8:89-97.

[26] Luk A, Ahn E, Soor GS, Butany J. Dilated Cardiomyopathy: A review. J Clin Pathol. 2008 Nov 18. [Epub ahead of print]

[27] Arbustini E, Morbini P, Pilotto A, Gavazzi A, Tavazzi L. Familial dilated cardiomyopathy: from clinical presentation to molecular genetics. Eur Heart J. 2000;21:1825-32.

[28] Price JF, Thomas AK, Grenier M, Eidem BW, O'Brian Smith E, Denfield SW, Towbin JA, Dreyer WJ. B-type natriuretic peptide predicts adverse cardiovascular events in pediatric outpatients with chronic left ventricular systolic dysfunction. Circulation. 2006;114:1063-9.

[29] Meune C, Van Berlo JH, Anselme F, Bonne G, Pinto YM, Duboc D. Primary prevention of sudden death in patients with lamin A/C gene mutations. N Engl J Med. 2006;354:209-10.

[30] van Berlo JH, de Voogt WG, van der Kooi AJ, van Tintelen JP, Bonne G, Yaou RB, Duboc D, Rossenbacker T, Heidbüchel H, de Visser M, Crijns HJ, Pinto YM. Meta-analysis of clinical characteristics of 299 carriers of *LMNA* gene mutations: do lamin A/C mutations portend a high risk of sudden death? J Mol Med. 2005;83:79-83.

[31] Pasotti M, Klersy C, Pilotto A, Marziliano N, Rapezzi C, Serio A, Mannarino S, Gambarin F, Favalli V, Grasso M, Agozzino M, Campana C, Gavazzi A, Febo O, Marini M, Landolina M, Mortara A, Piccolo G, Viganó M, Tavazzi L, Arbustini E. Long-term outcome and risk stratification in dilated cardiolaminopathies. J Am Coll Cardiol. 2008;52:1250-60.

Chapter 4

Massive parallel sequencing (MPS) and phenotype: a new diagnostic solution?

This chapter will deal with Next Generation Sequencing applied to Inherited Cardiovascular Diseases. As anticipated in the conclusions of the prior chapters, NGS is candidate to be a possibly solution for the genetic heterogeneity of several inherited cardiovascular diseases. I will focus on the rationale that supports our decision to achieve a NGS instrument, and the design of the studies we are developing with this new technology. Then, I will describe the pilot sequencing study that was performed to optimize this new methodology both in tecnical biological preparation and bioinformatics analysis.

CONCEPTS

- In 2009 the aim of research in inherited cardiovascular diseases was genotype-phenotype correlation.
- The main objective was to distinguish gene caused diseases (eg: *LMNA* defects leads to Laminopathies) associated with gene-specific phenotype features.
- Hereditary cardiovascular diseases are genetically heterogeneous, with more than 80 disease-genes known to date. Thus limited the possibility of testing all genes in all patients, leaving a percentage of diagnosis-orphan patients.
- In these latter cases, the only possibility was to study all exons and flanking regions of known candidate genes.
- In 2012 the goal of research in inherited cardiovascular diseases was to solve this heterogeneity using the best cost- and time-saving tools and technologies.

- Massive parallel sequencing (NGS testing at the same time a panel of selected genes in several patients) seemed to offer a possible solution.

4.1 From Sanger sequencing to NGS sequencing

4.1.1 The diagnostic model: critical points

In 2009 the aim of the research in inherited cardiovascular diseases was genotype-phenotype correlation, based on the idea that a single gene causes a single disease. In this context, the aim of the research was to identify clinical markers differently represented in different patients and families. Sanger sequencing was and still is currently the gold standard method to detect sequence variants within disease genes.

Two major critical points remained unresolved:

1. Technical limitations of Sanger-based methodology (missing large insertions/deletions) and requiring other detection techniques (ex: MLPA, RT-qPCR);

2. Time and resources are wasted in sequencing all genes in genetically heterogeneous diseases. Current diagnostic tests are certified on Sanger sequencing. We currently run 3 plates of 96 sequences per day (ABI Prism 3130xl) for 5 days a week (1440 sequences per week). Although methods are optimized and fast, the procedures are still too long for timely response to the progressivelly increasing requests.

NGS methods are dramatically improving, allowing the analysis of gigabases of sequence information in 1 single run, with lower costs and shorter time. For this reason in 2011 we achieved a 454 Roche platform (Genome Sequencer Junior System, Roche Applied Science, Mannheim, Germany), an NGS instrument based on a method coupling amplicon based gene capture with re-sequencing to identify mutations in cardiomyopathies or connective tissue diseases, analyzing all known disease genes in a single experiment. Although the biological technique at the basis of this method is different from the Sanger sequencing, MPS can analyze the entire panel of candidate genes in one or more patients in one week, decreasing dramatically the time for diagnosis.

4.1.2 Roche GS junior

The most widely used Next Generation Sequencing (NGS) platforms are the Genome Sequencer from Roche/454 Life Sciences, the Genome Analyzer from Illumina/Solexa, and the SOLiD System from Applied Biosystems. These platforms differ for the read length and the number of DNA molecules sequenced [1, 2].
For some applications, such as search for comprehensive rare variant or identification of private disease mutations, an even higher accuracy is warranted.

This higher accuracy is particularly needed by clinical applications that generally target only a small set of genes associated with genetically heterogeneous diseases [3]. Unfortunately, although greater accuracy can be achieved by increasing coverage depth, all current NGS platforms have capacities that exceed clinically sustainable needs and applications [4], leading to the production into a single run of an enormous amount of data difficult to translate in clinical testing.
This suggests the need for easy-to-tailor, smaller capacity next-generation sequencers that can accurately and rapidly sequence DNA for clinical applications. Roche has recently introduced the GS Junior platform as a next-generation bench-top DNA sequencing solution scaled to suit the needs of small projects requiring a rapid turnaround time [5, 6, 7].

Figure 4.1: Roche GS Junior platform

The 454 technology is based on pyrosequencing, a sequencing approach that uses chemiluminescent detection of pyrophosphate released during polymerase-mediated deoxynucleoside triphosphate incorporation. During sequencing a CCD camera-based imaging assembly is used to capture the pyrosequencing-derived light signal and to collect the readout data per flow, which is then used by a Genome Sequencer-specific base-caller to generate the sequence reads.

Data analysis

The AVA (Alignment Variation Analyzer) Roche application computes the alignment of reads from Amplicon libraries obtained on the Roche GS Junior, and identifies differences between the reads and a reference sequence. Variations are displayed both graphically with a histogram indicating positions of variation, and textually with a color-coded multiple alignment that emphasizes regions and bases of difference from the reference sequence. The software specifically reports the frequency of user-defined and software-identified

variants in a summary table, allowing for the high-throughput detection and quantization of known and putative variants in the samples sequenced. The AVA software features a Command Line Interface (CLI), by which python coded instructions can be launched. Using python, I can program AVA software automatically at starting of each run, only selecting the specific genes on the plate with the specific reference sequences, and the number of different patients. This number is fixed for each type of run, depending on the coverage desired and the number of genes sequenced.

In order to improve the technical expertise of biologists uncharged to prepare the run, and my expertise with the bioinformatical analysis of data, we decided to start with a pilot study. For this study we used a commercial kit (BRCA MASTR v2.1, Multiplicum) for the sequencing of *BRCA1* and *BRCA 2*. These two genes are involved in the development of familial Breast and Ovarian cancer (HBOC), which is addressed to our Center and currently diagnosed using Sanger sequencing.

4.1.3 Pilot study: the *BRCA* Multiplicum kit case

A Sequencing Master library of amplicons covering all the coding exons and splice sites of *BRCA1* and *BRCA2* was produced for 8 patients using the BRCAMASTR kit (Multiplicom) following manufacturer instructions. Briefly, 50 ng of cDNA were used as template in each of 8 multiplex PCR reactions for each patient. These reactions amplified the complete exonic and splice sites of *BRCA1* and *BRCA2*. A 1:1000 dilution of the purified PCR products were re-amplified using molecular identification (MID) adaptors for each patient. A *BRCA* amplicon library of each patient was generated and equivalent concentrations of the libraries were pooled to generate a Sequencing Master library. Pyrosequencing of the Master libraries were done in the sense and anti-sense strands with the 454 GS Junior (Roche) technology.

Data analysis was done with the GS Amplicon Variant Analyzer software (Roche) comparing results against genomic references NG_005905 and NG_012772 for *BRCA1* and *BRCA2*, respectively. The cDNA references utilized were NM_007294 and NM_000059 for *BRCA1* and *BRCA2*, respectively. The nomenclature is based on the cDNA sequence and is according to Human Genome Variation Society (http://www.hgvs.org/, hg19).
All the deleterious mutations found were controlled by Sanger sequencing of original patient blood DNA. The putative functional effects of missense variants were analyzed *in sylico* with PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/).

Direct sequencing of the purified PCR products, obtained with the same primers and PCR conditions described for amplicon library preparation, was performed in both directions (PE Big Dye Terminator Cycle Sequencing Kit)

on an ABI Prism 3310 genetic analyzer (PE Applied Biosystems, Forest City, CA, USA) and analyzed with the Sequencer software.

After mapping, 100% of the targeted region was covered in each patient. Mean percentage covered at >100-fold depth were 95%. A total of 356 variants were detected, with an average of 118 variants per patient within the 35966 bp targeted region.

In this pilot study, the gold standard Sanger method was used to distinguish real changes from false positives. Sanger sequencing confirmed 16 SNP and 4 mutations. To distinguish pathogenic mutations from polymorphisms, we compared results with literature and personal data and with public databases of known sequence variants (dbSNP). All confirmed variants corresponded to known polymorphisms.

Sanger sequencing was also used to evaluate the number of false-positive results and to determine indicative cut-off values to potentially use in a diagnostic setting. Sanger sequencing revealed that variations detected in a percentage <21% of forward and reverse sequence reads were all false positives of 454 technology. These sequence changes represent a high percentage of "not confirmed variations" (mean value 80%) and, therefore, this might be a useful signal for flagging possible false-positive results.

Furthermore, Sanger sequencing indicated that a small fraction of highly unbalanced data with one variation detected in a significant percentage (26 - 96%) of one sequence strand, but in a very low percentage (0 - 5%) of the other strand, should be considered as technical artifacts.

Finally, we found 18 variations located in polyT stretches that were not confirmed by Sanger sequencing. Drawbacks of pyrosequencing include the fact that the signal intensity must be correlated with the number of bases incorporated, which can be confounding when encountering homonucleotide regions that are greater than six bases in length. As these repeated regions can represent hotspots for disease causing mutations, accurate detection of insertions/deletions in these stretches is very important. Technical and/or bioinformatic improvements should, therefore, be provided for diagnostic application.

The GS Junior platform is more sensitive than Sanger sequencing in terms of random sequencing errors. For this reason the use of this platform in diagnostic laboratories requires the establishment of an analytic work-flow to select candidate pathogenic changes.
On the basis of the present study, I hypothesized the following flowchart for the detection of pathologic mutation (Figure 4.2):

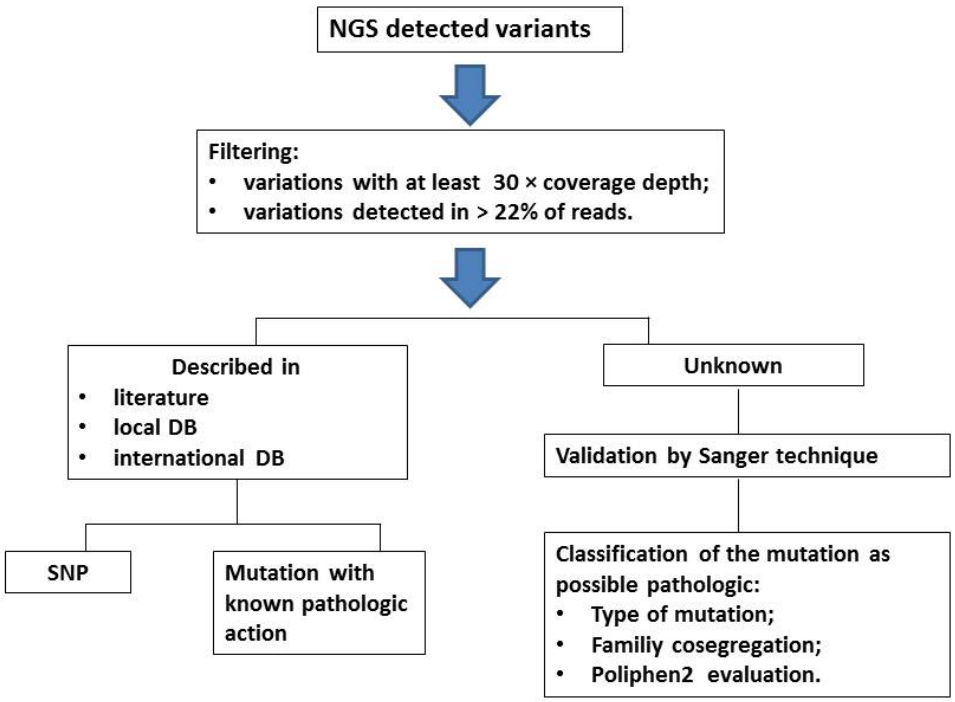


Figure 4.2: Flowchart illustrating the different steps to recognize pathologic variations detected by 454 technology in the pilot study.

1. Filtering:
   - of variations with at least 30 coverage depth;
   - of variations detected in >22% of reads.
2. Comparison with literature, known DBs (local, international) and individuation of described SNPs / mutations.
3. Validation at Sanger sequencing.
4. Classification of confirmed variation considering family co-segregation, type of mutation and Polyphen results.

Other runs are ongoing and will be analyzed in order to confirm this flowchart for the analysis.. Further development will be the creation of a web-based automatic analysis disease-specific system, which could show, in one single shot, variant candidate to Sanger validation for biologists.

4.2 Conclusions

Critical aspects of this new methodology include:

1. THE SIZE OF AMPLICONS that has to be smaller (smaller read length of $\tilde{3}$00-350 bp) than in Sanger sequencing (ũp to 700bp) so that the number of PCR is higher than with Sanger-based technology.
   RESOLUTION: all primers for the novel platform have been designed and, to date, first experimental runs have shown that the new primers work. Very few primers (6%) had to be redesigned and synthesized. A possible solution is the design of oligonucleotides containing adapters for plate preparation and MIDs for samples recognition, which allow to amplify the right size ($\tilde{3}$00-350 bp) of amplicons of a large number of genes. Novel commercial kits (i.e.: BRCA MASTR v2.1 used in pilot study) are under development based on multiplex PCR ready to use (at present the number of these kit are restricted to few diseases).

2. THE INCREASING NUMBER OF DISEASE GENES in genetically heterogeneous diseases (e.g. cardiomyopathies, TAAD, cancer, etc). The fast knowledge progression requires a real-time capacity of generating novel targeted panels considering the need of total coverage to detect rare variants. Each run will have to be progressively optimized by balancing the number of patients vs. the number of genes in each panel. The Pyrosequencing technology has its limitations for correct base-calling in homo-polymeric regions. Reads lacking one or more nucleotides or containing an insertion will complicate analysis.
   RESOLUTION: Specific expertise and optimal working conditions are necessary to overcome this problem. Collaborations with other labs. (MAMMAG, Irvine CA) using the NGS Pyrosequencing technology are ongoing and facilitating our learning about this problem [8].

3. FILTERING OUT A LARGE NUMBER OF CANDIDATE PATHOGENETIC VARIANTS (more than 30000) before identification of the mutation responsible for disease.
   RESOLUTION: Bioinformatic analysis tools as well as clinical support (extensive clinical family studies) will be the major contributors to the solutions of this problem.

To sum up, massive parallel sequencing is candidate to became a valid resource for the resolution of the genetic heterogeneity of these complex diseases, giving the possibility to create customized disease-specific panels of genes.
However, as for any new high technology, the learning period and the optimization of chemical procedures to obtain the better quality material to be sequenced takes time and money. For this reason, the real cost-saving advantages of this new techniques will be object of further benchmarking evaluations,

when the number of precise genetic diagnosis matching phenotype segregation in a large number of families will be available. The current way of presenting lower cost by NGS vs Sanger is based on bp numbers. In our view the way of measuring the true cost of the novel technology should be based on the number of cases really solved.
By Sanger sequencing of 40 genes in 100 patients with familial DCM, we identified about 20% of cases with double and triple mutations obligating clinicians to large family screening and monitoring to establish which of these mutations causes the disease. In fact we should not forget that clinical genetic rules exist and that more than 90% of inherited cardiovascular diseases are autosomal dominant.

A new pilot panel for cardiomyopathies studies has been designed on the basis of gene prevalence and will be synthesized (SeqCap EZ Design v3.0, Nimblegene) in the next few months.

Bibliography

[1] Margulies et al.Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437: 376-380.

[2] Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods. 2008; 5: 16-18.

[3] Harismendy et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol. 2009; 10: R32.

[4] Ng SB et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009; 461: 272-276.

[5] Ng et al. Massively parallel sequencing and rare disease. Hum Mol Genet. 2010; 19: R119-R124.

[6] Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010; 11: 31-46.

[7] Dames et al. Comparison of the Illumina Genome Analyzer and Roche 454 GS FLX for resequencing of hypertrophic cardiomyopathy-associated genes. J Biomol Tec. 2010; 21: 73-80.

[8] Zaragoza MV, Fass J, Diegoli M, Lin D, Arbustini E. Mitochondrial DNA variant discovery and evaluation in human Cardiomyopathies through next-generation sequencing. PLoS One. 2010;5:e12295.

Chapter 5

Further plans

For the short- and mid-term future my plans include a project for a global integration of multi-source data (genetic, imaging, clincial) from patients, families and controls.

I will also implement the massive parallel sequencing activities to support the overall scientific and clinical work of the Center and work for the indentification of novel genes for genetically-orphan familial diseases already collected and followed-up at the Center. To reach this aim we will use exome sequencing and linkage analysis in existing families.

In addition new inherited cardiovascular diseases are sistematically entering the activity of the center (i.e. familial atrial diseases including atrial fibrillation) and I am planning a novel model for the family screening of inherited cardiovascular diseases that may manifest episodically, rather than being chronically present.

List of publications

5.1 Submitted

- PREVALENCE OF LAMIN A/C gene mutations in a consecutive series of 1000 unrelated probands diagnosed with dilated cardiomyopathy. **Valentina Favalli**, Nupoor Narula, Andrea Pilotto, Maurizia Grasso, Alessandra Serio, Philippe Charron, Benjamin Meder, Yigal Pinto, Perry MElliott, Jens Mogensen, Eloisa Arbustini. JACC.

- Aortic Root 3D Morphological Model from 2D-Echo Images. Simone Morganti, **Valentina Favalli**, Adele Valentini, Alessandra Serio, Fabiana I. Gambarin, Danila Vella, Laura Mazzocchi, Massetti Massimo, Ferdinando Auricchio, Eloisa Arbustini. Global Heart

- Loeys-Dietz Syndrome represents a spectrum of malignant vascular phenotypes. Eloisa Arbustini, Eliana Disabella, **Valentina Favalli**, Alessandra Serio , Maurizia Grasso, Catherine Klersy ,Roberto Dore ,Attilio Odero , Savina Mannarino, Nupoor Narula, Marina Colombi, Shamik Bhattacharyya, Duke Cameron, Luca Vricella, Bart Loeys, Harry C Dietz. JACC

5.2 Accepted in press

- Establishing Age-Specific Reference Values for Carotid Arterial Stiffness. Tokuhisa Uejima, Frank D Dunstan, Eloisa Arbustini, Krystyna Åoboz-GrudzieÅ, Alun D Hughes, Scipione Carerj, **Valentina Favalli**, Francesco Antonini-Canterin, Olga Vriz, Dragos Vinereanu, Jose L Zamorano, Bogdan A Popescu, Arturo Evangelista, Patrizio Lancellotti, Georges Lefthériotis, Carlo Palombo, Alan G Fraser. Eur Heart J 2012, in press.

5.3 Published

- Familial DCM: a clinically and genetically heterogeneous disease. Alessandra Serio, Nupoor Narula, Takahide Kodama, **Valentina Favalli**, Eloisa Arbustini. In press Herz, 2012,37:822-829.

- Autosomal recessive atrial dilated cardiomyopathy with standstill evolution associated with mutation of Natriuretic Peptide Precursor A. Marcello Disertori, Silvia Quintarelli, Maurizia Grasso, Andrea Pilotto, Nupoor Narula, **Valentina Favalli**, Camilla Canclini, Marta Diegoli, Silvia Mazzola, Massimiliano Marini, Maurizio Del Greco, Roberto Bonmassari, Michela MasÃ¨, Flavia Ravelli, Claudia Specchia, Eloisa Arbustini.Circ Cardiovasc Genet, 2012 Dec 29 [Epub ahead of print].

- Quantitative Expression of the Mutated Lamin A/C Gene in Patients With Cardiolaminopathy. Narula N, **Favalli V**, Tarantino P, Grasso M, Pilotto A, Bellazzi R, Serio A, Gambarin FI, Charron P, Meder B, Pinto Y, Elliott PM, Mogensen J, Bolognesi M, Bollati M, Arbustini E. J Am Coll Cardiol. 2012; S0735-1097(12)04113-7.

- Structures of the lamin A/C R335W and E347K mutants: implications for dilated cardiolaminopathies. Bollati M, Barbiroli A, **Favalli V**, Arbustini E, Charron P, Bolognesi M. Biochem Biophys Res Commun. 2012;418:217-21.

- Information technology solutions to support translational research on inherited cardiomyopathies. Bellazzi R, Larizza C, Gabetta M, Milani G, Bucalo M, Mulas F, Nuzzo A, *Favalli V*, Arbustini E. Stud Health Technol Inform. 2011;169:907-11.

- Diagnostic work-up and risk stratification in X-linked dilated cardiomyopathies caused by dystrophin defects. Diegoli M, Grasso M, **Favalli V**, Serio A, Gambarin FI, Klersy C, Pasotti M, Agozzino E, Scelsi L, Ferlini A, Febo O, Piccolo G, Tavazzi L, Narula J, Arbustini E. J Am Coll Cardiol. 2011;58:925-34.

- Risk of dissection in thoracic aneurysms associated with mutations of smooth muscle alpha-actin 2 (*ACTA2*).Disabella E, Grasso M, Gambarin FI, Narula N, Dore R, **Favalli V**, Serio A, Antoniazzi E, Mosconi M, Pasotti M, Odero A, Arbustini E Heart. 2011;97:321-6.

- When should cardiologists suspect Anderson-Fabry disease? Gambarin FI, Disabella E, Narula J, Diegoli M, Grasso M, Serio A, **Favalli V**, Agozzino M, Tavazzi L, Fraser AG, Arbustini E. Am J Cardiol. 2010 Nov 15;106(10):1492-9.

- Transcriptomic and proteomic analysis in the cardiovascular setting: unravelling the disease? Marziliano N, Grasso M, Pilotto A, Porcu E,

Tagliani M, Disabella E, Diegoli M, Pasotti M, **Favalli V**, Serio A, Gambarin F, Tavazzi L, Klersy C, Arbustini E. J Cardiovasc Med (Hagerstown). 2009;10:433-42.

- Rationale and design of a trial evaluating the effects of losartan vs. nebivolol vs. the association of both on the progression of aortic root dilation in Marfan syndrome with *FBN1* gene mutations. Gambarin FI, **Favalli V**, Serio A, Regazzi M, Pasotti M, Klersy C, Dore R, Mannarino S, Viganó M, Odero A, Amato S, Tavazzi L, Arbustini E. J Cardiovasc Med (Hagerstown). 2009;10:354-62.

- Long-term outcome and risk stratification in dilated cardiolaminopathies. Pasotti M, Klersy C, Pilotto A, Marziliano N, Rapezzi C, Serio A, Mannarino S, Gambarin F, **Favalli V**, Grasso M, Agozzino M, Campana C, Gavazzi A, Febo O, Marini M, Landolina M, Mortara A, Piccolo G, Vigano' M, Tavazzi L, Arbustini E. J Am Coll Cardiol. 2008;52:1250-60.

Acknowledgments

I wish to thank my advisor, Prof. Eloisa Arbustini, and Prof. Riccardo Bellazzi for their guidance and for giving me the opportunity to pursue this scientic interest.

I want to thank all the staff of the Center for Inherited Cardiovascular Diseases for all their support.

I would like to thank my friends and collegues that accompanied me in the last three years.

In addition another special thank to Cecilia Trovati, Monica Concardi e Annamaria Tripi, for all the advice they gave me and for their valuable friendship; I'm honoured of this.

Last but not least I wish to thank my dad, my sister Silvia, my best friend Elisabetta (I'll always be grateful for all the work you've done for me), and Beppe, who never fails to support and believe in me.

List of Figures

List of Tables