Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Salute della Donna e del Bambino

CORSO DI DOTTORATO DI RICERCA IN MEDICINA DELLO SVILUPPO E SCIENZE DELLA PROGRAMMAZIONE SANITARIA

CURRICULUM EMATO-ONCOLOGIA, GENETICA, MALATTIE RARE E MEDICINA PREDITTIVA
CICLO XXXI

Implementation of NGS protocols for the diagnosis of rare genetic diseases and development of models for the validation of mutations

Coordinatore: Ch.mo Prof. Carlo Giaquinto

Supervisore: Ch.mo Prof. Leonardo Salviati

Dottorando : Dott.ssa Roberta Zordan

INDICE

RIASSUNTO

Presupposti dello studio. Il sequenziamento del DNA è un ambito della biologia molecolare in piena evoluzione. Nell'ultimo decennio le tecnologie di sequenziamento di nuova generazione – conosciute come Next Generation Sequencing (NGS) – sono diventate sempre più diffuse nella pratica clinica. Queste nuove tecnologie, per la loro caratteristica di fornire milioni di sequenze di DNA per reazione a prezzi contenuti, hanno completamente rivoluzionato l'approccio all'analisi del genoma umano.

Scopo della tesi. Lo scopo della tesi è verificare l'applicabilità di metodiche di sequenziamento di nuova generazione (NGS) da introdurre nella pratica diagnostica di laboratorio. A tal proposito sono state comparate due diverse strategie di sequenziamento massivo-parallelo basate sulle due diverse tecnologie *Illumina* e *Agilent*, che sono state adottate per lo studio dei pazienti in analisi presso la U. O. di Genetica ed Epidemiologia Clinica dell'Azienda Ospedaliera di Padova.

L'attività di ricerca si è focalizzata anche sullo sviluppo di modelli per la validazione della patogenicità e la caratterizzazione delle varianti geniche identificate mediante sequenziamento NGS.

A tal proposito, il sistema del minigene ibrido è stato utilizzato per verificare la capacità di alterare il meccanismo molecolare di *splicing* di una nuova variante nel gene *COQ4* identificata in un paziente affetto da sindrome da deficit del coenzima Q. E' stato inoltre sviluppato un sistema di complementazione funzionale in lievito S. *cerevisiae* del gene *ASS1* umano, al fine di analizzare e validare la patogenicità di alcune mutazioni missenso, identificate in pazienti affetti da citrullinemia di tipo 1, con lo scopo di ampliare in futuro l'analisi a possibili nuove varianti identificabili.

Risultati e conclusioni. L'analisi comparativa delle metodiche di sequenziamento NGS *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* e *HaloPlex HS Target Enrichment System (Agilent Technologies)* ha dimostrato come quest'ultima sia più performante in ambito diagnostico

considerando i paramentri di coverage medio, variabilità dei risultati, capacità di discriminare gli pseudogeni, tasso di falsi positivi, numero di ampliconi generabili per l'analisi. Per questi motivi dal 2017 stiamo utilizzado solo i pannelli *HaloPlex HS Target Enrichment System (Agilent).*

L'analisi mediante sequenziamento NGS di 85 pazienti afferenti alla U. O. di Genetica ed Epidemiologia Clinica dell'Azienda Ospedaliera di Padova risultati positivi allo *screening* metabolico neonatale per Mucopolisaccaridosi di tipo I (MPS I) o per Deficit di Biotinidasi o per Fenilchetonuria (PKU), ha rivelato come l'analisi molecolare permetta di distinguere i pazienti veri positivi dai falsi positivi allo *screening* metabolico neonatale, evitando terapie inutili e abbattendo notevolmente il costo dell'indagine.

Il sistema del minigene ibrido utilizzato per la validazione della nuova variante IVS4+1G>A nel gene *COQ4* si è rivelato essere un metodo efficace e relativamente semplice per dimostrare la patogenicità della mutazione in esame: essa si comporta come un allele ipomorfo in quanto provoca l'attivazione di un sito di *splicing* criptico a livello dell'esone 4 del gene *COQ4* producendo sia il trascritto *wild-type* sia un trascritto più piccolo di circa 100 bp.

Infine, il modello di lievito S. cerevisiae sviluppato per testare la patogenicità delle mutazioni missenso a carico del gene *ASS1*, ha permesso di classificare i mutanti in due classi: mutazioni di classe I che aboliscono completamente la crescita del lievito, mutazioni di classe II che consentono una crescita residua. Tale sistema è stato utile nel confermare la patogenicità delle mutazioni oggetto di studio ma non per stabilire correlazioni tra il genotipo e il fenotipo dei pazienti.

ABSTRACT

Background. DNA sequencing is a field of evolving molecular biology. In the last decade, new generation sequencing technologies - known as Next Generation Sequencing (NGS) - have become increasingly common in clinical practice. These new technologies, by their characteristic, are completely revolutionized to the analysis of the human genome.

Aim of the thesis. The aim of the thesis is to verify the applicability of the new generation sequencing methods (NGS). In this regard, the Illumina and Agilent different massively-parallel sequencing strategies were compared, which were adopted for the study of patients analyzed at the Genetics Unit of the Hospital of Padua.

The research activity is also focused on the development of models for the pathogenicity validation and the characterization of the genetic variants identified by NGS sequencing.

A new variant in the *COQ4* gene was identified in a patient with coenzyme Q deficiency syndrome and its potential ability to alter the splicing mechanism was studied by hybrid minigene system.

A functional complementation in S. *cerevisiae* yeast was performed in order to examine and validate the pathogenicity of some missense mutations in the *ASS1* gene, identified in patients with citrullinemia type 1, with the intent to extending the analysis to the potential new variants identifiable in the future.

Results and conclusions. The comparative analysis of the NGS technologies *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* and *HaloPlex HS Target Enrichment System (Agilent Technologies)* showed how the second one is more efficient in diagnostic, considering the followed parameters: coverage, results variability, pseudogenes and false positive discrimination, number of amplicons generated for the analysis. For these reason, from 2017 we use only the *HaloPlex HS Target Enrichment System (Agilent Technologies).*

In the study, 85 patients positive for neonatal metabolic screening for Mucopolysaccharidosis type I (MPS I) or for Biotinidase Deficiency or

Phenylketonuria (PKU), have been analyzed by Next Generation Sequencing: molecular analysis made it possible to distinguish true positive patients from false positives to neonatal metabolic screening (it would be important to review the cutoff of this screening), allowing to avoid unnecessary therapies and significantly reduce the cost of the investigation.

The hybrid minigene system used for the validation of the new IVS4+1G>A variant in *COQ4* gene was considered an effective and simple method to demonstrate the pathogenicity of the mutation: it behaves like a hypomorph allele causing the activation of a cryptic splicing site at exon 4 of the *COQ4* gene producing both the wild-type transcript and a smaller one less of about 100 bp.

Finally, the functional complementation in S. *cerevisiae* yeast system developed to test the pathogenicity of missense mutations in the *ASS1* gene, allowed to classify the mutants in the two classes: class I mutations that completely abolish yeast growth, class II mutations that use a residual growth. This system was useful to validate the pathogenicity of the mutations studied but not to made correlation between the genotype and the phenotype of the patients.

1. INTRODUCTION

3.1. DNA sequencing

DNA sequencing is a field of evolving molecular biology.

For years most applications have been based on the use of the Sanger method, or chain termination method (Sanger et al., 1977), which exploits modified nucleotides (didoxynucleotides triphosphates, ddNTPs) to interrupt the DNA synthesis reaction in specific positions. This concept allowed the development of automated sequencing systems based on this method (Smith et al., 1986) which, to date, represents the most accurate sequencing system available and the technique of choice in genetic diagnostics.

In the last decade, however, new generation sequencing technologies known as *Next Generation Sequencing* (NGS) (or, alternatively, *Massive Parallel Sequencing*, MPS, or *Second Generation Sequencing*, SGS) have become part of clinical practice, supplanting the methods used routinely. With the ability to provide, at reasonable prices, millions of DNA sequences by reaction, they completely revolutionized the way to approach the analysis of the human genome. The reduction in costs and times introduced by these tools is a great potential in the future of molecular biology.

3.1.1. Sanger sequencing

Sanger sequencing, also known as first-generation sequencing, was developed in 1977 by Frederick Sanger (Sanger et al., 1977) and is based on the use of modified nucleotides (dideoxynucleotides triphosphates, ddNTPs) to interrupt the synthesis reaction of DNA in specific positions.

From its proposal the method has undergone several evolutions that to date make the technique easier and faster, without the use of radioactive substances. Currently the Sanger sequencing is based on the clonal amplification of the DNA

mold by PCR and on the use of 4 ddNTPs labeled with different fluorochromes that allow to conduct the reaction in a single tube (Figure 1).

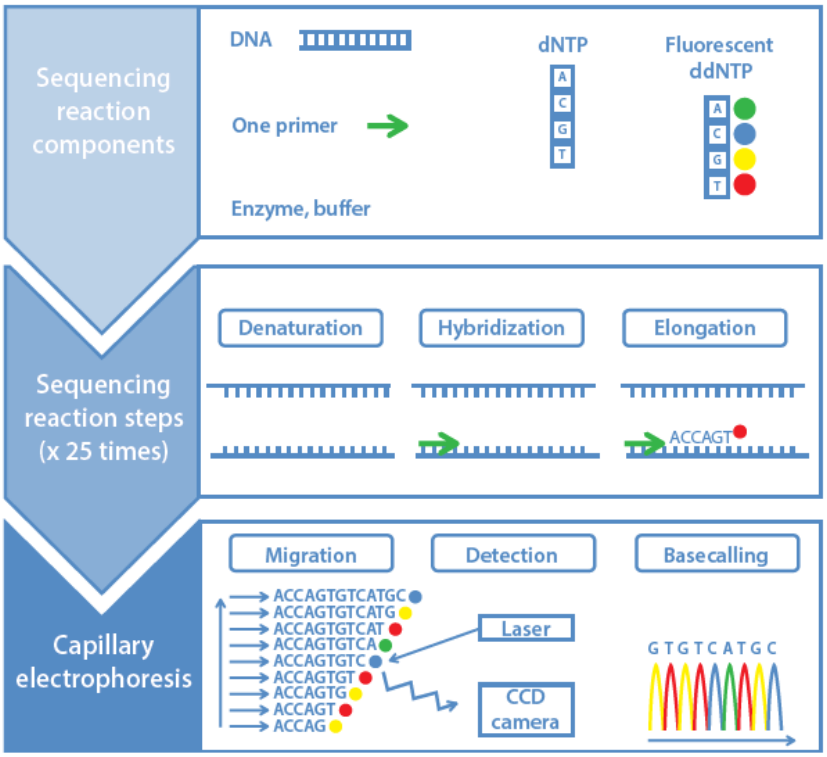


Figure 1. Method used by Sanger technology based on the use of fluorescent chain terminators (taken from Applied Biosystems website, https://tools.thermofisher.com/content/sfs/brochures/brochure-abgeneticanalyzers.pdf).

The concentration ratio between dNTPs ddNTPs is such that the DNA polymerase will statistically terminate the nascent chain in all possible positions, generating fragments of different lengths. This approach allowed the implementation of automated sequencing methods (Smith et al., 1986; Ansorge et al., 1987) based on a capillary electrophoresis system (Swerdlow and Gesteland, 1990) in which a polymeric matrix allows the migration of fragments of DNA at different rates depending on their length. After the excitation caused by a laser source present at the exit from the capillary, the separation of the fragments is followed by detecting the fluorescent emissions at different wavelengths of the four fluorochromes used as markers. The emissions are collected and analyzed by a CCD (*Charge-Coupled Device*) that processes the different fluorescence signals and returns them in the form of a graph called electropherogram, characterized by

a succession of peaks of four different colors corresponding to the fluorescent emissions of the four different fluorochromes. The raw data obtained from the sequencer are assembled and analyzed using dedicated software in order to obtain the nucleotide sequence.

The introduction of capillary electrophoresis for the separation of the marked fragments and the implementation of further improvements in Sanger technology (Carrilho, 2000, Metzker, 2005) have allowed a considerable increase in the processivity, allowing to obtain sequences of length between 600 bp and 1000 bp (Hert et al., 2008; Shendure et al., 2008) and sequencing up to 96 samples in parallel. The daily output of a current sequencer is around 6 million bases (6 Mb), with costs in the order of 500 dollars / Mb (Shendure and Ji, 2008, Kircher and Kelso, 2010).

The most frequently observed sequencing errors are essentially due to the errors introduced in the amplification step, to the contamination of the sample under examination or to the slippage of the polymerase in correspondence with homopolimeric sequences; nevertheless, the average error rate is around rather low values, in the order of 10-3.

3.1.2. Next Generation Sequencing (NGS)

DNA sequencing, based on the traditional Sanger method, the only method available until a few years ago, has undergone a sudden acceleration in the last decade thanks to the birth of the next generation sequencing technologies (NGS, *Next Generation Sequencing*). The development of platforms that exploit this new technology has led to a profound change in the nucleic acid sequencing methodology, opening the way to new opportunities in those areas in which DNA analysis is the basis of the investigation procedures.

NGS technologies, which implement a system with greater processivity for reading sequences in parallel, are able to produce large amounts of data in a short time: think, for example, that NGS *HiSeq X Series* platforms (*Illumina*) can produce up to 6 billion reads with a sequencing cycle of only 3 days. Parallel to this development in terms of throughput, there has been a drastic reduction in

terms of costs that led, for example, the sequencing of a Mb to cost less than 1 dollar against the approximately 10,000 dollars needed to achieve the same result in 2000 .

While NGS technologies have the considerable advantage of being able to produce millions of DNA sequences in a single reaction, on the other hand the huge amount of data produced still represents an important challenge for management and analysis. bioinformatics. Therefore, the implementation of systems capable of storing the ever-increasing quantity of data and, in parallel, the development of advanced bioinformatics algorithms that make it possible to efficiently analyze the data produced, simplifying the biological interpretation, have become necessary.

The NGS technologies allow, compared to the Sanger sequencing, to obtain a series of advantages, not only in terms of reduced costs and times, but also in relation to the possibility of processing several samples in parallel in a same sequencing reaction; the process, known as *multiplexing*, assigns to the fragments of each sample a short nucleotide sequence or barcode that, during data analysis, allows their identification and their univocal assignment to the sample to which they belong.

Another typical feature of NGS sequencing is the ability to accurately detect the presence of low-frequency alleles, allowing the identification of variations in mosaic samples or heterozygous deletions. Unlike the Sanger method, the number of times that a fragment of DNA is amplified and sequenced is proportional to the abundance of this segment in the original sample so that, through the use of specific algorithms in data analysis, it is possible to identify variations in the number of copies (CNVs) (Tucker et al., 2009b).

Given a number of positive aspects and the advantages that derive from it in the various areas of molecular biology, however, NGS technology has limitations that affect its application. The new generation sequencing platforms available today generate sequence reads in shorter meanings (35 - 400 bp) compared to those obtained with the traditional Sanger method (500 bp - 1 kb) (Hert et al., 2008). This aspect makes the assembly of sequences considerably difficult, especially in cases of unknown genomes (sequencing *de novo*) (Nagarajan and Pop, 2010) or with large areas repeated or containing rearrangements. The implementation of

paired-end or *mate-paired* sequencing, which involves the sequencing of DNA fragments from the extremities and in opposite directions, has allowed to exceed this limit almost completely, allowing the analysis of DNA fragments up to 5 times long. - 10 kb and being able to identify structural variants in the largest human genome of 3 kb (Korbel et al., 2007).

Moreover, compared to Sanger sequencing, the new generation technologies show an error rate in the call of the bases on a higher average: if in the case of capillary sequencers on average one reading in a thousand corresponds to an incorrect base (10-3), the The error rate of NGS platforms is around values between 0.1% and 1% (10-3 - 10-2) (Liu et al., 2012). This limit can be largely exceeded by increasing the reading depth (coverage or depth, indicating how many times a nucleotide is read during the sequencing), ie increasing the number of sequenced reads, so as to improve the quality of the experiment and make the product result more accurate. From the comparison between first generation and second generation sequencing technologies it is easy to understand how NGS techniques gave rise to a huge impulse in molecular biology studies thanks to their ability to produce a large amount of sequence data in extremely short time reduced, opening new fronts of research and allowing the achievement of results until a few years ago. The Sanger method, however, due to its high reliability, is still today the technique of choice in the clinical-diagnostic practice of nucleic acid sequencing, representing the most useful system in the validation of the results obtained by NGS technology.

3.1.3. NGS technologies

In the last decade many platforms have been developed and perfected based on the new generation sequencing system, giving rise to a real revolution in the field of nucleic acid sequencing, both for the large amount of data produced and for the speed with which they are generated.

Although they differ in the type of biochemistry at the base and in the methods of data acquisition and processing, the various NGS platforms share an operational *workflow* divided into four main phases:

- preparation of the DNA fragment library;
- clonal amplification of the fragments;
- sequencing;
- bioinformatic analysis of data.

Below are the different technologies used by second generation NGS platforms.

454 Genome Sequencer FLX (Roche Diagnotics Corporation). The *454* platform, marketed in 2005 by *454 Life Sciences* (now owned by *Roche Diagnostics Corporation*), was the first NGS instrumentation to be launched on the market (Margulies et al., 2005). It uses an alternative sequencing technology known as pyrosequencing (Ronaghi et al., 1996), based on the determination of the presence of the pyrophosphate released after the incorporation of a nitrogenous base in the nascent DNA chain. The pyrophosphate acts as a substrate for a series of reactions that, through the luciferase enzyme, produce an emission of light energy that is detected by a sensor.

In this system (Figure 2) the double-stranded DNA is fragmented and denatured and, at the ends of each fragment, short oligonucleotide sequences, called adapters, are required for the subsequent amplification and sequencing steps (Figure 2a).

The fragment library thus constructed is hybridized to 28 μm diameter agarose spheres which have complementary adaptive oligonucleotides on their surface. The clonal amplification of the fragments takes place by PCR in emulsion (emPCR, *PCR emulsion*) (Dressman et al., 2003): each bead, linked to a single type of fragments, is placed in the aqueous phase of an emulsion of water in oil containing PCR reagents (Figure 2b).

After amplification, the spheres are placed in a plate (PTP, *picotiter plate*) consisting of wells of diameter such as to be able to accommodate a single sphere (44 μm) (Figure 2c).

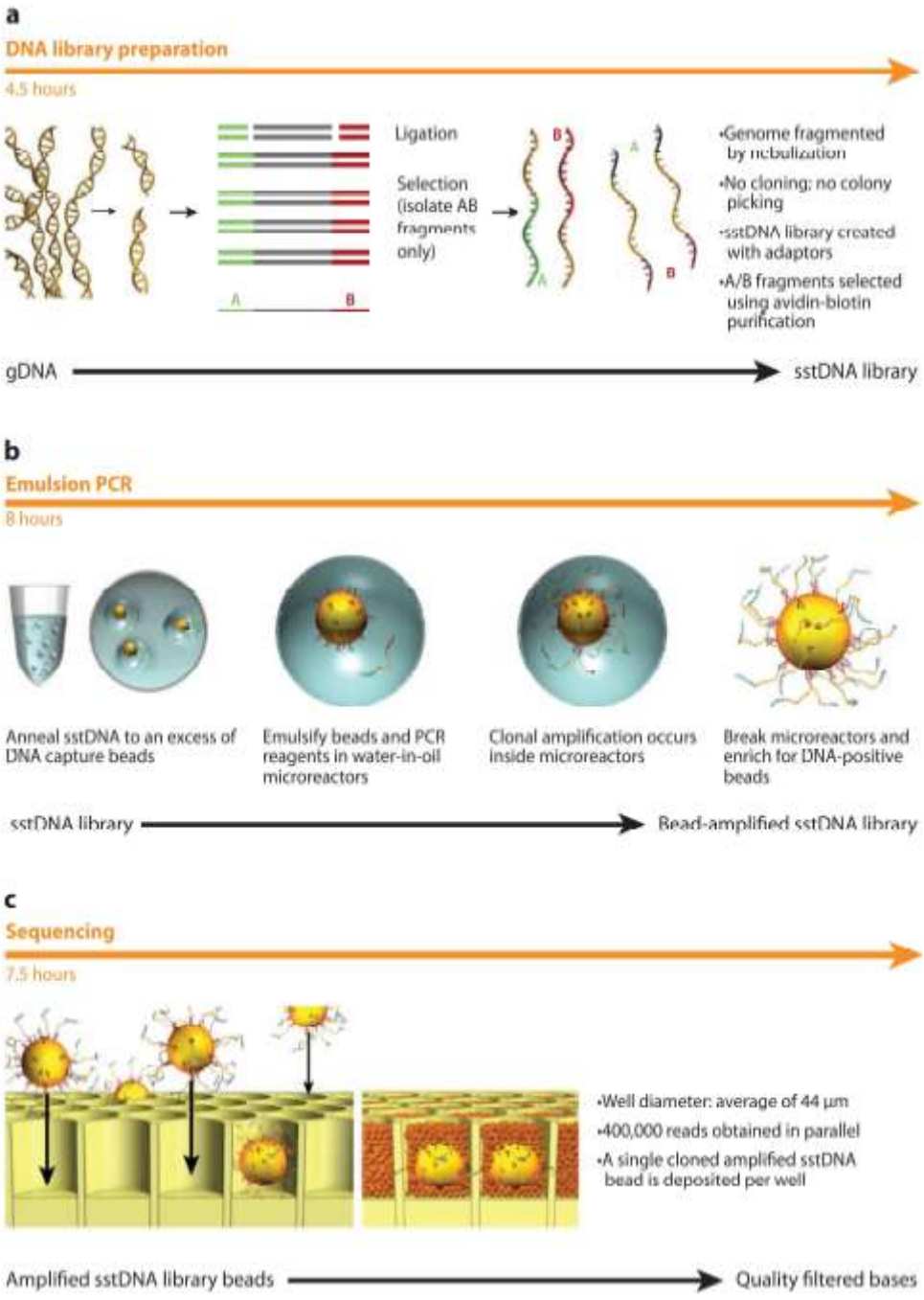


Figure 2. Method used by technology *454* (Mardis, 2008).

At this point smaller spheres are added linked to sulfurylase, luciferase, apirase and luciferin, enzymes and substrates necessary for the subsequent sequencing reaction. Pyrosequencing involves the iterative completion of the nascent DNA strand and the simultaneous reading of the signal emitted by the embedded nucleotides (SBS, *sequencing by synthesis*). If the dNTP is complementary to the mold strand, its incorporation causes the release of a pyrophosphate which,

following the action of sulfurylase, is converted into ATP (Figure 3). The produced ATP triggers the reaction of luciferase which, by converting luciferin into oxiluciferin, produces light energy in a proportional proportion to the present ATP and, consequently, to the number of nucleotides added to the filament. The light signal emitted is transmitted through the optical fibers of the plate and measured by a CCD (*Charge-Coupled Device*) which processes it to obtain the nucleotide sequence. Non-embedded dNTPs and ATP are degraded by the enzyme apirase and the system proceeds with sequencing of subsequent bases.

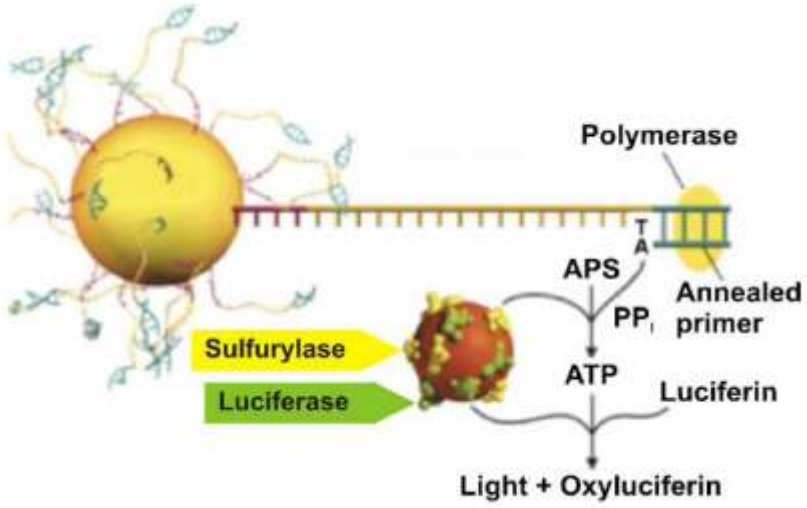


Figure 3. Schematic illustration of the Pyrosequencing reaction (Ansorge, 2009).

Pyrosequencing has disadvantages especially in the case of homopolymers (ie sequences consisting of the same base, for example AAAA): since there is no type of DNA chain terminator group, more equal bases can be incorporated at each sequencing cycle at the same time and their number can be deduced only from the intensity of the light emitted which, in some cases, can be misleading. As a consequence, the most frequently observed error using a platform 454 is the insertion or deletion of one or more bases, while the single base substitution errors are rather rare. Compared to other NGS technologies, this platform produces quite long reads, between 200 and 300 bp, which make it particularly suitable for some types of applications such as *de novo* sequencing and metagenomics (Shendure and Ji, 2008).

The latest version of this tool, known as *GS FLX Titanium XL*, can produce long reads up to 1000 bp with a throughput of 700 Mb per run (http://454.com/products/gs-flx-system/index .asp).

Despite the technological development promoted in recent years, this system has a rather high error rate (estimated around an average value of 1%) and not equally distributed. The main sources of error are the presence of homopolymers, the length of the sequences produced and the spatial location of the spheres in the PTP plates (Gilles et al., 2011).

***Genome Analyzer* (*Illumina*).** Commonly referred to as *Solexa* by the name of the company that introduced it into the market in 2006, this platform is the result of Turcatti's work and colleagues on reversible termination sequencing (Fedurco et al., 2006; Turcatti et al., 2008) and the merger of four companies - *Solexa, Lynx Therapeutics, Manteia Predictive Medicine* and *Illumina*.

The sequencing technology used, and which is currently adopted by most NGS platforms, is known as sequencing by synthesis (SBS, *sequencing by synthesis*); it allows, by using dNTPs labeled with four different fluorochromes and using reversible chain termination chemistry, the identification of the individual bases of a DNA fragment as they are incorporated into the synthesis process of a new complementary strand.

In the initial phase the DNA is fragmented and at the ends of each fragment oligonucleotide sequences (adapters) are attached for the subsequent amplification process. The fragments of DNA, once denatured, are bound, by one of the two ends, to the surface of a slide, called *flow-cell*, on which are present the oligonucleotic sequences complementary to the adapters. The *flow-cell*, made up of eight independent lanes, constitutes the physical support on which all the reactions of the instrument take place. The free end of each fragment, hybridizing to the complementary oligonucleotide present on the surface of the slide, creates a characteristic "bridge" structure that starts the amplification of the fragments. Each filament is then cloned through the so-called "bridge PCR" (Adessi et al., 2000; Fedurco et al., 2006) which generates, for each template, approximately 1000 identical copies of DNA organized in circumscribed microareas of the *flow-cell* called clusters (Figure 4a).

Once the clusters have been generated, the amplicons produced come denatured and linearised and a sequencing primer acts as a trigger for the actual sequencing reaction (Figure 4b). During each cycle, in addition to the DNA polymerase, the four suitably modified dNTPs are added to the reaction: each of them is bound to a different fluorescent marker and at the end 3'-OH presents a terminator group of the reversible chain, so as to to prevent more than one nucleotide being added at a time for each sequencing cycle.

At each cycle a laser excites the fluorescent group bound to the built-in dNTP generating a light emission that allows its identification; once the base has been assigned, the terminator group present at the 3'-OH and the fluorochrome are removed, so as to allow the pairing, and hence the sequencing, of the subsequent base. The entire process is repeated until the desired number of cycles is reached to be able to sequence the entire length of the fragment of interest.

The first versions of the platform were characterized by a non-negligible error rate and very short reads (around 35 bp), but were able to obtain high throughput at low cost. The changes made to the technology and the protocols have significantly reduced sequencing errors, even if there are still biases (Nakamura et al., 2011), and increased the length of the reads up to 150 bp. In contrast to platform 454, the most frequent error is the replacement of a single base (Hutchinson, 2007), while insertion or deletion errors are rare.

Only a few years after the launch of the first platform, *Illumina* implemented *paired-end* sequencing: each fragment of the library is sequenced from both ends and in opposite directions. In this way the technique is more sensitive and accurate than the *single-end* one, in which the fragments are sequenced starting from only one of the two ends, since it greatly facilitates subsequent alignment operations with the reference sequence.

The system proposed by *Illumina* is today the most widespread among NGS technologies, both thanks to the high accuracy of the data produced, and to the commercial availability of numerous sequencers suitable for different needs.

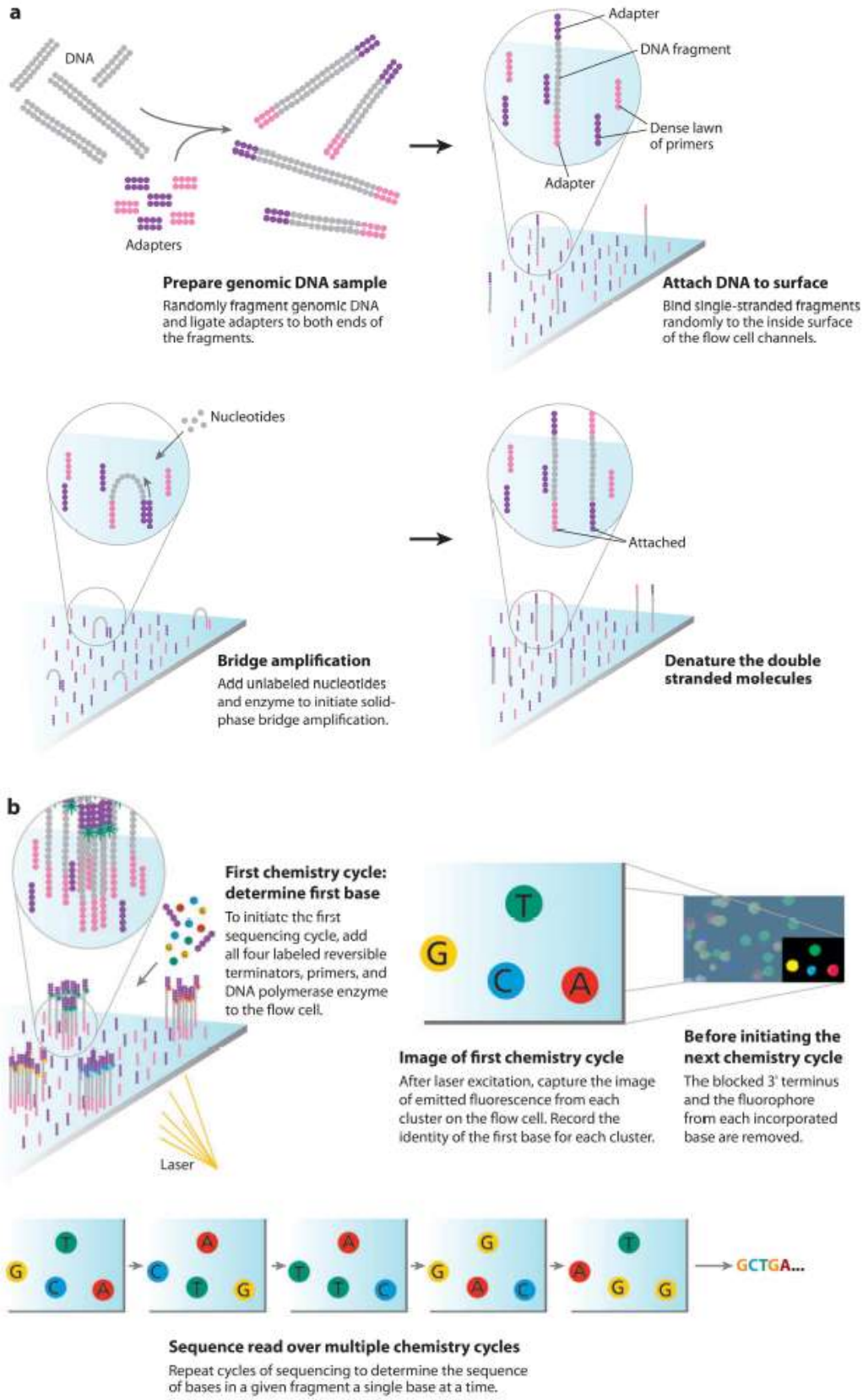


Figure 4. Sequencing technology by synthesis (Mardis, 2008).

SOLiD (Applied Biosystems). The *SOLiD* platform, introduced in the market since 2007, originates from the system described by Shendure and colleagues in 2005 (Shendure et al., 2005) and from the work of McKernan and colleagues (McKernan et al., 2006) at *Agencourt Personal Genomics* (acquired by *Applied Biosystems* in 2006).

As in system *454*, after the denaturing of the double helix and the addition of adapters at the ends, the DNA fragments are hybridized into small spheres by binding to complementary oligonucleotides that cover the surface. The balls, with a diameter of 1 μm, are immersed in an emulsion of water in oil together with the reagents for amplification; equally to the *454* system, the fragments are cloned by PCR in emulsion (Dressman et al., 2003).

Once amplification has occurred, the balls are deposited on the surface of a glass support which, unlike the PTP plate of *454*, does not have any wells: they covalently bind to the appropriately treated surface of the plate through the ends of the amplified filaments.

Unlike the previously described technologies that exploit the sequencing by synthesis, the *SOLiD* uses the sequencing by ligation (SBL) (Macevicz, 1998; McKernan et al., 2006; Shendure et al., 2005) from which the acronym *SOLiD* stems (*Sequencing by Oligonucleotide Ligation and Detection*). The sequencing reaction is triggered by the hybridization of a primer complementary to the adapter linked to the DNA mold (Figure 5) oriented in such a way as to allow the ligation to the 5 'end of appropriately constructed probes. Each probe consists of an octamer consisting of two specific bases followed by six degenerate bases with one of the four fluorescent markers bound to the 5 'end. The two non-degenerate bases correspond to one of the 16 possible combinations of base pairs. The last three octagon bases can be removed due to a cutting site between bases 5 and 6.

In the first sequencing cycle, by the action of the DNA ligase, the hybridization of the complementary probes to the DNA mold occurs and the consequent emission of a fluorescent signal that is detected. The marked portion of the ligated probes (last three bases) is cut away so as to regenerate a phosphate group at the 5 'end for bonding with the next probe (Figure 5). Seven ligation cycles are repeated in sequence to complete the extension of the adapter-bound primer. The newly synthesized filament is washed off and a new primer of phase shift sequencing of

a base (n-1) is hybridized with respect to the primer used in the previous step (n). At each passage new misaligned primers of a base are used with respect to the previous ones (n-2, n-3, etc ..., for a total of 5 different primers); this type of approach allows each base of the DNA mold to be sequenced twice.

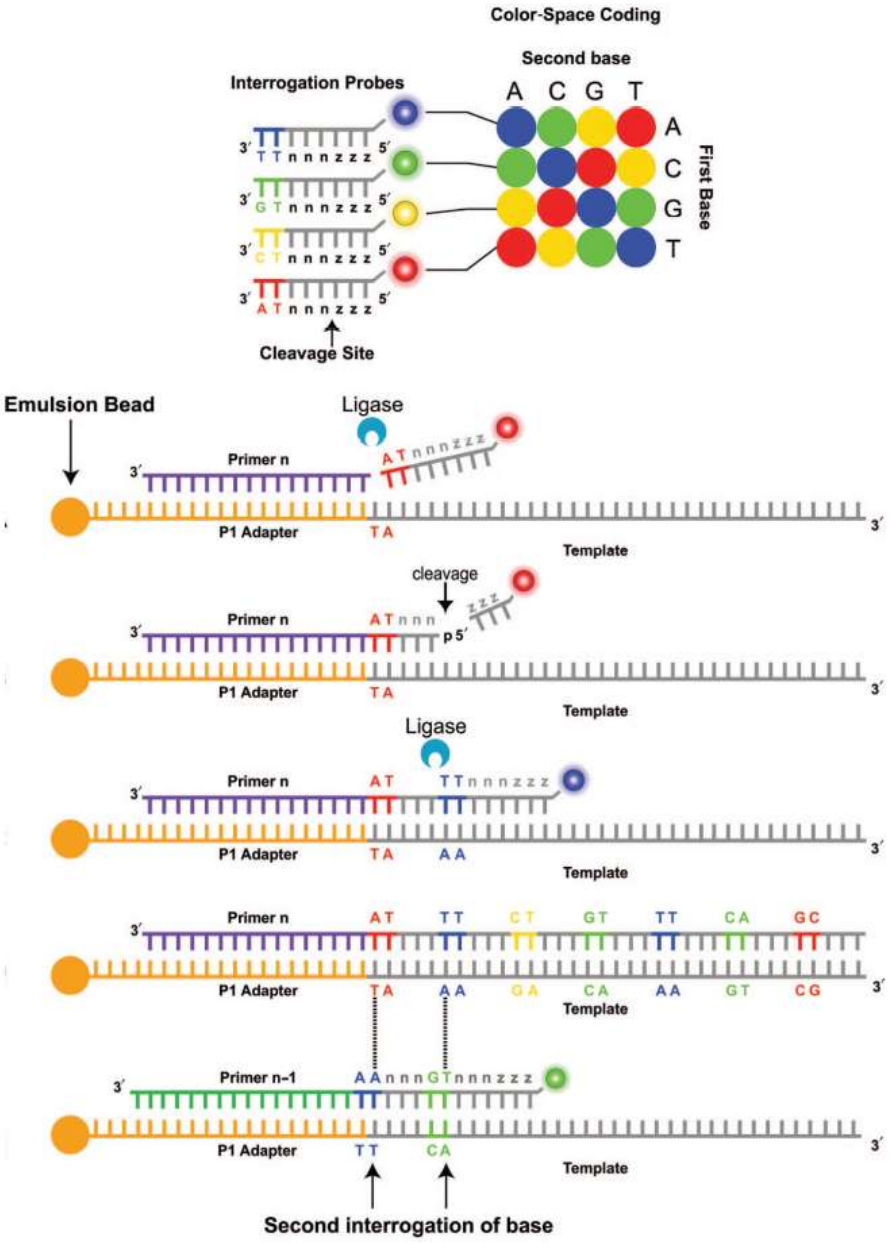


Figure 5. Method used by SOLiD technology (Voelkerding et al., 2009).

The nucleotide composition of the DNA template is interpreted on the basis of the emission of the fluorophores following the ligation of the various probes. The color emitted by each fluorophore allows to identify the couple of the first two bases, but it is not sufficient to distinguish one nucleotide from the other. In addition to the specific emission of each pair, it is necessary that one of the two bases of the sequence is known; this is possible because, thanks to the use of phase shift primers, different adapter bases are also sequenced. This information provides a starting reference sequence that is used in conjunction with a 2-base coding system (in which each pair of adjacent bases is correlated to a specific fluorophore) (McKernan et al., 2006) which allows to decipher algorithmically the DNA sequence mold (Figure 6).

In fact, given that each color represents four possible base pairs and for each of them the second nucleotide coincides with the first base of the next pair, knowing also only a base of the sequence allows to interpret the DNA sequence mold in its entirety (Figure 7).

The length of the reads generated by the *SOLiD* varies between 25 and 35 bp and each stroke of the instrument produces over 4 billion bases (4 Gb). Furthermore, thanks to two-base sequencing, accuracy is very high (around 99%).

Although the performance of the various technologies is constantly evolving and the data available are therefore very often conflicting, it is still possible to compare the main second-generation platforms available on the market based on certain parameters such as accuracy, length of reads and costs . Although each instrument presents unique characteristics suitable for certain applications, metagenomic experiments conducted with different platforms have shown that, within the limit of statistical errors, the results obtained are almost comparable (Luo et al., 2012).

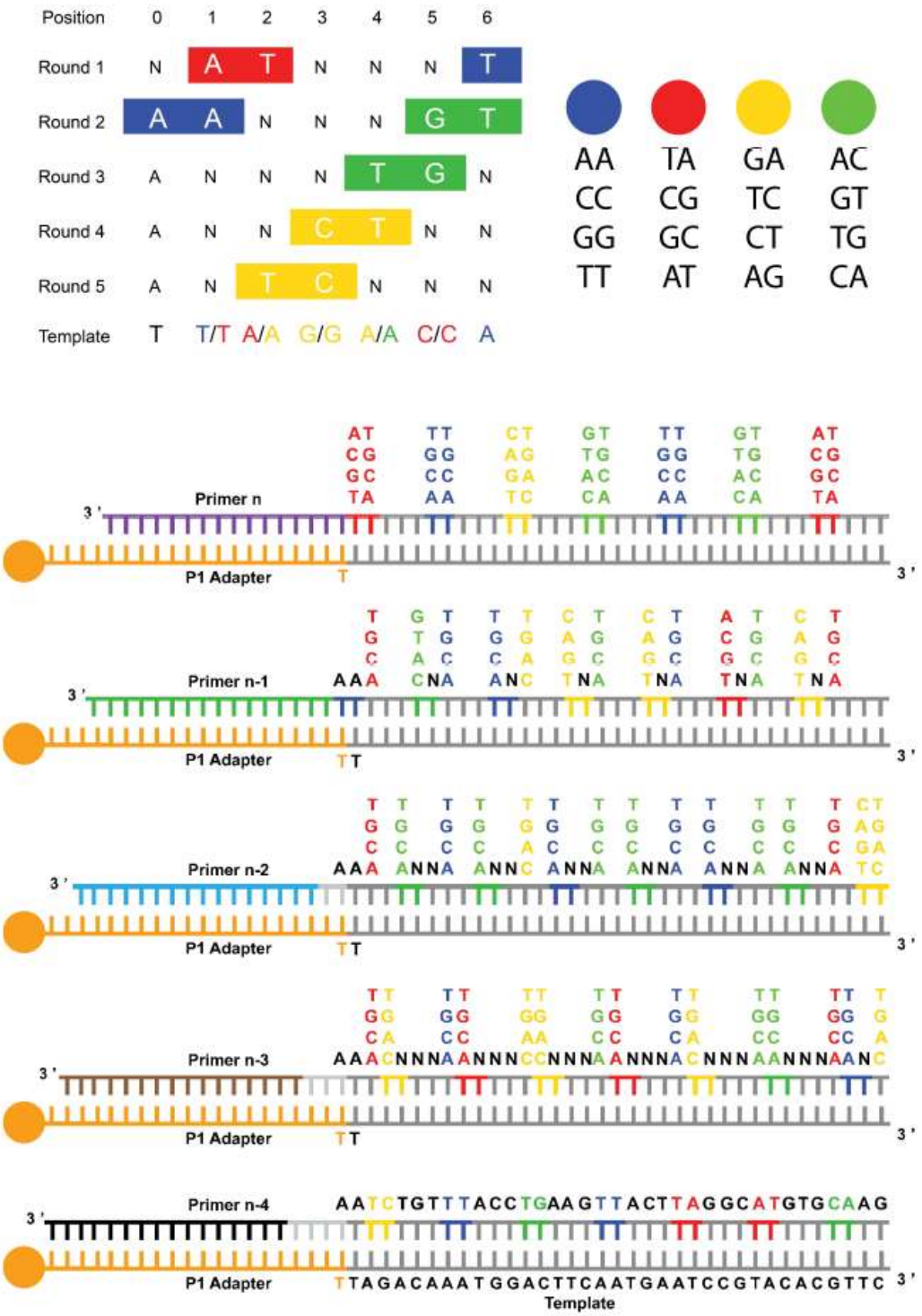


Figure 6. Sistema di codifica a due basi e ricostruzione statistica della sequenza di DNA stampo mediante tecnologia SOLiD. (**Voelkerding et al., 2009**).

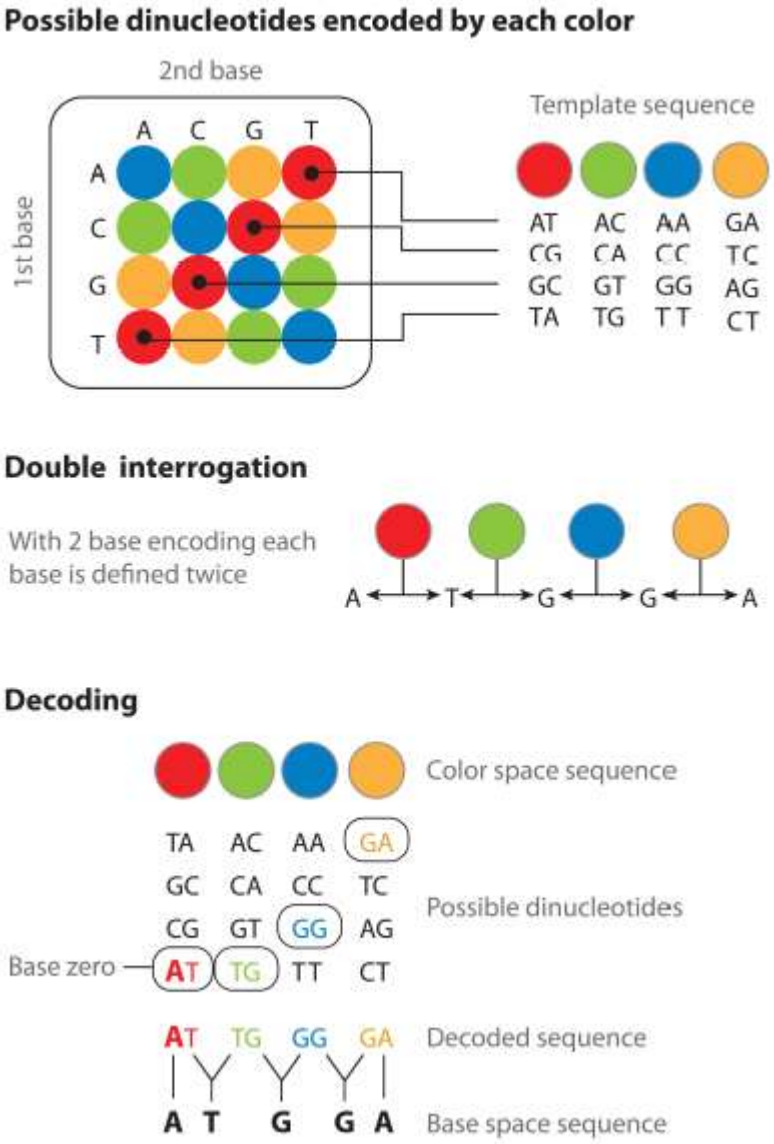


Figure 7. Two-base coding system.
The first base (A) and the order of the detected colors are supposed to be known. From the coding scheme it is clear that if the pair is identified by the red color and the first base is an A, then surely the second base will be a T. In the second pair, the first base is the newly identified T and the color of the pair is green: the second base will certainly be a G (Mardis, 2008).

3.1.4. Analysis of NGS data

The experiments conducted with the new generation sequencing platforms generate an unprecedented amount of information (in the order of 1012 bytes,

TB), representing an important challenge for the management, archiving and, above all, for the bioinformatics analysis of the data (Pop and Salzberg, 2008).

The NGS data analysis software has different characteristics depending on the specific technology of each platform, but on the whole they are based on bioinformatic analysis pipelines that convert luminescence or fluorescence images acquired by the instrument into nucleotide sequences, called sequence reads. The process, which takes the name of *base calling*, also includes the assignment of a quality score for each nucleotide of the read, called *Quality Score* (QS), which indicates the probability of error. The calculation of the *Quality Score* is based on the value of the historical *Phred quality score*, introduced in 1998 for the analysis of the sequence data of the Sanger method (Ewing et al., 1998; Ewing and Green, 1998). The *Phred quality score* (q) is calculated as $q = - 10 \bullet \log10 (e)$, where *e* represents the probability of incorrect identification of a base. Therefore a probability of identifying a base incorrectly equal to 0.1 (10%), 0.01 (1%) or 0.001 (0.1%) corresponds to a *Quality Score* of 10, 20 or 30 respectively .

The data concerning the nucleotide sequence of the reads and the *Quality Score* values associated to each nucleotide are organized into an output file in FASTQ format and are used in the subsequent analysis phase consisting in aligning the sequence reads produced by the instrument to the sequence of a reference genome. The alignment process is rather complex and, in order to be efficient and avoid incorrect alignments, requires sequence reads longer than 30 bp. It has been shown that while 97% of the E. *coli* genome can be uniquely aligned with 18 bp reads, 30 bp reads uniquely align only with 90% of the human genome (Whiteford et al., 2005; Warren et al., 2007). The alignment phase is possible thanks to the development and diffusion on the market of numerous software, each based on a different alignment algorithm; Examples include BWA (*Burrows-Wheeler Aligner*) software (Li and Durbin, 2009), *Bowtie*, MAQ (*Mapping and Assembly with Quality*) (Li et al., 2008), SOAP (*Short Oligonucleotide Analysis Package*) and SHRiMP (*SHort Read Mapping Package*). They also allow the subsequent assembly of the reads, in which the read and aligned sequences are combined to form the original gene sequence of the input sample. The results obtained from the alignment and assembly operations are stored in output files in SAM (*Sequence Alignment Map*) and BAM (*Binary Alignment Map*) format.

Once the gene sequence of the test sample has been reconstructed, the gene positions where the reads differ from it are identified from the comparison with the reference genome. The process, known as *variant calling*, identifies all the variants present in the sample under examination, reporting not only polymorphisms or mutations, but also any variants deriving from sequencing errors. One of the major problems of a *variant calling* operation is the ability of the software to discriminate between variants actually present in the sample under examination and variants due to sequencing artifacts or errors in the alignment phase.

The most widely used software for identifying and analyzing variants is the GATK (*Genome Analysis Toolkit*), distributed by the Broad Institute and used in two major projects such as the "*1000 Genome Project*" (http: //www.internationalgenome. org /) and the "*The Cancer Genome Atlas*" (https://cancergenome.nih.gov/).

Other software widely used in the *variant calling* phase are *SOAPsnp, VarScan* and *Atlas 2*.

The variants identified by the variant calling process are stored in a VCF (*Variant Call Format*) file which, in addition to reporting all the identified variants (single nucleotide variants (SNVs), insertions / deletions (indels) and other structural variants) , contains information related to the chromosomal position, the reading depth, the allele frequency, etc, of each of them.

Since an NGS analysis leads to the identification of a large number of variants (especially in the case of whole genome or exome sequencing experiments), it is necessary to determine which of these variants, not due to sequencing errors, could have a pathogenic meaning, thus reducing the amount of data to be analyzed. The *filtering* process makes use of some rather broad selection criteria, but effective in removing variants of little biological significance. Some criteria used in filtering operations are:

- the allelic frequency (the variants with a frequency equal to or higher than 1% are probably neutral polymorphisms and therefore can be excluded from the analysis);
- the mode of segregation of variants in the pedigree (all variants compatible with a given transmission model are selected);

- the comparison between healthy tissue and tumor tissue (if we want to identify the somatic mutations developed only in tumor cells, we proceed to the exclusion of the variants present in both tissues).

Once the filtering phase is completed, which has the advantage of considerably reducing the amount of data to be analyzed, we move on to the functional *annotation* phase of the variants: the analysis software provides all the data available in the databases and/or in literature, indicating whether they are polymorphisms, known pathogenetic mutations or entirely new variants and therefore of uncertain significance. The annotation phase is rather complex and, although it may be automatic, it is often still partially or totally manual; it takes into account numerous parameters such as, for example, the results of *in silico* predictions, the possible influence of the surrounding genomic context, any modulatory effects on the phenotype even in the absence of a direct pathogenic effect and, if existing, the data already published in literature.

The most commonly used algorithms for the annotation of the variants are SIFT (*Sorting Intolerant from Tolerant*) (Kumar et al., 2009), Polyphen-2 (*Polymorphism Phenotyping v2*), *ANNOVAR* (Wang et al., 2010), *SnpEff*, *SNPeffect* and VAT (*Variant Annotation Tool*).

3.1.5. Applications of NGS technologies

After the first platform was put on the market, NGS technologies have greatly accelerated the growth of various sectors of molecular biology, allowing to carry out experiments that were previously not technically possible or cost-effective from an economic point of view.

The main applications of NGS technology are described below.

De novo sequencing. *De novo* sequencing is the determination of the sequence of a genome for which no reference data is available.

The ability of NGS technologies to perform high-throughput analysis has been exploited to sequence entire genomes, from those of microorganisms to human ones (Sundquist et al., 2007; Chaisson and Pevzner, 2008; Cokus et al., 2008;

Durfee et al. ., 2008). As it is understandable, genomics is the area that has most benefited from the increase in throughput and the reduction of costs achieved with the new generation sequencing techniques. Suffice it to say that today the sequencing of the human genome would take a few days against the 10 years that were necessary for obtaining the first draft with the Sanger method (International Human Genome Sequencing Consortium, 2001).

Sequencing of the transcriptome (RNA-Seq) or microRNAs. RNA sequencing provides quantitative information on differences in gene expression levels in different tissues or in the same tissue analyzed under different experimental conditions (for example, subjected to different treatments) or in different stages of development. Compared to microarray of gene expression that allow the analysis of only known gene sequences, the *RNA-Seq* technology is advantageous because it allows to characterize the transcripts of a gene without a priori knowledge of the sites of the beginning of the transcription (Wang et al., 2009). The technique also shows a better ability to distinguish the different RNA isoforms, to determine the allelic expression and to detect sequence variations.

This technique has many applications and has been used to study the transcriptomic profile of stem cells (Cloonan et al., 2008), to study alternative splicing in human cells (Sultan et al., 2008), as well as for transcriptomic studies in numerous organisms, including Saccharomyces *cerevisiae* and Arabidopsis *thaliana* (Lister et al., 2008; Nagalakshmi et al., 2008).

Resequencing. It is the re-sequencing of an entire genome already sequenced (or of a specific portion thereof) that is carried out in order to identify possible sequence variations, such as point mutations, insertions, deletions and variations in the number of copies (CNVs, *Copy Number Variants*). This technique has been successfully applied to the study of human diseases, allowing the identification of new gene *loci* involved in their etiopathogenesis and the understanding of the mechanisms of disease development.

In addition to the whole genome (*whole genome sequencing*, WGS), NGS platforms can be used to selectively sequence specific genomic regions or specific

genes (through *targeted resequencing* or *amplicon sequencing* techniques) or to obtain information on the whole exome (via *whole exome sequencing*, WES).
This latter approach has proved particularly useful for the search for candidate genes in polygenic and multifactorial diseases.

Metagenomics. Metagenomics is a branch of genomics that deals with the study of microbial communities directly in their natural environment, thus avoiding the problem of the cultivation of individual microbial species on selective media.
NGS technologies have had a huge impact on metagenomics studies, allowing access to information that is enormously greater than previously obtained. The possible applications are innumerable and include the characterization of human microbial communities for clinical purposes and the study of environmental microbial populations in order to identify any genes useful for commercial purposes such as the production of biofuels or other pharmaceutical or agrochemical compounds (Schloss and Handelsman, 2003; Jaenicke et al., 2011).

ChIP Sequencing (ChIP-Seq). The characterization of the regulatory proteins associated with DNA has long availed itself of the use of a technique based on chromatin immunoprecipitation (ChIP, *Chromatin ImmunoPrecipitation*) followed by hybridization to a microarray (chip) (*ChIP-on-chip*) (Ren et al., 2000).
Thanks to the introduction of the new generation sequencing platforms, *ChIP-on-chip* technology has been progressively replaced by *ChIP-Seq* approaches, in which the immunoprecipitated DNA is converted into a fragment library subjected to NGS sequencing (Barski et al., 2007; Schones and Zhao, 2008). The *ChIPSequencing* studies conducted to date indicate that this type of technology has many advantages compared to the classical *ChIP-on-chip* approach because it allows to obtain a better resolution, as evidenced by the identification of new binding sites, and why, like *RNA-Seq*, it does not require a priori knowledge of the genomic regions in which proteins are bound (Johnson et al., 2007, Robertson et al., 2007).

3.2. Mitochondrial diseases

Mitochondrial diseases are a group of inherited disorders that may derive either from alterations in mitochondrial DNA (mtDNA) (Wallace et al., 1999) (Figure 8) or from mutations in nuclear genes encoding proteins that participate in oxidative phosphorylation.

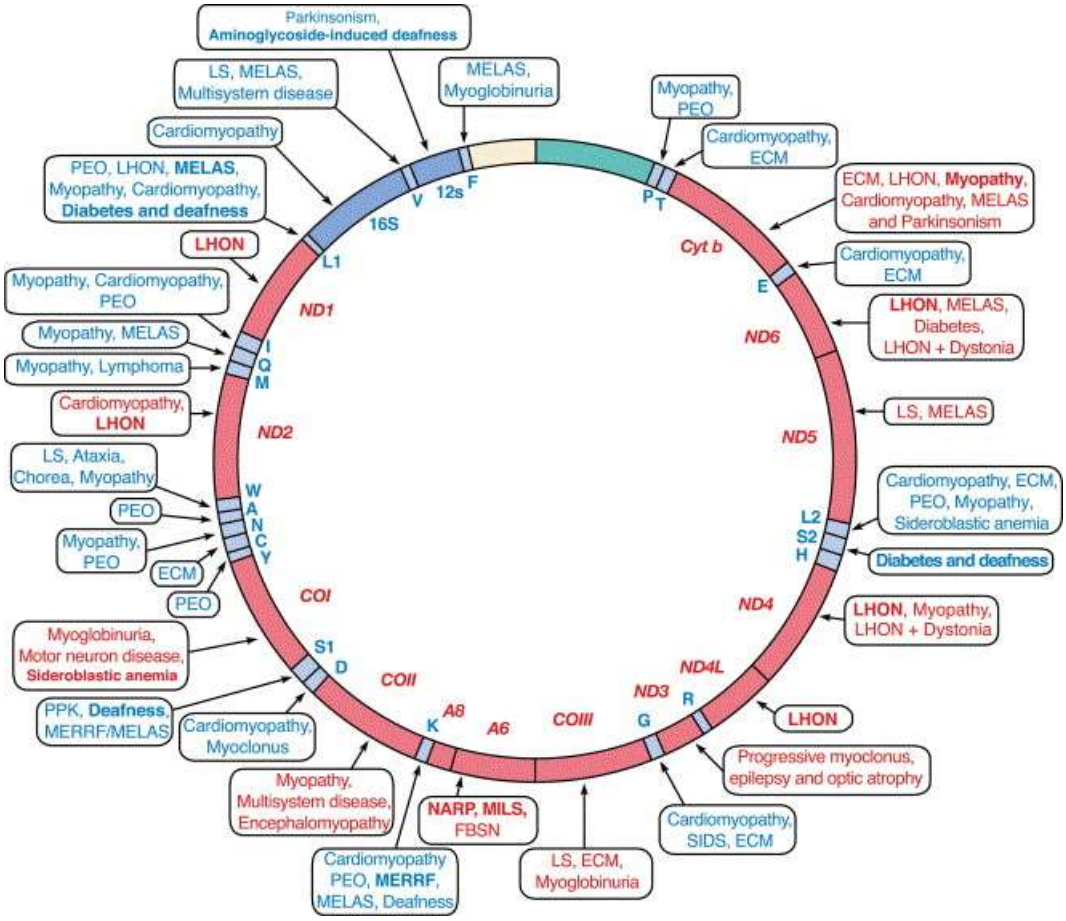


Figure 8. Map of human mitochondrial DNA. Diseases due to mutations that impair mitochondrial protein synthesis are shown in blue; diseases due to mutations in protein-coding genes are shown in red (Di Mauro, 2004).

These pathologies are very heterogeneous from the clinical point of view; since mitochondria are present in all tissues, mitochondrial diseases can affect any organ but the most affected are the tissues with the greatest energy demand, including muscle, heart, brain and retina. Unlike nuclear DNA (nDNA), mutations in mtDNA are inherited through the maternal route and the high variability of the clinical phenotype between individuals of the same family or between different

families carrying the same mutation is mostly due to phenomenon of heteroplasmy, or the coexistence of wildtype and mutated mtDNA molecules.

Phenotypic changes, at the level of proteins, ribosomes or tRNAs encoded by the mitochondrial genome, can potentially alter cellular energy production. Such point mutations are believed to lead to a faulty ("leaky") phenotype, in which the altered protein or tRNA still function, albeit partially (Scheffler, 1999).

However, the reduction in efficiency of oxidative phosphorylation (OXPHOS) can have significant medical consequences that can occur differentially over time (some point mutation diseases are, in fact, at a late onset) or between tissues (Scheffler, 1999). Variations in the metabolic structure can in turn induce apoptosis of the cells, since the mitochondrial permeability plays a central role in the pathway of apoptotic cell death (Liu et al, 1997).

There are over one hundred mitochondrial genetic diseases, defined, or associated or implicated in mtDNA mutations (Kogelnik et al, 1998). The MitoMap website contains a list, which is continuously updated and implemented, of mutations of the disease-associated / implicated mtDNA (http://www.mitomap.org).

Human mitochondrial DNA (mtDNA) is a circular double-strand molecule of about 16.6 kb. It encodes 2 rRNAs, 22 tRNAs, and 13 polypeptides that are all subunits of complexes of the respiratory chain/oxidative phosphorylation system that drives oxidative energy metabolism.

3.2.1. Coenzyme Q deficiency syndrome

Mitochondrial diseases are gathered in sindrome. One of them is the coenzyme Q deficiency syndrome.

It was first described in 1989 but only in the last decade the molecular bases of this disorder have been elucidated (Doimo et al., 2010). Patients with this biochemical phenotype may be classified in two groups, those with primary deficiency harbour mutations in one of the genes involved in the biosynthesis of CoQ10 while with secondary deficiency is associated to mutations in genes unrelated to the CoQ10 biosynthetic pathway or to non-genetic causes (Acosta et al., 2016).

Coenzyme Q (ubiquinone, Q or CoQ) is a lipophilic molecule ubiquitously present in cell membranes, but especially abundant in mitochondria (Turunen et al., 2004). It is comprised of a quinone group and of a polyisoprenoid tail of variable length in different species: yeast have six units (CoQ6), mice nine (CoQ9) and humans ten (CoQ10).

In mammalian the precursor of the quinone ring is only 4HB, which is derived from tyrosine through an uncharacterized set of reactions. The isoprenoid tail is synthesized through the mevalonate pathway, which is common also to cholesterol biosynthesis (Bentinger et al., 2010). The mevalonate pathway comprises the reactions that starting from acetyl-CoA produce farnesyl pyrophosphate (FPP). The terminal steps in CoQ biosynthesis are thought to be rate limiting for the process in eukaryotes and take place in the mitochondrial matrix (Figure 9).

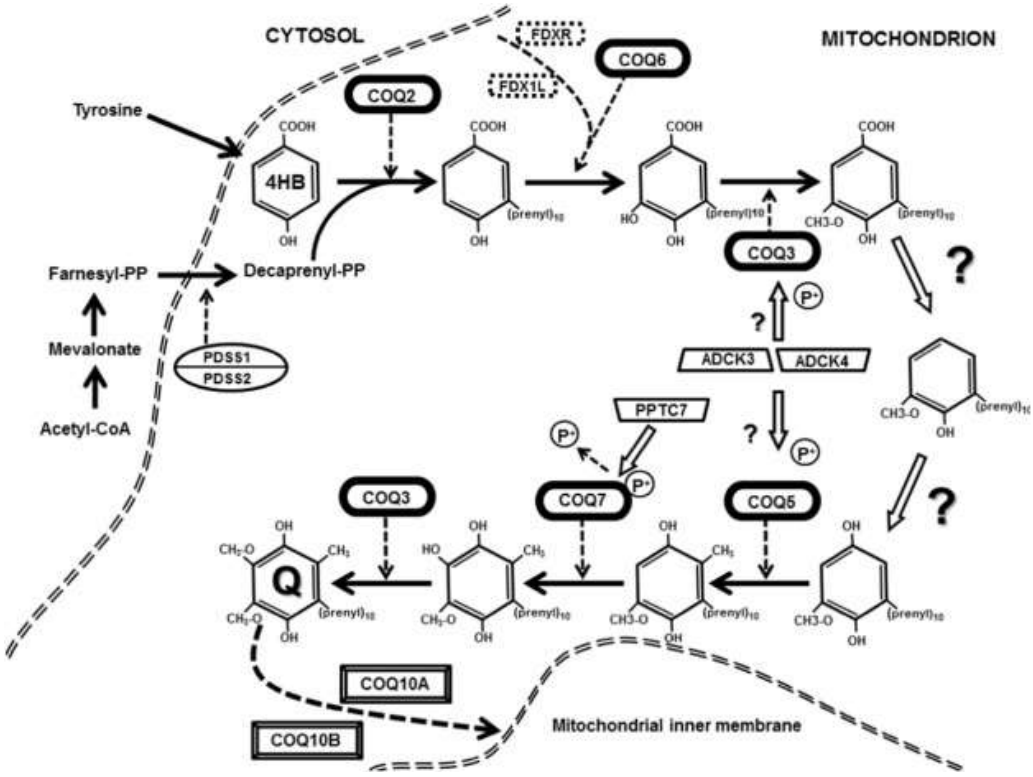


Figure 9. The coenzyme Q biosynthetic pathway in mammalian cells (the question marks indicate the still uncharacterized enzymatic steps)(Acosta et al., 2016).

Coenzyme Q plays an essential role in energy metabolism due to its redox properties. In the respiratory chain it transfers electrons from NADH: coenzyme Q

reductase (complex I) and succinate: coenzyme Q reductase (complex II) to coenzyme Q: cytochrome c reductase (complex III). It is also an electron acceptor starting from the β-oxidation of fatty acids (Frerman, 1988), from the synthesis of uridine (Nagy at al., 1992) and from the oxidation of sulfide to sulfate (Hildebrandt et al., 2008).

The deficit of coenzyme Q is associated with a heterogeneous group of predominantly infantile onset diseases. It is difficult to propose a comprehensive clinical classification, not only because of the marked clinical differences among different genes, but also because of the extremely wide spectrum of clinical manifestations among patients with mutations in individual genes. However, it is possible to to classify the genetic defects of the CoQ10 biosynthetic pathway within three different groups (Acosta et al., 2016):

- *PDSS1*, *PDSS2, COQ2, COQ6* and *ADCK4* genes mutations: these defects share common features and are associated with glomerular renal involvement manifesting as steroid resistant nephrotic syndrome (SRNS);.
- *COQ4, COQ7* and *COQ9* genes mutations: these patients never display SRNS, but the main clinical feature is encephalomyopathy. Other manifestations include hypertrophic cardiomyopathy, lactic acidosis, and (if renal involvement is present) tubulopathy (Emma et al., 2016; Brea-Calvo et al., 2015; Chung et al., 2015);
- *ADCK3* gene mutations: the main clinical feature is cerebellar ataxia. Interestingly these patients have essentially central nervous system involvement (other common symptoms are seizures, dystonia, cognitive impairment), but have virtually no other extra-CNS manifestations, despite a reduction of CoQ10 also in other tissues (Lagier-Tourenne et al., 2008; Mollet et al., 2008).

The treatment of the symptomatology provides for the administration of oral supplements of coenzyme Q: 10-30 mg/Kg/day for child and 1000-3000 mg/day for adults (Horvath et al, 2012). In most cases the treatment is carried out with the oxidized form of coenzyme Q which has a low bioavailability. But more recently a treatment with reduced form has been achieved, where the reduced form has shown greater bioavailability.

Deficit of coenzyme Q associated with diseases such as Parkinson's have been described (Shults at al., 2005). In this disease and in other neurodegenerative diseases such as Alzheimer's, Huntington's disease or Friedreich's ataxia, oxidative damage and mitochondrial dysfunction occur, so the use of coenzyme Q as an antioxidant is becoming increasingly relevant (Beal et al.,2004).

A decrease in the amount of coenzyme Q causes the increasing of the production of ROS and oxidative stress that is known is associated with aging (Beyer et al., 1985; Kalen et al., 1989), so it is believed that a supplementation with Q enzyme favors the delay of senescence.

3.2.1.1. *COQ4* gene

The *COQ4* gene encodes is located on chromosome 9q34.13, and is transcribed into a 795 base-pair open reading frame, encoding a 265 amino acid (aa) protein (Isoform 1) with a predicted N-terminal mitochondrial targeting sequence. COQ4 protein has no known enzymatic function, but may be a core component of multisubunit complex required for CoQ biosynthesis.

The human transcript is about 1.4 kb and is expressed ubiquitously, but at high levels in liver, lung, and pancreas. Transcription initiates at multiple sites, located 333–23 nucleotides upstream of the ATG. A second group of transcripts originating inside intron 1 of the gene encodes a 241 aa protein, which lacks the mitochondrial targeting sequence (isoform 2). Expression of GFP-fusion proteins in HeLa cells confirmed that only isoform 1 is targeted to mitochondria (Casarin et al., 2008). The functional significance of the second isoform is unknown.

3.2.1.2. Splicing regulation

The pre-mRNA splicing is a critical event for the eukaryotic genes expression: it is a complex process that allows the junction of the proteins coding parts (exons), which are separated by non-coding sequences (introns).

It is considered one of the main mechanisms that allows the generation of different proteins from the same gene.

Knowledge about complexity of the splicing process revealed the existence of important elements of regulation (Matlin et al., 2005). Variations of these elements, which are found in both coding and non-coding regions of genes, can manifest themselves with deleterious effects on the pre-mRNA splicing.

The correct maturation of the mRNA requires the assembly of a ribonucleoprotein complex (the spliceosome) resulting from the coordinated action of 5 small snRNPs (small nuclear ribonucleoproteins: U1, U2, U4, U5 and U6) and more than 60 polypeptides (Will et al., 2011).

The splicing regulatory elements are:

- in cis elements: splicing site at 5' (donor site), site of splicing at 3' (acceptor site), branch site, enhancers (ESEs), silencers (ESSs), intronic-splicing processing element (ISPE)(Figure 10);
- in trans elements: UsnRNPs (small nuclear ribonucleoproteins), SR (serine/arginine rich proteins), HnRNPs (heterogeneous nuclear ribonucleoproteins) (Wang and Burge, 2008).

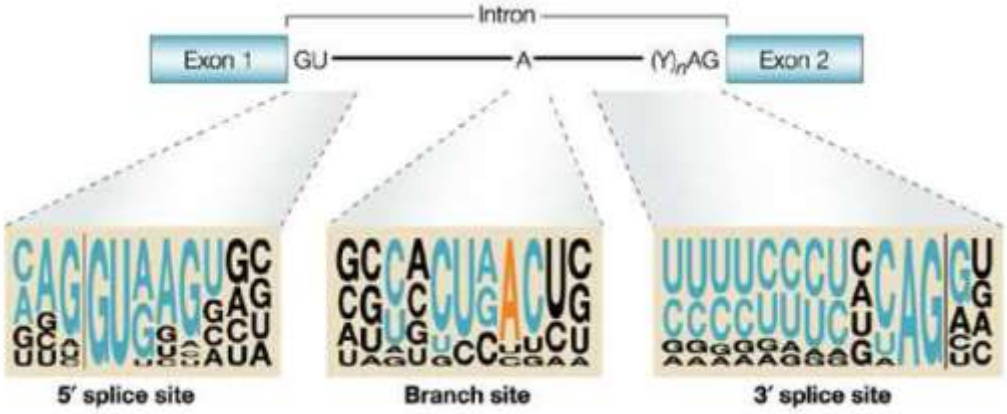


Figure 10. The GU bases of the donor site and AG bases in the acceptor site are highly conserved; the base A of the branch site is instead variable.
The nucleotides belonging to the are highlighted in blue
more frequent consensus sequences (Cartegni et al., 2002).

A wrong recognition of the splicing sites between exon and intron or failure removal of one or more introns leads to the formation of aberrant mRNA that can be unstable and therefore be degraded or can lead to the formation of defective protein isoforms or even harmful.

Summarizing, the correct separation of the exons from the introns into a determined mRNA depends on two factors:

- the intrinsic strength of splicing sites determined by sequences flanking GU and AG;
- the combination of positive and negative effects due to presence of enhancers and silencers.

It is known that point mutations can cause alterations of the splicing. One of the most frequent effects is the skipping of the exon in proximity of the mutation. In other cases, the mutation can determine the retention of an intronic fragment or activate a "cryptic site" of splicing causing profound modifications of the transcript.

3.3. Hereditary metabolic diseases and neonatal screening

Neonatal screening is a public health activity aimed at early detection of some congenital diseases. The pathologies subject of neonatal screening, if not recognized early, can cause damage often irreversible, especially in the central nervous system with consequent severe disability. The identification of these diseases in the first days of life is essential to intervene in time and to avoid serious consequences on health in the newborn. In fact, the early diagnosis of these diseases allows a pharmacological and/or dietetic therapeutic intervention aimed at preventing possible harm to the baby's organism.

Neonatal screening is one of the most important public preventive medicine programs. For many of these diseases the treatments carried out after the appearance of clinical signs and after episodes of metabolic decompensation are not effective and are not able to normalize the clinical picture.

The neonatal screening program provides that all newborns are submitted free of charge to simple tests performed on a few drops of blood taken from the heel of the newborn in the first days after birth. Both the basic screening for 3 pathologies, currently mandatory by law, and the screening extended to more than 40 diseases can be performed on the same blood sample. The withdrawal is done

by the Birth Center, before discharge, and the blood sample is sent to the reference Neonatal Screening Centers.

Diseases subject to neonatal screening are inherited endocrine and metabolic diseases which, although singularly rare, reach an incidence of about 1:2000 births. Hereditary metabolic diseases are genetic diseases that cause the alteration of a protein involved in the metabolic processes indispensable for the life of the cell such as the production of energy, the elimination of toxic compounds, the biosynthesis of compounds essential for its functioning. The onset and severity of symptoms depend mainly on the extent of the enzyme defect and on the toxicity of the accumulated metabolites. So it must be considered that hereditary metabolic diseases are complex diseases and the phenotypic expression for some of them can cover a spectrum ranging from mild to more severe forms with a variability of onset, if not treated, which can occur in the early months of life or at a more advanced age.

The decision-making criteria for determining whether a disease has the characteristics to be included in neonatal screening still refer to those defined in 1968 by Wilson and Jungner, but the discoveries made in recent decades in the field of medical genetics make this topic still open and debated.

In Italy neonatal screening is currently mandatory throughout the country for 3 diseases: phenylketonuria, congenital hypothyroidism and cystic fibrosis. The introduction of extended neonatal screening (SNE, also known as enlarged neonatal screening), marked an epoch-making turning point unthinkable until the 90s, years in which tandem mass spectrometry was introduced in neonatal screening: it was then passed from the basic screening for 3 pathologies to the screening for more than 40 pathologies. This profound change has also marked a turning point in the clinical history of the pathologies involved, which have finally found an instrument of effective treatment in early diagnosis.

3.3.1. Mucopolysaccharidosis type 1

Mucopolysaccharidosis type 1 (MPS1) is a rare autosomal recessive disease caused by mutations in the *IDUA* gene, coding for alpha-L-iduronidase, which

causes gradual accumulation of glycosaminoglycans (GAGs) in all organs and tissues (Lorne et al, 2016). This accumulation leads to multisystem clinical manifestations of varying severity.

Historically we recognize three main syndromes that represent the whole spectrum of the disease: Hurler (OMIM #607014) (more severe clinical phenotype), Hurler-Scheie (OMIM #607015) (intermediate clinical phenotype) and Scheie (OMIM #607016) (phenotype less severe clinical). These distinctions were made on the basis of the age of appearance of the first symptoms, of the rapidity of progression of the pathology and of the pre-eminent symptomatology.

The incidence is 1/100,000: Hurler's syndrome represents 57%, Hurler-Scheie's syndrome 23% and Scheie's syndrome 20%. In severe form (Hurler syndrome or MPS I-H), the main symptoms are skeletal deformities and psychomotor retardation. It begins at the age of 6-8 months. Other clinical signs are corneal opacity, organomegaly, cardiac malformations, short stature, hernias, facial dysmorphism and hirsutism. Radiological examination of the skeleton shows typical signs of multiple dysostosis. Patients with the intermediate form (Hurler-Scheie syndrome or MPS I-H/S) have normal or near-normal intelligence, but have varying degrees of physical disability (Wraith et al, 1987). The first diagnosis is based on the finding of an increase in heparansulphate and dermatan sulfate levels in the urine and is confirmed by the measurement of the activity of the enzyme alpha-L-iduronidase in the blood.

Genetic analysis is also available, with a search for mutations in the *IDUA* gene. In pregnancies from couples in which the genetic defect in the parents has already been identified, prenatal diagnosis can be performed by villocentesis or amniocentesis. The earlier the diagnosis is, the greater the chances that the available therapeutic interventions will be effective: for this reason, pilot studies for the neonatal screening of this pathology are already under way in Italy (Burlina et al, 2018).

3.3.2. Biotinidase deficiency

Biotinidase deficiency (BTD) (OMIM # 253260) is a congenital defect of biotin metabolism that, in the classical form is characterized by convulsions, respiratory disorders, hypotonia, skin rash , alopecia, hearing loss and developmental delay (Sweetman et al, 1981). It is a rare disease with autosomal recessive transmission, due to mutations in the *BTD* gene which cause a (virtually) complete inactivation of the enzyme. The incidence of the clinical deficit of biotinidase is estimated at 1/61,000 and the frequency of carriers in the general population is about 1/120.

The treatment of the symptomatology provides for the administration of oral supplements of biotin in free form, not related to proteins (Suormala et al, 1990). This treatment is given throughout life and no significant side effects are known. Some signs, such as optic atrophy, hearing loss or developmental delay, are not reversible with biotin treatment. The prognosis of patients with BTD deficiency is very good, if they are treated before the onset of symptoms and adhere to the intended treatment.

A more frequent form is characterized by a partial defect. These patients usually are compound heterozygotes for a severe *BTD* mutation and for a relatively common p.Asp444His polymorphism (MAF 4% in Europeans). These patients display a reduction of BTD (in the range of 20-25% residual activity), are usually asymptomatic, but during periods of stress, such as illness, fever or fasting, they may develop symptoms similar to those of patients with the classical form. One of the main open issues is whether these patients should be treated with biotin or not. Individuals homozygous for the p.Asp444His polymorphism do not display symptoms and do not require therapy.

BTD deficiency is identified with neonatal screening when available. The other cases are diagnosed based on clinical signs and symptoms and are confirmed by demonstrating the defect of BTD activity in serum. The search for mutations in the BTD gene is also possible.

3.3.3. Phenylketonuria

Phenylketonuria (PKU) (OMIM # 261600) is the most common disease of amino acid metabolism which, due to lack of treatment, causes mental retardation of varying degrees (from mild to severe).

The incidence of PKU has marked geographical differences and in Europe it is estimated to be 1/10,000 live births with a higher rate in some countries such as Ireland and Italy. The incidence is particularly high in Turkey where it has been calculated in 1/4,000 live births. PKU is much rarer in Finnish, African and Japanese populations.

PKU is an autosomal recessive transmission pathology caused by a wide range of mutations in the PAH gene (12q22-q24.2), which codes for phenylalanine hydroxylase, an enzyme that catalyzes the hydroxylation of phenylalanine to tyrosine, a key enzymatic reaction of the phenylalanine catabolism.

The many variants of classical PKU are due to the heterogeneity of mutant alleles, since many patients are compound heterozygotes rather than homozygous for a particular mutant allele (Bartholome at al, 1984). The reduced or absent activity of this enzyme results in a neurotoxic effect from hyperphenylalaninemia (Zurfluh et al, 2008).

The normal levels of phenylalanine in the blood are 58 +/- 15 micromoles/liter in adults, 60 +/- 13 micromoles/liter in adolescents and 62 +/- 18 micromoles/liter (mean +/- SD) in infancy. In the newborn, the upper limit of the norm is 120 micromoles/liter (2 mg/dl) (Scriver et al., 1985; Gregory et al., 1986). In the untreated classical PKU, blood levels of up to 2.4 mM/liter can also be found. Pediatric patients also have low plasma tyrosine levels (Hanley et al, 2000).

The treatment of choice for the forms that require therapy consists in the diet with a low content of phenylalanine and mixed amino acids.

3.3.4. Urea cycle disorders

Urea cycle disorders (UCDs) have an incidence of about 1: 30,000 live births and are caused by changes in hepatocyte metabolism induced by a deficiency of one of the enzymes involved in the transfer of nitrogen from ammonia to urea, through the UC. UDCs are genetic diseases inherited in an autosomal recessive manner,

with the exception of the OTC deficit which is an X-linked disease. These defects may present an acute picture characterized by rapidly progressive neonatal onset that leads to coma and death if a targeted treatment is not promptly started; the milder forms can manifest themselves in childhood; asymptomatic patients in adulthood have also been described.

As with most enzyme defects, the pathogenicity of UCDs (Table 1) can be attributed to the accumulation of metabolites upstream of the block and / or the lack of essential nutrients downstream of the block (Mian and Lee, 2002), but also to the lack of different intermediate compounds.

Urea cycle disorder	Gene
CPS deficit	*CPS*
OTC deficit	*OTC*
Citrullinemia	*ASS1*
Argininosuccinic aciduria	*ASL*
Argininemia	*ARG*

Table 1. UCDs and involved genes.

However, recent data indicate that the pathogenesis of these diseases is more complex.

Some of the urea cycle enzymes (ASS1 and ASL) are ubiquitously expressed and are involved in the production of arginine (Figure 11). Arginine is the precursor of important biologically active molecules, including nitroxide (NO), polyamines, creatine, proline and glutamate. NO has different physiological functions: it is a potent vasodilator, it is involved in neurogenesis and nerve transmission. Alterations of the L-arginine-NO pathway may result in the reduction of NO levels. In some UCDs, for example in argininosuccinic aciduria, it has been shown that the pathogenesis of some chronic complications is related not only to hyperammonemia, but also to a tissue-specific deficiency of arginine and therefore NO (Nagamani et al., 2012).

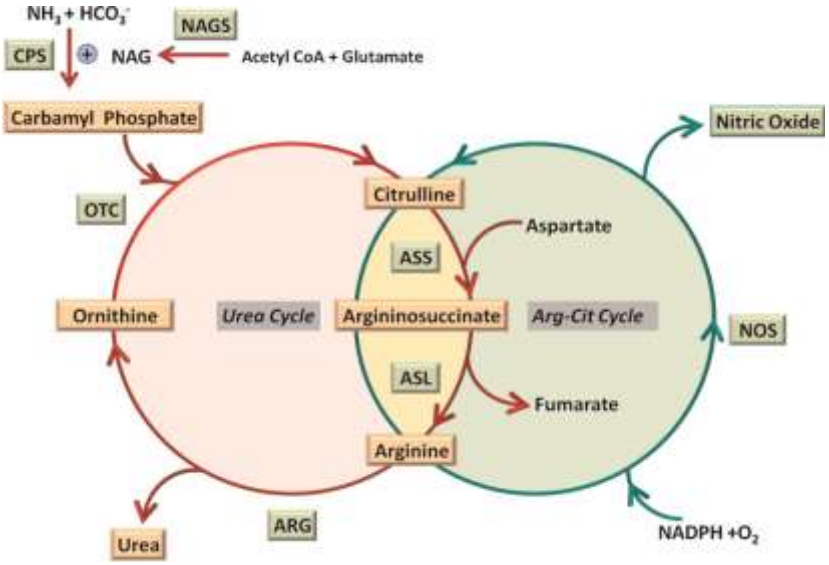


Figure 11. Arginine-citrulline cycle. (Erez et al., 2011).

3.3.4.1. Citrullinemia type 1

Citrullinemia includes two different conditions: citrullinemia type 1 (CTLN1; MIM #215700) and citrullinemia type 2 (CTLN2; MIM #603471).

CTLN1 is due to deficiency of the enzyme argininosuccinate synthase due to mutations in the *ASS1* gene, otherwise CTLN2 is caused by Citrin deficiency, a mitochondrial carrier aspartate/glutamate Ca_2^+-dependent (Kobayashi et al., 1999), due to defects in the *SLC25A13* gene (7q21.3).

Citrin forms a complex with ASS1 and the other two cytoplasmic soluble enzymes of the urea cycle: mutations in the carrier can cause destabilization and / or degradation of this complex, with consequent reduction of the hepatic levels of argininosuccinate synthase (Saheki et al. , 2004).

The classical form of citrullinemia, known as type 1, is an autosomal recessive transmission pathology caused by mutations affecting the *ASS1* gene and is the most common defect among UCDs with an incidence between 1: 44,300 and 1: 200,000 born live (Woo et all., 2012).

The clinical picture of this pathology is highly heterogeneous, both as regards the age of onset and the severity of the symptoms. It is possible to distinguish two main phenotypes: an acute neonatal onset and a later onset subclinical one. Arginosuccinate synthase deficiency causes variable degree of hyperammonemia,

increased citrulline levels in body fluids (> 1,000 μmol / L, normally <50-60 μmol / L), orotic aciduria, increased and decreased levels plasma glutamine and arginine respectively (Woo et al., 2012). The clinical manifestations of the acute form are characterized by: loss of appetite, vomiting, tachypnoea, hyperammonaemia, lethargy, hypothermia and convulsions with subsequent progression to coma and death, if not treated. The late form may instead present with nonspecific symptomatology, which includes psychomotor retardation, convulsions, sometimes slowing of growth, irritability and behavioral disturbances.

The introduction in many countries of expanded neonatal metabolic screening programs with mass spectrometry has also allowed to identify asymptomatic patients with high levels of citrulline (Sander et al., 2003); in these cases the molecular analysis identified mutations in both alleles of the *ASS1* gene (Haberle et al., 2003). The diagnosis of type 1 citrullinemia is based on biochemical criteria and is established during the acute phase of the disease by quantifying the plasma levels of citrulline (> 1000 μmol / L, normally <60 μmol / L), ammonium and glutamine (elevated), arginine (low), argininosuccinic acid (absent) and increased urinary acid excretion in the urine (Haberle et al., 2003). The clinical signs described above are indicative of the enzymatic defect of argininosuccinate synthase. The biochemical measurement of ASS1 activity can be performed by indirect enzymatic assay on cutaneous fibroblasts or hepatic tissue by incubating cells with 14C-citrulline and measuring the incorporation of radioactivity into proteins via arginine (Kleijer et al., 1984).

Molecular genetic analysis is performed by sequencing the *ASS1* gene, a fundamental test also for prenatal diagnosis and for genetic counseling for the carrier individuals. Rapid diagnosis and immediate treatment are essential elements to prevent neurological damage induced by prolonged hyperammonaemia. Therapeutic strategies aim to limit the exogenous and endogenous sources of nitrogen-containing compounds and to create alternative pathways for its excretion. This result is obtained by following a diet with reduced protein intake and with the prevention of catabolic stress that involves the mobilization of amino acids from the muscle. An alternative therapy is based on the use of scavenger molecules, such as sodium benzoate or sodium phenylbutyrate, capable of complexing with glycine and glutamate (nitrogen

donors in the urea cycle) favoring their renal excretion (Auron et al., 2012). In patients with ASS1 deficiency, the diet is also supplemented with arginine.

3.3.4.2. *ASS1* gene

The human gene *ASS1* (MIM #603470) is located on chromosome 9q34.11 where it occupies about 63 kb. It includes 16 exons: the start codon is located in the 5' region of exon 3 and the termination codon in exon 16.

Analysis of human genomic DNA demonstrated the presence of 14 copies of the argininosuccinate synthase gene, of which 13 are processed pseudogenes, distributed in 11 chromosomes including sex chromosomes (Husson et al., 2003).

The *ASS1* promoter extends for 800 bp and contains a TATA box region, 6 potential Sp1 binding sites (GC boxes) and a potential AP-2 site.

Three GC regions act synergistically and are required for complete activation of the promoter.

It is known that the expression of *ASS1* is regulated by hormones such as glucagon, insulin and glucocorticoids but, surprisingly, in the promoter region no regulatory elements have been identified as C/EBP or GRE binding sites (glucocorticoid responsive elements) .

The existence of a 10 kb cAMP response element upstream of the transcription initiation codon has been demonstrated (Guei et al., 2008) confirming the regulatory action of TSE1 (tissue-specific extinguisher locus 1), a hepatic enhancer -specific for the expression of *ASS1* encoding the regulatory subunit Riα of PKA (Boshart et al., 1991).

3.3.4.3. Arginine succinate synthetase enzyme

The enzyme argininosuccinate synthase (ASS1, EC 6.3.4.5) is a homotetrameric protein, each monomer comprises 412 amino acids and has a molecular weight of 47 KDa.

ASS1 is associated with the outer mitochondrial membrane on the cytoplasmic side (Karlberg et al., 2007) and expressed ubiquitely, although mainly in the liver.
In periportal hepatocytes it catalyzes the conversion of citrulline and aspartate into argininosuccinate and pyrophosphate by the hydrolysis of ATP, an essential step in the process of detoxification of ammonia through the urea cycle (Figure 12).
ASS1 also performs other functions: as mentioned above, it is involved in the de novo synthesis of arginine, in the synthesis of nitric oxide, polyamides and creatine (Haines et al., 2011) (Figure 13).
In the intestinal-renal pathway (Morris et al., 2002), citrulline is produced at the level of the small intestine as a product of the metabolism of glutamine and, following its release into the bloodstream by the enterocytes, is captured by the cells of the renal proximal tubule where it is converted to arginine.
Different types of non-hepatic cells have the ability to convert citrulline into arginine resulting in the generation of nitric oxide by the enzyme nitric oxide synthase (NOS).
It has been shown that the regulation of the expression of *ASS1* is tissue-specific: in the liver, where the enzyme is expressed at high levels, the main regulation is linked to the action of hormones (glucagon, insulin and glucocorticoids) and nutrients . More precisely, the expression is induced by glucocorticoids, cAMP and glucagon, while insulin and growth hormone act by modulating the effects of glucocorticoids.
In NO-producing cells, instead, ARP (argininosuccinate synthase regulatory protein) acts which inhibits the expression of ASS1 at the translation level: only 10% of the enzyme transcripts contain a uORF encoding this 4.4 kDa inhibitor (Haines et al., 2010).
The expression of argininosuccinate synthase in humans increases progressively, from 53% (compared to adults) to the thirteenth week to 90% to the thirty-sixth (Husson et al., 2003). Arginine, in fact, is not an essential amino acid in the fetal age and in the newborn but it becomes it in adulthood.

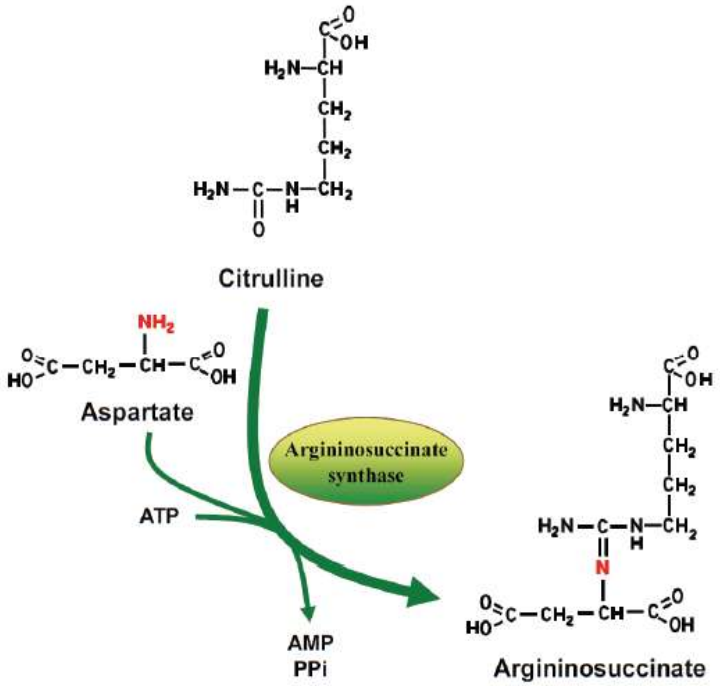


Figure 12. Reaction catalyzed by argininosuccinate synthase (Haines et al., 2011).

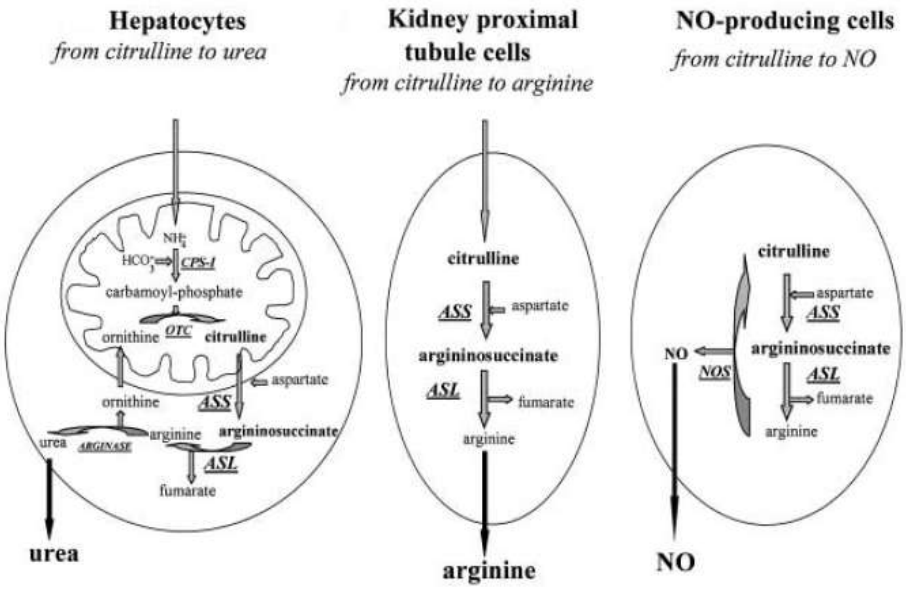


Figure 13. Schematic representation of the three main functions of ASS1 in mammals (Husson et al., 2003).

The crystalline structure of the human monomer ASS1, complexed with citrulline and aspartate, was determined by Karlberg and colleagues in 2007 (Figure 14). The protein is a homotetramer formed from two identical dimers with sequence

and structure highly conserved in the different organisms. Each monomer comprises two main regions, a globular component that contains a nucleotide binding domain (ATP) and an active site with synthase activity, and an α-helix C-terminal tail involved in oligomerization. The monomers are assembled according to the head-tail scheme.

Figure 14. Crystallized structure of the human ASS1 protein (www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?Dopt=s&uid=106372).

As can be seen from the topological diagram of a moninero of the Argininosuccinate synthase of E. *coli* (Curis et al., 2005) (Figure 15), the first N-terminal domain consists of β-sheets alternating with some α -elices and includes a P-loop strongly involved in the interaction with ATP.

The second domain is responsible for the binding of citrulline and the aspartate; one of its α-helices is inserted in the middle of the first domain, facilitating the dimerization of the monomers at the level of the catalytic site.

The dimerization takes place mainly at the end of the second domain and determines the confinement of the α-helices in a cavity formed by the β-sheets: this cavity at the center of the tetramer is occupied by the solvent.

The structure is stabilized by salt bridges, hydrogen bonds and C-terminal tails that embrace the other monomer.

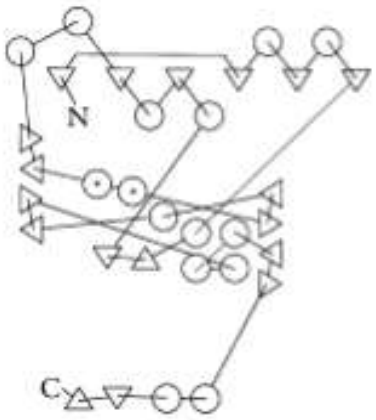


Figure 15. Topological diagram of the protein: the triangle represents a β-sheet structure while the circle represents an α-helix (tratto da Curis et al., 2005).

In the active site of each monomer, interaction with a molecule of citrulline and an aspartate molecule takes place.

Citrulline is strongly anchored, by means of salt bridges and hydrogen bonds, to the side chains of the residues Glu270, Tyr282, Arg127, Asn123 and Tyr87, the two nitrogen atoms bind Ser189 and the oxygen atom interacts with aspartate. Aspartate interacts with the residues Asn123 and Thr119 (Figure 16).

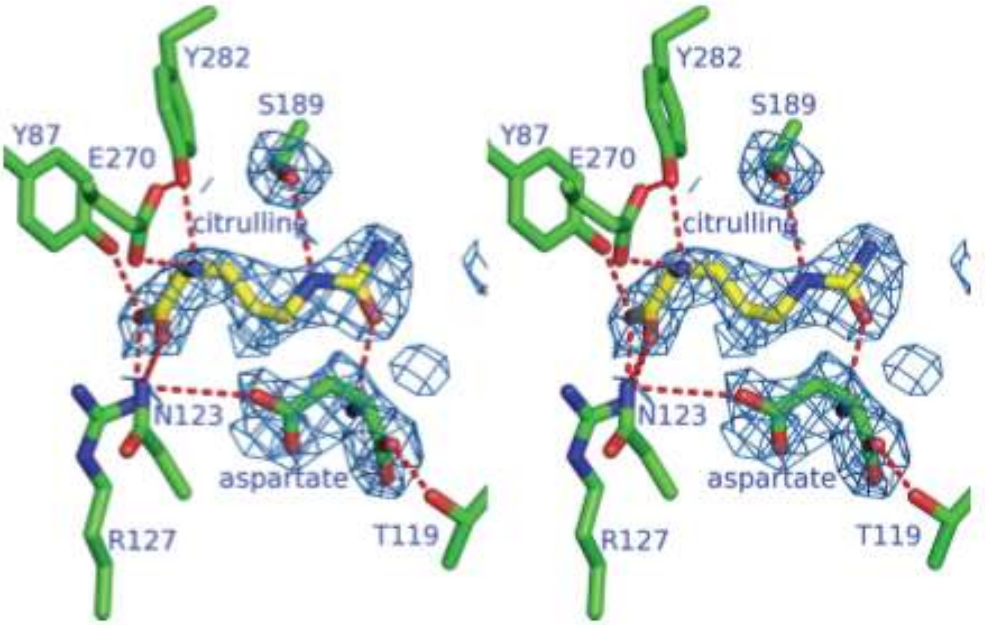


Figure 16. Citrulline and aspartate bonds in the ASS1 active site (Karlberg et al., 2007).

In E. *coli* it has been shown that argininosuccinate synthase undergoes a conformational change following ATP binding for the reduction of the distance between the phosphate groups of the ATP and the citrulline molecule from 7.9 Å to 5.8 Å, the domain of ATP binding approaches the active site, making it possible to catalysis (Lemke et al., 2002).

In humans the distance between citrulline and the phosphate groups of ATP is equal to 4.8 Å: this difference indicates that a considerably lower conformational change is necessary for the catalysis in ASS1 compared to that demonstrated for the bacterial enzyme (Figure 17) .

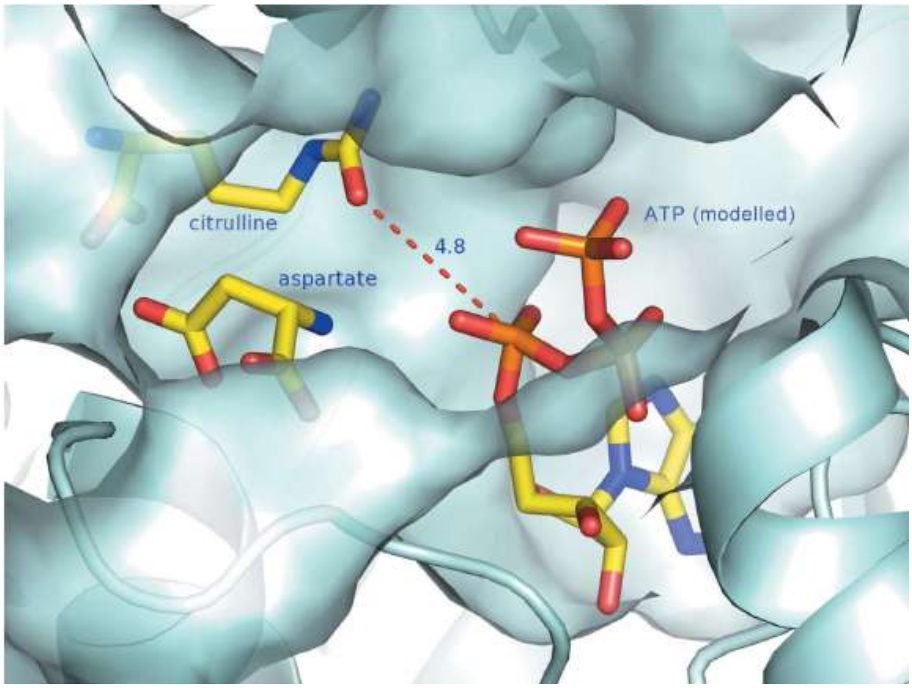
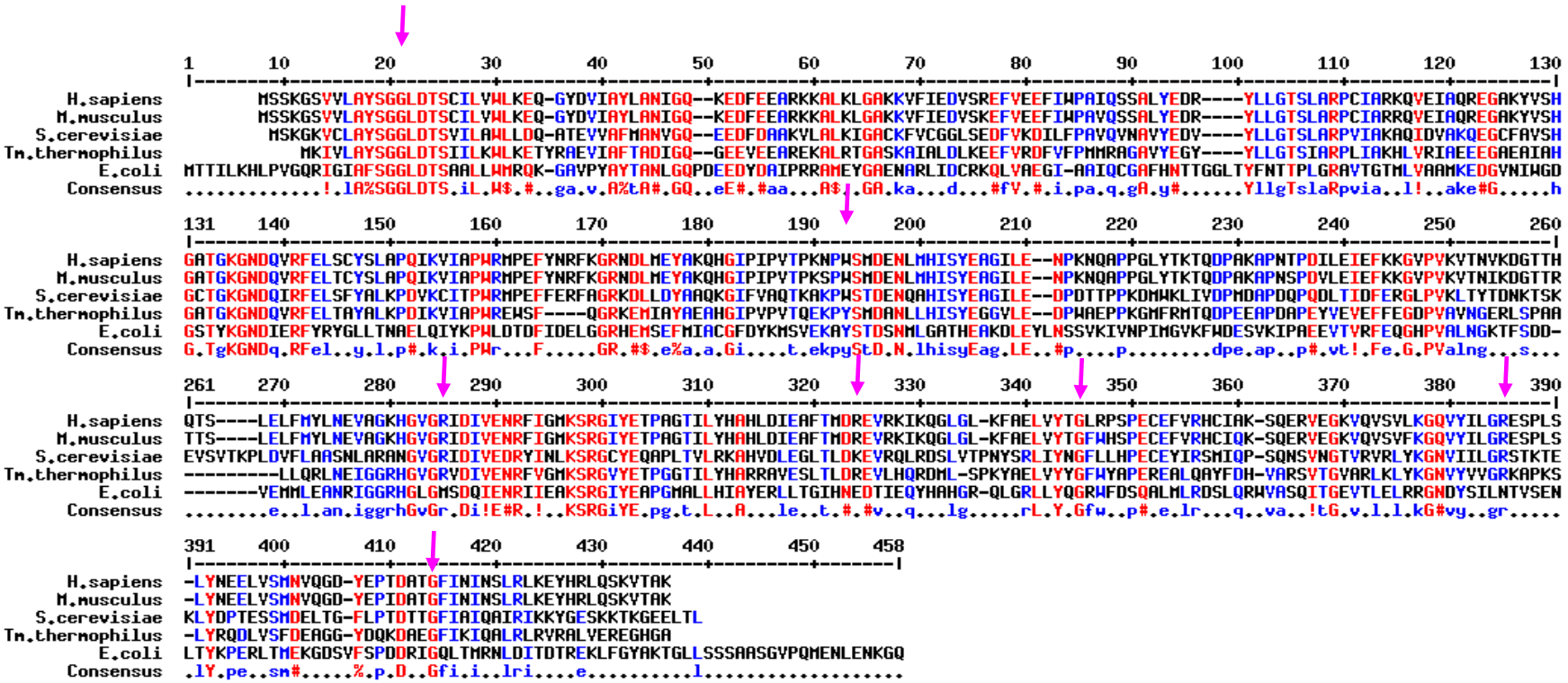


Figure 17. ATP-citrulline interaction model in ASS1 (Karlberg et al., 2007).

Figure 18. (Figure on the next page) Alignment of the sequence of the argininosuccinate synthase protein of different species (H.sapiens, M.musculus, S.cerevisiae, T.termophilus, E.coli) using the Multalin software (http: //multalin.toulouse. inra.fr/multalin/multalin.html). The arrows indicate the conserved residues involved by the mutations under study (in red the highly conserved regions are indicated and the partially conserved regions in the different species are in blue).


1 10 20 30 40 50 60 70 80 90 100 110 120 130
H.sapiens MSSKGSVVLAYSGGLDTSCILVWLKEQ-GYDVIAYLANIGQ--KEDFEEARKKALKLGAKKVFIEDVSREFVEEFIWPAIQSSALYEDR----YLLGTSLARPCIARKQVEIAQREGAKYVSH
M.musculus MSSKGSVVLAYSGGLDTSCILVWLKEQ-GYDVIAYLANIGQ--KEDFEEARKKALKLGAKKVFIEDVSKEFVEEFIWPAVQSSALYEDR----YLLGTSLARPCIARRQVEIAQREGAKYVSH
S.cerevisiae MSKGKVCLAYSGGLDTSVILAWLLDQ-ATEVVAFMANVGQ--EEDFDAAKVLALKIGACKFVCGGLSEDFVKDILFPAVQVNAVYEDV----YLLGTSLARPVIAKAQIDVAKQEGCFAVSH
Tm.thermophilus MKIVLAYSGGLDTSIILKWLKETYRAEVIAFTADIGQ--GEEVEEAREKALRTGASKAIALDLKEEFVRDFVFPMMRAGAVYEGY----YLLGTSIARPLIAKHLVRIAEEEGAEAIAH
E.coli MTTILKHLPVGQRIGIAFSGGLDTSAALLWMRQK-GAVPYAYTANLGQPDEEDYDAIPRRAMEYGAENARLIDCRKQLVAEGI-AAIQCGAFHNTTGGLTYFNTTPLGRAVTGTMLVAAMKEDGVNIWGD
Consensus .....!.lA%SGGLDTS.iL.W$.#..ga.v.A%tA#.GQ..eE#.#aa...A$..GA.ka...d...#fV.#.i.pa.q.gA.y#.....YllgTslaRpvia..l!..ake#G.....h
131 140 150 160 170 180 190 200 210 220 230 240 250 260
H.sapiens GATGKGNDQVRFELSCYSLAPQIKVIAPWRMPEFYNRFKGRNDLMEYAKQHGIPIPVTPKNPWSMDENLMHISYEAGILE--NPKNQAPPGLYTKTQDPAKAPNTPDILEIEFKKGVPVKVTNVKDGTTH
M.musculus GATGKGNDQVRFELTCYSLAPQIKVIAPWRMPEFYNRFKGRNDLMEYAKQHGIPIPVTPKSPWSMDENLMHISYEAGILE--NPKNQAPPGLYTKTQDPAKAPNSPDVLEIEFKKGVPVKVTNIKDGTTR
S.cerevisiae GCTGKGNDQIRFELSFYALKPDVKCITPWRMPEFFERFAGRKDLLDYAAQKGIFVAQTKAKPWSTDENQAHISYEAGILE--DPDTTPPKDMWKLIVDPMDAPDQPQDLTIDFERGLPVKLTYTDNKTSK
Tm.thermophilus GATGKGNDQVRFELTAYALKPDIKVIAPWREWSF----QGRKEMIAYAEAHGIPVPVTQEKPYSMDANLLHISYEGGVLE--DPWAEPPKGMFRMTQDPEEAPDAPEYVEVEFFEGDPVAVNGERLSPAA
E.coli GSTYKGNDIERFYRYGLLTNAELQIYKPWLDTDFIDELGGRHEMSEFMIACGFDYKMSVEKAYSTDSNMLGATHEAKDLEYLNSSVKIVNPIMGVKFWDESVKIPAEEVTVRFEQGHPVALNGKTFSDD-
Consensus G.TgKGNDq.RFel..y.l.p#.k.i.PWr...F.....GR.#$.e%a.a.Gi....t.ekpyStD.N.lhisyEag.LE..#p....p.........dpe.ap..p#.vt!.Fe.G.PValng...s...
261 270 280 290 300 310 320 330 340 350 360 370 380 390
H.sapiens QTS----LELFMYLNEVAGKHGVGRIDIVENRFIGMKSRGIYETPAGTILYHAHLDIEAFTMDREVRKIKQGLGL-KFAELVYTGLRPSPECEFVRHCIAK-SQERVEGKVQVSVLKGQVYILGRESPLS
M.musculus TTS----LELFMYLNEVAGKHGVGRIDIVENRFIGMKSRGIYETPAGTILYHAHLDIEAFTMDREVRKIKQGLGL-KFAELVYTGFWHSPECEFVRHCIQK-SQERVEGKVQVSVFKGQVYILGRESPLS
S.cerevisiae EVSVTKPLDVFLAASNLARANGVGRIDIVEDRYINLKSRGCYEQAPLTVLRKAHVDLEGLTLDKEVRQLRDSLVTPNYSRLIYNGFLLHPECEYIRSMIQP-SQNSVNGTVRVRLYKGNVIILGRSTKTE
Tm.thermophilus ---------LLQRLNEIGGRHGVGRVDIVENRFVGMKSRGVYETPGGTILYHARRAVESLTLDREVLHQRDML-SPKYAELVYYGFWYAPEREALQAYFDH-VARSVTGVARLKLYKGNVYVVGRKAPKS
E.coli -------VEMMLEANRIGGRHGLGMSDQIENRIIEAKSRGIYEAPGMALLHIAYERLLTGIHNEDTIEQYHAHGR-QLGRLLYQGRWFDSQALMLRDSLQRWVASQITGEVTLELRRGNDYSILNTVSEN
Consensus ....e..l.an.iggrhGvGr.Di!E#R.!..KSRGiYE.pg.t.L..A...le..t.#.#v..q...lg.....rL.Y.Gfw..p#.e.lr...q..va..!tG.v.l.l.kG#vy..gr.....
391 400 410 420 430 440 450 458
H.sapiens -LYNEELVSMNVQGD-YEPTDATGFININSLRLKEYHRLQSKVTAK
M.musculus -LYNEELVSMNVQGD-YEPIDATGFININSLRLKEYHRLQSKVTAK
S.cerevisiae KLYDPTESSMDELTG-FLPTDTTGFIAIQAIRIKKYGESKKTKGEELTL
Tm.thermophilus -LYRQDLVSFDEAGG-YDQKDAEGFIKIQALRLRVRALVEREGHGA
E.coli LTYKPERLTMEKGDSVFSPDDRIGQLTMRNLDITDTREKLFGYAKTGLLSSSAASGVPQMENLENKGQ
Consensus .lY.pe..sm#.....%.p.D..Gfi.i..lri....e..........l.....................

3.3.4.4. Arginine biosynthesis in yeast

Some of the urea cycle enzymes are involved in arginine biosynthesis and are highly conserved (Figure 19).

The yeast homologues enzymes of the human ornithine-transcarbamylase (OTC), argininosuccinate synthase (ASS1) and argininosuccinate lyase (ASL), are ARG3, ARG1 and ARG4, respectively.

The yeast does not need to eliminate the ammonium ions in the form of urea.

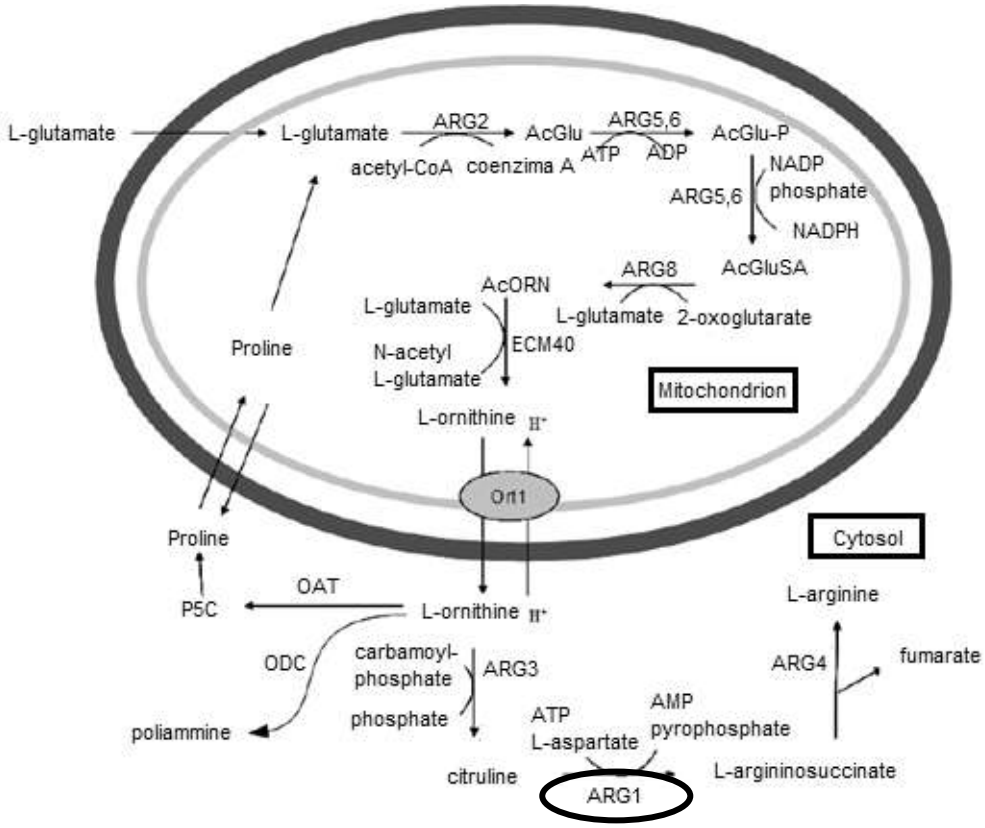


Figure 19. Arginine biosynthesis in S. cerevisiae.

Similarly to humans, also in yeast the first steps of arginine biosynthesis take place within the mitochondria and determine the formation of ornithine that is exported to the cytoplasm by the transporter Ort1p.

In the cytoplasm L-ornithine is converted to L-arginine by three reactions catalyzed by the enzymes ARG3, ARG1 and ARG4.

In yeast the transcription of these genes is induced under conditions of amino acid deficiency by the transcription factor Gcn4p and is inhibited, in the presence of

arginine, by the ArgR/Mcm1p complex comprising Arg80p, Arg81p, Arg82p, and Mcm1p (Natarajan et al., 2001).

The Gcn4p activator interacts with the subunits of the ArgR/Mcm1p repressor allowing proper transcriptional regulation in response to the availability of arginine (Hinnebusch et al., 1992). Like the human enzyme, Arg1p has tetrameric structure: each monomer includes 420 amino acids and has a molecular weight of 49 kDa.

The *ARG1* gene is located in the XV chromosome, occupies about 1.2 kb and is devoid of intronic sequences.

3.3.4.5. Saccharomyces *cerevisiae* as a model system for mutation validation

S. *cerevisiae* is a yeast belonging to the phylum of the Ascomycetes.

It is one of the most used organisms in molecular and cellular biology studies thanks to its characteristics: economy, ease of handling, rapidity of growth and easy to isolate.

The S. *cerevisiae* genome has a length of 1.3x10 bp and contains 16 chromosomes ranging in size from 200 to 2200 kb. The total DNA sequence was published in 1996 and reports 6183 open reading frames (ORF) (Goffeau et al., 1996).

Although the human genome is 100 times longer, the number of genes is only 3-5 times greater; this discrepancy is due to the high number of interposed sequences and non-coding regions present within the human genome and absent in the yeast one. In fact, 30% of yeast genes have a high degree of similarity with human genes and 2/3 share at least one conserved domain (Botstein et al., 1997; Walberg, 2000).

S. *cerevisiae* meets cell division by budding from the mother cell and exists in both haploid and diploid form. The mating type is determined by the two non-homologous alleles MATa and MATα, located in the *MAT* locus of chromosome 3.

The two haploid states produce two different pheromones called factor "a" and "α": each of the two cell types expresses on its surface the specific receptor for the

pheromone produced by the opposite mating type, allowing contact and fusion between the two different haploids and the consequent formation of the diploid line. The process is reversible, in fact in conditions of nutrient deficiency the diploid cell undergoes sporulation by releasing four haploid progenitors, each containing two MATa and two MATα cells (Haber et al., 2012).

2. AIM OF THE THESIS

Today's molecular diagnostics makes use of numerous techniques for the identification of causative mutations in genes responsible for genetic diseases. The most commonly used method and the gold standard f or the identification of mutations is the automatic Sanger sequencing, preceded by the amplification of the coding sequences of the candidate gene by PCR. However, if the gene is large and mutations are distributed along the gene randomly due to the lack of mutational hot spots, the Sanger method is rather laborious and requires extremely long times, allowing the sequencing of only 96 reactions in parallel.

The emergence of new generation sequencing methods (NGS) has completely revolutionized the fields of molecular biology and genetics; thanks to the high processivity, new technologies have given a new impulse to clinical practice, facilitating the search for mutations in large genes and reducing costs and analysis times.

The aim of the thesis is to verify the applicability of new generation sequencing methods (NGS) to be introduced in the practice of laboratory diagnostics. In this regard, two different massively-parallel sequencing strategies were compared based on the two different *Illumina* and *Agilent Technologies*, which were adopted for the study of patients in analysis at the Clinical Genetics Unit of the Hospital of Padua. The two approaches used for the analysis of metabolic genetic diseases involve the use of genomic DNA as a starting material for the construction of libraries and the use of the *MiSeq Dx (Illumina)* platform for the sequencing of samples.

The research activity has also focused on the development of models for the validation and characterization of the gene variants identified by NGS sequencing. From the diagnostic point of view, the improvement of the sequencing techniques such as the NGS have allowed to identify an ever increasing number of variants with the consequent need to study their potential pathogenetic effects.

For this reason we chose to analyze some variants with uncertain meaning (VUS) and potentially able to alter the splicing mechanism using the hybrid minigene system. In this study it has been evaluated the possible involvement of the new

IVS4+1G>A mutation in the *COQ4* gene splicing mechanism, allowing to determine its pathogenicity. Since several years the Clinical Genetics Unit invests in the study of the molecular mechanisms involved in the coenzyme Q biosynthesis pathway and in the characterization of the new mutations identified in patients with coenzyme Q deficiency syndrome.

In this study it was also developed a functional complementation system in S. *cerevisiae* yeast of the human *ASS1* gene, in order to analyze and validate the pathogenicity of some missense mutations, identified in patients affected by type 1 citrullinemia. The criteria used to determine the pathogenicity of new sequence variants identified in patients affected by monogenic diseases are defined by the Mutations Database Initiative/Human Genome Variation Society guidelines, but do not provide any functional information on the effects of these mutations; for this reason, the development of a system based on S.*cerevisiae* to evaluate in vivo the pathogenicity of the mutated alleles was found to be effective.

Previous studies (Trevisson et al., 2009; Doimo et al., 2012) have revealed how the yeast complementation allows the evaluation of residual activity of the mutant forms of the enzyme argininosuccinate lyase, also belonging to the urea cycle: for this reason we have decided to apply the same method also for the identification of residual activity of the ASS1 mutants, in order to establish possible genotype-phenotype correlations.

The *ARG1* yeast gene (used as a positive control), the human homologous *ASS1* wild-type gene and the versions containing the mutations under study (inserted by site-specific mutagenesis) were cloned into the pCR8 TOPO TA vector and then transferred, via recombination LR, in the yeast expression vector pYES.2. With these constructs the yeast strain containing a deletion of the *ARG1* gene was transformed, allowing to obtain a model with which to analyze its residual growth capacity in selective medium. Western blot analysis allowed to verify the stability of mutant proteins expressed in yeast.

3. MATERIALS AND METHODS

3.1. Part I: NGS sequencing analysis

3.1.1. Patients

The family history and subsequent diagnosis of the patients under examination were performed at the U. O. of Genetics and Clinical Epidemiology of the Hospital of Padua, according to the diagnostic protocols used in the Operative Unit.

Each patient was asked to provide informed consent for the execution of the diagnostic investigation.

3.1.2. Genomic DNA extraction

Genomic DNA was extracted from leukocyte cells of peripheral blood collected in tubes containing EDTA using the *MagPurix Blood DNA Extraction kit 200* (*Resnova*) commercial kit on *MagPurix* automatic extractor (*Resnova*). The extraction took place from 400 µl of whole blood and the extracted DNA was eluted in 100 µl of Tris-HCl buffer and stored at -20 ° C.

The quality of the extracted DNA was evaluated by 1.5% agarose gel electrophoresis and *NanoDrop 2000* spectrophotometer (*Thermo Scientific*).

The amount of DNA extracted was measured using the *QuantiFluor ONE dsDNA System (Promega)* kit on the *Quantus Fluorometer (Promega)* fluorometer.

3.1.3. Mitochondrial genome amplification

The mitochondrial genome has a double-stranded circular structure and in humans it includes 16569 base pairs encoding 37 genes (encoding 13 polypeptides

synthesized from the mitochondrial ribosome, 22 tRNA and 2 rRNA), involved in the production of proteins necessary for cellular respiration.

To proceed with NGS sequencing, the entire mitochondrial genome was amplified by two overlapping PCRs of approximately 9 kb each using specific pairs of primers for each amplified. Primers have been found in the literature (Gunnarsdottir et al, 2015).

Table 2 shows the sequences of primer pairs (forward and reverse) with the respective annealing temperatures and the amplified dimensions.

Amplicon	Primer FOR (5'->3')	Primer REV (5'->3')	Bp	Ta
1	gacgggctcacatcaccccataa	gcgtacggccagggctattggt	8338	68.5 °C
2	gccacaactaacctcctcggactcct	ggtggctggcacgaaattgacc	8647	66 °C

Table 2. Primers used for the amplification of the human mitochondrial genome, with the respective amplified dimensions and annealing temperatures.

The PCR amplification was performed in a volume of 25 µl according to the protocol shown in Table 3, using the *Expand Long Template Taq (Roche)* enzyme and subjecting the reaction to the temperature cycle described in Table 4. The presence of products PCR was verified by 1.5% agarose gel electrophoresis.

Reaction mix	Volume
H_2O milliQ	29.25 µl
Buffer 1 10X	5 µl
dNTPs 10 mM	7 µl
Primer forward 10 µM	1.5 µl
Primer reverse 10 µM	1.5 µl
Long Template Taq	0.75 µl
DNA	5 µl
Total	**50 µl**

Table 3. Mix of the amplification reaction (PCR) of human mitochondrial genome.

Temperature	Time	N° of cycles
94 °C	2'	1
94 °C	10"	10
Ta	30"	
68 °C	8'	
94 °C	15"	25
Ta	30"	
68 °C	8' + 20"/ciclo	
68 °C	7'	1

Table 4. Thermal profile of the amplification reaction (PCR) of human mitochondrial genome.

3.1.4. Construction of NGS libraries according to the *Nextera XT DNA Library Prep Kit (Illumina)*

NGS sequencing of mitochondrial DNA was performed according to the *Nextera XT DNA Library Prep Kit (Illumina)* commercial kit protocol. The protocol uses PCR amplicons (> 300 bp) as starting material and provides for the fragmentation of DNA and the simultaneous "tagging" of the fragments obtained with oligonucleotide sequences (adapters) necessary for the primers to appear in the next step. This is followed by PCR amplification of the obtained fragment library which inserts, at the ends of the fragments, both the adapters for the generation of clusters on the flow-cell and the index sequences for the identification of each sample.

Below is a description of the subsequent steps provided for the construction of the bookcases.

Quantification and dilution of mitochondrial DNA amplicons. The mitochondrial DNA amplicons of each patient were quantified using the *QuantiFluor ONE dsDNA System (Promega)* kit on the *Quantus Fluorometer (Promega)* fluorometer and, taking into account the size of each amplified, their molarity was calculated. Each amplicon was diluted to 5 nM and, after combining all the amplicons of each patient in a single tube, a pool dilution of 0.2 ng/µl was prepared.

Tagmentation. 5 µl of each amplicon pool at 0.2 ng/µl (1 ng) were incubated with 10 µl of *Tagment DNA Buffer* and 5 µl of *Amplicon Tagment Mix* at 55 °C for 5 minutes and cooled to 10 °C. The transposome was immediately inactivated by adding 5 µl of *Neutralize Tagment Buffer* to the reaction and incubating at room temperature for 5 minutes.

Library amplification. At the reaction of the previous step, 15 µl of *Nextera PCR Master Mix*, 5 µl of primer index 1 and 5 µl of *primer index 2* were added. The samples were subjected to a PCR program defined as follows:

- 72 ° C for 3 minutes
- 95 ° C for 30 seconds
- 12 cycles of:
    - 95 ° C for 10 seconds
    - 55 ° C for 30 seconds
    - 72 ° C for 30 seconds
- 72 ° C for 5 minutes
- 10 ° C ∞

Purification of amplified libraries. 30 µl of magnetic beads (*AMPure XP beads, Beckman Coulter*) were added to the 50 µl of PCR of the previous step and incubated at room temperature for 5 minutes. The plate was transferred onto a magnetic support and, after the migration of the marbles, the supernatant was removed and eliminated. The beads were washed in two successive steps adding 200 µl of 80% fresh prepared ethanol. Once the ethanol was removed, the magnetic beads were allowed to dry at room temperature for 15 minutes. Then the beads were resuspended by adding 52.5 µl of *Resuspension Buffer* and incubated at room temperature for 2 minutes. The plate was again transferred onto a magnetic support and, after the migration of the beads, 50 µl of supernatant were removed and transferred to a new plate.

Library control. The quality of the libraries obtained was evaluated by subjecting 1 µl of each capillary electrophoresis library to *2200 TapeStation (Agilent Technologies)* using the chips and reagents of the *High Sensitivity D1000 Reagents*

(Agilent Technologies). Good quality libraries show fragments between 250 bp and 1000 bp.

Manual normalization of libraries. The normalization of the libraries was performed manually rather than through passages with magnetic beads as required by the protocol. The libraries were quantified using the *QuantiFluor ONE dsDNA System (Promega)* kit on the *Quantus Fluorometer (Promega)* fluorometer and, taking into account the average fragment size of each library calculated from the *Bioanalyzer*, were diluted to 3 nM in a final volume of 10 µl.

Preparation of libraries for loading on MiSeq Dx. 5 µl of each 3 nM library were collected and merged into an equimolar pool. 10 µl of the library pool were denatured with 10 µl of 0.2 N NaOH at room temperature for 5 minutes. The pool (1.5 nM) was further diluted in *Hybridization Buffer* (HT1) in two steps up to the final molarity of 10 pM.
The libraries were uploaded to *MiSeq Dx* (*Illumina*) along with a library of control (*PhiX Control v3, Illumina*) at 1%.

3.1.5. PCR-RFLP analysis for the confirmation of mitochondrial mutations

The confirmation of mitochondrial mutations identified with the *Nextera XT DNA Library Prep Kit (Illumina)* was performed by PCR-RFLP analysis (PCR amplification followed by enzymatic digestion) that allows to identify the presence of the mutation even at low percentages of heteroplasmy.
The target regions of the mitochondrial genome were amplified by PCR, each using pairs of specific primers for each amplified. The primers were designed using the *Primer3* software (http://primer3.ut.ee/) available online.
Table 5 shows the sequences of primers pairs (forward and reverse) with the respective amplified size of the main mutations identified by the mitochondrial genome (of which m.15170G>A has been identified for the first time). For all reactions the annealing temperatures was 55°C.

Mutation	Primer FOR (5'->3')	Primer REV (5'->3')	Bp
m.3243A>G	CCAGGTCGGTTTCTATCTAC	AGAGTTTTATGGCGTCAGCG	386
m.3460G>A	TCCTAATGCTTACCGAACGA	TCTTTGGTGAAGAGTTTTATCG	121
m.11778G>A	CCATCTGCCTACGACAAACA	TCCTTGAGAGAGGATTATGACG	223
m.14484T>C	ACTCACCAAGACCTCAACC	GATTGTTAGCGGTGTGGTCG	176
m.15170G>A	TGAAACTTCGGCTCACTCCT	TTTTATCGGAATGGGAGGTG	562

Table 5. Primers used for the amplification of the target regions for the main mutations of the human mitochondrial genome, with the respective amplified dimensions and annealing temperatures.

The PCR amplification was performed in a volume of 25 µl according to the protocol reported in Table 6, using the enzyme *Taq Gold (Applied Biosystem)* and subjecting the reaction to the temperature cycle described in Table 7. The presence of products of PCR was verified by 1.5% agarose gel electrophoresis.

Reaction mix	Volume
H_2O milliQ	14.85 µl
Buffer 10X	2.5 µl
$MgCl_2$	1.5 µl
dNTPs 10 mM	0.5 µl
Primer forward 10 µM	1.75 µl
Primer reverse 10 µM	1.75 µl
Taq Gold	0.15 µl
DNA	2 µl
Total	**20 µl**

Table 6. Amplification reaction (PCR) mix of target regions for main mutations of the human mitochondrial genome.

Temperature	Time	N° of cycles
95 °C	10'	1
95 °C	30"	35
Ta	40"	
72 °C	1'	
72 °C	10'	1

Table 7. Thermal profile of the amplification reaction (PCR) of the target regions for the main mutations of the human mitochondrial genome.

After amplification, an enzymatic digestion was performed using the restriction enzymes reported in Table 8.

Mutation	Restriction enzyme	T of incubation	BSA	Inactivation at 65 °C for 20'
m.3243A>G	ApaI	25 °C	Yes	Yes
m.3460G>A	BstUI	60 °C	No	No
m.11778G>A	BstUI	60 °C	No	No
m.14484T>C	BccI	37 °C	No	Yes
m.15170G>A	AcuI	37 °C	No	Yes

Table 8. Restriction enzymes used for enzymatic digestion of the target regions for the main mutations of the human mitochondrial genome.

The enzymatic digestion reaction (RFLP) was performed in a volume of 20 µl (Table 9) using the specific restriction enzyme and incubating the reaction to the respective incubation T (Table 8) overnight. The presence of RFLP products was verified by 12% acrylamide gel electrophoresis.

Reaction mix	Volume
H_2O milliQ	To 20 µl
Buffer CutSmart 10X	2 µl
BSA 10X	2 µl
Restriction enzyme	1 µl
PCR product	500 ng

Table 9. Enzymatic digestion reaction mix.

3.1.6. Construction of NGS libraries according to the *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)*

The analysis of hereditary metabolic diseases based on NGS sequencing is performed using the two different *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* and *HaloPlex HS Target Enrichment System (Agilent*

Technologies) technologies. These are two amplicon-based methods based on the generation of DNA amplicons from small amounts of the patient's genomic DNA, using customized oligonucleotides as primers for amplification, for capture and enrichment of regions of interest.

TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina), after the hybridization of the DNA of each patient to the custom probe pool, provides a step of extension and ligation of the portion of DNA between the two probes that leads to the formation of products containing the target regions of interest flanked by short sequences complementary to the primers of the subsequent amplification step. PCR, in addition to cloning regions of interest, adds adapters for the generation of clusters on the flow-cell and the specific index sequences for each sample.

Below is a description of the subsequent steps envisaged for the construction of the bookcases.

Design the probe pool. The probes of our gene panels were previously designed using the *DesignStudio software (Illumina).*

Each panel generates an average of 600 amplicons long on average 250 bp and has an average coverage of 99.71%.

Quantification and dilution of genomic DNA. DNA samples were quantified using the *QuantiFluor ONE dsDNA System (Promega)* kit on the *Quantus Fluorometer (Promega)* fluorometer and diluted to the final concentration of 25 ng/µl in 10 mM Tris-HCl in a final volume of 25 µl. 1 µl of each DNA was aliquoted in plate and diluted by adding 3 µl of *Resuspension Solution 1* and 1 µl of *Sample Stabilization Solution 1*. A positive control, provided by the kit, and a negative control were also added to the plate. 2 µl of *2800M control DNA* was diluted with 2 µl of *Resuspension Solution 1* and 1 µl of *Sample Stabilization Solution 1* while the negative control consists in 5 µl of *Resuspension Solution 1*.

Hybridization of the probe pool. To each sample were added 5 µl of the custom probe pool (*Custom Amplicon Oligo Tube*) previously diluted with *Resuspension Solution 1*. To the positive control were added 5 µl of a specific probe pool

(*Control Oligo Pool kit*) previously diluted with *Resuspension Solution 1*. 15 µl of *Oligo Hybridization for Sequencing 2* were added to all samples. The samples were incubated in thermomixer at 95 °C for 1 minute and the temperature was subsequently lowered to 40 ° C.

Removal of unbound probes. 25 µl of *Sample Purification Beads* were added to the reaction of the previous step and incubated at room temperature for 5 minutes. After transferring the plate onto a magnetic support, the supernatant was removed and discarded. The beads were washed three times with 80 µl of *Stringent Wash 1* and then with 80 µl of 60% prepared fresh ethanol. The supernatant was removed and the ethanol was evaporated at room temperature for 5 minutes.

Extension and ligation of the bound probes. To each sample were added 22 µl of a mixture of *Extension-Ligation Enzyme* and *Extension-Ligation Buffer* previously prepared. The samples were subjected to the following thermal extension and ligation program:

- 37 ° C for 45 minutes
- 70 ° C for 20 minutes
- 4 ° C ∞

Library amplification. At the reaction of the previous step were added 4 µl of *Index 1* (i7), 4 µl of *Index 2* (i5) and 20 µl of a mix of *Enhanced Master Mix* and *Enhanced DNA Polymerase*.

The samples were subjected to a PCR program defined as follows:

- 95 ° C for 3 minutes
- 28 cycles of:
    - 98 ° C for 20 seconds
    - 67 ° C for 20 seconds
    - 72 ° C for 40 seconds
- 72 ° C for 1 minute
- 10 ° C ∞

Purification of amplified libraries. After centrifuging, 45 µl of supernatant was taken from each sample and transferred to a new plate. 36 µl of *Sample Purification Beads* were added and the plate was incubated first at 1800 rpm for 2 minutes and then at room temperature for 5 minutes. The plate was transferred onto a magnetic support and, after the migration of the beads, the supernatant was removed and discarded. The beads were washed in two successive steps adding 200 µl of 80% fresh prepared ethanol. Once the ethanol was removed, the beads were allowed to dry at room temperature for 5 minutes. To resuspend the beads, 25 µl of Resuspension Buffer were added and the plate was incubated first at 1800 rpm for 2 minutes and then at room temperature for 2 minutes. After relocating the plate on a magnetic support to allow the migration of the beads, 20 µl of supernatant were removed and transferred to a new plate.

Library control. The quality of the libraries obtained was evaluated by an electrophoretic run on 4% agarose gel. The PCR products of the samples have a size of about 350 bp, as expected for amplicons of 250 bp.

Manual normalization of libraries. The normalization of the libraries was done manually. The libraries were quantified using the *QuantiFluor ONE dsDNA System (Promega)* kit and the *Quantus Fluorometer (Promega)* fluorometer.
Then the molarity was calculated using the expected length of PCR products (350 bp). Finally, the libraries were diluted to 3 nM in a final volume of 100 µl.

Preparation of libraries for loading on MiSeq Dx. 5 µl of each 3 nM library were collected and mixed into an equimolar pool. 10 µl of the library pool were denatured with 10 µl of 0.2 N NaOH at room temperature for 5 minutes. The pool (1.5 nM) was further diluted in *Hybridization Buffer* (HT1) in two steps up to the final molarity of 10 pM. 600 µl of the pool was loaded on *MiSeq Dx (Illumina)* with a control library (*PhiX Control v3, Illumina*) at 1%.

3.1.7. Construction of NGS libraries according to the *HaloPlex HS Target Enrichment System (Agilent Technologies)*

The other amplicon-based NGS technology that has been used for the analysis of inherited metabolic diseases is *HaloPlex HS Target Enrichment System (Agilent Technologies).*

An enzymatic digestion step of the genomic DNA of each patient precedes the hybridization to the custom probe pool.

Below is a description of the subsequent steps envisaged for the construction of the bookcases.

Design of the probe pool. The probes of our gene panels were previously designed using the *SureDesign software (Agilent Technologies).*

Two types of panels have been designed: medium-sized panels that generate an average of 4500 amplicons, and large panels that produce an average of 16,000 amplicons. In both cases the amplicons are 150 bp long and the panels have an average coverage of 99.8%.

Quantification and dilution of genomic DNA. DNA samples were quantified using the *QuantiFluor ONE dsDNA System (Promega)* kit and the *Quantus Fluorometer (Promega)* fluorometer. DNA were diluted to the final concentration of 1.8 ng/μl in 10 mM Tris-HCl in a final volume of 32 μl. The 32 μl of each genomic DNA were transferred to the respective well of an 8-well strip (*DNA Sample Strip*), in which there was also transferred 32 μl of ECD (*Enrichment Control DNA*) provided by the kit.

Preparation of Restriction Enzyme Master Mix. 16 different restriction enzymes are provided in two 8-well strips (*Green Enzyme Strip* and *Red Enzyme Strip*), named from A to H. Using a multichannel pipette 0.35 μl of each enzyme of the *Green Enzyme Strip* and *Red Enzyme Strip* have been aliquoted in the respective wells of the *Restriction Enzyme Master Mix Strip*, to which were then added 24.6 μl of *RE buffer* and 0.64 μl of *BSA solution*.

Digestion of genomic DNA with restriction enzymes. Using a 3.5 µl multichannel pipette, enzymes were aliquoted from the *Restriction Enzyme Master Mix* into each column of a 96-well plate (*Restriction Digest Reaction Plate*). Each column contains 3.5 µl per well of the same combination of restriction enzymes.
3.5 µl of each genomic DNA previously aliquoted in the *DNA Sample Strip*, were distributed in the respective column of the *Restriction Digest Reaction Plate*. In the last column of the plate, 3.5 µl of the ECD were aliquoted.
The plate was briefly centrifuged and then transferred to the thermocycler for the digestion reaction:

- 37 ° C for 30 minutes
- 80 ° C for 5 minutes
- 4 ° C ∞

Digestion validation. Enzymatic digestions were validated by electrophoretic analysis of ECD reactions with the *TapeStation 2200 (Agilent Technologies)* instrument and using *High Sensitivity D1000 ScreenTape (Agilent Technologies*) and *High Sensitivity D1000 Reagents (Agilent Technologies).*
ECD includes genomic DNA and a 800 bp PCR product that contain restriction sites for all enzymes used in digestion reactions. The results of the electrophoretic run were visualized with the *2200 TapeStation Analysis (Agilent Technologies*) software: the 8 digestions must show three predominant bands at about 125, 225 and 450 bp.

Hybridization of digested DNA to HaloPlex HS probes. The 8 digestion reactions of each patient were transferred to the respective tube for hybridization with the custom probes. To each tube were added 5 µl of the specific *HaloPlex HS Probe*, 34 µl of *Hybridization Solution* and 5 µl of *HaloPlex HS Indexing Primer* (unique for each patient).
The tubes were briefly centrifuged and then transferred to the thermocycler for the hybridization reaction:

- 95 ° C for 5 minutes
- 58 ° C for 2 hours

Elimination of the hybridization buffer. For each sample to be purified a mix was prepared by combining 20 µl of *HS Hybridization Stop Solution* and 80 µl of *AMPure XP (Beckman Coulter)* beads. 100 µl of the mix were added to each sample which was then incubated at room temperature at 1300 rpm for 5 minutes.
The tubes were transferred onto a magnetic support and, after the migration of the beads, the supernatant was removed and eliminated. The beads were washed in two successive steps adding 200 µl of fresh prepared 70% ethanol. Once the ethanol was removed, the beads were allowed to dry at room temperature for 5 minutes.

Ligation of the captured fragments. The beads were resuspended by adding 50 µl of the ligation mix obtained by combining, for each sample, 10 µl of *HS Ligation Solution*, 0.6 µl of 1 mM *rATP* and 39.4 µl of H2O nuclease free.
The tubes were incubated at room temperature for 2 minutes to allow the DNA to elute from the beads and then transferred back to a magnetic support. 47.5 µl of supernatant were taken and transferred to new tubes which, after adding 2.5 µl of the enzyme *HS DNA Ligase*, were incubated at 55 ° C for 10 minutes in the thermocycler.

Capture the target DNA. 40 µl of *Dynabeads MyOne Streptavidin T1* magnetic beads (*Thermo Fisher Scientific*) diluted in *HS Capture Solution* were added to each sample. The samples were then incubated at room temperature for 15 minutes. The tubes were transferred onto a magnetic support and, after the migration of the beads, the supernatant was removed and eliminated.
Subsequently 100 µl of Wash 1 mix was added, combining 90 µl of *HS Wash 1 Solution* and 10 µl of fresh 1 M NaOH. The tubes were incubated at room temperature for 1 minute and then transferred back to a magnetic support to remove the Wash 1 mix. After adding 150 µl of *HS Wash Solution 2* to each sample, the tubes were transferred to the magnetic media and the supernatant was removed.

Library amplification. The target libraries captured in the previous step were amplified by PCR adding 100 µl of PCR master mix to the washed beads as

follows: 53.2 µl of nuclease-free H2O, 30 µl of *Herculase II Reaction Buffer*, 0.8 of 100 mM *dNTPs*, 4 µl of *Primer 1*, 8 µl of *Primer 2*, 4 µl of *Herculase II Fusion DNA Polymerase.*

The samples were submitted to the following PCR program:

- 98 ° C for 2 minutes
- N cycles (the number is variable for each panel) of:
    - 98 ° C for 30 seconds
    - 60 ° C for 30 seconds
    - 72 ° C for 1 minute
- 72 ° C for 10 minutes
- 8 ° C ∞

When the PCR program was completed the tubes were briefly centrifuged and transferred onto a magnetic support. 40 µl of supernatant were taken and transferred to new purification tubes.

Purification of amplified libraries. To the 40 µl of each sample, 100 µl of *AMPure XP* beads *(Beckman Coulter)* and 40 µl of nuclease-free H2O were added. The tubes were incubated at room temperature for 5 minutes and then transferred onto a magnetic support to remove the supernatant. The beads were washed in two successive steps adding 200 µl of fresh 70% ethanol. Once the ethanol was removed, the beads were allowed to dry at room temperature for 5 minutes. To resuspend the beads 45 µl of *HS Elution Buffer* were added. The beads were incubated at room temperature for 2 minutes. After relocating the plate on a magnetic support to allow the migration of the beads, 40 µl of supernatant were removed and transferred into new tubes.

Library validation. The quality of the libraries obtained was measured by electrophoretic analysis with the *TapeStation 2200 (Agilent Technologies)* instrument and using *High Sensitivity D1000 ScreenTape (Agilent Technologies)* and *High Sensitivity D1000 Reagents (Agilent Technologies).*

The PCR products of the samples are between 175 and 625 bp.

Preparation of libraries for loading on MiSeq Dx. 5 µl of each 4 nM library were collected and mixed into an equimolar pool. 10 µl of the library pool were denatured with 10 µl of 0.2 N NaOH at room temperature for 5 minutes. The pool (2 nM) was further diluted in *Hybridization Buffer* (HT1) in two steps up to the final molarity of 10 pM. 600 µl of the final pool were loaded on *MiSeq Dx* (*Illumina*) with a control library (*PhiX Control v3, Illumina*) at 1%.

3.1.8. NGS sequencing and data analysis

The libraries of each sample were sequenced by paired end method (2x150 bp) on *MiSeq Dx* platform with chemistry v2 (*MiSeq Reagent Kit v2, Illumina*) with an expected coverage not less than 500X.

The minimum threshold of the *Quality Score* has been set to 30 (QS30): the instrument has performed the "call of the bases" (base calling) with an accuracy of 99.9% (probability of 1/1000 to call the wrong base) .

For libraries made according to the *Nextera XT DNA Library Prep Kit (Illumina)*, during the sample sheet preparation phase was chosen the "small genome resequencing" analysis pipeline. To simplify the analysis, the reads produced by the instrument were aligned not to the entire human genome (GRCh37/hg19), but to the reference sequence of the human mitochondrial genome (human_mtDNA_16569_bp_andersen).

The "small genome resequencing" pipeline uses BWA software (*Burrows-Wheeler Aligner*) (Li and Durbin, 2009) and GATK (*Genome Analysis Toolkit*) for the alignment and for call phases of the variants, respectively.

For libraries made according to the *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* a "targeted resequencing" approach was adopted: the sequencing of the target regions and the subsequent alignment of the reads to the sequence reference was made thanks to the information contained in the "manifest file", a specific file containing the indications relating to the chromosomal position of the amplicons designed, and the choice of the human genome (Homo sapiens, hg19) as the reference genome. This pipeline exploits the *Smith-*

Waterman algorithm for alignment of the reads and the GATK (*Genome Analysis Toolkit*) software for the call of the variants.

The secondary analysis of libraries generated by *Illumina* protocols, the *MiSeq Reporter v2.5.1.3 (Illumina)* and *VariantStudio v2.2 (Illumina)* software were used in combined with *IGV v2.1.2* software (*Integrative 61 Genomics Viewer - Broad Institute*) to allows the graphical display of aligned reads and variants.

For libraries made according to the *HaloPlex HS Target Enrichment System (Agilent Technologies)* protocol, the samplesheet was set up to generate only FASTQ files, a text format used to represent sequences. FASTQ files are the main inputs for alignment. The analysis was performed using the *SureCall* software (*Agilent Technologies*) that uses the BWA software (*Burrows-Wheeler Aligner*) (Li and Durbin, 2009) for alignment and the *SAMtools* software for the call of the variants.

The run parameters of the *MiSeqDx* instrument, generated by the *RTA v1.18.54.0 (Real Time Analysis)* software integrated into the instrument, were evaluated in real time using the *Sequencing Analysis Viewer software v1.8.37.0 (Illumina).*

Once the variants of interest were identified, a search was carried out in the literature and in the various databases available online (LOVD, Leiden Open Variation Database, http://www.lovd.nl/3.0/home, and HGMD, Human Gene Mutation Database, Institute of Medical Genetics, Cardiff, http://www.hgmd.org/) to verify if mutations had a pathogenic clinical significance already described.

3.1.9. MLPA analysis

Patients analyzed by NGS sequencing which shows only one mutation in heterozygosis in the *PAH* gene, were tested by MLPA analysis (*Multiplex Ligation-dependent Probe Amplification*) to check the presence of deletions/duplications of one or more exons of the *PAH* gene.

MLPA analysis was conducted according to the protocol of the *SALSA MLPA P055* commercial kit (*MRC-Holland*).

In the first step 3 µl of DNA of each patient (50-100 ng / µl) were denatured at 98 °C for 5 minutes. After cooling the samples at 25 °C, 0.75 µl of *SALSA Probe Mix*

and 0.75 µl of *SALSA MLPA Buffer* were added to each tube. The probes were hybridized by incubating the samples at 95 °C for 1 minute and at 60 °C for about 16-20 hours (overnight). Keeping the temperature of the thermal cycler at 54 °C, 16 µl of the following mix was added to the reaction:

- 12.5 µl of filtered H2O MilliQ
- 1.5 µl of Ligase Buffer A
- 1.5 µl of Ligase Buffer B
- 0.5 µl of SALSA Ligase-65.

The ligation of the probes was performed by incubating the reaction at 54 °C for 15 minutes and then the enzyme was inactivated at 98 ° C for 5 minutes. The samples were finally cooled at 20 °C.

For the subsequent amplification reaction, a mix containing 1.875 µl of filtered H2O MilliQ, 0.5 µl of *SALSA PCR Primer Mix* and 0.125 µl of *SALSA Polymerase* was prepared for each sample. The mix was combined with 10 µl of ligation reaction of the previous step and was subjected to the following temperature program:

- 40 cycles of:
    - 90 ° C for 30 seconds
    - 60 ° C for 30 seconds
    - 72 ° C for 1 minute
- 72 ° C for 20 minutes.

1.5 µl of each PCR product were combined with 8 µl of formamide (*Hi-Di Formamide, Applied Biosystems*) and 0.5 µl of *500 ROX* markers (*GeneScan 500 ROX dye Size Standard, Applied Biosystems*) and denatured at 94 °C for 3 minutes. The samples were subjected to electrophoresis on a capillary sequencer *3500Dx Genetic Analyzer (Applied Biosystems).* The results were visualized and analyzed using the *Coffalyser v140721.1958 (MRC-Holland)* software.

3.2. Part II: B-globine minigene system for *COQ4* IVS4+1G>A mutation validation

To study the possible splicing alteration caused by the new IVS4+1G>A mutation in *COQ4* gene, it was used an approach based on hybrid minigene. In particular, the β-globin minigene previously generated in ours laboratory (Forzan et al., 2009).

It was chosen to use *β-globin* because it is a gene consisting only of three exons and therefore easy to amplify into one reaction. The amplified was obtained by PCR starting from the human genomic DNA and cloned, using the sites of NheI and ApaI restriction enzymes, into the pcDNA 3.1 Hygro vector (*Invitrogen*), which allows the expression of the gene of interest under the guidance of the CMV promoter. To facilitate the subsequent cloning steps, in intron 2 of *β-globin* gene has been cloned a polylinker bearing three sites of restriction (XhoI, NotI, HindIII), as shown in Figure 20.

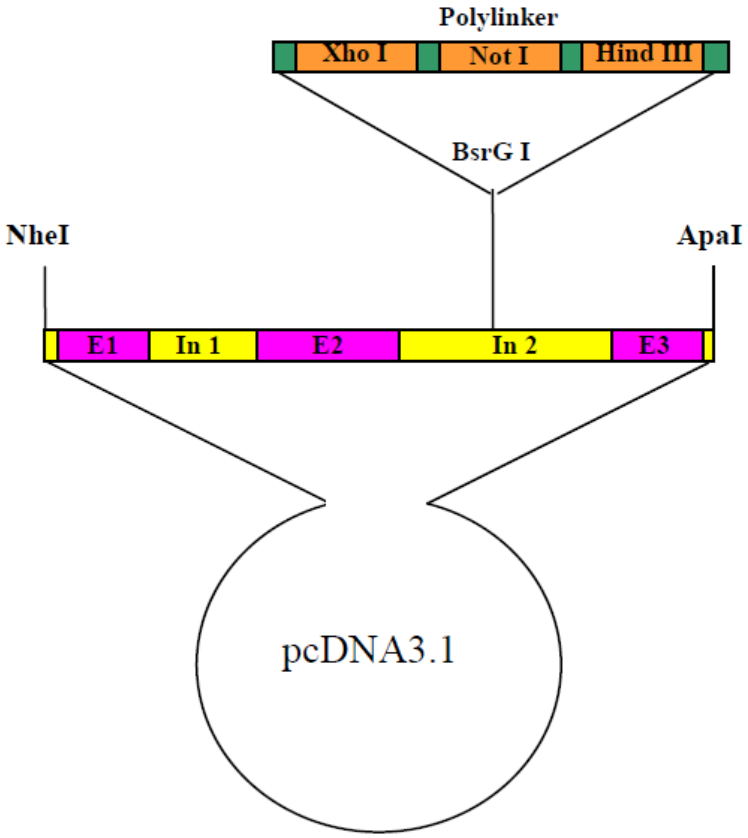


Figure 20. Graphic representation of the insertion of the *β-globin* minigene into the pcDNA 3.1 Hygro expression vector (*Invitrogen*).

3.2.1. Patient genomic DNA amplification

The primers used for the amplification of patient genomic DNA have been designed by using *Primer3* software. Primers contain the restriction sites of XhoI and HindIII restriction enzyme (Table 10).

The DNA region amplified contained the exon adjacent to the mutation in analysis with a exon's upstream and downstream portions of at least 100 bp.

Intron	Forwad primer (5'-3')	Reverse primer (3'-5')	Bp	Tm (°C)
4	cttCTCGAGgccagttgtaggtgctccat	cttGCGGCCGCctttcctcagggtgtcatgc	341	55

Table 10 . Primers used for the amplification of the target region of the patient genomic DNA, with the respective amplified dimension and melting temperature.

The PCR reaction is catalyzed by the *Phusion DNA-polymerase HF* (*Thermo Scientific*): at about 300 ng of DNA were added 5X buffers, 200 µM dNTPs, 5% Betaine, 200 nm of each primer and 2.5U of enzyme to the final volume of 50 µl.

The reactions occurred under the following conditions:

- 98 ° C for 30 seconds
- 35 cycles of:
    - 98 °C for 10 seconds
    - 55 °C for 30 seconds
    - 72 °C for 1 minute
- 72 °C for 7 minutes.

The presence of products of PCR was verified by 1.5% agarose gel electrophoresis.

3.2.2. *pCR8* and *pCDNA3.1 Hygro* $^{+}$ vectors

In this study were used the vector *pCR8 TOPO TA (Invitrogen*) and the mammalian expression vector *pCDNA3.1 Hygro*$^{+}$ (*Invitrogen*).

The vector *pCR8 TOPO TA* (*Invitrogen*) is a cloning vector that contains the TOPO recognition sites with overhang -TT ends used for the ligation of the

construct of interest, the *attL1* and *attL2* sites for recombination and the gene that confers resistance to spectinomycin for selection in E. *coli* (Figure).

The mammalian expression vector *pCDNA3.1 Hygro*$^{+}$ (*Invitrogen*) contains a human cytomegalovirus immediate-early (CMV) promoter for high-level expression in a wide range of mammalian cells, the gene that confers resistance to ampicillin for selection in E. *coli* and the hygromycin resistance gene for selection of stable cell lines (Figure 21).

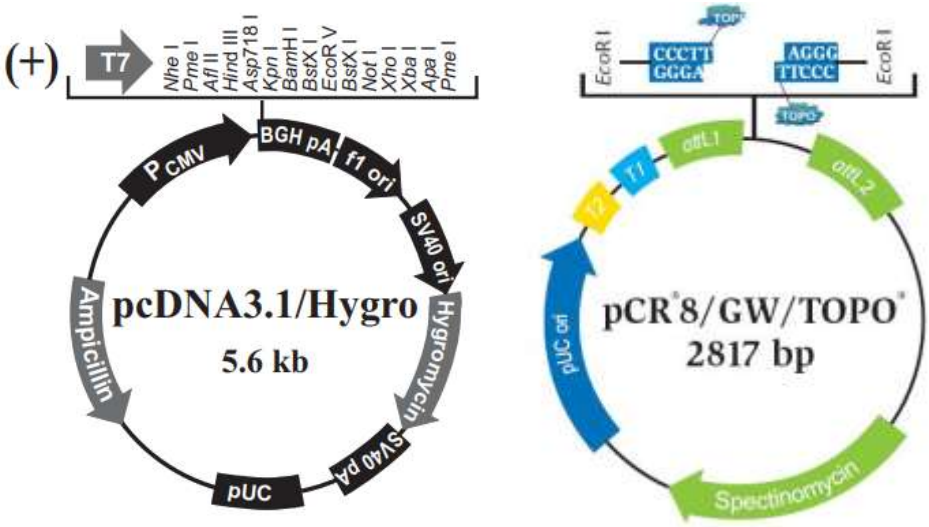


Figure 21. Maps of *pCDNA3.1 Hygro*$^{+}$ and *pCR8* (*Invitrogen*) vectors.

3.2.3. *pCR8* TOPO TA cloning

The amplification products were incubated at 72 ° C for 10 minutes with 1 unit of *Taq DNA polymerase (Euroclone)*, the specific buffer 10X and dATPs 0.2 mM, in order to reconstruct the overhang poly-A ends. Subsequently at 4 µl of PCR product, 1 µl of salt solution and 1 µl of vector *pCR8 TOPO TA (Invitrogen)* were added. The mixture was incubated for 30 minutes at room temperature.

Using 3 µl of the ligation reaction, 50 µl of E. *coli* DH5α bacteria were transformed by heat shock:

- on ice for 30 minutes
- 42 °C for 45 seconds
- on ice for 2 minutes.

After adding 250 µl of SOC medium, the bacteria were incubated at 37 °C for 1 hour. They were subsequently plated in LB medium with spectinomycin for negative selection, and incubated overnight at 37 °C.

The colonies that incorporated the vector with the correct orientation insert were selected by screening PCR: the 10X buffer, dNTPs 200 µM, 200 nm of the primers (primers M13f and Reverse primers were used; Table **14**) and 2.5U of the enzyme *EconoTaq DNA-polymerase (Lucigen)* were added to each colony, in a final volume of 19 µl. PCR amplification reactions occurred at the following conditions:

- 94 °C for 12 minutes
- 35 cycles of:
    - 94 °C for 1 minute
    - 55 °C for 1 minute
    - 72 °C for 1 minute
- 72 °C for 7 minutes.

The plasmid DNA of the selected colonies was purified with the *Spin Miniprep Kit (Qiagen)* and quantified at the *Nanodrop*.

3.2.4. Miniprep and sequencing

The bacteria inoculum was carried out in Falcon tubes containing 2 ml of LB culture medium and 2 µl of spectinomycin antibiotic. The culture was incubated at 37 ° C overnight at 300-400 rpm of oscillation.

The plasmid DNA was extracted from each inoculums (Miniprep) and then they were sequenced by Sanger method: the wild-type clone and the mutated clone have been selected.

3.2.5. Digestion of *pCR8* vector and gel extraction of DNA insert

To transfer the insert from the *pCR8* vector to the *β-globin* minigene, 6 µg of both plasmids (*pCR8* and *pCDNA3.1 Hygro*$^{+}$*-β-globin*) was digested with HindIII and XhoI enzymes able to recognize the restriction sites into the primers sequence.

A double digestion reaction was performed, incubating at 37 °C for 3 hours and then at 80 °C for 20 minutes (to inactivate the enzyme).

The digestion product was separated by electrophoresis on agarose gel and the corresponding size band of the fragment of interest has been cut off. The DNA insert was purified with the *Qiagen* kit and finally it was quantified at the *Nanodrop*.

3.2.6. Ligation

The DNA insert and the *pCDNA3.1 Hygro*$^{+}$*-β-globin* vector obtained from the digestion reaction were subjected to ligation reaction.

The amount of insert to be used in the reaction was calculated based on the followed formula:

Insert (ng) = [6 x construct size] / [7294 x vector (ng)]

In our study it has been used a relationship

insert:vector of 6: 1.

The digestion products were incubated at 16 °C overnight with 1 µl of *T4 DNA ligase (BioLabs)*, the specific buffer 10X and ATP 10X.

Transformation, screening, inoculation and plasmid purification (of the wild-type and mutated vectors) were then performed as reported in the previous paragraphs with the difference that the spectinomycin has been replaced by ampicillin, whose resistance is conferred by the pcDNA3.1 vector.

3.2.7. Transient transfection of HEK 293 cells

1 μg of the wild-type and the mutated *COQ4 pCDNA3.1 Hygro*$^{+}$*-β-globin* was used for transfect 7x10^{5} HEK 293 cell cultures (Human Endothelial Kidney that do not express *β-globin)* in 250 μl of Opti-MEM 1X (*Gibco*) antibiotic and serum-free medium.

Within the Falcon tubes were also aliquoted 250 μl of Opti-MEM 1X and 1 μl of Lipofectamine 2000 (*Invitrogen*), a cationic lipid formulation that it is able to complex with negative DNA charges facilitating its entry through the cell membrane, and were left to room temperature for 20 minutes.

Finally, the 500 μl obtained were added to each culture and after 6 hours the ground was changed replacing Opti-MEM with 2 ml of DMEM 10% FCS (*Gibco*), a medium with 10% serum in which normally cells are grown.

3.2.8. RNA extraction from HEK 293 cells

Total human RNA was extracted and purified using the *Trizol* kit (*Invitrogen*).

Cells were washed with PBS, lysed with 0.05% trypsin and resuspended in 1 ml of Trizol. After adding 200 μl of chloroform, the cells were centrifuged at 12,500 g for 15 minutes. The upper phase containing the RNA was recovered and subsequently precipitated in 0.5 ml of isopropanol and incubated at -80 °C overnight.

After centrifugation, the pellet was washed with 70% cold ethanol and resuspended in DEPC water (treated with diethyl pyrocarbonate, a powerful RNAase inhibitor).

3.2.9. Reverse transcription

The purified RNA was quantified by spectrophotometer at 260 nm.

Reverse transcription was performed using the *Superscript II Reverse Transcriptase II Kit (Invitrogen).*

Briefly: 100 μg of random primers, 1μl of 0.5 mM dNTPs and DEPC water were added to 1 μg of RNA, up to a final volume of 10.5 μl. The reaction mixture was incubated at 65 °C for 5 minutes. Subsequently were added: Buffer 5X, DTT 0.01 M, 40U of RNase OUT and the mixture was incubated at 25 °C for 2 minutes. 200U of the *SuperScript II RT* enzyme was added and then the reaction was incubated at the following temperature program:

- 25 °C for 10 minutes
- 42 °C for 50 minutes
- 70 °C for 15 minutes.

3.2.10. cDNA amplification

The cDNAs obtained by the RNA reverse transcription was amplified by PCR with specific primers for the coding regions of the *β-globin* gene, in particular the primer forward within exon 2 and the reverse primer inside exon 3 of the gene (ex2 *β-globin* forward primer: TCTGTCCACTCCTGATGCTG and ex3 *β-globin* reverse primer: CACTGGTGGGGTGAATTCTT). To confirm the result it has been performed also an amplification reaction using the *COQ4* exon 4 internal primer (ex4 *COQ4* forward primer: cttCTCGAGgccagttgtaggtgctccat and ex3 *β-globin* reverse primer: CACTGGTGGGGTGAATTCTT). It has been used the *EconoTaq DNA polymerase (Lucigen)*as described in section 3.2.3.

The thermal cycler has been set as follows:

- 94 °C for 3 minutes
- 35 cycles of:
    - 94 °C for 1 minute
    - 55 °C for 1 minute
    - 72 °C for 1 minute
- 72 °C for 7 minutes.

The PCR fragments thus obtained were analyzed by electrophoretic 2% agarose gel to separate the bands of smaller dimensions.

3.3. Part III: Saccharomyces *cerevisiae* as a model system for *ASS1* mutation validation

3.3.1. Microorganisms used in the study

For the validation and characterization of the *ASS1* gene variants identified by NGS sequencing were used the bacterial and yeast strains shown in Table 11:

Strain	Genotype	Reference
E. *coli* ccdB Survival	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG tonA::Ptrc-ccdA	Invitrogen
E. *coli* DH5α	F- recA1 endA1 hsdR17(rk-, mk+) supE44 λ- thi-1 gyrA96 relA1	Invitrogen
S. *cerevisiae* BY4741	MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0	Euroscarf
S. *cerevisiae* BY4741 Δ*ARG1*	Mat a; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YOL058w::kanMX4	Euroscarf

Table 11. Strains and relative genotype of the microorganisms used in the study.

E. *coli ccdB Survival* bacteria are competent for transformation and are used for negative selection as they are a strain resistant to the *ccdB* gene, that encodes a protein that interacts and inhibits DNA-gyrase, leading to cell death (Bahassi et al., 1998).

The DH5α bacteria are E. *coli* competent for transformation but not resistant to the *ccdB* gene.

3.3.2. Culture media used in the study

For the bacterial transformation the liquid medium SOC (Super *Optimal Broth with Catabolite repression*) was used (2% Tryptone, 0.5% yeast extract, 10mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10mM MgSO4 and 20 mM glucose).

The LB medium (Triptone 10 g/l, Yeast extract 5 g/l, NaCl 10g/l, pH 7, Agar 15g/l for solid medium) was used to grow the transformed bacterial cells, added the different antibiotics (Table 12) depending on the gene resistance of the vector used in the transformation.

Antibiotic	Final concentration
Ampicilline	100 µg/ml
Chloramphenicol	30 µg/ml
Spectinomicine	100 µg/ml

Table 12. Antibiotics and their concentration used in the study.

Different media have also been used for yeast cells.

YPDA (1% yeast extract, 1% peptone and 2% glucose) is a rich medium used to propagate and maintain wild and mutated strains; the corresponding solid medium contains 2% of agar.

SM (0.17% yeast base nitrose without amino acids, 0.5% ammonium sulfate and 2% glucose or galactose, or 2% raffinate and 0.1% galactose) is a minimal medium used to select and maintain yeasts transformed with the different constructs, for functional complementary tests and for tests on the requirements necessary for the growth of the yeast strain.

The corresponding solid media contains 2.3% of agar.

To these minimum medium are added the amino acids for which the yeast strain is auxotrophic, at the concentrations reported in Table 13:

Amino acid	µl for plate (20 ml)	µl for ml of medium	Final concentration
Histidine 20 mg/ml	10	0.5	10 mg/L
Methionine 10 mg/ml	40	2	20 mg/L
Leucine 20 mg/ml	60	3	60 mg/L
Arginine 10 mg/ml	40	2	20 mg/L

Table 13. Amino acids at the concentration used in the study.

3.3.3. *pCR8* and *pYES.2* vectors

In this study two different vectors were used: the yeast expression vector *pYES.2* (*Invitrogen*) present in high number of copies in the cell and the vector *pCR8 TOPO TA* (*Invitrogen*) (see the describtion at the section 3.2.2) for the recombination through the *Gateway* system.

The *pYES.2* vector has the origin of 2µ replication and the pGAL1 promoter that allows to modulate the expression of the gene of interest by varying the carbon source added to the culture medium: galactose favors high levels of expression, the raffinose maintains the basal expression levels and glucose inhibits expression. It contains the gene that confers resistance to ampicillin for selection in E. *coli* and the URA3 marker for yeast selection (Figure 22).

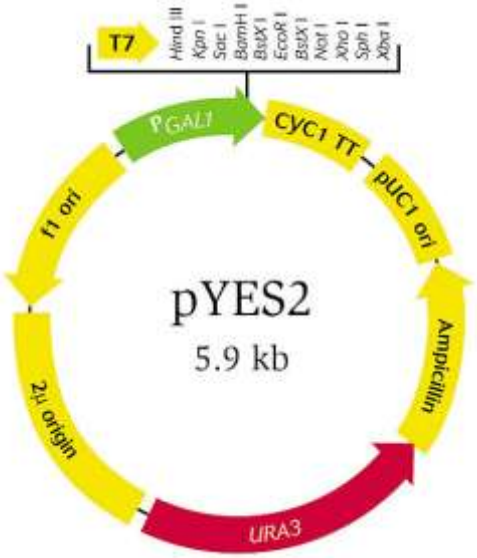


Figure 22. Map of *pYES.2* (*Invitrogen*) vector.

3.3.4. Extraction of yeast genomic DNA

An yeast inoculum in 2 ml of YPDA liquid medium was incubated at 30 °C overnight.

The pellet was recovered by centrifugation and incubated for 10 minutes at -80 °C. Next were added glass beads and 200 µl of phenol:chloroform:isoamyl alcohol in a ratio of 25:24:1. The cells were vortexed 5 times for 1 minute and

incubated for 1 minute in ice. After centrifuging for 13 minutes at 13,000 g, 200 µl of TE buffer were added.

To recovered the supernatant containing DNA, 400 µl of chloroform were added and the mixture was again centrifuged for 10 minutes at 13,000 g.

The aqueous phase was recovered and 2.5 volumes of 100% cold ethanol were added. The mixture was left to precipitate for 30 minutes at -80 °C and then centrifuged for 15 minutes at 13,000g.

The supernatant was removed and the pellet was washed with 1 ml of 70% cold ethanol. Centrifuge again for 15 minutes at 13,000g and remove the supernatant.

The pellet was resuspended in DEPC water, incubated for 10 minutes at 37 ° C and then left overnight to hydrate.

3.3.5. RNA extraction from human fibroblasts

Total human RNA was extracted and purified from a culture of cutaneous control fibroblasts using the *Trizol* kit (*Invitrogen*). See the describtion at paragraph 3.2.8.

3.3.6. Reverse transcription

The purified RNA was quantified by spectrophotometer at 260 nm.

Reverse transcription was performed using the *Superscript II Reverse Transcriptase II Kit (Invitrogen).* See the describtion at paragraph 3.2.9.

3.3.7. *Gateway* system

The *Gateway* technology (*Invitrogen*) allows to perform an *LR recombination* reaction between an entry clone in which the gene of interest is cloned and a destination vector containing the *RfA cassette*. The advantage of this technique lies in allowing a rapid and efficient transfer of a specific DNA sequence in several expression vectors, keeping the correct orientation and reading frame.

The reaction is catalyzed by the enzyme *LR clonase* (*Invitrogen*) which recognizes the corresponding *attL* and *attR* sequences (Figure 23).

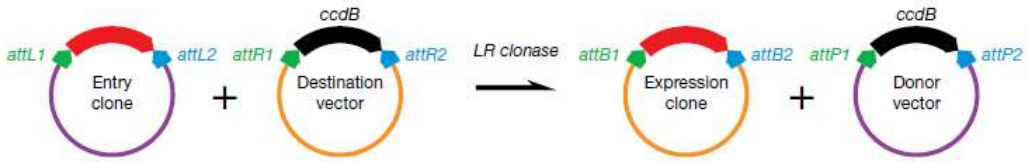


Figure 23. Schematic representation of *LR recombination* reaction (*Invitrogen*).

3.3.8. Gene amplification and TOPO TA cloning

ASS1 gene and the corresponding yeast homologue *ARG1* gene were amplified starting respectively from human retrotranscribed cDNA and yeast genomic DNA. The PCR reaction is catalyzed by the *Phusion DNA-polymerase HF* (*Thermo Scientific*): at about 300 ng of DNA were added 5X buffers, 200 µM dNTPs, 5% Betaine, 200 nm of each primer and 2.5U of enzyme to the final volume of 50 µl. The reactions occurred under the following conditions:

- 98 ° C for 30 seconds
- 35 cycles of:
    - 98 °C for 10 seconds
    - 55 °C for 30 seconds
    - 72 °C for 1 minute
- 72 °C for 7 minutes.

Primers used for the amplification of the coding region of the two homologues and the respective Tm are shown in Table 14.

The amplification products were checked by electrophoretic run in 1% agarose gel.

The amplified *ASS1* and *ARG1* genes were then cloned into the recombination vector *Gateway pCR8 TOPO TA (Invitrogen).*

The amplification products were incubated at 72 ° C for 10 minutes with 1 unit of *Taq DNA polymerase (Euroclone),* the specific buffer 10X and dATPs 0.2 mM, in order to reconstruct the overhang poly-A ends. Subsequently at 4 µl of PCR

product, 1 µl of salt solution and 1 µl of vector *pCR8 TOPO TA (Invitrogen)* were added. The mixture was incubated for 30 minutes at room temperature.

Using 3 µl of the ligation reaction, 50 µl of E. coli DH5α bacteria were transformed by heat shock:

- on ice for 30 minutes
- 42 °C for 45 seconds
- on ice for 2 minutes.

After adding 250 µl of SOC medium, the bacteria were incubated at 37 °C for 1 hour. They were subsequently plated in LB medium with spectinomycin for negative selection, and incubated overnight at 37 °C.

The colonies that incorporated the vector with the correct orientation insert were selected by screening PCR: the 10X buffer, dNTPs 200 µM, 200 nm of the primers (primers M13f and ASS1_1285r or ARG1_1289r were used; Table 14) and 2.5U of the enzyme *EconoTaq DNA-polymerase (Lucigen)* were added to each colony, in a final volume of 19 µl. PCR amplification reactions occurred at the following conditions:

- 94 °C for 12 minutes
- 35 cycles of:
    - 94 °C for 1 minute
    - 55 °C for 1 minute
    - 72 °C for 1 minute
- 72 °C for 7 minutes.

The plasmid DNA of the selected colonies was purified with the *Spin Miniprep Kit (Qiagen)* and quantified at the *Nanodrop*.

To verify the correct insertion of the *ASS1* and *ARG1* genes, a restriction analysis was performed. The appropriate enzymes were chosen by consulting the vector restriction map (available at www.NEBcutter.com). To 500 ng of plasmid are added 10U of digestion enzyme, Buffer 10X, BSA 10X, to a final volume of 10 µl. The reaction was incubated at 37 °C for 1 hour.

Subsequently 5 µl of the digestion were loaded on 1% agarose gel.

The sequences were finally checked by sequencing (*BMR Genomics*): 500 ng of plasmid and 6.4 pmol of each primer were dried at 65 °C. The sequences were aligned with the reference ones using the *BLAST tool*.

3.3.9. Destination vector preparation

The *pYES.2* vector was used as destination vector: for this purpose the *RfA cassette* was inserted into the MCS.

The vector was linearized by enzymatic digestion and the digestion product was purified by electrophoretic run in 1% agarose gel. The restriction enzyme used creates overhang ends, so it is necessary to make them blunt by adding 1U of *Klenow DNA polymerase (Thermo Scientific)*, the specific 10X buffer and the dNTPs. *Klenow DNA polymerase (Thermo Scientific)* must then be inactivated by incubating it at 75 °C for 10 minutes.

The linearized vector was phosphorylated by incubating it with the *SAP enzyme (Shrimp Alkaline Phosphatase)* and purified through the columns of the *Amicon Ultra 100K device Kit (Millipore).*

The *RfA cassette* was ligated into the vector using overnight incubation at 16 °C with 1U of the enzyme *T4 ligase (Thermo Scientific).* The *RfA cassette* contains resistance to chloramphenicol, the toxin coding *ccdB* gene and it is flanked by the *attR1* and *attR2* recombination sites (Figure 24).

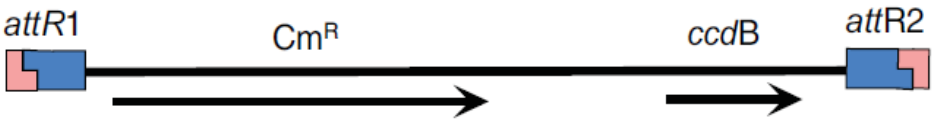


Figura 24. Scheme of the *RfA cassette (Invitrogen).*

E. *coli* ccdB bacteria (as described previously) were resistant to the ccdB toxin, plated in LB medium to which ampicillin and chloramphenicol were added.

The presence and the orientation of the *RfA cassette* in the colonies were screened by screening PCR (as described above) using the T7 and ccdBr primers, shown in Table 14.

The plasmid DNA of the selected colonies was purified with the *Spin Miniprep Kit (Qiagen)* and quantified at the *Nanodrop*.

The sequences were controlled by sequencing (*BMR Genomics*).

3.3.10. LR ricombination

The recombination reaction was performed between the entry clone containing the gene of interest and the destination vector with the *RfA cassette*.

2 µl of the pCR8_RfA vector (75 ng/µl) were added to 1 µl of pYES.2_ASS1 vector or to 1 µl of pYES.2_ARG1 (150 ng / µl) vector.

5 µl of Buffer TE and 2 µl of enzyme *LR clonase* (*Invitrogen*) were then added. The reaction mixture was incubated at 25 °C overnight. The reaction was stopped by adding 1 µl of proteinase K and incubating it at 37 °C for 10 minutes.

2 µl of the recombination reaction were used to transform 50 µl of DH5α bacteria, then plated in LB medium containing ampicillin and incubated overnight at 37 °C. The plasmid DNA of the selected colonies was purified with the *Spin Miniprep Kit (Qiagen)* and quantified at the *Nanodrop*.

The correct integration of the gene of interest in the pYES.2 vector was verified by PCR (as described above but by performing an initial denaturation at 94 °C for 5 minutes) using the primers T7 and ASS1_1285r / ARG1_1289r (Table 14).

3.3.11. Site-specific mutagenesis

The mutants were generated by site-specific mutagenesis of the pCR8_ASS1 constructs using the *Site-Directed Mutagenesis II QuickChange II Kit (Stratagene)*. The primers containing the specific mismatches, designed by the *Stratagene Primer Design* software (www.genomics.agilent.com) are shown in Table 14. The amplification reaction was performed by adding to the 10 ng plasmid: 125 ng of each primer, 1 µl of dNTPs, the reaction buffer 1X, 2.5U of the enzyme *PfuUltra H* to the final volume of 50 µl. The following conditions were used for the PCR amplification reaction:

- 95 °C for 30 seconds
- 14 cycles of:
    - 95 ° C for 30 seconds
    - 55 ° C for 1 minute
    - 68 ° C for 9 minutes
- 68 ° C for 7 minutes.

The amplification product was incubated at 37 °C for 2 hours with 10U of the restriction enzyme *DpnI* (*Thermo Scientific*), which specifically digests the methylated and hemimethylated DNA (parental DNA).

The chemo-competent *DH5α* bacteria (*Invitrogen*) were transformed with 5 µl of the digestion product and then plated into LB medium containing spectinomycin for the positive selection of the transformants that acquired the plasmid.

The presence of point mutations was confirmed by automatic sequencing after amplification with *Taq DNA polymerase (Roche)* under standard conditions.

The vectors containing the correct sequences and the desired mutations were extracted and purified with the *Spin Miniprep Kit (Qiagen*).

The *ASS1* gene containing the different missense mutations was subsequently transferred to the expression vector in yeast pYES.2 by *LR recombination*, as described in § 2.14.3.

Primers used in the study (*Integrated DNA Technologies*) are shown in Table 14:

Reaction	Primer sequence	T_m
***ASS1* amplification from cDNA**	**hASS_cDNA-27f** (5'-gaactcacgcctccaatcc-3')	64°C
	hASS_cDNA_1285r (5'-gggatctgcaaattgaggag-3')	
***ARG1* amplification from genomic DNA**	**yARG1-10f** (5'-gcataaaataatgtctaaggg-3')	56°C
	yARG1_1289r (5'-tatcttgaggcgatgaacta-3')	
Mutagenesis	**G14S** (5'-ggcctacagtggcagcctggacacctc-3')	78.0°C
	G14S_antisense (5'-gaggtgtccaggctgccactgtaggcc-3')	
	W179R (5'-ctcccaagaacccgaggagcatggatgag-3')	78.81°C
	W179R_antisense (5'-ctcatccatgctcctcgggttcttgggag-3')	
	R265H (5'-gcatggcgtgggccatattgacatcgtgg-3')	78.81°C
	R265H_antisense (5'-ccacgatgtcaatatggcccacgccatgc-3')	
	R304W (5'-gccttcaccatggactgggaagtgcgcaaaa-3')	78.98°C
	R304W_antisense (5'-ttttgcgcacttcccagtccatggtgaaggc-3')	

Reaction	Primer sequence	T_m
Mutagenesis	**G324S** (5'-gagctggtgtataccagtttctggcacagcc-3')	78.98°C
	G324S_antisense (5'-ggctgtgccagaaactggtatacaccagctc-3')	
	R363L (5'-tacatcctcggcctggagtccccactg-3')	76.0°C
	R363L_antisense (5'-cagtggggactccaggccgaggatgta-3')	
	G390R (5'-agccaactgatgccaccaggttcatcaacatcaat-3')	78.10°C
	G390R_antisense (5'-attgatgttgatgaacctggtggcatcagttggct-3')	
Validation PCR	**T7f** (5'-taatacgactcactataggg-3')	47.50°C
	pYES.2r (5'-ttcggttagagcggatgtgg-3')	57.0°C
	M13f (5'-gtaaaacgacggccag-3')	50.70°C
	M13r (5'-caggaaacagctatgac-3')	47.0°C
	***ASS1*_1285r** (5'-gggatctgcaaattgaggag-3')	53.40°C
	***ARG1*_1289r** (5'-tatcttgaggcgatgaacta-3')	50.30°C
	ccdBf (5'- ggtaagcacaaccatgcaga-3')	56.0°C
	ccdBr (5'- acctgcagactggctgtgta-3')	58.0°C

Table 14. List of used primers.

3.3.12. Yeast transformation

Yeast cells were transformed by PEG-Lithium acetate method as reported by Chen et al. in 1992.

Yeast cells were allowed to grow overnight at 28 °C in 1.5 ml of YPDA medium. The pellet has been recovered and then 5 µl of denatured carrier DNA (*Clontech*), 750 ng of plasmid DNA and 100 µl of transforming solution (40% PEG 4000, 0.2 M of LiAc at 7.5 pH and 0,1 M of DDT) were added. The cells were resuspended by vortex, incubated for 30 minutes at 45 °C and then plated in minimal medium

with 2% glucose (SM GLU2% HMLR) to which were added the amino acids for which the strain is auxotrophic, but not the amino acid that allows the plasmid selection (Uracil for pYES.2).

The transformed yeasts were propagated for two generations in SM liquid medium added with H, M, L, R with 2% galactose in order to induce the expression of the *ASS1* gene, which is under the control of the *pGAL1* promoter.

3.3.13. Phenotypic growth test in plate (Drop test)

The phenotypic growth analysis was performed starting from steady-state cultures of transformed yeasts, in a minimum medium containing 2% galactose and the aminoacids for which the strain is auxotrophic, including arginine (SM GAL2% HMLR).

A volume of cells was taken with OD 600 nm of 1, the pellet was washed 2 times in 500 µl of H_2O milliQ by centrifuging at 5000 g for 5 minutes and then resuspended in 1 ml of H_2O milliQ.

5 serial dilutions were made: 10^0, 10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}.

5 µl of each dilution were sown in 3 different minimum media:

- 2% galactose medium (SM GAL2% HML);
- 2% raffinose medium with 0.1% of galactose, to which were added the amino acids for which the strain is auxotrophic but not arginine (SM RAF2% -GAL0.1% HML);
- medium containing arginine as a control (SM GAL2% HMLR).

The plates were incubated at 28 °C for the time necessary for growth (between 6 and 10 days).

3.3.14. Extraction of proteins from yeast

Yeast total proteins were extracted by mechanical lysis and precipitation with TCA (trichloroacetic acid).

Starting from a steady-state culture in minimal medium containing 2% galactose and the amino acids for which the strain is auxotrophic (SM GAL2% HMLR), a volume of cells with an OD of 600 nm of 5 was taken.

The pellet was resuspended in 100 µl of 20% TCA. The mechanical lysis of the cells was performed using a vortex after adding a volume of glass beads. After adding 500 µl of 5% TCA, the supernatant was transferred into a new eppendorf and centrifuge at 14000 rpm for 5 minutes at 4 °C, to obtaining a precipitate containing the proteins. The pellet was then resuspended in 200 µl of Laemmli buffer (Tris-Cl 0.1 M pH 6.8, SDS 2%, DTT 50 Mm, glycerol 10%, bromophenol blue 0.01%). Subsequently 20 µl of 1.5 M Tris at pH 8.8 were added to turn the sample to the initial violet color. The sample was denatured by incubating it at 95 °C for 5 minutes.

3.3.15. Western blot

Yeast protein extract was separated by electrophoretic run on acrylamide gel (*Biorad* apparatus).

Then were prepared a stacking gel (4% solution of acrylamide-bis acrylamide, 0.1M Tris pH 6.8, 0.1% SDS, 0.1% APS 0.1% and 0.1% TEMED) and a running gel (12% solution of acrylamide-bis acrylamide, 0.4 M Tris pH6.8, SDS 0.1%, APS 0.1% and 0.05% TEMED). After charging equal amounts of protein extract, a constant voltage of 130 V was applied for 90 minutes in the presence of the Tris-Glycine-SDS reaction buffer (25 mM Tris, 192 mM glycine, 0.1% SDS).

At the end of the run, the samples were transferred onto the PVDF (GE *Healthcare*) membrane by applying a constant voltage of 100 V for 90 minutes in the presence of the Tris-Glycine reaction buffer (25 mM Tris, 192 mM glycine, 10% methanol).

The membrane was activated by incubating it with methanol for 1 minute and then saturated by incubating it for 1 hour with 3 % BSA (bovine serum albumin) in Tween-TBS (Tris-HCl 50 mM pH 7.5, NaCl 150 mM, NaN_3 0.01 % and Tween 0.1%).

Subsequently, the membrane was incubated in overnight agitation at 4 °C with the monoclonal primary anti-ASS1 antibody diluted 1:500 in 3% TBA. The membrane was washed 3 times with Tween-TBS for 10 minutes and incubated for 1 hour at room temperature with the immunoperoxidase-conjugated antibody diluted in 3% TBA. The signal was detected using the *ChemiDocTM XRS*$^{+}$ *Molecular Imager apparatus (Biorad)* after adding *Lite Ablot TURBO* reagent *(Euroclone)* which acts as a substrate for peroxidase.

As a reference for the quantification of proteins, the membrane was also incubated with the primary anti-yPorina antibody (*Mitoscience*) diluted 1:1000 in 3% TBA and then with the mouse anti-goat secondary antibody (*Santa Cruz Biotechnology*) diluted 1: 2000 in 3% TBA.

4. RESULTS AND DISCUSSION

4.1. Identification of a new renal phenotype associated with the m.15170G>A mitochondrial DNA mutation in the *MT-CYB* gene

Nextera XT DNA Library Prep Kit (Illumina) technology is used for the analysis of the mitochondrial genome, which is amplified in two superimposed amplicons of 9 Kb each. It is a fast and economical system that allows to obtain a high coverage and is highly sensitive (i.e. able to detect percentages of heteroplasmy up to 1-5%).

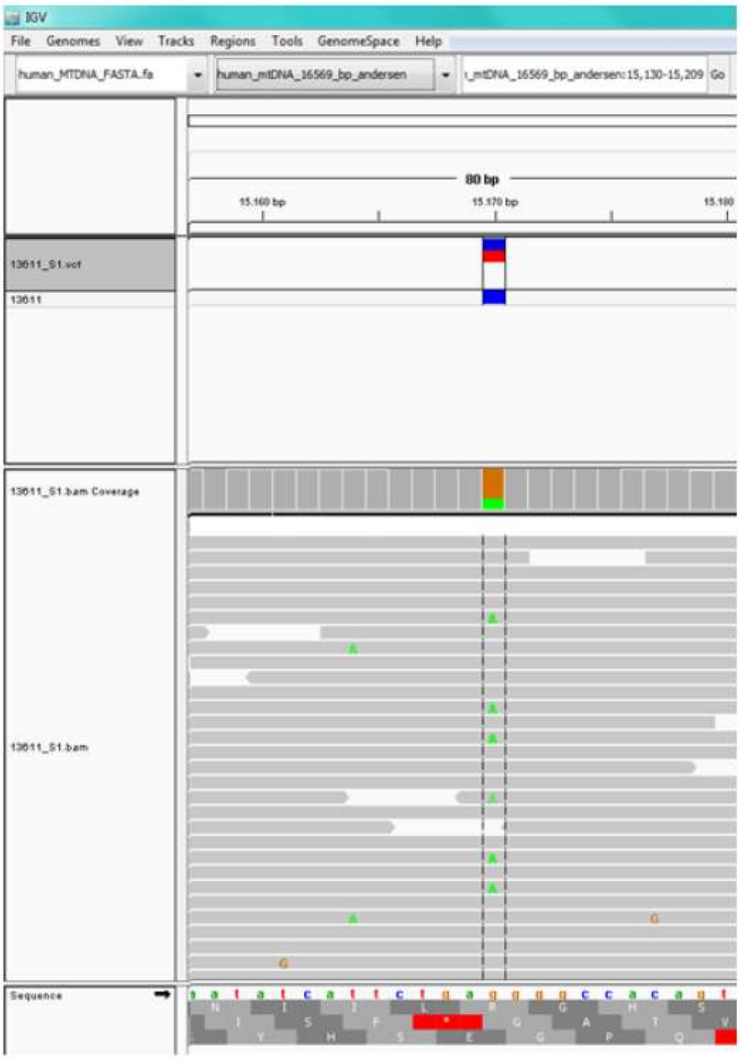
We studied a patient with a multisystem disorder that presented at age 20 years with proteinuria and microhematuria. At age 25 she developed cataracts and sensorineural hearing loss. Alport syndrome was suspected and a renal biopsy showed a slight increase in mesangial matrix, some scattered areas of fibrosis along the interstitial tubule with some infiltrates, tubular atrophy, some microcystic formations, and mild thickening of the vascular walls. EM was unremarkable ruling out Alport syndrome. Renal function progressively worsened and at age 34 she started dialysis. At age 37 she developed hypertrophic cardiomyopathy. A brain CT scan was essentially normal. She was evaluated by our service at age 38. Physical eamination was unremarkable except for diffuse muscle atrophy but she did not complain of cramps or of excersise intolerance.

In the suspicion of a mitochondrial disorder the entire mtDNA, extracted from peripheral white blood cells, was sequenced. The analysis identified the m.15170G>A p.Gly142* heteroplasmic nonsense mutation in the *MT-CYB* gene: the mutation affected approximately 22% of mitochondrial genomes (Figure 25).

PCR-RFLP analysis confirmed the mutation in blood and revealed its presence also in DNA extracted from other tissues as hair follicles, buccal cells, and in urinary sediment cells (heteroplasmy in these cells was around 60%) (Figure 26).

The mutation was absent in tissues from the healthy mother and from a control individual.

A



B

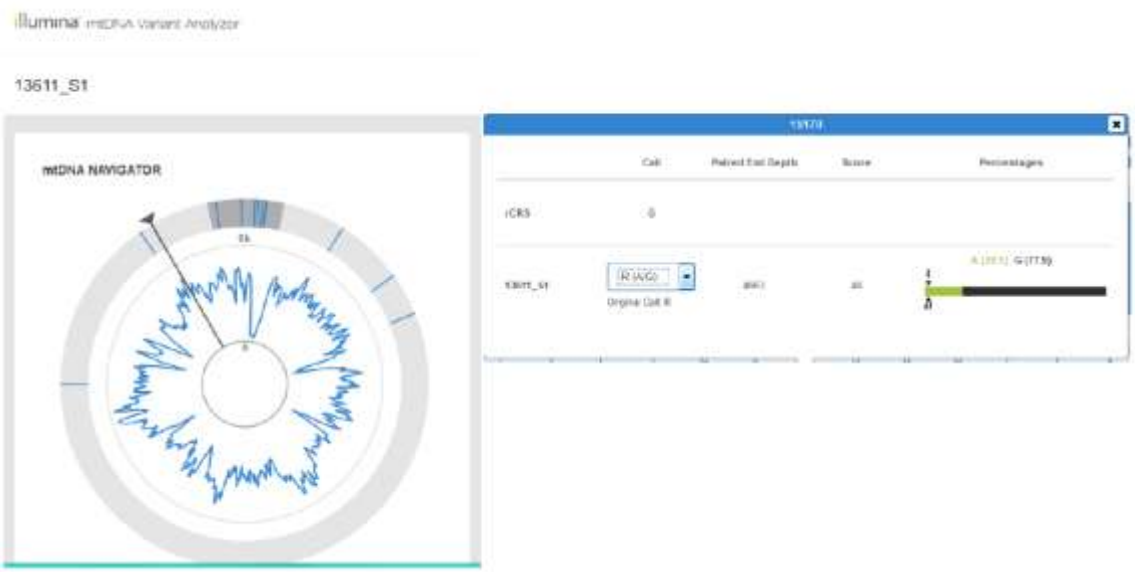


Figure 25. Analysis of the patient's entire mtDNA extracted from peripheral blood leukocytes using the *IGV* (*Illumina*) (A) and *mtDNA Variant Analyzer* (*Illumina*) software (B).

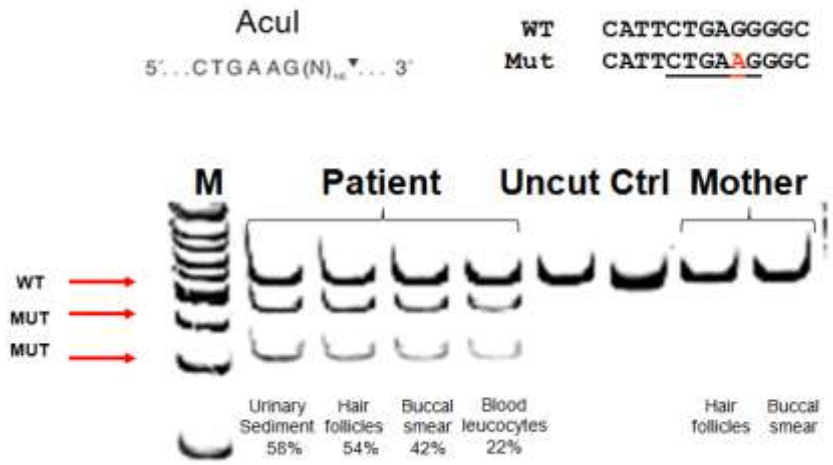


Figure 26. Heteroplasmy levels of the m.15170G>A p.Gly142 * mutation in the *MT-CYB* gene in different tissues of the patient. The analysis was made by PCR-RFLP amplifying using primers flanking the mutation and digesting with the enzyme *AcuI* (the mutation creates an AcuI site). The electrophoretic run was made using a 12% acrylamide gel, using the *1Kb DNA Ladder marker (Sigma-Aldrich).*

The mutation introduces a premature STOP codon which abolishes translation of almost 2/3 of the protein, its pathogenicity is obvious, and accounts for the phenotype of the patient.

This case highlights several important points. Nonsense MT-CYB mutation are usually restricted to skeletal muscle and cause exercise intolerance and myoglobinuria. Our findings show that they may cause systemic phenotypes affecting also the kidney. Interestingly the renal tubule was not affected as in other autosomal recessive defects of complex III, but instead we observed glomerular involvement, similar to what is seen with the 3243A>G MELAS mutation. This is the first defect in a mtDNA protein coding gene associated with glomerulopathy. In patients with glomerular involvment, in the suspicion of a mitochondrial disorder, the analysis should not be restricted to the recurring m.3243A>G mutation but it should include the entire mtDNA sequence.

4.2. Comparative analysis of NGS methods *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* and *HaloPlex HS Target Enrichment System (Agilent Technologies)* used to study human genomic DNA

In the Clinical Genetics Unit of the Hospital of Padua the NGS technology is used for the diagnosis of various rare genetic diseases. A disease is considered to be rare when it affects no more than one individual per 2000 inhabitants.

The main diseases analyzed are: metabolic diseases (from neonatal and non-neonatal screening), mitochondrial diseases from nuclear genes and mitochondrial DNA, RASopathies, neurofibromatosis type 1 and type 2, schwannomatosis, Marfan syndrome and related syndromes, long QT syndrome and related syndromes, craniofacial anomalies, peripheral neuropathies, spinal muscular atrophy, hypertrophic cardiomyopathies, deafness (syndromic and non), porphyrias.

The laboratory has a *MiSeqDx* sequencer (*Illumina*), optimized for diagnostic applications, a flexible and convenient system for small runs. Unlike the Sanger method, NGS technology allows a high degree of multiplexing, it is able to analyze millions of DNA fragments sequences of different patients in a single run, realizing a clear reduction in costs and times needed to generate sequence data. The sequencing technology that the *MiSeqDx* tool uses and which is currently adopted by most NGS platforms is known as sequencing by synthesis (SBS); it uses dNTPs labeled with four different fluorochromes and a reversible chain termination chemistry.

The NGS technologies *TruSeq Custom Amplicon Sequencing Protocols Low Input Library Prep Kit (Illumina)* and *HaloPlex HS Target Enrichment Systems (Agilent Technologies)* are two amplicon-based methods based on the generation of DNA amplicons from small amounts of the patient's genomic DNA, using customized oligonucleotides as primers for the amplification.

For the comparison between the two methods the following parameters were considered.

Coverage: The HaloPlex *HS Target Enrichment System (Agilent Technologies)* protocol provides greater coverage, in fact the percentage of gene regions with a coverage higher than 20X is greater than the *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina).*

The main difference between the two methods consists in the approach used to generate libraries of DNA fragments.

TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina) uses a pool of custom probes to capture and enrich the target regions of interest, while the *HaloPlex HS Target Enrichment System (Agilent Technologies)* technology provides a previous step of enzymatic digestion with 8 different pairs of restriction enzymes that increase the heterogeneity of the generated DNA fragments. This heterogeneity ensures greater coverage, overcoming the problem of mutations located in critical positions (e.g. the binding sites of the custom probes). That cause the loss of coverage of that specific target region. Allele dropout alone does not explain loss of coverage. The problem is intrinsic in the amplification protocol.

With *HaloPlex HS Target Enrichment System (Agilent Technologies)* each nucleotide is covered by multiple amplicons, with a different beginning and end, and if a mutation is located in a restriction site, it alters only few fragments (one or two) but not all the others; with *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* each base is covered by a single type of amplicon and if a mutation is located in the binding site of the primer the result is the complete dropout of the target region (the proportion of target regions with coverage less than 20X is 3-4 times higher than *Agilent Technology* method) (Table 15).

Gene	Disease	Actual coverage (% of coding region)		Missed Targets	
		HaloPlex HS system	TruSeq system	HaloPlex HS system	TruSeq system
GAA	Pompe disease	100	96	0	2
GBA	Gaucher disease	99.6	97	0	2*

Gene	Disease	Actual coverage (% of coding region)		Missed Targets	
		HaloPlex HS system	TruSeq system	HaloPlex HS system	TruSeq system
CPOX	Coproporphyria Harderoporphyria	99.3	94	0	1
NPC1	Niemann-Pick disease	100	97	0	2
CFTR	Cystic fibrosis	99.3	93	1	4
ATP7B	Wilson disease	100	96	0	2
NF1	Neurofibromatosis	99.1	96	1	3*
IDUA	MPS1	99.5	98.5	0	1

Table 15. Comparative analysis of NGS methods *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* and *HaloPlex HS Target Enrichment System (Agilent Technologies)* about the number of proportion of target regions with coverage less than 20X.

Results variability: *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* demonstrated a marked variability in the performance between one sequencing run and another, while for the *HaloPlex HS Target Enrichment System (Agilent Technologies)* panels the results obtained are highly reproducible.

Pseudogenes and false positives (mutations not confirmed by Sanger sequencing) discrimination: *NF1* gene mutations cause the Neurofibromatosis type 1 disease, an autosomal dominant disease whose main feature is the predisposition to the development of tumors involving the peripheral and central nervous system. *NF1* analysis has shown that the major problems of pseudogenes discrimination is obtained using the *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina).* In patients, a large number of variants are identified in the *NF1* gene (most of them are false positives) due mainly to the presence in the human genome of pseudogenes that show more than 90% of identity with the sequence of *NF1* gene.

Amplicon generation: A further limitation of the Truseq technology is the maximum size of the panels. Illumina recommends not to use panels larger than 500-600 amplicons because there is a drop in the performance for larger designs. This corresponds to 25-30 genes of average size (10 exons gene - 2 amplicons/exon). Instead, the Haloplex technology allows to design panels of more than 100 genes (for the same price of a 25-30 gene panel).

For all these reasons since 2017, we have employed only the *HaloPlex HS Target Enrichment System (Agilent Technologies)* panels.
We have optimized these panels for the novel V2 Reagent kits. Currently we are using two sets of panels "small" and "large".
Small panels comprise on average 30-40 genes, for a total sequenceable design of about 250-260 kb. These panels allow to run 10 patients on a V2 micro kit (Table) and 30 patients on a standard kit with an average coverage of at least 500x.
Such high coverage is essential to ensure that >98% of the design has at least 20x coverage. Large panels include 90-120 genes with a total sequenceable design of about 750-820 kb. These panels allow to run 10 patients on a standard kit (Table 16) with a similar performance.
The designs covered more than 99.7% of the target region and more than 98% of patients had less than 2% of the analyzable target with coverage inferior to 20x (on average this value was 99%).
It is important to note different panels can be included in the same run.

Reagent kit	N° of reads	N° of cycles	Output
MiSeq Reagent Standard Kit V2	15 M	300	4.5 Gb
MiSeq Reagent Micro Kit V2	4 M	300	1.2 Gb

Table 16. *MiSeq Reagent Kit V2 (Illumina)* commercially available with relative output values.

Each panel includes common disease genes and rarer disease genes, even of unrelated diseases, allowing to provide the results also of the most rare diagnoses quickly.

As of September 2018 this system covers more than 1300 genes and we have studied more than 1700 patients (considering only those analyzed with Haloplex assays).

Until January 2018 all mutations were confirmed by Sanger sequencing. However we realized that we ha no false positive results (provided that coverage and quality of the sequences were adequate) whereas we had 5 cases of false negative results on Sanger. All these cases were due to allele dropout because of rare polymorphisms on the PCR primer binding sites, which could not be avoided due to insufficient data in the different variant database (EVS, ExAC, and in parto also GnomAD). For these reasons, we are stopped confirming by Sanger sequencing those variants that have satisfactory quality parameters.

4.3. Neonatal metabolic screening by *HaloPlex HS Target Enrichment System (Agilent Technologies)* panels

Among all patient referred to our center to confirm the results of neonatal metabolic screening, we selected for this work 106 patients who were found positive for either Mucopolysaccharidosis type I (MPS I), Biotinidase Deficiency, or Phenylketonuria (PKU).

IDUA deficiency

Very little data is available in the literature on the molecular characterization of patients identified by neonatal screening. Some of our preliminary data were recently published (Burlina et al, 2018). We have shown that the majority of positive cases actually present pseudodeficiency, a situation caused by the presence of alleles with reduced IDUA enzymatic activity, but still compatible with a normal phenotype. Standard biochemical assays are usually not sensitive enough to discriminate these cases from true IDUA deficiency (especially from the mild forms). Several common pseudodeficiency alleles have been reported: p.Ala79Thr, p.His82Gln, and p.Asp223Asn (Pollard et al, 2017). It is important to correctly classify these individuals because they do not need enzyme replacement

therapy. We now present our complete series of patients. NGS sequencing of the *IDUA* gene in 17 patients (Table 17) revealed that 13 of them (76%) had a genotype compatible with pseudodeficiency (they carried at least one polymorphism associated with pseudodeficiency), 2 harbored known pathogenic mutations and therefore were diagnosed with MPS1 (11%), 1 patient had a novel missense change in the homozygous state which was classified as "variant of unknown significance", 1 had a single heterozygous pathogenic mutation, and in 1 patient no mutations were detected. In the 13 pseudodeficiency patients (total 26 alleles) we found 16 p.Ala79Thr alleles (61%) in 5 cases in the homozygous state, in 2 in trans with a pathogenic *IDUA* mutation, in 2 in trans with a VUS, and in 2 cases in trans with another pseudodeficiency polymorphism. Virtually all cases harboring the p.Ala79Thr variant were of African origin, consistent with the high MAF (4%) in this population (in Europeans MAF is less tha 1:15000).

Interestingly, patients 7a and 7b are dizygotic twins who both tested positive for IDUA deficiency at neonatal screening but had discordant genotypes. Both harbored the p.Ala79Thr polymorphism in the heterozygous state, patient 7a had the p.(Leu237Phe) VUS, while her sister had the p.Asp223Asn pseudodeficiency allele. After identification of patient 7a, her sister was screened by Sanger sequencing, that detected only the p.Ala79Thr variant. Because of the clear biochemical phenotype patient 7b was studied by NGS and the second allele was identified. Given the high frequency in the population (especially of African origin) of these pseudodeficiency alleles, relatives of pseudodeficiency cases found positive at newborn screening should be screened for mutation in the whole gene since they could harbor a different genotype than the index case.

Providing a clear diagnosis is indispensable to reassure the parents in case of pseudodeficiency, and to institute enzyme replacement therapy in cases with a definite diagnosis of MPS1. Overall, our NGS protocol allowed to obtain a definite result in 15/17 patients (88%), while in 2 (12%) the analysis was inconclusive. Considering that the cost of NGS is relatively low (less than 120€/sample in terms of reagents) it represents a convenient diagnostic technique for patients identified by neonatal screening and is essential to avoid unnecessary treatments (which are very expensive) in individuals with pseudodeficiency.

Pt	NGS analysis results	Hom/Het	Clinical significance
1	c.235G>A p.Ala79Thr + c.1081G>A p.Ala361Thr in cis	het	pseudodeficiency
	c.1743C>G p.(Tyr581*)	het	pathogenic
2	c.235G>A p.Ala79Thr + c.1081G>A p.Ala361Thr in cis	hom	pseudodeficiency
3	c.46_57del p.(Ser16_Ala19del)	het	pathogenic
	c.246C>G p.His82Gln	het	pseudodeficiency
4	c.235G>A p.Ala79Thr	hom	pseudodeficiency
5	c.235G>A p.Ala79Thr	hom	pseudodeficiency
6	c.1598C>G p.Pro533Arg	hom	pathogenic
7a	c.235G>A p.Ala79Thr	het	pseudodeficiency
	c.709C>T p.(Leu237Phe)	het	uncertain significance
7b	c.235G>A p.Ala79Thr	het	pseudodeficiency
	c.667G>A p.Asp223Asn	het	pseudodeficiency
8	c.1577T>C p.(Leu526Pro)	hom	uncertain significance likely pseudodeficiency (high MAF)
9	c.235G>A p.Ala79Thr	hom	pseudodeficiency
10	c.46_57del p.(Ser16_Ala19del)	het	pathogenic
	c.603C>G p.(Tyr201*)	het	pathogenic
11	c.1501T>C p.(Phe501Leu)	het	uncertain significance
	c.235G>A p.Ala79Thr	het	pseudodeficiency
12	c.1882C>G p.(Arg628Gly)	hom	uncertain significance
13	c.235G>A p.Ala79Thr	hom	pseudodeficiency
14	c.235G>A p.Ala79Thr	hom	pseudodeficiency

Pt	NGS analysis results	Hom/Het	Clinical significance
15	c.235G>A p.Ala79Thr	het	pseudodeficiency
	c.667G>A p.Asp223Asn	het	pseudodeficiency
16	No mutations detected		
17	c.1598C>G p.Pro533Arg no second mutation identified	het	pathogenic

Table 17. List of patients (Pt) positive at neonatal metabolic screening for Mucopolysaccharidosis type I. Patients marked in grey harbored known pathogenic mutations and therefore were diagnosed with MPS1; in patient marked in pink no mutations were detected; patients marked in yellow had a novel missense change in the homozygous state or had a single heterozygous pathogenic mutation.

BTD deficiency

NGS analysis of *BTD* gene of 32 patients showed that 5 patients (16%) had a genotype compatible with the classical form of the disease. 21 (68%) had partial deficiency (they were compound heterozygotes for a severe mutation and for the mild variant p.Asp444His); 3 patients were heterozygotes for a single pathogenetic mutation, 1 was homozygote for p.Asp444His, and in 1 patient two variants were found (p.Ala171Thr and p.Asp444His) but their phase could not be confirmed (Table 18). In 1 patient no variants were detected.

All patients with a severe deficiency were correctly genotyped, as well as the majority of those with partial deficiency. We do not have follow up data for all our cases but at least in two among those with an inconclusive analysis the deficiency was not confirmed at follow up.

Most of the variants are located in exon 4 of the gene therefore we considered the possibility of performing as initial screening for these patients Sanger sequencing of exon 4. This would have allowed a correct diagnosis in 18/32 patients (56%). However we found the p.Asp444His variant in cis with a pathogenic mutation in the 5' region of the gene in at least 2 different complex alleles, in a total of 3 patients, who would have been incorrectly diagnosed using this strategy. For this reason we rejected this strategy, and mutation screening is still performed by

NGS. Identifying the patient's genotype is essential for a correct treatment. We are currently treating only those patients with 2 severe mutations.

Pt	NGS analysis results	Hom/Het	Clinical significance	BTD deficiency
1	c.1368A>C p.Gln456His	het	pathogenic	severe
	c.1439G>A p.(Gly480Glu)	het	pathogenic	
2	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1595C>T p.(Thr532Met)	het	pathogenic	
3	c.278A>G p.Tyr93Cys	het	pathogenic	partial
	c.1330G>C p.Asp444His	het	polymorphism	
4	c.1330G>C p.Asp444His	hom	polymorphism	partial
5	[c.133G>A +c.980dup] [p.Gly45Arg + p.Asn327Lysfs*27] hetero	het	pathogenic	partial
	c.1330G>C p.Asp444His	het	polymorphism	
6	c.511G>A p.Ala171Thr	het	pathogenic	partial
	c.1330G>C p.Asp444His (phase unknown)	het	polymorphism	
7	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
8	No mutations detected			partial
9	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
10	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
11	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	

Pt	NGS analysis results	Hom/Het	Clinical significance	BTD deficiency
12	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
13	c.631C>T p.(Arg211Cys) no second mutation identified	het	pathogenic	partial
14	c.98_104delinsTCC p.Cys33Phefs*36	het	pathogenic	partial
	c.1330G>C p.Asp444His	het	polymorphism	
15	c.98_104delinsTCC p.Cys33Phefs*36 no second mutation identified	het	pathogenic	partial
16	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1466A>G p.(Asn489Ser)	het	pathogenic	
17	c.382T>G p.(Phe128Val)	het	pathogenic	partial
	c.1330G>C p.Asp444His	het	polymorphism	
18	c.1368A>C p.Gln456His	het	pathogenic	severe
	c.1489C>T p.(Pro497Ser)	het	pathogenic	
19	c.[1024delG ; 1330G>C] [p.Val442Serfs*59; p.Asp444His]	het	pathogenic	severe
	c.1368A>C p.Gln456His	het	pathogenic	
20	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
21	c.594_596delCGT p.Val199del	het	pathogenic	partial
	c.1330G>C p.Asp444His	het	polymorphism	
22	c.1595C>T p.(Thr532Met) no second mutation identified	het	pathogenic	partial
23	c.1368A>C p.Gln456His	hom	pathogenic	severe

Pt	NGS analysis results	Hom/Het	Clinical significance	BTD deficiency
24	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
25	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
26	c.[511G>A ; 1330G>C] p.[Ala171Thr; Asp444His]	het	pathogenic	partial
	c.1330G>C p.Asp444His	het	polymorphism	
27	c.98_104delinsTCC p.Cys33Phefs*36	het	pathogenic	partial
	c.1330G>C p.Asp444His	het	polymorphism	
28	c.98_104delinsTCC p.Cys33Phefs*36	het	pathogenic	partial
	c.1330G>C p.Asp444His	het	polymorphism	
29	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
30	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	
31	c.334G>A p.Glu112Lys	het	pathogenic	severe
	c.515A>G p.Asn172Ser	het	pathogenic	
32	c.1330G>C p.Asp444His	het	polymorphism	partial
	c.1368A>C p.Gln456His	het	pathogenic	

Table 18. List of patients (Pt) positive for neonatal metabolic screening for Biotinidase Deficiency. Patients marked in grey harbored known pathogenic mutations and therefore were diagnosed with BTD deficiency; patients marked in yellow had a single heterozygous pathogenic mutation or had two variants but their phase could not be confirmed; in patient marked in pink no mutations were detected.

PAH deficiency

The NGS sequencing protocol for the PAH gene could not adequately cover exon 7 of the gene, therefore all patients with a single (or without) mutations were also analysed by Sanger sequencing of this exon. We studied a total of 57 patients (Table 19).

Fortythree patients (75%) had two mutarions (8 novel mutations). All novel mutations were considered likely pathogenic based on MAF, conservations, and molecular modelling data.

Eleven patients (19%) carried only one mutation in heterozygous state. Eight (for the other 3 no adequate DNA was available) of them were also studied by MLPA analysis: 2 are heterozygous for the exon 3 deletion while in the other 6 no deletions were found. It is important to note that we found one patient with a homozygous deletion of exon 3, confirmed by MLPA, indicating that such rearrangements are not rare in our population.

In 2 patients no variants were detected and they were also negative for MLPA analysis.

Pt	NGS analysis results	Region with coverage <20X	Hom/Het	Clinical significance	MLPA
1	No mutations detected	No mutations detected			negative
2	c.898G>T p.Ala300Ser		het	pathogenic	
	c.1208C>T p.Ala403Val		het	pathogenic	
3	c.824C>G p.(Pro275Arg)		het	likely pathogenic	
	c.1222C>T p.Arg408Trp		het	pathogenic	
4	c.168+5G>C		het	pathogenic	negative
		no second mutation identified			
5	c.143T>C p.Leu48Ser		hom	pathogenic	
6	c.842C>T p.Pro281Leu		het	pathogenic	exon 3 deletion (het)
		no second mutation identified			
7	c.1208C>T p.Ala403Val		het	pathogenic	
	c.1315+1G>A		het	pathogenic	

Pt	NGS analysis results	Region with coverage <20X	Hom/Het	Clinical significance	MLPA
8	c.898G>T p.Ala300Ser		het	pathogenic	no adequate DNA
		no second mutation identified			
9		c.734T>C p.Val245Ala	het	pathogenic	
		c.755G>A p.(Arg252Gln)	het	pathogenic	
10	c.526C>T p.(Arg176*)		het	pathogenic	
	c.898G>T p.Ala300Ser		het	pathogenic	
11	c.1163T>C p.(Val388Ala)		hom	likely pathogenic	
12	c.1243G>A p.(Asp415Asn)		het	pathogenic	negative
		no second mutation identified			
13	c.1315+1G>A		het	pathogenic	negative
		no second mutation identified			
14	c.473G>A p.Arg158Gln		het	pathogenic	
	c.526C>T p.(Arg176*)		het	pathogenic	
15	c.842+3G>C		het	pathogenic	
	c.1208C>T p.Ala403Val		het	pathogenic	
16	c.473G>A p.Arg158Gln		het	pathogenic	
	c.1315+1G>A		het	pathogenic	
17	c.1208C>T p.Ala403Val		het	pathogenic	
		c.1066-6 G>A	het	likely pathogenic	
18	c.842+3 G>C		het	pathogenic	
	c.898G>T p.Ala300Ser		het	pathogenic	
19	c.143T>C p.Leu48Ser		het	pathogenic	
	c.1223G>A p.Arg408Gln		het	pathogenic	
20	c.1315+1G>A		hom	pathogenic	

Pt	NGS analysis results	Region with coverage <20X	Hom/Het	Clinical significance	MLPA
21	c.842C>T p.Pro281Leu		het	pathogenic	
	c.1208C>T p.Ala403Val		het	pathogenic	
22	c.473G>A p.Arg158Gln		hom	pathogenic	
23	c.782G>A p.Arg261Gln		het	pathogenic	
		c.734T>C p.Val245Ala	het	pathogenic	
24	exon 3 deletion		hom	pathogenic	exon 3 deletion
25	c.1139C>T p.Thr380Met		het	pathogenic	negative
		no second mutation identified			
26	c.506G>A p.(Arg169His)		het	likely pathogenic	negative
		no second mutation identified			
27	c.1222C>T p.Arg408Trp		het	pathogenic	
	c.1315+1G>A		het	pathogenic	
28	c.632C>T p.(Pro211Leu)		hom	likely pathogenic	
29	c.782G>A p.Arg261Gln		het	pathogenic	
	c.842C>T p.Pro281Leu		het	pathogenic	
30	c.1089del p.(Lys363Asnfs*37)		het	pathogenic	
	c.1315+1G>A		het	pathogenic	
31	c.863T>G p.(Ile288Trp)		het	likely pathogenic	
	c.1315+1G>A		het	pathogenic	
32	c.842C>T p.Pro281Leu		het	pathogenic	
	c.898G>T p.Ala300Ser		het	pathogenic	
33	c.434A>T p.Asp145Val		het	pathogenic	
	c.1054delC p.(Gly352Valfs*48)		het	pathogenic	
34	c.842+3G>C		het	pathogenic	negative
		no second mutation identified			

Pt	NGS analysis results	Region with coverage <20X	Hom/Het	Clinical significance	MLPA
35	c.473G>A p.Arg158Gln		het	pathogenic	
	c.1315+1G>A		het	pathogenic	
36		c.734T>C p.Val245Ala	het	pathogenic	
	c.1208C>T p.Ala403Val		het	pathogenic	
37	c.143T>C p.Leu48Ser		het	pathogenic	
	c.1159T>C p.(Tyr387His)		het	likely pathogenic	
38	c.673C>A p.(Pro225Thr)		het	pathogenic	
	c.764T>C p.(Leu255Ser)		het	pathogenic	
39	No mutations detected	No mutations detected			negative
40	c.529G>A p.Val177Met		het	pathogenic	
	c.1222C>T p.Arg408Trp		het	pathogenic	
41	c.194T>A p.(Ile65Asn)		het	likely pathogenic	
	c.1045T>C p.(Ser349Pro)		het	pathogenic	
42	c.842C>T p.Pro281Leu		het	pathogenic	exon 3 deletion (het)
		no second mutation identified			
43	c.782G>A p.Arg261Gln		het	pathogenic	
	c.1208C>T p.Ala403Val		het	pathogenic	
44		c.734T>C p.Val245Ala	het	pathogenic	
	c.1208C>T p.Ala403Val		het	pathogenic	
45	c.1208C>T p.Ala403Val		hom	pathogenic	
46	c.284_286delTCA p.(Ile95del)		het	pathogenic	
	c.505C>T p.(Arg169Cys)		het	likely pathogenic	
47	c.473G>A p.Arg158Gln		het	pathogenic	
	c.898G>T p.Ala300Ser		het	pathogenic	

Pt	NGS analysis results	Region with coverage <20X	Hom/Het	Clinical significance	MLPA
48	c.782G>A p.Arg261Gln		het	pathogenic	
		c.734T>C p.Val245Ala	het	pathogenic	
49	c.782G>A p.Arg261Gln		het	pathogenic	no adequate DNA
		no second mutation identified			
50	c.143T>C p.Leu48Ser		hom	pathogenic	
51	c.898G>T p.Ala300Ser		het	pathogenic	no adequate DNA
		no second mutation identified			
52	c.842C>T p.Pro281Leu		het	pathogenic	
	c.1208C>T p.Ala403Val		het	pathogenic	
53	c.158G>A p.(Arg53His)		het	likely pathogenic	
	c.1055delG p.(Gly352Valfs*48)		het	pathogenic	
54	c.1222C>T p.Arg408Trp		hom	pathogenic	
55	c.1222C>T p.Arg408Trp		hom	pathogenic	
56	c.158G>A p.(Arg53His)		het	likely pathogenic	
	c.782G>A p.Arg261Gln		het	pathogenic	
57	c.158G>A p.(Arg53His)		het	likely pathogenic	
	c.611A>G p.(Tyr204Cys)		het	pathogenic	

Table 19. List of patients (Pt) positive for neonatal metabolic screening for Phenylketonuria. Patients marked in grey were diagnosed with PKU; patients marked in yellow had a single heterozygous pathogenic mutation; in patient marked in pink no mutations were detected. Novel mutations are reported in red.

We plan to complete the MLPA analysis in the remaining patients and to analyse cDNA in all negative/single mutation patients.

We are still gathering clinical data, in particular phenylalanine levels at neonatal screening and during follow up. This will allow to establish an algorhythm for those patients lacking a molecular diagnosis.

4.4. Identification of the IVS4+1G>A new mutation in the *COQ4* mitochondrial nuclear gene using *HaloPlex HS Target Enrichment System (Agilent Technologies)* technology and its validation by hybrid minigene system

Since several years the Clinical Genetics Unit invests in the study of the molecular mechanisms involved in the coenzyme Q biosynthesis pathway and in the characterization of the new mutations identified in patients with coenzyme Q deficiency syndrome.

Using the NGS sequencing method *HaloPlex HS Target Enrichment System (Agilent Technologies)* it was identified a new *COQ4* gene mutation that involving the first nucleotide after exon 4 of the gene.

To verify the possible impact of that nucleotide variation, it was first analyzed by a predictive bioinformatics software able to predict the probability of variant to cause alterations in the splicing mechanism.

The software used they were *Human Splicing Finder* (www.umd.be/HSF3) and *Mutation Taster* (www.mutationtaster.org). Bioinformatics analysis predicted an alteration of splicing.

To validate the pathogenicity of IVS4+1G>A variant with uncertain significance has been amplified by PCR the genomic DNA of the patient (a fragment comprising the exon of the *COQ4* gene with the mutation and a portion of at least 100 nucleotides of the flanking intronic regions). The product of PCR was subsequently cloned into the minigene vector at the level of the *β-globin* intron 2 (Figure 27).

After verified by direct sequencing the correctness of the constructs, the selected plasmids (one wild-type, one mutated) were used to transfect HEK 293 cells. From the latter it was extracted the total RNA that has been retrotranscribed in cDNA and amplified using specific primers. The products of amplification were then separated and displayed on agarose gel at 2%.

From the electrophoretic run it emerged that the wild-type minigene expressed in HEK 293 cells it determines a correct splicing, while the expression of the mutated minigene produces the same band of the wild-type construct and a smaller one (Figure 28).

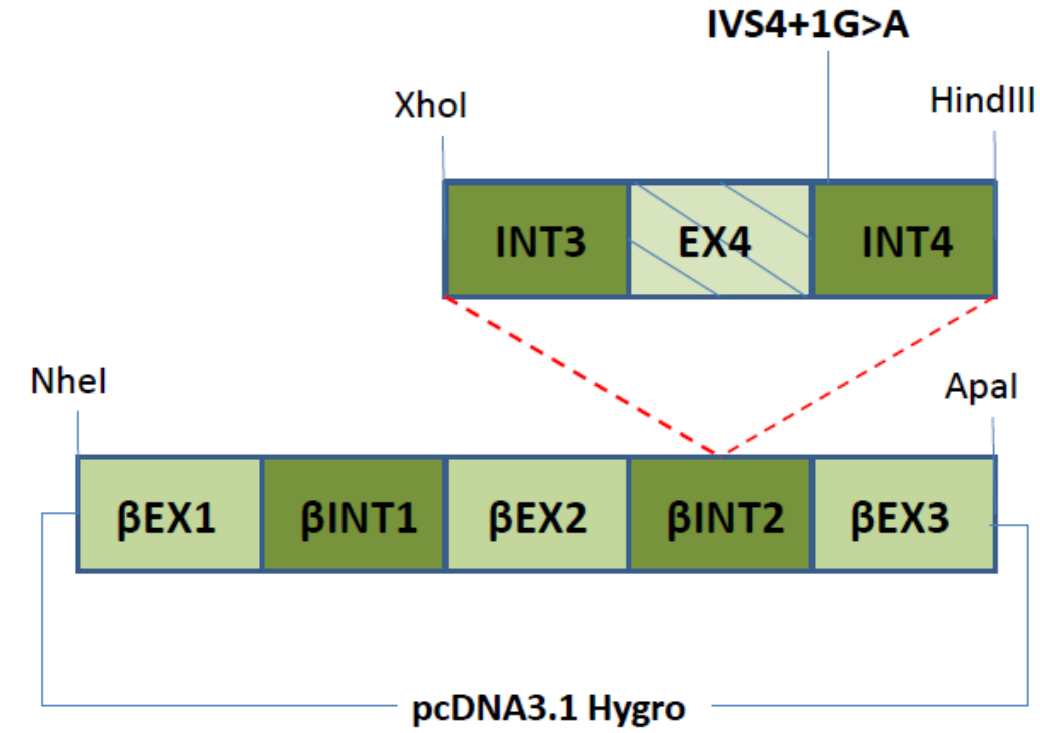
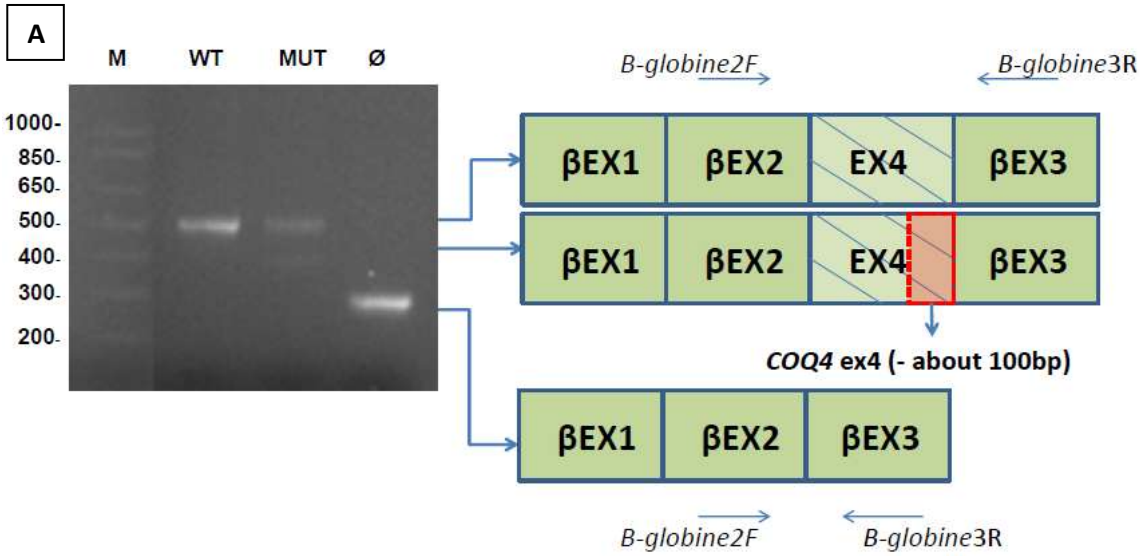


Figure 27. IVS4+1G>A mutation cloning scheme.



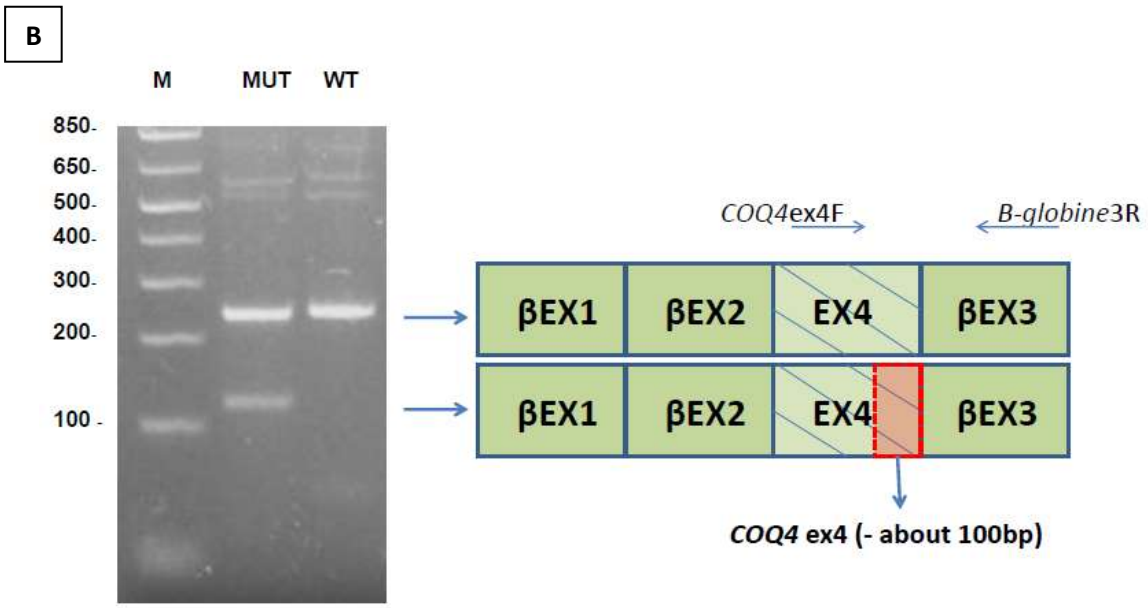


Figure 28. 2% agarose gel electrophoresis analysis. [A] Amplification performed using *β-globine* exon 2 forward primer and *β-globine* exon 3 reverse primer. The band of the wild-type construct (WT) has dimensions equal to 505 bp, the band of the mutated (MUT) construct loses about 100 nucleotides of the exon 4 downstream of the mutation and the band of the empty vector is 284 bases. [B] To confirm the result it has been performed also an amplification using *COQ4* exon 4 forward primer and *β-globine* exon 3 reverse primer. The band of the wild-type construct (WT) has dimensions equal to 250 bp while the band of the mutated (MUT) construct loses about 100 nucleotides.

This result suggests that this variant behaves like a hypomorph allele as it allows to produce part of the wild-type transcript.

The result obtained with the minigene shows that the mutation causes the activation of a cryptic splicing site of the exon 4 and agrees with what is hypothesized through the prediction software.

The hybrid-minigene technique used in this study proved to be an effective and relatively simple method for the validation of variants potentially capable of altering splicing, in the absence of RNA of the patient.

On the other hand this technique shows some limitations given by the fact that requires the use of an artificial hybrid gene as a point of departure and that some anomalies in splicing could be tissue-specific. These problems are actually partly solvable: in the first case it is the technique itself that allows to evaluate the correct construction of the artificial hybrid gene through the comparison with the wild-type construct (if this is correctly processed it means in fact that the minigene respects all the necessary regulatory elements to the spliceosoma for a

correct maturation of the transcript and that the alterations detected in the mutated construct are attributable exclusively to the presence of the mutation). Regarding instead the possible tissue-specific expression, it is possible to use for the analysis the cell lines derived from tissues in which the mutation exerts its pathogenic effects more.

4.5. Saccharomyces *cerevisiae* as a model for the validation of *ASS1* gene mutations identified in patients with citrullinemia type I

As mentioned previously, we designed a *HaloPlex HS Target Enrichment System (Agilent Technologies)* panel for the analysis of the genes involved in hereditary metabolic diseases, including also urea cycle diseases.

Previous studies (Trevisson et al., 2009; Doimo et al., 2012) have revealed how the yeast complementation allows the evaluation of residual activity of the mutant forms of the enzyme argininosuccinate lyase, belonging to the urea cycle: for this reason we have decided to apply the same method also for the identification of residual activity of the ASS1 mutants, in order to establish possible genotype-phenotype correlations.

S. *cerevisiae* model system was used to verified that the heterologous expression of *ASS1* human gene was able to correct the growth defect in arginine-free medium of a yeast strain bearing a deletion of the corresponding homologue gene *ARG1* (*ΔARG1*).

Subsequently, various missense mutations (G14S, W179R, R265H, R304W, G324S, R363L and G390R) were introduced in *ASS1* human gene by site-specific mutagenesis. The ability of the various *ASS1* mutant alleles to restore the growth of the ΔARG1 yeast strain was then analyzed with functional complementary experiments, testing growth in selective medium under different conditions of expression.

Table 20 shows the genotype of patients taken from literature with the mutations analyzed and the corresponding clinical picture.

Patient	Genotype	Clinic features	Activity % ASS1	Reference
1	G14S/G324S	asymptomatic	n.a.	Haberle 2003
2	G14S/ c.1128-6_1188dup67	asymptomatic	n.a.	Lee 2013
3	W179R/W179R	asymptomatic	26 7 20	Haberle 2002 Haberle 2002 Dimmock 2008
4	W179R/W179R	mild	7.5	Haberle 2002
5	W179R/R279X	asymptomatic *	n.a.	Dimmock 2008
6	W179R/R307C	asymptomatic	n.a.	Haberle 2003
7	R265H/R265H	mild	n.a.	Gao 2003
8	R265H/R279X	mild	n.a.	Gao 2003
9	R265H/952delG	mild	n.a.	Gao 2003
10	R304W/R304W	neonatal	n.a.	Kakinoki 1997 Lee 2013
11	R304W/ c.1128-6_1188dup67	neonatal	n.a.	Lee 2013
12	R304W/IVS-6-2A>G	neonatal	n.a.	Kakinoki 1997
13	R304W/ΔExon7	neonatal	2.1 7	Kobayashi 1994
14	R304W/R363L	neonatal	<5	Kobayashi 1994
15	G324S/G324S	neonatal	n.a.	Gao 2003
16	G324S/G324S	asymptomatic **	n.a.	Haberle 2003
17	G324S/ IVS-6-2A>G	neonatal	n.a.	Lee 2013
18	G324S/V263M	asymptomatic	0.5	Dimmock 2008
19	G324S/R363W	Neonatal	n.a.	Kobayashi 1990
20	R363L/R304W	neonatal	<5	Kobayashi 1994
21	G390R/G390R	neonatal	<1	Gao 2003 Haberle 2003 Vilaseca 2001
22	G390R/G117D	mild	n.a.	Gao 2003 Vilaseca 2001
23	G390R/Y190D	asymptomatic	n.a.	Haberle 2003
24	G390R/W179R	mild	n.a.	Gao 2003
25	G390R/R272H	asymptomatic	n.a.	Dimmock 2008
26	G390R/Y291S	mild	0	Dimmock 2008
27	G390R/T389I	mild	<10	Gao 2003

Table 20. List of missense mutations analyzed, associated clinical phenotypes and activity of the ASS1 enzyme measured by an indirect assay with 14C-citrulline according to Kleijer et al. 1984.
* Patient 5 was treated with a restricted protein diet.
** Patient 16 was subjected to a restricted protein diet, supplementation of arginine, isoleucine and treated with phenylbutyrate.

The mutations studied involve residues with different conservation degree in the species (Figure 29).

The 3D structure of ASS1 protein was analyzed *in silico* in order to understand the effects of the mutations studied: they can interfere with the active site, with the interaction between monomers or destabilize the protein peripheral loops.

In our study we considered the crystallized structure of the bacterial argininosuccinate synthase enzyme of Thermus *thermophilus*, a orthologian of the human ASS1 with a homology degree of 53% (Karlberg et al., 2008). It was crystallized as a tetramer with its substrates , including ATP (Goto et al., 2002).

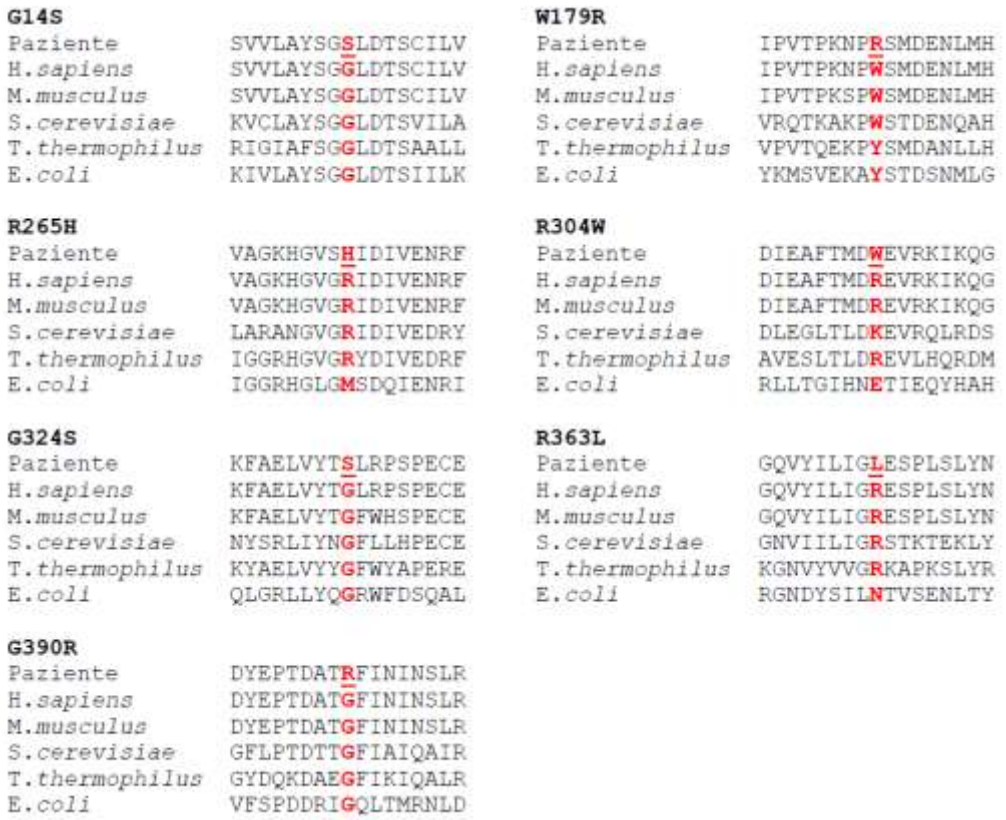


Figure 29. Alignments of patients's sequences with the mutations studied, with the protein's sequences in the different species.

Table 21 summarizes the possible effects of missense mutations analyzed, hypothesized according to molecular modeling and to what reported by Gao et al., 2003.

Mutation	Mutation locus/possible effect of the mutation
G14S	Active site/ATP binding
W179R	β-*sheet* 9, close to active site/citrulline binding
R265H	close to active site / Folding of the synthesis domain
R304W	Dimer-dimer interactions
G324S	Elica α10 breaker/Dimer-dimer interactions
R363L	β-*sheet* 16 close to the dimer's interface/Dimer-dimer interactions
G390R	C-terminal α-helix/oligomerization

Table 21. Mutations studied and their localization in the protein.

4.5.1. Functional complementation of ASS1 in yeast

The studies of functional complementation in model organisms are useful in order to analyze complex metabolic pathways and to validate the pathogenicity of missense mutations present in patients affected by genetic diseases.

Functional complementation occurs when a gene product (in our study the human gene *ASS1*) is able to compensate for the lack of a gene (in our study the yeast gene *ARG1*).

To be able to use this approach it is necessary that the heterologous expression allows to restore the activity of the gene deleto, and that the corresponding phenotype is easily measurable (in our case growth/non-growth of the yeast strain for *ARG1* in selective medium without arginine). This creates a simple model that can be used to study the pathogenicity of the different mutants, evaluating their residual activity and determining possible genotype-phenotype correlations.

The protein encoded by the *ASS1* gene is highly conserved in the different species and has a 51% identity with the corresponding orthologian encoded by the yeast *ARG1* gene (Figure 18).

With the functional complementation test we have shown that *ASS1* is able to partially complement the haplous yeast strain bearing a deletion of the *ARG1* gene and consequently auxotrophic for arginine.

The coding region of the *ASS1* gene and the S. *cerevisiae ARG1* homologue gene were subcloned in the galactose-inducible pYES.2 expression vector containing the selection marker *URA3*.

The deaf yeast strain for the *ARG1* gene was transformed with the two constructs (pYES.2_*ASS1* and pYES.2_*ARG1*) and with the empty vector (pYES.2 ø) as a negative control.

After inoculating the yeasts transformed into minimal medium containing 2% galactose and the amino acids for which the strain is auxotrophic (SM GAL2% HMLR), they were incubated overnight at 30 ° C to promote the over-expression of the recombinant genes .

The growth test (Drop test) was performed in minimum medium containing 2% galactose, or 2% raffinose added with 0.1% galactose, to which were added the amino acids for which the strain is auxotrophic but not the arginine (SM GAL2% HML and SM RAF2% -GAL0.1% HML). As a control the same test was performed also in the presence of arginine (SM GAL2% HMLR).

The growth occurred in these conditions is indicative of the successful functional complementation, which in the specific case of the *ASS1* gene appears to be partial: the restoration of the growth defect of the *ASS1*-delete strain is observed after about 10 days and is evident until the fourth dilution (10^{-3}) under over-expression (Figure30A) and up to the second dilution (10^{-1}) at baseline expression conditions (Figure 30B). As a further check, the test was performed also in the medium to which arginine was added (Figure 30BC).

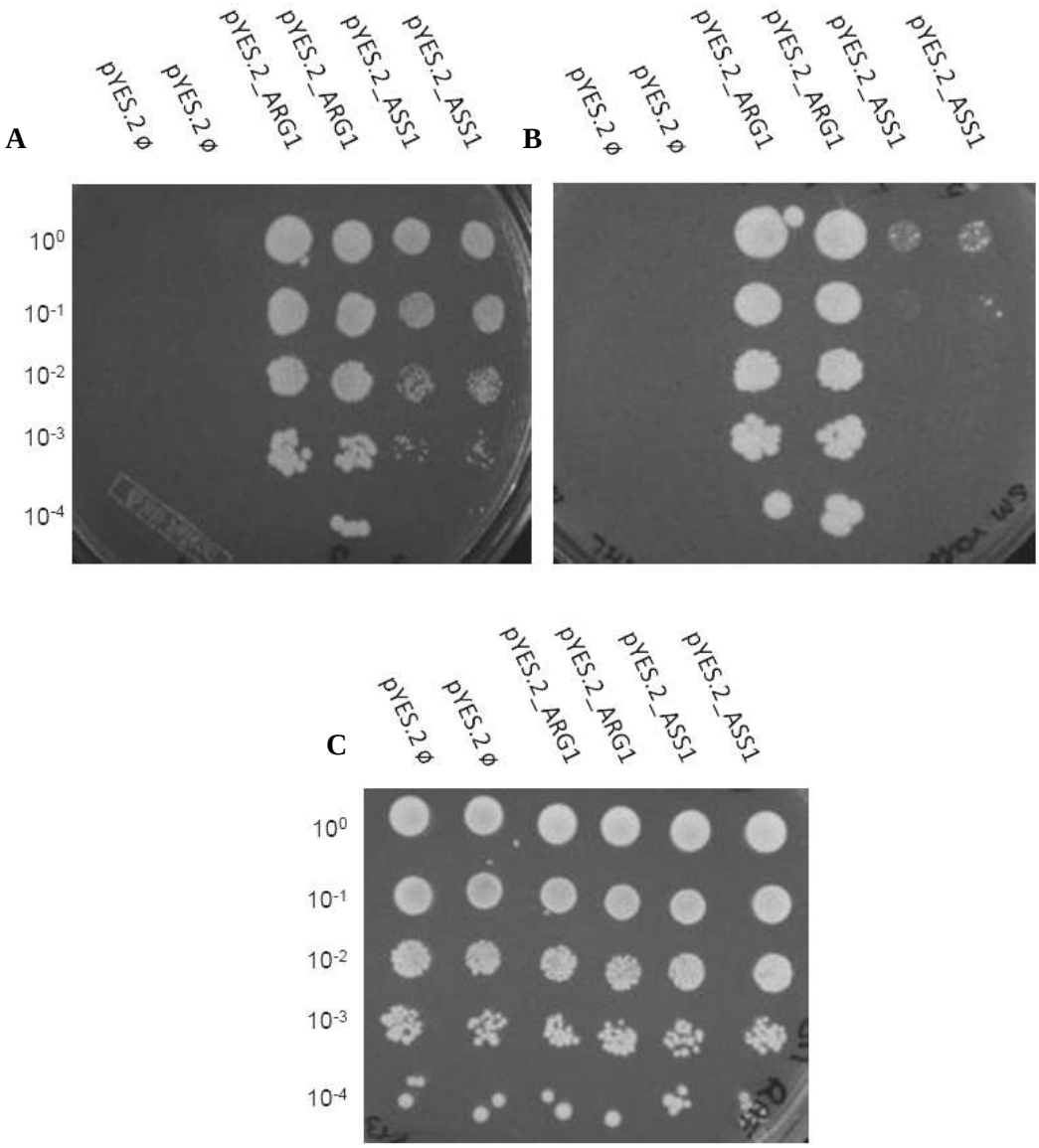


Figure 30. Functional complementation test in arginine-free medium containing 2% galactose (SM GAL2% HLM) [A] and in arginine-free medium containing 2% raffinose and 0.1% galactose (SM RAF2% GAL 0. 1% HLM) [B], to evaluate the complementary capacity of the transformed strains. The vector containing the ARG1 yeast gene was used as a positive control and the empty pYES.2 vector was a negative control. [C] The same experiment was repeated in the presence of arginine. The plates were incubated at 28 ° C for about 10 days.

4.5.2. ASS1 has no toxic effect in yeast

It was then verified whether the partial complementation of *ASS1* achieve any toxic effects of human gene expression in yeast.

For that reason, the BY4741 *MatA* wild type yeast strain was transformed with the pYES.2_*ASS1* and pYES.2_*ARG1* vectors and the growth test (Drop test) was performed under the different conditions used in the previous functional complementation test (Figure 31).

The wild type strain transformed with the empty pYES.2 vector was used as a control.

The growth of the *ASS1* transformed strain with is comparable to the control and to *ARG1*, this excludes that the partial complementation of the human gene in yeast is due to toxicity phenomena.

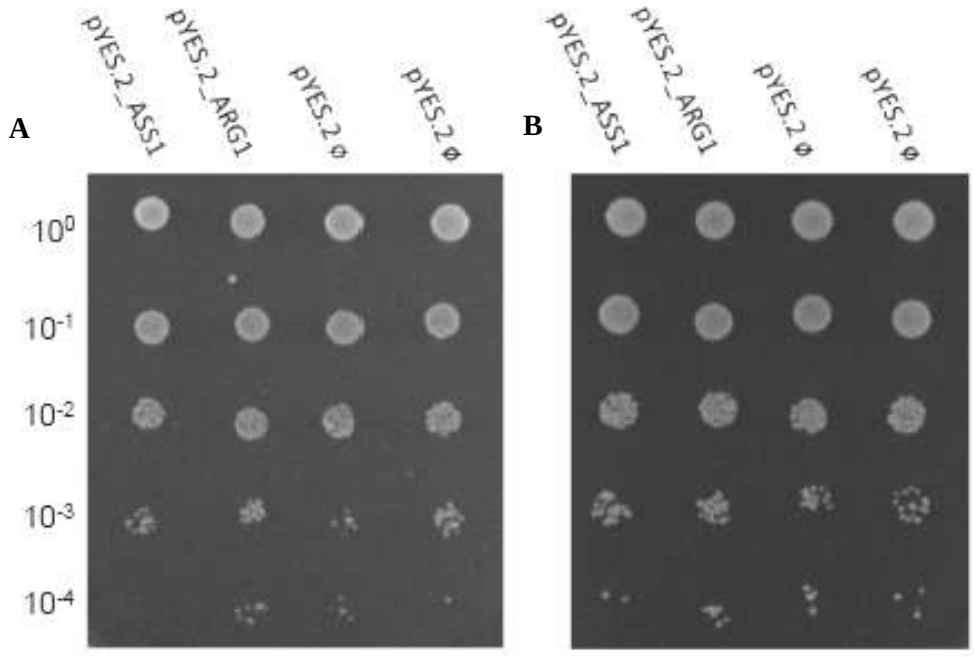


Figure 31. Functional complementation test in arginine-free medium containing 2% galactose (SM GAL2% HLM) [A] and in arginine-free medium containing 2% raffinose and 0.1% galactose (SM RAF2% GAL0. 1% HLM) [B] to evaluate the complementary capacity of the BY4741 *MatA* wild-type strains transformed with the indicated constructs. The plate was incubated for 6 days at 28 ° C.

4.5.3. Effect of *ASS1* gene mutations on yeast complementation

Functional complementation experiments were then performed using the *ΔARG1* yeast strain to study the growth phenotype of some missense mutations identified in patients with CTLN1. This test is based on the fact that the growth of the *ARG1* delete yeast strain in minimal arginine-free medium can occur only in the presence of residual enzymatic activity of ASS1.

The seven mutations reported in this study were introduced into the pCR8 TOPO TA vector by site-specific mutagenesis and were subsequently transferred into the pYES.2 expression vector by LR recombination reaction (*Gateway* system). These constructs were used to transform the *ΔARG1* yeast strain and the ability of the transformants to grow in the absence of arginine under different expression conditions was analyzed (2% galactose or 2% raffinose added with 0.1% galactose). As controls, the delete strain was also transformed with the yeast *ARG1* gene, with the human wild type *ASS1* gene and with the empty vector.

In these conditions, different growth profiles were observed (Figure 32) that allow to distinguish the alleles analyzed in two classes:

- Class I mutations: absence of growth in minimal medium, both in over-expression conditions (Figure 32A) and in basal conditions (Figure 32B). The G14S, W179R, R363L and G390R mutations belong to this class.
- Class II mutations: residual growth in minimal medium under over-expression conditions (Figure 32A) and marked reduction/abolition of growth at basal conditions (Figure 32B). The R265H, R304W and G324S mutations belong to this class.

The test was performed also in the presence of arginine (Figure 32E).

To evaluate the yeast phenotype transformed with the G324S mutant expressing construct, which showed a weak growth at the first dilution, the plates were incubated at 28 °C for 5 additional days. As shown in Figure 32, this allele allows residual growth of the yeast (evident until the second dilution) only after induction with galactose (Figure 32C), but not under basal conditions (Figure 32D).

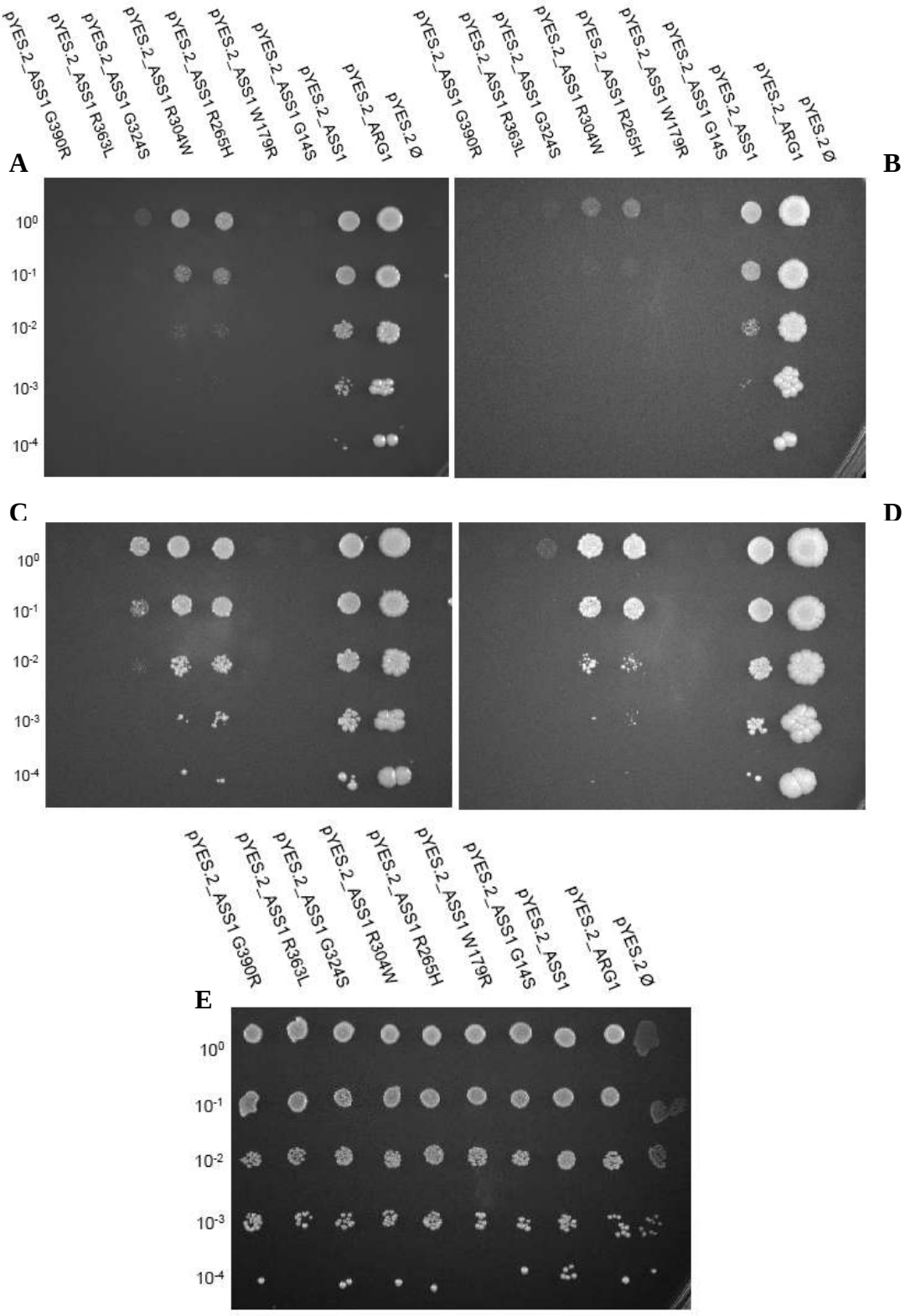


Figure 32. Functional complementation test in arginine-free medium containing 2% galactose (SM GAL2% HLM) [A] [C] and in arginine-free medium containing 2% raffinate and 0.5% galactose (SM RAF2 % GAL0.1% HLM) [B] [D] to evaluate the complementary capacity of *ASS1* mutated transformed strains. *ASS1* wild-type was used as a positive control, and pYES.2 empty as negative control. The plate was incubated at 28 °C for 9 [A] and [B] or for 15 days [C] and [D]. The same experiment was repeated in the presence of arginine [E], incubating the plates at 28 °C for about 10 days.

4.5.4. Effect of ASS1 mutations on protein stability

After the complementation analysis of the various mutations in different conditions of expression (using media containing galactose and raffinose), we examined the levels of ASS1 protein produced by the different mutant alleles, by western blot analysis (Figure 33).

The densitometric analysis performed shows that the protein encoded by the R304W allele is stable, those expressed by the alleles R265H and G324S are partially stable, while the mutations G14S, W179R, R363L and G390R considerably destabilize the protein, as demonstrated by the absence or the markedly reduced levels of protein shown by the immunoblot.

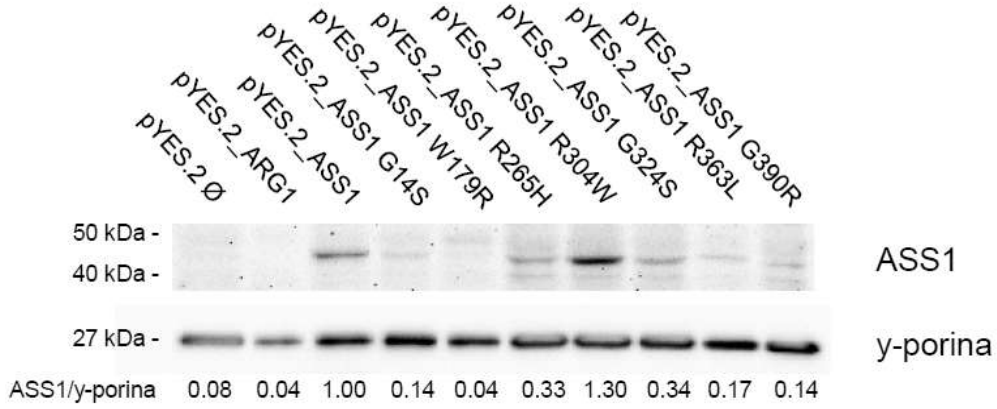


Figure 33. Western blot analysis to measure the stability of wild-type and mutated ASS1 protein. The proteins were extracted from the respective yeasts and subsequently detected using the corresponding antibody. The densitometric analysis was performed using the *ImageJ* program, normalizing the data by referring to pYES.2_*ASS1* wild-type.

5. CONCLUSIONS

In the last 3 years in the Genetic Unit of Hospital of Padua we have worked to introduce the Next Generation Sequencing technologies into clinical routine.

The comparative analysis of the NGS technologies *TruSeq Custom Amplicon Low Input Library Prep Kit (Illumina)* and *HaloPlex HS Target Enrichment System (Agilent Technologies)* showed how the second one is more efficient in diagnostic, considering the followed parameters: coverage, results variability, pseudogenes and false positive discrimination, number of amplicons generated for the analysis. For these reason, from 2017 we use only the *HaloPlex HS Target Enrichment System (Agilent Technologies).*

The main diseases analyzed in our laboratory are: metabolic diseases (from neonatal and non-neonatal screening), mitochondrial diseases from nuclear genes and mitochondrial DNA, RASopathies, neurofibromatosis type 1 and type 2, schwannomatosis, Marfan syndrome and related syndromes, long QT syndrome and related syndromes, craniofacial anomalies, peripheral neuropathies, spinal muscular atrophy, hypertrophic cardiomyopathies, deafness (syndromic and non), porphyrias. The new generation sequencing technologies represent a valid strategy for the analysis of these diseases, allowing to significantly reduce the time and costs involved in the analysis, in fact the NGS technology turns out to be the most rapid approach for the analysis of rare genetic diseases because it allows to study several variants in genes (that could be also of large dimensions) related to the pathology at the same time. In contrast at the conventional Sanger sequencing approach in these cases would be rather complex due to the large number of genes potentially causing the observed phenotype.

This technology has proved useful also for the molecular analysis of patients positive at the neonatal metabolic screening. In this study, we analyzed 85 patients positive at the neonatal metabolic screening for Mucopolysaccharidosis type I (MPS I) or for Biotinidase Deficiency or Phenylketonuria (PKU), by Next Generation Sequencing: molecular analysis made it possible to distinguish true positive patients from false positives to neonatal metabolic screening, allowing to avoid unnecessary therapies and significantly reduce the cost of the investigation.

The increasing number of variants identified in the post-genomic era, in particular following the introduction of Next Generation Sequencing technology, has been accompanied by significant difficulties in the interpretation of the same, making fundamental the development of methods for their functional characterization.
Notable difficulties of interpretation are encountered especially for the intronic variants, especially if they are outside from the canonical splicing sites. Molecular analyzes based on mRNA have also highlighted how some exonic mutations (missense, synonymous and nonsense) are incorrectly classified as they actually interfere with the splicing mechanism (Messiaen and Wimmer, 2008).
In the absence of functional tests, to evaluate the pathogenicity of a variant we take in consideration some parameters including the absence of other pathogenic variants in the entire gene (including deletions or duplications), absence of the variant in healthy controls, the degree of conservation of the amino acid involved and the type of change, the genotype-phenotype segregation (absence of variant in unaffected parents and presence in all affected family members) (Messiaen and Wimmer, 2008). The use of predictive software can also help in the evaluation of pathogenicity (PolyPhen, SIFT, etc.) (Adzhubei et al., 2013).
In this study we used an experimental approach for the validation of a new variant identified in the *COQ4* gene able to altering the splicing, this method consists in the analysis of transient expression in human cells of hybrid minigens: these are constructs formed by recombinant DNA artificially designed, in which the coding sequence of the gene of interest is associated with regulatory elements present in the vector, which allow the expression of the insert and the production of the desired protein.
Our hybrid-minigene validation system has shown that this mutation behaves like a hypomorphic allele, in fact the analysis with hybrid minigen has shown that the splicing leads to the production of both the wild-type transcript and an abnormal transcript with loss of about 100 bp of *COQ4* exon 4.

In this study it was also studied a missense mutations of the *ASS1* gene, identified in patients affected by CTLN1 with heterogeneous clinical phenotypes, by functional complementation in yeast model system.

After verifying that the heterologous expression of the human gene *ASS1* was able to correct the growth defect in arginine-free medium of a *ΔARG1* yeast strain, it was analyzed the ability of the various *ASS1* mutant alleles to restored the *ΔARG1* yeast growth.

The phenotypic analysis allowed:

1) to confirm the pathogenicity of all the mutations under study;

2) to classify the mutants into two classes:

- class I mutations that completely abolish yeast growth;
- class II mutations that allow a residual growth.

The effect of missense mutations on the stability of the protein was evaluated by western blot analysis.

The results of the study are summarized in Table 22.

Allele	Clinical picture of patients	Protein localization	Yeast growth ability	Protein stability
G14S	asymptomatic	Active site	-	- / +
W179R	asymptomatic, mild	Active site	-	-
R265H	mild	Active site	++	++
R304W	neonatal	Dimer-dimer interactions	++	+++
G324S	asymptomatic, neonatal	Dimer-dimer interactions	+	++
R363L	neonatal	Dimer-dimer interactions	-	- / +
G390R	asymptomatic, mild, neonatal	Oligomerization	-	- / +

Table 22. Summary of the characteristics of missense mutations analyzed in this study.

The results demonstrate that: the mutations G14S, W179R, R363L and G390R abolish the function of the human gene in yeast and have a marked effect on the

stability of the protein, abolishing (W179R) or significantly reducing (G14S, R363L and G390R) the levels of protein; the mutations R265H, R304W and G324S, on the other hand, allow a residual growth of the *ΔARG1* yeast strain, proving to maintain a greater stability of the protein.

These results confirm the absence of a clear correlation between the residual enzymatic activity (measured indirectly by analysis of functional complementation in yeast) and the clinical picture presented by the patients (age of onset, severity of the manifestations). In fact, the presence of high levels of residual activity of ASS1 can not be considered a reliable marker of a mild/asymptomatic clinical course.

In particular, patients with mild late-onset or asymptomatic forms with a biochemical phenotype, may present class I mutations (for example G14S or W179R in homozygosity or in compound heterozygosis) while, conversely, the presence of a class II mutation (for example R304W or G324S) is not sufficient to guarantee a mild phenotype.

In this last case it is possible to hypothesize that the nucleotide substitution involves an aberrant effect on the *ASS1* gene splicing, with the consequent formation of an null alelle. It will be interesting to test this hypothesis for the two class II mutations listed above (c.970G> A p.G324S and c.910C> T p.R304W) using an in vivo expression system based on hybrid minigens (Trevisson et al., 2007).

In conclusion, the yeast model developed was useful for testing the pathogenicity of *ASS1* gene missense mutations, but not for establishing correlations between the genotype and the phenotype of patients.

6. PUBLICATIONS

Acosta M. J., Fonseca L. V., Desbats M. A., Cerqua C., Zordan R., Trevisson E., Salviati L.: Coenzyme Q biosynthesis in health and disease. *Bioch. Et Biop. Acta.* 2016; 1857:1079-1085.

Burlina A. B., Polo G., Salviati L., Duro G., Zizzo C., Dardis A., Bembi B., Cazzorla C., Rubert L., Zordan R., Desnick R. J., Burlina A. P.: Newborn screening for lysosomal storage disorders by tandem mass spectrometry in North East Italy. *J Inherit Metab Dis.* 2018; 41(2):209-219.

7. BIBLIOGRAPHY

Adessi C., Matton G., Ayala G., Turcatti G., Mermod J. J., Mayer P., Kawashima E.: Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res*. 2000; 28 (20): E87.

Acosta M. J., Fonseca L. V., Desbats M. A., Cerqua C., Zordan R., Trevisson E., Salviati L.: Coenzyme Q biosynthesis in health and disease. *Bioch. Et Biop. Acta.* 2016; 1857:1079-1085.

Adzhubei I., Jordan D. M., Sunyaev S. R.: Predicting functional effect of human missenso mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; 7: 7.20.

Ansorge W., Sproat B., Stegemann J., Schwager C., Zenke M.: Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.* 1987; 15 (11): 4593-602.

Auron A., Brophy P. D.: Hyperammonemia in review: pathophysiology, diagnosis and treatment. *Pediatr. Nephrol*. 2012; 27:207-222.

Bahassi E. M., O'Dea M. H., Allali N., Messens J., Geller M., Couturier M.: Interactions of CcdB with DNA gyrase. *J. Biol. Chem.* 1998; 274:10936-10944.

Barry W.: Biotinidase Deficiency. *GeneReviews*. 2016

Bartholome K., Olek K., Trefz F.: Compound heterozygotes in hyperphenylalaninaemia. *Hum. Genet*. 1984; 65: 405-406.

Beal M. F.: Mitochondrial dysfunction and oxidative damage in Alzheimer's and Parkinson's diseases and coenzyme Q10 as a potential treatment. *Journal of Bioenergetics and Biomembranes.* 2004; 36(4): 381-386.

Berning C, Bieger I, Pauli S, Vermeulen T, Vogl T, Rummel T, Höhne W, Koch HG, Rolinski B, Gempel K, Häberle J.: Investigation of citrullinemia type I variants by in vitro expression studies. *Hum Mutat.* 2008;29:1222-7.

Bentinger M., Tekle M., Dallner G.: Coenzyme Q–biosynthesis and functions *Biochem. Biophys. Res. Commun*. 2010; 396:74-79.

Beyer R. E., Burnett B. A., Cartwright K. J., Edington D. W., Falzon M. J., Kreitman K. R., Kuhn T. W.: Tissue coenzyme Q (ubiquinone) and protein concentrations over the life span of the laboratory rat. *Mechanisms of Ageing and Development.* 1985; 32(2-3): 267-281.

Boshart M., Weih F., Nichols M., Schutz G.: The tissue-specific extinguisher locus TSE1 encodes a regulatory subunit of cAMP dependent protein kinase. *Cell.* 1991; 66:849-859.

Botstein D., Chervitz S. A., Cherry J. M.: Yeast as a model organism. *Science*. 1997; 277:1259-1260.

Brea-Calvo G., Haack T. B., Karall D., Ohtake A., Invernizzi F., Carrozzo R., Kremer L., Dusi S., Fauth C., Scholl-Bürgi S., Graf E., Ahting U., Resta N., Laforgia N., Verrigni D., Okazaki Y., Kohda M., Martinelli D., Freisinger P., Strom T. M., Meitinger T., Lamperti C., Lacson A., Navas A., Mayr J. A., Bertini E., Murayama K., Zeviani M., Prokisch H., Ghezzi D.: COQ4 mutations cause a broad spectrum of mitochondrial disorders associated with CoQ10 deficiency. *Am. J. Hum. Genet*. 2015; 96:309-317.

Burlina A. B., Polo G., Salviati L., Duro G., Zizzo C., Dardis A., Bembi B., Cazzorla C., Rubert L., Zordan R., Desnick R. J., Burlina A. P.: Newborn screening for lysosomal storage disorders by tandem mass spectrometry in North East Italy. *J Inherit Metab Dis.* 2018; 41(2):209-219.

Carrilho E.: DNA sequencing by capillary array electrophoresis and microfabricated array systems. *Electrophoresis* 2000; 21 (1): 55-65.

Casarin A., Jemenez-Ortega J. C., Trevisson E., Pertegato V., Doimo M., Ferrero-Gomez M. L., Abbadi S., Artuch R., Quinzii C., Hirano M., Basso G., Ocana C. S., Navas P., Salviati L.: Functional characterization of human COQ4, a gene required for Coenzyme Q10 biosynthesis. *Bioch. And Biop. Research Comm.* 2008; 372:35-39.

Chung W. K., Martin K., Jalas C., Braddock S. R., Juusola J., Monaghan K. G., Warner B., Franks S., Yudkoff M., Lulis L., Rhodes R. H., Prasad V., Torti E., Cho M. T., Shinawi M.: Mutations in COQ4, an essential component of coenzyme Q biosynthesis, cause lethal neonatal mitochondrial encephalomyopathy. *J. Med. Genet*. 2015; 52:627-635.

Curis E., Nicolis I., Moinard C., Osowska S., Zerrouk N., Benazeth S., Cynober L.: Almost all about citrulline in mammals. *Amino Acids* 2005; 29:177-205.

Di Mauro S.: Mitochondrial diseases. *Biochimica et Biophysica Acta* 2004; 1658:80-8.

Dimmock D. P., Trapagne P., Feigenbaum A., Keegan C. E., Cederbaum S., Gibson J., Gambello M. J., Vaux K., Ward P., Rice G. M., Wolff J. A., O'Brien W. E., Fang P.: The role of molecular testing and enzyme analysis in the management of hypomorphic citrullinemia. *Am. J. Med. Genet*. 2008; 146A:2885-2890.

Doimo M., Desbats M. A., Cerqua C., Cassina M., Trevisson E., Salviati L.: Genetics of coenzyme Q 10 deficiency. *Mol. Syndromol*. 2010; 5:156-162.

Doimo M., Trevisson E., Sartori G., Burlina A., Salviati L.: Yeast complementation is sufficiently sensitive to detect residual activity of ASL alleles

associated with mild forms of argininosuccinic aciduria. *J. Inherit. Metab. Dis*. 2012; 3:557-558.

Dressman, D., Yan, H., Traverso, G., Kinzler, K. W., Vogelstein, B.: Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA* 2003; 100 (15): 8817-22.

Emma F., Montini G., Parikh S. M., Salviati L.: Mitochondrial dysfunction in inherited renal disease and acute kidney injury. *Nat. Rev. Nephrol.* 2016

Engel K., Hohne W., Haberle J.: Mutations and polymorphism in the human argininosuccinate synyhetase (*ASS1*) gene. *Hum. Mut*. 2009; 3:300-307.

Erez A., Nagamani S. C. S., Lee B.: Argininosuccinate lyase deficiency-argininosuccinic aciduria and beyond. *Am. J. Med. Genet. Part C.* 2011; 157:45-53.

Ewing, B., Hillier, L., Wendl, M. C., Green, P.: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*. 1998; 8 (3): 175-85.

Ewing B., Green P.: Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 1998; 8 (3): 186-94.

Fedurco M., Romieu A., Williams S., Lawrence I., Turcatti G.: BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* 2006; 34 (3): e22.

Forzan M, Salviati L, Pertegato V, Casarin A, Bruson A, Trevisson E, Di Gianantonio E, Clementi M.: Is CFTR 621+3 A>G a cystic fibrosis causing mutation? *J Hum Genet*. 2009; 1-4.

Frerman F. E.: Acyl-CoA dehydrogenases, electron transfer flavoprotein and electron transfer flavoprotein dehydrogenase. *Biochem Soc Trans.* 1988; 16(3): 416-418.

Gao H. Z., Kobayashi K., Tabata A., Tsuge H., Iijima M., Yasuda T., Kalkanoglu H.S., Dursun A., Tokatli A., Coskun T., Trefz F. K., Skaldal D., Mandel H., Seidel J., Kodama S., Shirane S., Ichida T., Makino S., Yoshino M., Kang J. H., Mizuguchi M., Barshop B. A., Fuchinoue S., Seneca S., Zeesman S., Knerr I., Rodes M., Wasan P., Yoshida I., Meirleir L., Jalil M. A., Begun L., Horiuchi M., Katunuma N., Nakagawa S., Saheki T.: Identification of 16 novel mutations in the argininosuccinate synthetase gene and genotype-phenotype correlation in 38 classical citrullinemia patients. *Hum. Mut*. 2003; 22:24-34.

Gilles A., Meglecz E., Pech N., Ferreira S., Malausa T., Martin J. F.: Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* 2011; 12: 245.

Gregory D. M., Sovetts D., Clow C. L., Scriver C. R.: Plasma free amino acid values in normal children and adolescents. *Metabolism* 1986; 967-969.

Goffeau A.: 1996: a vintage year for yeast and Yeast. *Yeast*. 1996; 12:1603-1605.

Goto M, Nakajima Y, Hirotsu K. Crystal structure of argininosuccinate synthetase from Thermus thermophilus HB8. Structural basis for the catalytic action. *J Biol Chem*. 2002; 277:15890-6.

Guei T. R., Liu M. C., Yang C. P., Su T. S.: Identification of a liver-specific cAMP response element n the human argininosuccinate synthetase gene. *Biochem. Biophys. Res. Commun*. 2008; 377:257-261.

Haber J. E.: Mating-type genes and MAT swithcing in S. cerevisiae. *Genetics* 2012; 191:33-64.

Haberle J., Pauli S., Linnebank M., Kleijer W. J., Bakker H. D., Wanders R. J. A., Harms E., Koch H. G.: Structure of the human argininosuccinate synthetase gene and an improved system for molecular diagnostics in patients with classical and mild citrullinemia. *Hum. Genet*. 2002; 110:327-333.

Haberle J., Pauli S., Schmidt E., Schulze-Eilfing B., Berning C., Koch H. G.: Mild citrullinemia in caucasians in an allelic variant of argininosuccinate synthetase deficiency (citrullinemia type 1). *Mol. Genet. Metab*. 2003; 80:302-306.

Haines R. J., Pendleton L. C., Eichler D. C.: Argininosuccinate synthase: at the center of arginine metabolism. *Int. J. Biochem. Bio. Mol*. 2011; 2:8-23.

Hanley W. B., Lee A. W., Hanley A. J. G., Lehotay D. C., Austin V. J., Schoonheyt W. E., Platt B. A., Clarke J. T. R.: 'Hypotyrosinemia' in phenylketonuria. *Molec. Genet. Metab*. 2000: 286-294.

Hert D. G., Fredlake C. P., Barron A. E.: Advantages and limitations of next generation sequencing technologies: a comparison of electrophoresis and nonelectrophoresis methods. *Electrophoresis* 2008; 29 (23): 4618-26.

Hildebrandt T. M., Grieshaber M. K.: Three enzymatic activities catalyze the oxidation of sulfide to thiosulfate in mammalian and invertebrate mitochondria. *FEBS J*. 2008; 275(13): 3352-3361.

Hinnebusch A.: General and pathway-specific regulatory mechanisms controlling the synthesis of amino acid biosynthetic enzymes in Saccharomyces cerevisiae. *The Molecular and Cellular Biology of the Yeast Saccharomyces*. 1992; 319-414.

Horvath R., Czermin B., Gulati S., Demuth S., Houge G., Pyle A., Dineiger C.: Adult onset cerebellar ataxia due to mutations in CABC1/ADCK3. *Journal of Neurology, Neurosurgery and Psychiatry*. 2012; 83(2): 174-178.

Husson A., Brasse-Lagnel C., Fairand A., Renouf S., Lavoinne A.: Argininosuccinate synthetase from the urea cycle to the citrulline-NO cycle. *Eur. J. Biochem.* 2003; 270, 1887-1899.

Hutchinson, C. A.: DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res*. 2007; 35 (18): 6227-37.

Kakinoki H., Kobayashi K., Terazono H., Nagata Y., Saheki T.: Mutations and DNA diagnoses of classical citrullinemia. *Hum. Mut.* 1997; 9:250-259.

Kalen A., Appelkvist E. L., Dallner G.: Age-related changes in the lipid compositions of rat and human tissues. *Lipids.* 1989; 24(7): 579-584.

Karlberg T., Collins R., van den Berg S., Flores A., Hammarstrom M., Hogbom M., Schiavone L. H., Uppenberg J.: Structure of human argininosuccinate synthetase. *Acta Cryst*. 2008; D64:279-286.

Kircher M., Kelso J.: High-throughput DNA sequencing--concepts and limitations. *Bioessays* 2010; 32 (6): 524-36.

Kleijer W. J., Blom W., Huijmans J. G., Mooyman M. C., Niermeijer M. F.: Prenatal diagnosys of citrullinemia: elevated levels of citrulline in the amniotic fluid in the three affected pregnancies. *Prenat. Diagn*. 1984; 4:113-118.

Kleijer W. J., Garritsen V.H., van de Sterre M. L. T., Berning C., Haberle J., Huijmans J. G. M.: Prenatal diagnosis of citrullinemia and argininosuccinic aciduria: evidence for a transmission ratio distortion in citrullinemia. *Prenat. Diagn*. 2006; 26:242-247.

Kobayashi K., Jackson M. J., Tick D. B., O'Brient W. E., Beaudet A. L.: Heterogeneity of mutations in argininosuccinate synthetase causing human citrullinemia. *The Journal of Biological Chemistry* 1989; 265:11361-11367.

Kobayashi K., Shaheen N., Terazono H., Saheki T.: Mutations in argininosuccinate synthetase mRNA of japanese patients, causing classical citrullinemia. *Am. J. Hum. Genet*. 1994; 55:1103-1112.

Kobayashi K., Sinasac D. S., Iijima M., Boright A. P., Begum L., Lee J. R., Yasuda T., Ikeda S., Hirano R., Terazono H., Crackower M. A., Kondo I., Tsui L. C., Scherer S. W., Saheki T.: The gene mutated in adult-onset type II citrullinemia encodes a putative mitochondrial carrier protein. *Nat. Genet*. 1999; 22:159-163.

Kogelnik A. M., Lott M. T., Brown M. D., Navathe S. B., Wallace D. C.: MITOMAP: a human mitochondrial genome database--1998 update. *Nucleic Acids Res.* 1998; 26: 112-5.

Korbel J. O., Urban A. E., Affourtit J. P., Godwin B., Grubert F., Simons J. F., Kim P. M., Palejev D., Carriero N. J., Du L., Taillon B. E., Chen Z., Tanzer A., Saunders A. C., Chi J., Yang F., Carter N. P., Hurles M. E., Weissman S. M., Harkins T. T., Gerstein M. B., Egholm M., Snyder M.: Paired-end mapping reveals extensive structural variation in the human genome. *Science* 2007; 318 (5849): 420-6.

Kumar P., Henikoff S., Ng P. C.: Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 2009; 4 (7): 1073-81.

Lagier-Tourenne C., Tazir M., López L. C., Quinzii C. M., Assoum M., Drouot N., Busso C., Makri S., Ali-Pacha L., Benhassine T., Anheim M., Lynch D. R., Thibault C., Plewniak F., Bianchetti L., Tranchant C., Poch O., DiMauro S., Mandel J. L., Barros M. H., Hirano M., Koenig M.: ADCK3, an ancestral kinase, is mutated in a form of recessive ataxia associated with coenzyme Q10 deficiency. *Am. J. Hum. Genet*. 2008; 82:661-672.

Lamperti C., Naini A., Hirano M., De Vivo D. C., Bertini E., Servidei S., Valeriani M.: Cerebellar ataxia and coenzyme Q10 deficiency. *Neurology.* 2003; 60(7): 1206-1208.

Lee B. H., Kim J. M., Heo S. H., Kim G. H., Choi I. H., Lee B. S., Kim E. A. R., Kim K. S., Jhang W. K., Park S. J., Yoo H. W.: High prevalence of neonatal presentation in korean patients with citrullinemia type 1, and their shared mutations. *Mol. Genet. Metab*. 2013; 108:18-24.

Lemke C. T., Howell P. L.: Substrate induced conformational changes in argininosuccinate synthetase. *J. Biol. Chem*. 2002; 277:13074-13081.

Li H., Durbin R.: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25 (14): 1754-60.

Li H., Ruan J., Durbin R.: Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008; 18 (11): 1851-8.

Lin A. E., Birch P. H., Korf B. R., Tenconi R., Niimura M., Poyhonen M., Armfield Uhas K., Sigorini M., Virdis R., Romano C., Bonioli E., Wolkenstein P., Pivnick E. K., Lawrence M., Friedman J. M.: Cardiovascular malformations and other cardiovascular abnormalities in neurofibromatosis 1. *Am. J. Med. Genet.* 2000; 95 (2): 108-17.

Liu Y., Cortopassi G., Steingrimsdottir H., Waugh A. P., Beare D. M., Green M. H., Robinson D. R., Cole J.: Correlated mutagenesis of bcl2 and hprt loci in blood lymphocytes. *Environ. Mol. Mutagen*. 1997; 29: 36-45.

Liu L., Li Y., Li S., Hu N., He Y., Pong R., Lin D., Lu L., Law M.: Comparison of next-generation sequencing systems. J. *Biomed. Biotechnol.* 2012: 251364.

Lorne A. C.: Mucopolysaccharidosis Type I. *Gene Reviews* 2016.

Luo C., Tsementzi D., Kyrpides N., Read T., Konstantinidis K. T.: Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample. *PLoS One* 2012; 7 (2): e30087.

Macevicz S. C.: DNA sequencing by parallel oligonucleotide extensions. *US patent* 1998; 5750341.

Mardis, E. R.: Next-generation DNA sequencing methods. Annu. Rev. *Genomics Hum. Genet.* 2008; 9: 387-402.

Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka J., Braverman M. S., Chen Y. J., Chen Z., Dewell S. B., Du L., Fierro J. M., Gomes X. V., Godwin B. C., He W., Helgesen S., Ho C. H., Irzyk G. P., Jando S. C., Alenquer M. L., Jarvie T. P., Jirage K. B., Kim K. B., Knight J. R., Lanza J. R., Leamon J. H., Lefkowitz S. M., Lei M., Li J., Lohman K. L., Lu H., Makhijani V. B., McDade K. E., McKenna M. P., Myers E. W., Nickerson E., Nobile J. R., Plant R., Puc B. P., Ronan M. T., Roth G. T., Sarkis G. J., Simons J. F., Simpson J. W., Srinivasan M., Tartato K. R., Tomasz A., Vogt K. A., Volkmer G. A., Wang S. H., Wang Y., Weiner M. P., Yu P., Begley R. F., Rothberg J. M.: Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; 437 (7057): 376-80.

Matlin A. J., Clark F., Smith C. W.: Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 2005; 6: 386-398.

McKernan K., Blanchard A., Kotler L., Costa G.: Reagents, methods, and libraries for bead-based sequencing. US patent application 2006; 20080003571.

Messiaen L. M., Wimmer K.: NF1 mutational spectrum. In: Neurofibromatoses. *Monogr Hum Genet. Kaufmann ed Karger Publ, Basel*. 2008; 16: 63-77.

Metzker M. L.: Emerging technologies in DNA sequencing. *Genome Res.* 2005; 15 (12): 1767-76.

Mian A., Lee B.: Urea-cycle disorders as a paradigm for inborn errors of hepatocyte metabolism. *Trends Mol. Med*. 2002; 8:583-859.

Mollet J., Delahodde A., Serre V., Chretien D., Schlemmer D., Lombes A., Boddaert N., Desguerre I. P. de Lonlay, H.O. de Baulny, Munnich A., Rotig A.: CABC1 gene mutations cause ubiquinone deficiency with cerebellar ataxia and seizures. *Am. J. Hum. Genet*. 2008; 82:623-630.

Morris S. M.: Regulation of enzymes of the urea cycle and arginine metabolism. *Annu. Rev. Nutr*. 22:87-105 (2002).

Musumeci O., Naini A., Slonim A. E., Skavin N., Hadjigeorgiou G. L., Krawiecki N., Weissman B. M.: Familial cerebellar ataxia with muscle coenzyme Q10 deficiency. *Neurology.* 2001; 56(7): 849-855.

Nagamani S. C. S., Campeau P.M., Shchelochkov O. A., Premkumar M. H., Guse K., Brunetti-Pierri N., Chen Y., Sun Q., Tang Y., Palmer D., Reddy A. K., Li L., Slesnick T. C., Feig D. I., Caudle S., Harrison D., Salviati L., Marini J. C., Bryan N. S., Erez A., Lee B. : Nitric-oxide supplementation fot treatment of long-term complications in argininosuccinic aciduria. *Am. J. Hum. Genet.* 2012; 90:836-846.

Nagarajan N., Pop M.: Sequencing and genome assembly using nextgeneration technologies. *Methods Mol. Biol.* 2010; 673: 1-17.

Nagy M., Lacroute F.: Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts. *PNAS*. 1992; 89(19): 8966-8970.

Nakamura K., Oshima T., Morimoto T., Ikeda S., Yoshikawa H., Shiwa Y.,

Ishikawa S., Linak M. C., Hirai A., Takahashi H., Altaf-UI-Amin M., Ogasawara N., Kanaya S.: Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res*. 2011; 39 (13): e90.

Natarajan K., Meyer M. R., Jackson B. M., Slade D., Roberts C., Hinnebusch A. G., Marton M. J.: Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol. Cell. Biomol*. 2001; 21:4347-4368.

Ogasahara S., Engel A. G., Frens D., Mack D.: Muscle coenzyme Q deficiency in familial mitochondrial encephalomyopathy. *Proceedings of the National Academy of Sciences of the United States of America.* 1989; 86(7): 2379-2382.

Pollard L. M., Braddock S. R., Christensen K. M., Boylan D. J., Smith L. D., Heese B. A., Atherton A. M., Lawson C. E., Strenk M. E., Willing M., Manwaring L., Wood T. C.: Three apparent pseudo-deficiency alleles in the *IDUA* gene identified by newborn. *63rd Annual Meeting of the American Society of Human Genetics, Boston, MA.* October 22–26, 2013. Available at: (Accessed April 2017).

Pop, M., Salzberg, S. L.: Bioinformatics challenges of new sequencing technology. *Trends Genet*. 2008; 24 (3): 142-9.

Ronaghi M., Karamohamed S., Pettersson B., Uhlen M., Nyren P.: Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem*. 1996; 242 (1): 84-9.

Rotig A., Appelkvist E. L., Geromel V., Chretien D., Kadhom N., Edery P., Lebideau M.: Quinone-responsive multiple respiratory-chain dysfunction due to widespread coenzymeQ10 deficiency. *Lancet.* 2000; 356(9227): 391-395.

Saheki T., Kobayashi K., Iijima M., Horiuchi M., Begum L., Jalil M. A., Li M. X., Lu Y. B., Ushikai M., Tabata A., Moriyama M., Hsiao K. J., Yang Y.:

Adult-onset type II citrullinemia and idiopathic neonatal hepatitis caused by citrin deficiency: involvement of the aspartate glutamate carrier for urea synthesis and maintenance of the urea cycle. *Mol. Genet. Metab*. 2004; 1:S20-26.

Sanger F., Nicklen S., Coulson A. R.: (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 1977; 74 (12): 5463-7.

Scheffler I. E.: Mitochondria. *New York: Wiley-Liss, Inc*. 1999.

Scriver C. R., Gregory D. M., Sovetts D., Clow C. L., Tissenbaum G.: Normal plasma free amino acid values in adults: the influence of some common physiological variables. *Metabolism* 1985; 868-873.

Shendure J., Porreca G., J., Reppas N. B., Lin X., McCutcheon J. P., Rosenbaum A. M., Wang M. D., Zhang K., Mitra R. D., Church G. M.: Accurate multiplex polony sequencing of an evolved bacterial genome. Science 2005; 309 (5741): 1728-32.

Shendure J., Ji H.: Next-generation DNA sequencing. *Nat. Biotechnol*. 2008; 26 (10): 1135-45.

Shults C. W.: Therapeutic role of coenzyme Q10 in Parkinson's disease. *Pharmacology and Therapeutics*. 2005; 107(1): 120-130.

Smith L. M., Sanders J. Z., Kaiser R. J., Hughes P., Dodd C., Connell C. R., Heiner C., Kent S. B., Hood L. E.: Fluorescence detection in automated DNA sequence analysis. *Nature* 1986; 321 (6071): 674-9.

Suormala T. M., Baumgartner E. R., Wick H., Scheibenreiter S., Schweitzer S.: Comparison of patients with complete and partial biotinidase deficiency biochemical studies. *J. Inherit. Metab. Dis.* 1990; 13: 76-92.

Swango K. L., Demirkol M., Huner G., Pronicka E., Sykut-Cegielska J., Schulze A., Mayatepek E., Wolf B.: Partial biotinidase deficiency is usually due to the D444H mutation in the biotinidase gene. *Hum Genet*. 1998; 102:571–5.

Sweetman L.: Two forms of biotin-responsive multiple carboxylase deficiency. *J. Inherit. Metab. Dis.* 1981; 4: 53-54.

Swerdlow H., Gesteland R.: Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res.* 1990; 18 (6): 1415-9.

Trevisson E, Salviati L, Baldoin MC, Toldo I, Casarin A, Sacconi S, Cesaro L, Basso G, Burlina AB.: Argininosuccinate lyase deficiency: mutational spectrum in Italian patients and identification of a novel ASL pseudogene. *Hum Mutat.* 2007;28:694-702.

Trevisson E., Burlina A., Doimo M., Pertegato V., Casarin A., Cesaro L., Navas P., Basso G., Sartori G., Salviati L.: Funcional complementation in yeast allows molecular characterization of missense argininosuccinate lyase mutations. *J. Biol. Chem.* 2009; 284:28926-28934.

Tucker T., Friedman J. M., Friedrich R. E., Wenzel R., Funsterer C., Mautner V. F.: Longitudinal study of neurofibromatosis 1 associated plexiform neurofibromas. *J. Med. Genet.* 2009a; 46 (2): 81-5.

Turcatti G., Romieu A., Fedurco M., Tairi A. P.: A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res.* 2008; 36 (4): e25.

Turunen M., Olsson J., Dallner G.: Metabolism and function of coenzyme Q Biochim. Biophys. *Acta*. 2004; 1660:171-199.

Van Maldergem L., Trijbels F., Di Mauro S., Sindelar P. J., Musumeci O., Janssen A., Delberghe X.: Coenzyme Q-responsive Leigh's encephalopathy in two sisters. *Annals of Neurology*. 2002; 52(6): 750-754.

Vilaseca M. A., Kobayashi K., Briones P., Lambruschini N., Campistol J., Tabata A., Alomar A., Rodes M., Lluch M., Saheki T.: Phenotype and genotype heterogeneity in mediterranean citrullinemia. *Mol. Genet. Metab*. 2001; 74:396-398.

Voelkerding K. V., Dames S. A., Durtschi J. D.: Next-generation sequencing: from basic research to diagnostics. Clin. Chem. 2009; 55 (4): 641-58.

Walberg M. W.: Appicability of yeast genetics to neurologic disease. *Arch. Neurol.* 2000; 57:1129-1134.

Wallace D. C., Brown M. D., Lott M. T.: Mitochondrial DNA variation in human evolution and disease. *Gene* 1999; 238: 211-30.

Wang Z., Burge C. B.: Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* 2008; 14: 802-813.

Wang, K., Li, M., Hakonarson, H.: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38 (16): e164.

Warren R. L., Sutton G., G., Jones S. J., Holt R. A.: Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* 2007; 23 (4): 500-1.

Whiteford N., Haslam N., Weber G., Prugel-Bennett A., Essex J. W., Roach P. L., Bradley M., Neylon C.: An analysis of the feasibility of short read sequencing. *Nucleic Acids Res.* 2005; 33 (19): e171.

Will C. L., Lührmann R.: Spliceosome structure and function. *Cols Spring Harb Perspect Biol* 2011; 3: a003707.

Woo H. I., Ki C. S., Lee S. Y., Kim J. W., Song J., Jin D. K., Park W. S., Lee D. H., Park H. D.: Mutation spectrum of the *ASS1* gene in korean patients with citrullinemia type 1. *Clinical Bioch*. 2013; 46:209-213.

Wraith J. E., Rogers J. G., Danks D. M.: The mucopolysaccharidoses. *Aust Paediatr J.* 1987; 23(6):329-34.

Zurfluh M. R., Zschocke J., Lindner M., Feillet F., Chery C., Burlina A., Stevens R. C., Thony B., Blau N.: Molecular genetics of tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency. *Hum. Mutat.* 2008; 167-175.