**Universidade de Aveiro** Departamento de Química

**2017**

**Inês Mariana Cardoso Eulálio**

# DESENVOLVIMENTO DE UM MÉTODO DE DETEÇÃO DE CNVs ATRAVÉS DE qPCR

**Inês Mariana Cardoso Eulálio**

# Desenvolvimento de um método de deteção de CNVs através de qPCR

*Development of a CNVs detection method through qPCR*

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biotecnologia, realizada sob a orientação científica da Doutora Gabriela Maria Ferreira Ribeira de Moura, Professora Auxiliar do Departamento de Ciências Médicas da Universidade de Aveiro e coorientação da Dra. Maria de Fátima e Costa Torres, Especialista em Genética Humana, Coordenadora de área do Laboratório de Diagnóstico Molecular do CGC Genetics.

Para os meus pais. Por me terem dado a vida, dou-vos agora o melhor de mim.

**Júri**

| | |
|---|---|
| Presidente | Doutora Mara Guadalupe Freire Martins<br>Equiparado a Investigador Coordenador da Universidade de Aveiro |
| Arguente | Prof.ª Doutora Patrícia Espinheira Sá Maciel<br>Professora Associada da Escola de Medicina da Universidade do Minho |
| Orientador | Prof.ª Doutora Gabriela Maria Ferreira Ribeira de Moura<br>Professora Auxiliar do Departamento de Ciências Médicas da Universidade de Aveiro |

**Agradecimentos**

Ao Dr. Jorge Pinto-Basto, que me deu a oportunidade de pertencer ao CGC. À professora Gabriela Moura, que acompanhou o meu trabalho. À Dra. Fátima Torres, um agradecimento muito especial, pois mesmo com todo o trabalho e responsabilidades que já possuía, aceitou orientar-me e apoiou-me ao longo de todo este processo. À Liliana, em particular, assim como à Xana, Marisa e Sofia que me aturaram durante o meu tempo de estágio. Obrigada por partilharam comigo o vosso cantinho, conversas, computadores e tudo mais. Aos colegas do CGC que me proporcionaram um ótimo ambiente de trabalho.

Aos meus pais, que mesmo vendo a filha a fugir, sempre apoiaram a minha decisão de sair do conforto e aventurar-me pela Invicta. Obrigada pela vida que me deram, pelo apoio, pelos conselhos, pelas discussões e castigos. Obrigada, por mesmo passando por tantas aventuras e desventuras, fazerem todos os esforços para que possa estar onde estou agora. À minha irmã, que é a maior chata do mundo, que me inferniza a vida e irrita profundamente, mas que me acolheu, mesmo que apertadinhas, no cantinho dela. Sabes bem o quanto gosto de ti, mesmo com os nossos constantes desatinos. À minha família, sem a qual não sei o que seria de mim. A vocês que me criaram, me educaram, tantas vezes me ampararam, um muito obrigada do tamanho da felicidade que me dão.

Ao Rodrigo, que ao longo destes anos me tem aturado o melhor que pode, mesmo com o meu "feitiozinho" difícil e resmunguice constante. Sem o teu apoio, compreensão, tantas vezes ombro para chorar quando desesperei, não seria definitivamente a mesma pessoa. À Catarina, uma amiga que, mesmo à distância, sempre foi um pilar. Ao Mário que já perdi a conta aos anos que me atura.

**Palavras-chave**

*Array Comparative Genomic Hybridization* (aCGH); *Copy Number Variations* (CNVs); Deep Sequencing; Diagnóstico Molecular; *Fluorescent in situ Hybridization* (FISH); Multiplex *Ligation-Dependent Probe Amplification* (MLPA); PCR quantitativo em tempo real (qPCR); *Single Nucleotide Polymorphism Arrays*; SYBR-Green I; Validação

## Resumo

Os *copy number variations* (CNVs) consistem em segmentos de DNA de uma kilobase ou mais, que se encontram num número variável de cópias, em comparação com um genoma de referência. A deteção de CNVs é convencionalmente realizada através de técnicas de citogenética, como *fluorescence in situ hybridization* e *array comparative genomic hybridization*, ou com base em PCR, como *multiplex ligation-dependent probe amplification*, SNP *arrays* ou *deep sequencing*. Porém, a evolução da técnica de PCR quantitativo em tempo real (qPCR) permitiu que fosse, actualmente, considerada o método *gold standard* para a detecção de CNVs devido, sobretudo, ao elevado rendimento, sensibilidade, precisão e versatilidade. O presente trabalho descreve o desenvolvimento e validação de um método de deteção de CNVs através da técnica de qPCR. A metodologia adotada provou ser um método preciso e sensível para a detecção de CNVs em regiões específicas.

**Keywords** Array Comparative Genomic Hybridization (aCGH); Copy Number Variations (CNVs); Deep Sequencing; Molecular Diagnosis; Fluorescent *in situ* Hybridization (FISH); Multiplex Ligation-Dependent Probe Amplification (MLPA); real time quantitative PCR (qPCR); Single Nucleotide Polymorphism Arrays; SYBR-Green I; Validation

## Abstract

Copy number variations (CNVs) consist in DNA segments of one kilobase or larger, that are present in variable copy number, in comparison to a reference genome. CNVs detection is conventionally performed through cytogenetic, such as fluorescence *in situ* hybridization and array comparative genomic hybridization, or PCR-based techniques, like multiplex ligation-dependent probe amplification, SNP arrays or deep sequencing. However, the evolution of quantitative PCR (qPCR) allows it to be considered the gold standard for CNVs detection, mainly due to its high throughput, precision and versatility. The present work describes the development and validation of a qPCR method for CNVs detection. This method proved to be an accurate and sensitive method for CNVs detection in targeted regions.

## Contents

Resumo ..... v
Abstract ..... vi
Contents ..... vii
Figures ..... viii
Tables ..... ix
Abbreviations ..... x
Introduction ..... 1
Copy Number Variations ..... 1
Fluorescence in situ Hybridization ..... 2
Comparative Genomic Hybridization and Array Comparative Genomic Hybridization ..... 3
Single Nucleotide Polymorphism Arrays ..... 6
Deep Sequencing ..... 7
Multiplex Ligation-dependent Probe Amplification ..... 7
Polymerase Chain Reaction ..... 11
1) Real-time Quantitative PCR ..... 12
2) Experimental Design ..... 14
3) Normalization and Reference Genes ..... 15
4) Amplicon Detection ..... 16
5) Quantitation Methods ..... 18
Objectives ..... 19
Methods ..... 20
Samples ..... 20
Selection of Regions and Primers Design ..... 20
Empirical Validation of Primers ..... 22
qPCR Assays for CNVs Detection ..... 23
Data Analysis and Statistics ..... 24

Results and Discussion ..... 25
Method Validation ..... 25
1) The *MECP2* gene ..... 26
2) The *NF1* gene ..... 27
3) The *PARK2* gene ..... 28
4) Results comparison ..... 30
Reference Genes ..... 31
Primers validation ..... 31
Scoring limits ..... 34
Relative Quantitation for CNVs Detection ..... 35
1) Heterozygous deletion in *COL6A1* gene ..... 36
2) Homozygous deletion in the *BCKDHA* gene ..... 38
3) Heterozygous deletion in the *ABCA3* gene ..... 39
4) CNVs search in genes from the X chromosome ..... 41
Conclusion ..... 45
References ..... 46
ANNEX I – Samples ..... 55
ANNEX II -Primers ..... 60
ANNEX III - Method Validation – Results Comparison ..... 72
ANNEX IV – Primers Validation ..... 76
ANNEX V – Relative Quantitation Results ..... 79

## Figures

Figure 1 – Overview of the reference design for array comparative genomic hybridization. ..... 4
Figure 2 – Schematic representation of the multiplex ligation-dependent probe amplification process. ..... 9
Figure 3 – The basic steps in the polymerase chain reaction. ..... 11

Figure 4 – Model of a single amplification plot illustrating the nomenclature commonly used in qPCR. ..... 13
Figure 5 – Model of an amplification plot illustrating the mathematical basis of the $2^{-\Delta\Delta Ct}$ method. ..... 18
Figure 6 - Efficiency of the primers pairs for the *MECP2*, *NF1* and *PARK2* genes, used in the methodology validation. ..... 25
Figure 7 - Relative quantitation of the *MECP2* gene, during method validation. ..... 26
Figure 8 - Relative quantitation of the *NF1* gene, during method validation. ..... 28
Figure 9 - Relative quantitation of the *PARK2* gene, during method validation. ..... 29
Figure 10 - Evaluation of the accuracy, specificity and sensitivity of the developed methodology. ..... 30
Figure 11 - Efficiency of the primers for the reference genes, using the cycling programs with melting temperatures of 95ºC and 99ºC. ..... 31
Figure 12 – Efficiencies of the primers pairs of the tested genes. ..... 33
Figure 13 - Relative quantitation for the *COL6A1* gene. ..... 37
Figure 14 - Relative quantitation for the *COL6A2* gene ..... 37
Figure 15 - Relative quantitation of the *COL6A3* gene. ..... 38
Figure 16 - Relative quantitation for the *BCKDHA* gene. ..... 39
Figure 17 - Relative quantitation for the *ABCA3* gene. ..... 41
Figure 18 - Relative quantitation of the *BTK* gene. ..... 43
Figure 19 - Relative quantitation of the *NAA10* gene. ..... 44

## Tables

Table 1 - Efficiency, $R^2$ and PCR conditions of the not validated primers. ..... 32
Table 2 - Scoring limits for scoring a deletion, duplication or normal copy number sample, based on the mean haploid copy numbers and standard-deviation of the test and control samples. ..... 35

Table S1 - Samples identification, matrix of extraction, CNV status and studied genes for the validation and development of the methodology. ..... 55
Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design. ..... 60

Table S3- Results comparison between qPCR and previous results for methodology validation ..... 72
Table S4 - Primers efficiency, $R^2$ and PCR characteristics for the tested genes. ..... 76
Table S5 - Relative quantitation results for the samples used as normal controls. ..... 79
Table S6 - Relative quantitation results for the test samples scored as normal..... 81
Table S7 - Relative quantitation results for the test samples scored as deleted and duplicated. ..... 83

## Abbreviations

$2^{-\Delta\Delta Ct}$ method – Delta Delta Ct method
aCGH – array comparative genomic hybridization
BAC - bacterial artificial chromosomes
BCKD – branched-chain alpha-keto acid dehydrogenase
BM – Bethlem myopathy
bp – base pair
CGH – comparative genomic hybridization
CNVs – Copy Number Variations
Cq – quantification cycle
Ct – threshold cycle
dbGaP – Database of Genotype and Phenotype
DGV – Database of Genomic Variants
DGV – Database of Genomic Variants
DNA – deoxyribonucleic acid
dsDNA – double stranded DNA
EDTA – Ethylenediamine tetraacetic acid
FISH – fluorescence *in situ* hybridization
FRET – fluorescence resonance energy transfer
GWAS – genome-wide association studies
HGMD – Human Gene Mutation Database
HGMD – The Human Gene Mutation Database
ISCA – International Standards for Cytogenomic Array
kb – kilobase

LMS – Lenz microphthalmia syndrome

LPO – left probe oligonucleotide

Mb – megabase

MLPA – multiplex ligation-dependent probe amplification

MSUD – Maple Syrup Urine Disease

NCBI - National Center for Biotechnology Information

NGS – Next-generation sequencing

NM – Nucleotide NCBI Reference Sequence

NTC – No-template control

PCR – polymerase chain reaction

qPCR – real-time quantitative polymerase chain reaction

RPO – right probe oligonucleotide

SNP – single nucleotide polymorphism

Ta – annealing temperature

Tm – melting temperature

UCMD – Ullrich congenital muscular dystrophy

VOUS – variants of unknown significance

XLA – X-linked agammaglobulinemia

# Introduction

## Copy Number Variations

Copy number variations (CNVs) are defined by Feuk and colleagues as DNA segments of one kilobase (1kb) or larger that are present at a variable copy number, in comparison to a reference genome[1]. CNVs can be simple in structure, such as gains (tandem duplications or insertional transpositions) and losses (deletions), or may involve complex rearrangements of the genome, such as gains or losses of homologous sequences at multiple sites in the genome[2,3]. Over the years, new variations in the repetitive regions of DNA have been identified, suggesting that CNVs might be as important to genomic diversity as single nucleotide polymorphisms (SNPs)[4,5]. In fact, CNVs might account for about 1% of the inter-individual variation, over the 0,1% accounted for SNPs[6].

The HapMap Project revealed that regions with copy number variation cover approximately 12% of the human genome, and other studies found out that approximately one-third of them are unique to the human species. Regarding evolution, through multiple generations, deletions are usually selected against, while duplications may experience positive selection, as suggested by the existence of gene families[4]. Large CNVs usually exist due to *de novo* events, since they typically suffer negative selection[5]. CNVs <1kb in size are of particular importance to genetic medicine, since they are more numerous than the larger CNVs[7]. Also, CNVs generally are located outside of coding sequences and ultra-conserved regulatory elements, which are sequences of at least 200bp that are 100% conserved across several species[2]. CNVs can involve one or multiple genes, being presented as a recessive or dominant allele, due to variations in gene expression through gene dosage effects. Also, gene expression can be altered through position effects due to the disruption of the coding region or deletions or duplications outside the coding sequences[2,7].

Regarding functional categories, genes involved in cell adhesion, the sensory perception of smell, neurophysiological processes, brain development, and responses to chemical stimuli, such as immune responses and drug metabolism, are the most enriched in CNVs[4,7]. On the other hand, genes involved in cell signalling and cell proliferation, as well as in regulation of protein phosphorylation are the less represented[4]. While some CNVs have no apparent influence on phenotype, others have been implicated in the aetiology of

monogenetic and complex disorders[4,7]. Also, their interaction with additional genetic or environmental factors may influence detection of their phenotypic effect [4].

Routine screening in clinical diagnostics passes by the measurement and analysis of CNVs, which can assist in subsequent prognostic monitoring. The currently used techniques applied in CNV studies are fluorescence *in situ* hybridization (FISH), comparative genome hybridisation (CGH), single nucleotide polymorphism (SNP) arrays, deep sequencing, multiplex ligation-dependent probe amplification (MLPA), and real-time quantitative PCR (qPCR)[8,9].

**Fluorescence *in situ* Hybridization**

Fluorescence *in situ* Hybridization (FISH), is a molecular cytogenetic technique that allows direct examination of chromosomal DNA molecules and the visualization of specific sequences in the nucleus and/or on metaphase chromosomes[7,10]. FISH uses fluorescently labelled DNA probes that are hybridized to metaphase chromosome spreads or interphase nuclei, immobilized on a glass slide[7,10]. Thus, it allows to locate the physical position of specific DNA sequences that hybridize to the probes on the chromosomes, through examining of the preparation with a microscope[10,11].

There are several FISH-based techniques that can be performed, depending on the application. The visualization and analysis of genetic changes affecting one or a few specific genes or loci in individual interphase nuclei can be performed through FISH[12]. Fiber-FISH has the highest resolution, ranging from 5 to 500kb, being the preferred method to precisely determine the genomic structure of complex CNVs[7]. Also, FISH can be performed with multiple labelled probes with different fluorochromes, being widely used in clinical diagnosis in order to confirm the presence of CNVs, for example[7,10].

However, FISH has several limitations. FISH analysis requires the availability of the specific probe for the gene or locus of interest, with the appropriate size and location and so, it cannot be used to screen unknown genetic aberrations[12]. Additionally, locus-specific probes are expensive and the procedure is time-consuming and labour intensive[7]. Also, smaller copy number changes (less than 10kb) are too small to be detected by FISH, requiring other molecular methods[12]. Moreover, only a limited number of chromosomal loci can be screened in a single experiment[7].

## Comparative Genomic Hybridization and Array Comparative Genomic Hybridization

Comparative genomic hybridization is a cytogenetic molecular technique, introduced around the 1990s, that allows genome-wide screening for CNVs[6,13]. It provides alternative means for cytogenetic analysis such as banding patterns (G-banded karyotyping) and FISH. In CGH, a test and a control genome are differentially labelled and competitively hybridized to metaphase chromosomes. The fluorescent signal intensity of the labelled test DNA relative to that of the reference DNA can be linearly plotted across each chromosome, allowing for the identification of copy number changes. CGH can be used to quickly scan an entire genome for imbalances, and does not require cells that are undergoing division. However, the resolution of the CGH has been limited to alterations of approximately 5-10Mb for most clinical applications[14].

Array CGH (aCGH) combines the principles of CGH with the use of microarrays, in order to overcome some limitations of conventional CGH. It uses slides arrayed with small segments of DNA as the targets for analysis, instead of metaphasic chromosomes, allowing to test a smaller sample size[14,15]. These microarrays are created by the immobilization of small amounts of DNA (probes) on a solid support, in an ordered manner[14]. Probes vary in size, from large genomic clones, such as the earlier bacterial artificial chromosomes (BAC), to synthetized oligonucleotides[6]. The BAC-based arrays allow the detection of single-copy losses or gains, usually being targeted to subtelomeric and pericentromeric regions and areas within the genome known to be associated with diseases, such as microdeletion and microduplication syndromes[6,16,17]. The more frequently used though are oligonucleotide aCGH, which allows genome-wide interrogation and more precise identification of breakpoints, providing high accuracy and reproducibility[6,18]. The oligonucleotides represent areas of interest, being more flexible in terms of probe selection and facilitating high probe density and customization of array content[14,16]. Arrays designed specifically for copy number analysis use longer oligonucleotide probes (~60-mer), that provide a better signal-to-noise ratio than the shorter oligonucleotide probes (~22-mer) used in some SNP-detecting array platforms[17]. The resolution of aCGH (up to 1kb) is

higher than that of conventional CGH and is determined by considering both probe size and the genomic distance between DNA probes (probe density)[14,15].

The basic methodology for aCGH analysis is independent of the probe type[14]. The experimental design will depend on the objective of the study, as well as on the sources of variability[19]. Most studies apply the reference design (Figure 1), where initially, DNA is extracted from a test sample, followed by the labelling with a fluorescent dye of a specific colour[14,19]. On the other hand, DNA extracted from a normal control sample, usually from several healthy donors, is labelled with a dye of a different colour[14,18]. Both DNAs are then mixed together in equal proportions and applied to a microarray[15]. The single-stranded denatured DNA from both the sample and control competes to hybridize with the array single-stranded probes, and any unbound DNA is stringently removed by washing[14,18]. However, this design uses half of the available resources to analyse a reference sample that is not of biological interest, not being cost effective. Alternatively, the dye-swap design, that also needs an on-chip reference sample, reduces artifacts due to the probe-specific dye biases. In the dye-swap design, each sample is run in duplicate, with one of the replicates labelled with the Cy3 dye and the other on a separate array with the Cy5 dye. Nevertheless, this requires the

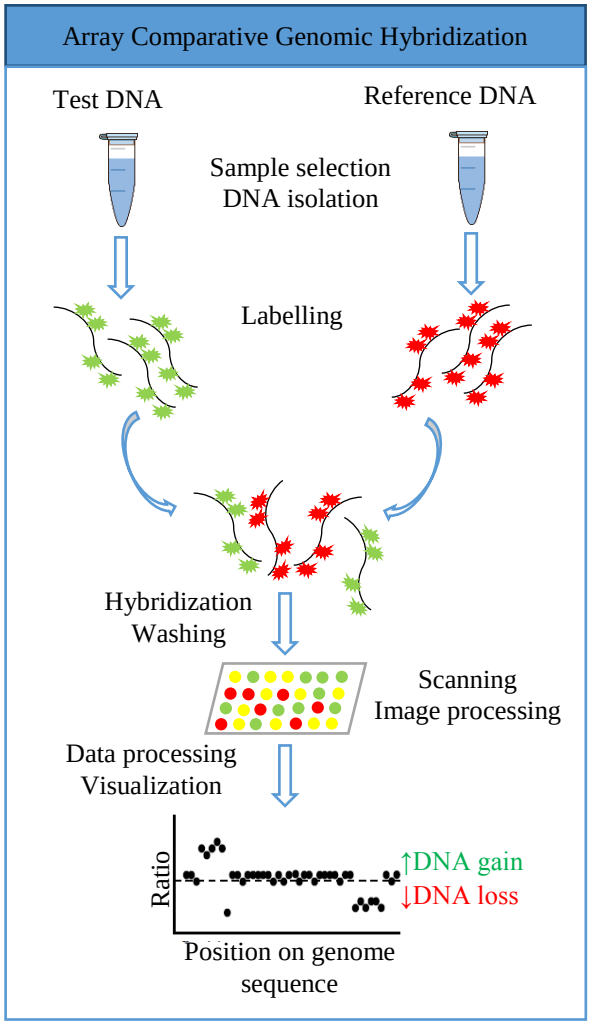


*Figure 1 – Overview of the reference design for array comparative genomic hybridization.*
*The extracted test and reference DNA are labelled with different fluorescent dyes and mixed together in equal proportions. Then, they are applied to a microarray and both DNAs compete to hybridize with the array probes. The relative fluorescence intensities are captured and quantified by digital systems and the fluorescence ratio between the two signals is determined at different positions along the genome. Adapted from Szuhai, K. & Vermeer, M. Microarray Techniques to Analyze Copy-Number Alterations in Genomic DNA: Array Comparative Genomic Hybridization and Single-Nucleotide Polymorphism Array. J. Invest. Dermatol. 135, 1–5 (2015).*

doubling of every sample hybridization, increasing the cost of the study in comparison to the reference design. Another set of design does not require on-chip reference samples, since off-chip references can be utilized. The use of an off-chip reference processed on the same day possesses similar precision in copy-number estimates as an on-chip reference. In a balanced-block design, two groups of interest can appear on each assay simultaneously. If there are more than two groups of interest, not every group can appear at the same time on each array, resulting in a balanced incomplete-block design. Although balanced-block and incomplete-block designs still require that each group is half labelled with Cy3 fluorescent dye and the other half labelled with Cy5 fluorescent dye, they are more efficient than a reference design. Also, these designs allow investigators to nearly double their sample size while using the same number of arrays as required for a reference design. This increases the statistical power, which is very important in genome-wide research studies exploring the association between copy number and disease[19]. Regardless of the previously chosen design, the relative fluorescence intensities of the labelled DNA probes that have hybridized to each target are captured and quantified by digital imaging systems. The fluorescence ratio of the test and reference hybridization signals is determined at different positions along the genome. This provides information on the relative copy number of sequences in the test genome as compared to the normal genome[14,15].

The main advantage of aCGH is its capacity to simultaneously detect aneuploidies, deletions, duplications, and/or amplifications of any locus represented on an array, also being less labour-intense and expense-saving, with a faster turnaround time[14,15]. The application of aCGH on the high resolution measurement of gains and losses of DNA sequence was developed by Pinkel and colleagues[7,20]. Since then, multiple whole genome CNVs studies in normal populations have emerged, now being a commonly ordered clinical genetic test[16,17]. It is a powerful tool for the detection of sub-microscopic chromosomal abnormalities in individuals with idiopathic intellectual disability and various birth defects[14]. The sensitivity of aCGH is determined by the number and density of the probes and their resolution, and it must accurately detect imbalances smaller than 5Mb in order to be advantageous over G-band karyotyping[16]. Also, an international consensus statement recommended aCGH as a first-line genetic test, instead of conventional karyotype analysis, for patient with unexplained developmental delay/intellectual disability, autism spectrum disorders, and multiple congenital

anomalies[17,21]. The use of oligo arrays have improved aCGH resolution, which allows better detection and mapping of localized changes, such as microdeletions and microduplications[16,22]. Moreover, aCGH testing can be subjected to automation, high-throughput objective analysis and quality control[15,16].

However, aCGH can only potentially detect "unbalanced chromosomal changes", failing to detect balanced genomic rearrangements, such as translocations or inversions, since the amount of DNA within the sample does not differ from the control DNA, as well as low-level mosaicism[15–17]. Also, polyploidies are not detected because there is no loss or gain of DNA content at a scale smaller than the chromosome[15]. In addition, the identification of novel variants of unknown significance (VOUS) through aCGH may cause difficulties in clinical management, since they cannot be clearly classified as benign or pathogenic[15,16]. Therefore, the development of CNVs databases, such as the International Standards for Cytogenomic Array (ISCA) Consortium from within the Database of Genotype and Phenotype (dbGaP) at the National Center for Biotechnology Information and the Database of Genomic Variants (DGV) facilitates the resolution of VOUS into benign or pathogenic variants[16,17]. The key to overcome this disadvantage is to increase sensitivity and specificity, through the selection of a microarray that is capable of detecting the majority of pathological CNVs without detecting too many VOUS[16]. Up to date, aCGH is more expensive than conventional cytogenetic testing, since higher probe density increases the cost per analysed sample[15,18].

**Single Nucleotide Polymorphism Arrays**

Single Nucleotide Polymorphism (SNP) Arrays are widely used for genome-wide association studies (GWAS), which contributed greatly for population genetics and genetic epidemiology[23]. Nevertheless, the widespread availability of SNP array data contributed to its application in CNV discovery and mapping[7,23]. Nowadays, commercial SNP arrays offer high genomic coverage, at somewhat affordable prices and improvements in genomic regions previously problematic[23].

SNP arrays are based on hybridizations with SNP marker probes designed specifically for particular genomic locations, typically targeting biallelic SNPs. For each SNP, an array platform includes two types of hybridization probes specific to two types of

known alleles. Thus, the SNP genotype can be determined by the ratios of the hybridization intensities for each probe, which can be altered due to CNVs. Computational methods use hybridization intensities and allele frequencies from SNP markers, in order to detect common or rare CNVs[24].

The main limitations of SNP arrays are that they are still relatively expensive and not ideal when thousands of samples need to be screened. Moreover, the power of SNP arrays is limited by their resolution by a lower detection limit of ~10–25 kb[7].

**Deep Sequencing**

Next-generation sequencing technologies (NGS) are based on various implementations of cyclic-array sequencing, redefining the study of CNVs. NGS platforms allow for the sequencing of millions of short DNA fragments simultaneously and may process a whole human genome in a relatively short period of time and at a reasonable price[25].

Nowadays, rapidly developing NGS technologies provide a sensitive and accurate alternative approach for accessing genomic variations[26]. Short reads from NGS platforms are randomly sampled from the entire genome, as opposed to array-based approaches, where probes are predefined for limited genomic regions. The NGS approach also has the advantages of possessing a higher coverage and resolution, more accurate estimation of copy numbers, more precise detection of breakpoints, and higher capability to identify novel CNVs[27]. However, the use of NGS for CNV identification has been limited by a lack of available and effective statistical approaches[26].

**Multiplex Ligation-dependent Probe Amplification**

Multiplex ligation-dependent probe amplification (MLPA) was first described in 2002 by Schouten and colleagues and developed for the detection of large mutations in disease-related genes[28,29]. It is a PCR-based technique that has been applied to the investigation of a wide variety of genetic disorders, expression analysis, methylation identification, multi-colour labelling as well as stuffer-free and multiple-oligo probe design[29,30]. MLPA is used for the large-scale screening of samples for the presence of copy

number variations at the genomic level (gains or losses) in a test DNA, compared to a control[31,32]. Several studies have demonstrated that MLPA allows sensitive and accurate detection of disease-specific clinically significant copy number changes[30]. Furthermore, two variations of MLPA have been developed for mRNA profiling (RT-MLPA) and methylation profiling (MS-MLPA)[33].

In MLPA, probe pairs are designed against genomic regions of interest (test probes), along with another set of probes that serve as reference probes, being able to utilize up to 45 specific probes for different DNA sequences[32,34]. Each MLPA probe (Figure 2) consists of two oligonucleotide half-probes, with target-specific sequences, that anneal to adjacent sites on DNA, and a generic primer pair attached to the upstream half-probe (also known as short probe and/or left probe oligonucleotide, LPO) at its 5' end and to the downstream half-probe (also known as right probe oligonucleotide, RPO) at its 3' end[32–34]. Moreover, one or both half-probes contain a stuffer sequence, which allows the differentiation of the reference and test probes length during electrophoresis, and, therefore, the size of the amplification product[32,34]. The complementary regions of the MLPA probes range from 19 to 43 nucleotides in length, and the stuffer sequence ranges from 19 to 370 nucleotides, creating a final ligated probe sequence ranging from 130 to 480 nucleotides, which are separated in size by 6 to 9 nucleotides[33]. The reference probes, used for normalization of the amount of PCR products from different reactions, hybridize to copy-number stable regions that are not subjected to variations, such as SNPs or structural variations[29,32].

The MLPA reaction (Figure 2) begins with denaturation of genomic DNA, at 98ºC for 5min, followed by incubation with a mixture of probes that hybridize to single-stranded DNA, at 60ºC for 16 to 20 hours, to allow complete hybridization. Then, only if the two half-probes hybridize to contiguous sequences, without a single gap, they are joined together by NAD-dependent DNA ligase, i.e., the ligation reaction occurs, at 54ºC[32–34]. The following dosage-dependent PCR amplification is carried out using only one pair of primers, complementary to generic sequences, whose 5' ends are fluorescently labelled[32,34]. Thus, PCR does not amplify the target sequences, but the ligated probes[34]. Therefore, using the same PCR primers, with equal efficiency, different target sequences can be amplified, mitigating amplification issues due to different primer efficiencies[33]. Afterwards, PCR products are separated by size and detected through capillary

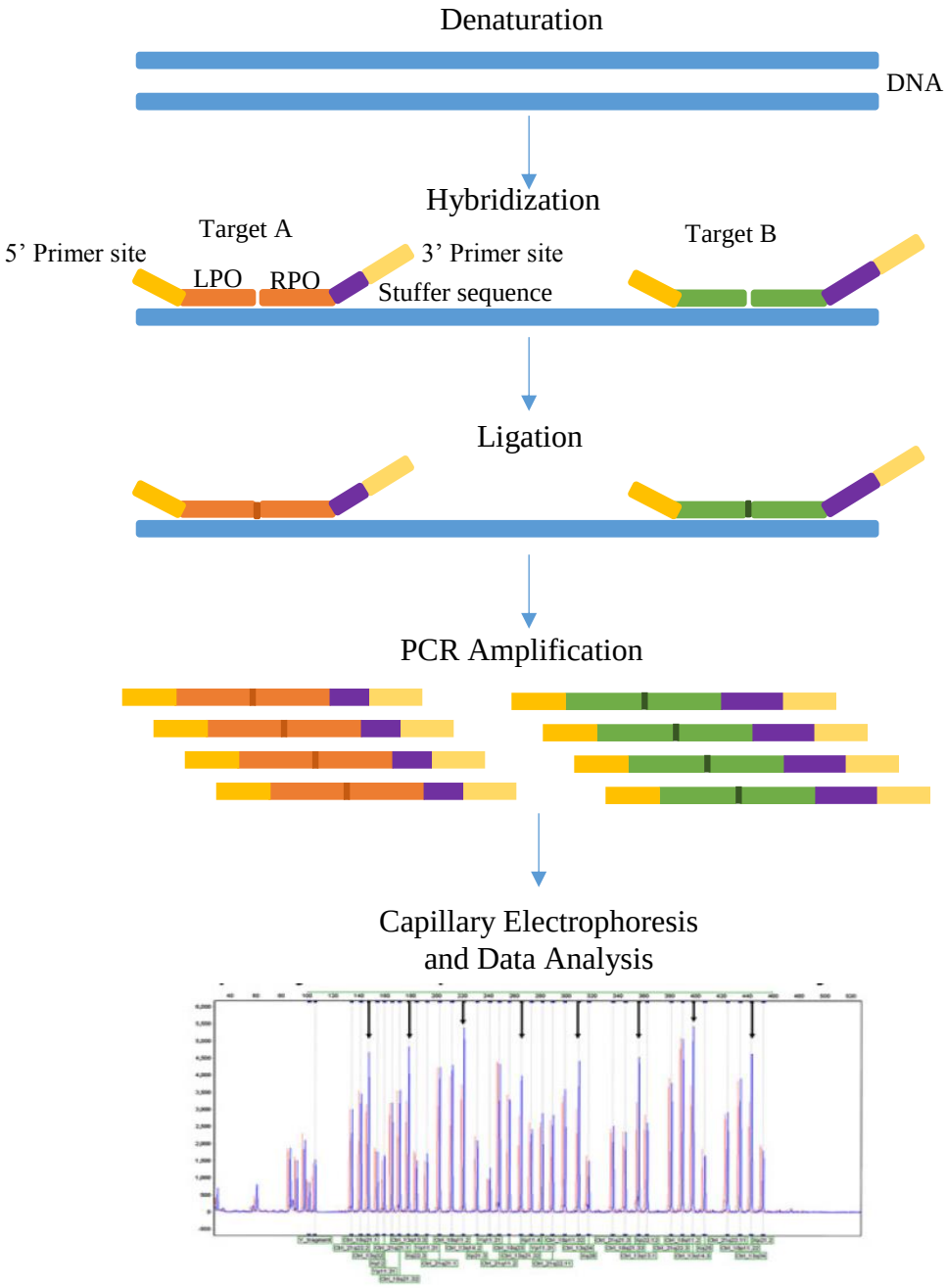


*Figure 2 – Schematic representation of the multiplex ligation-dependent probe amplification process.*
*The MLPA reaction begins with the denaturation of the DNA, followed by the immediate incubation with the MLPA probes, in order to occur hybridization. Then, the ligation reaction occurs, joining the two half-probes. The subsequent PCR reaction occurs using only PCR primers pairs, amplifying the MLPA probes. Afterwards, the PCR products are separated by size and detected by capillary electrophoresis, followed by normalization and data analysis. Adapted from Willis, A. S., Veyver, I. Van Den & Eng, C. M. Multiplex ligation-dependent probe amplification (MLPA) and prenatal diagnosis. Prenat. Diagn. 32, 315–320 (2012).*

electrophoresis[29,32]. Each specific probe migrates according to its molecular weight and the resulting electropherograms show PCR-derived-fluorescence-specific peaks that

correspond to each probe[31]. The intensities (height or area) of the peaks are measured, quantifying the amount of PCR product, followed by normalization and comparison to control DNA samples[31,34]. This indicates the relative amount of target DNA sequence in the input DNA sample, with peak ratios between 0,7 and 1,3 considered normal, ratios above 1,3 or below 0,7 indicating the presence of a gain or a loss of the target sequence, respectively[33,34].

The interpretation of the MLPA results is a key step in the application of this methodology for molecular diagnosis of deletions and/or duplications. Results interpretation can be challenging due to different PCR efficiencies among the different probes and sample-to-sample variations. Different strategies for data analysis, such as Coffalyser, have been developed to allow correct interpretation of raw data[34].

The main advantages of MLPA are its low cost, reliability, high-throughput and speed, since it can be completed within 24 to 48 hours. Besides, it is relatively simple to implement, requiring minimal training for an experienced technician[31,33]. Moreover, MLPA can detect small changes, at the level of a single exon. Comparing to another methods, MLPA requires less DNA, is less technically demanding, is faster than Southern blotting, and yields fewer results of unclear clinical significance than quantitative PCR[33].

However, like any dosage technique, MLPA is sensitive to DNA quality, being sensitive to contaminants and DNA degradation, and requiring 100-200ng of DNA for reliable and reproducible results. Furthermore, MLPA is sensitive to the kind of sample that is used for DNA extraction[34]. It is recommended to compare different MLPA analysis only if they come from the same DNA extraction method, from the same tissue, with control DNA ideally from the same type of source material, and extracted in the same manner as the test sample, in order to reduce variability[33,34]. MLPA cannot detect copy neutral loss of heterozygosity and may experience problems with mosaicism, tumour heterogeneity, or contamination with normal cells[34]. The DNA ligase used in the ligation step is sensitive to the presence of mismatches near the ligation site, however, mismatches further away from the ligation site may have little to no effect on the ligation[33]. If a pathogenic point mutation or polymorphism not affecting gene function occurs within an MLPA probe-binding site, it might hamper the correct hybridization of the probe, resulting in failure of ligation that will appear as a deletion of a whole exon[34]. Sometimes, due to

sequence constraints within the exon, probes may target nearby intronic sequences and signal changes observed for these probes may not reflect the copy number of the exon itself[33]. Therefore, an independent method must be applied to verify apparent single exon deletions detected by MLPA[34].

## Polymerase Chain Reaction

The polymerase chain reaction (PCR) technique was developed in 1983, by Kary Mullis, awarded with the Nobel Prize in Chemistry in 1993[35]. PCR is based on the principle that, through a set of reagents, one can exponentially amplify a chosen region of a DNA molecule *in vitro*, mimicking the naturally occurring cellular process[36,37]. This mixture of reagents includes a template DNA, multiple copies of a pair of synthetized primers, a heat-tolerant DNA polymerase and a supply of nucleotides[37]. The chosen primers determine the region to be amplified. The most widely used DNA polymerase is *Taq* DNA Polymerase I, from the bacteria *Thermus aquaticus*, that is naturally occurring of thermal vents and hot springs. This enzyme is thermostable and can resist to high temperatures, where DNA polymerases from other species would denature[10,37].

The PCR comprises a sequence of steps that occur in a thermal cycler at different temperatures (Figure 3). It initiates with the denaturation of the template DNA at high temperature, usually around 95ºC. The hydrogen bonds that bind the two strands of the double-stranded DNA break and single-stranded DNA molecules are formed[10,37]. This step is followed by the annealing, i.e. the hybridization, of the oligonucleotide primers to their complementary sequences in the DNA template, at a lower temperature, around 50

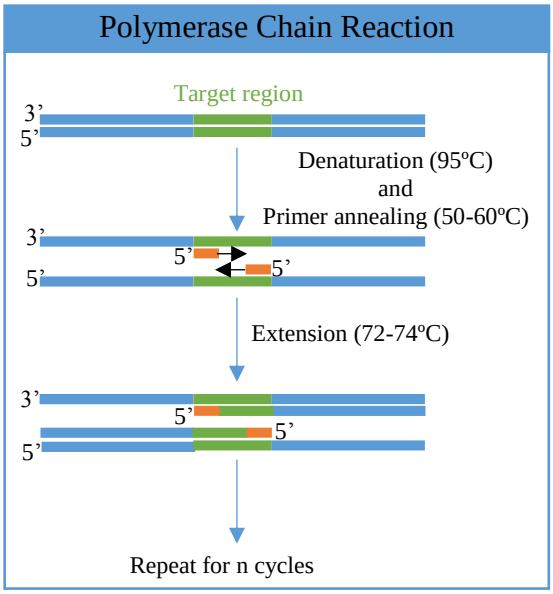


*Figure 3 – The basic steps in the polymerase chain reaction.*

*PCR consists in the repetition of the cycle that comprises three main phases: denaturation of the dsDNA at high temperatures; annealing of the primers to their complementary sequence in the DNA template; and extension of the chains through DNA polymerase synthesis activity. Adapted from Griffiths AJF. Introduction to Genetic Analysis. 10th ed. Winslow S, editor. W. H. Freeman & Company; 2012.*

to 60ºC. Each one of the primers binds to a different end of the gene or region to be amplified, in opposite DNA strands with their 3' ends pointing at each other[37]. At this temperature, the single stranded DNA could be re-joined, but the excess of the primers, that anneal specifically to each strand, prevents this[10]. The temperature is raised until the optimal temperature of the *Taq* polymerase in use, around 72-74ºC[37]. The *Taq* polymerase binds to the ends of the primers, allowing for synthesis of the complementary new DNA strands from the primer, the extension phase[10,37]. At this point, the generated molecules have identical 5' ends, but different 3' ends, due to the randomness of the DNA synthesis finish, forming the "long products"[10]. Then, this cycle repeats itself many times, exponentially increasing the number of DNA molecules, denaturing the "long products" and forming "short products", where the 5' and 3' ends are defined by the primers annealing positions. These "short products" are identical to each other and ideally double at each cycle, while the reagents are being depleted.

### *1) Real-time Quantitative PCR*

The number of DNA molecules present in the starting mixture directly affects the quantity of product synthetized in a given number of PCR cycles. Eventually, the exponential rate of amplification of the PCR reaction stops due to inhibitors of the polymerase reaction in the template, reagent limitation or accumulation of pyrophosphate molecules, generating variable amounts of PCR product. At this point the quantitation is not reliable, therefore the extrapolation should be made at the exponential phase of the reaction[35,38]. Initially, these comparisons were made through agarose gel electrophoresis that, although simple and easy to perform, is imprecise. Nowadays, the quantitation is carried out by real-time quantitative PCR (qPCR), that allows the measurement of the product synthesis over time, during the PCR cycles[35].

In 1993, the first publication describing qPCR was published[35]. Quantitative PCR was developed for DNA quantitation, allowing specific and sensitive quantitation of nucleic acids[39,40]. It is advantageous over other methods of gene expression analysis, since: (a) it does not require post-PCR manipulation, reducing contamination risk; (b) it has low instrumentation and consumable costs, a fast turnaround and assay development time, and an open format; and (c) it is highly sensitive, requiring minimal amounts of input DNA[41,42].

It is widely used in biomedical research and diagnosis, particularly in mRNA expression studies, DNA copy number measurements in genomic or viral DNAs, allelic discrimination assays, expression analysis of specific splice gene variants, as well as other scientific fields, such as water, food and feed testing[38,40]. Furthermore, qPCR has been used to validate aCGH results, such as the ones regarding CNVs, in several studies[22,43–46].

During qPCR, the thermocycler collects fluorescence emission data that are analysed by the computer software, which constructs amplification plots, in the form of a sigmoidal amplification curve (Figure 4)[38,41]. When analysing these plots, it is important to recognize some important nomenclature related to PCR. First of all, the baseline is the PCR cycles where the fluorescence signal remains beneath the limits of detection of the thermocycler, usually until cycle 15. The software calculates the threshold based on the standard-deviation of the average signal of the baseline signal, although it can be manually altered. It is considered a real signal when a fluorescent signal is above the threshold, and it can be used to define the threshold cycle (Ct) for a sample[38,42]. The Ct, or the quantification cycle (Cq), is the fractional cycle at which the reporter fluorescence is higher than the threshold, being essential for accuracy and reproducibility. It is inversely proportional to the initial amount of input DNA, and it always occurs during the exponential phase of the amplification, in the early cycles of the reaction. As the PCR continues, the amplification rate decreases until it reaches the plateau phase, where the fluorescence stabilizes[38,41].

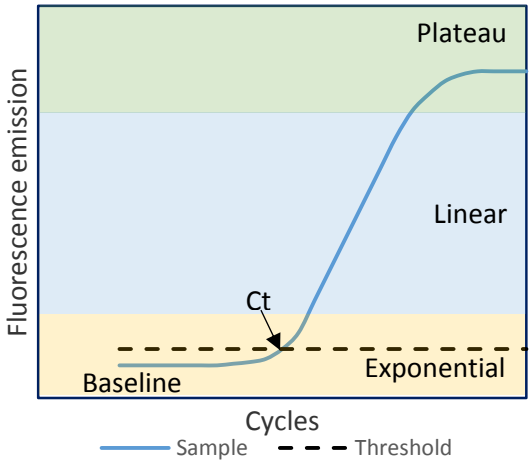


*Figure 4 – Model of a single amplification plot illustrating the nomenclature commonly used in qPCR.*
*The PCR goes through three phases, the exponential, the linear and the plateau phases. The PCR product increases throughout the cycles, increasing the fluorescence emission and generating a sigmoidal curve. Adapted from Arya, M., Shergill, I. S., Williamson, M. & Gommersall, L. Basic principles of real-time quantitative PCR. Expert Rev Mol Diagn 5, 209–219 (2005).*

### *2) Experimental Design*

Although it might seem a simple technique, the PCR experiment design must be well thought and planned, with all the assays run under the same conditions[47]. All the steps can limit efficiency, sensitivity and accuracy of the technique. In qPCR the aim is to minimize the interference with the doubling of the target amplicon at each cycle, in order to maximize efficiency[39]. PCR efficiency, often expressed in percentage (%), is defined by Bar and colleagues as "the fraction of double-stranded DNA molecules that is copied at a given cycle", with an efficiency of 100% representing a perfect doubling of all DNA molecules[40]. PCR efficiency is often affected by $MgCl_2$, primer and probe concentrations[47]. qPCR assays must be appropriately designed in order to amplify target DNA with at least 90% efficiency[48].

The optimization of the amplification step is essential and can be achieved through modifications of the nucleic acids preparation protocol and the choice of adequate primers, buffers and cycling conditions[39]. The nucleic acids purity of samples, frequently obtained from fixed archival, fresh cell culture or blood products, is highly important in order to have reproducibility. In order to determine DNA purity and the presence of PCR inhibitors, if possible, each DNA sample should be tested with a standard curve[47].

The primers design is of extreme importance for the success of a PCR, because incorrect primers can result in no amplification at all or in non-specific amplification of multiple fragments. The primers sequence must flank and be complementary to the target region on the DNA template, with their 3' ends pointing towards one another. The size of the primer is also of great importance to achieve specific hybridization. On one hand, a primer that is too short might hybridize to non-target sites and produce undesired fragments. On the other hand, a long primer hybridizes at a slow rate, and complete hybridization does not occur in adequate time, reducing the efficiency of the PCR. Thus, primer's optimal length is between 15 and 20 nucleotides, with a GC content of 20-80%, and the melting temperature ($T_m$) for both primers of the pair should be similar ($\Delta Tm \leq 2^oC$). In order to maximize the specificity of primers, it is recommended for each primer to have only one or two Gs or Cs within the last 5 nucleotides, at the 3' end. The generated DNA fragment, the amplicon, ideally should be less than 1kb in length, in order to maintain the amplification efficiency, obtain consistent results and be tolerant to reaction conditions. When the utilization of probes is required, probe design also follows a set of

specifications. In order to guarantee that the probe is fully hybridized during primer extension, it is desirable that the probe Tm is 8-10ºC higher than the Tm of the primer[10,38].

Regarding to the temperature conditions, the annealing temperature ($T_a$) is the most important, once it can affect the specificity of the reaction. At high temperatures, the hybridization cannot occur, and the primers and template DNA remain dissociated. In contrast, at low temperatures mismatched hybrids are frequent and stable, producing non-target amplification. Hence, the annealing temperature must be low enough to allow hybridization, but high enough to prevent the formation of mismatched hybrids. In order to estimate this "perfect" temperature, one should calculate the $T_m$ of the primer-template hybrid, at which they dissociate, that is 1-2ºC higher than the annealing temperature. Therefore, the primers should have identical $T_m$ so the annealing temperature is adequate for both. The $T_m$ can be determined experimentally or, more frequently, it can be calculated from the formula:

$$T_m\ (ºC) = 4 \times (G + C) + 2 \times (A + T)$$

in which (G + C) is the number of G and C nucleotides in the primer sequence, and (A + T) is the number of A and T nucleotides[10]. The hot start consists of starting the PCR reaction at high temperature, in order to eliminate unwanted hybridization that is frequent in the first cycle. This increases sensitivity and reproducibility of the amplification and quantitation[39].

As previously described, quantitation should be performed in the exponential phase of the PCR, frequently somewhere between the 15$^{th}$ and the 30$^{th}$ cycles. These are usually associated with the linear range of amplification, where the quantitated amount of amplified target is proportional to the initial amount of target molecules. Often, this linear range of amplification is defined through the testing of control samples or standards[39].

### *3) Normalization and Reference Genes*

During qPCR experiments, specific errors are introduced due to minor differences in the starting amount, the quality of DNA, or the PCR amplification efficiency. In order to minimize these errors and correct non-specific variation, data from qPCR should be submitted to appropriate normalization. Usually, reference genes in CNVs studies, or

housekeeping genes in the case of expression studies, are selected in the surrounding area or on the same chromosome as the disease gene[42]. While a housekeeping gene should be theoretically expressed at a constant level in different tissues of an organism, at all stages of development, and in different experimental conditions, the reference gene should not have CNVs described in the literature. Generally, the housekeeping genes are indispensable for the cellular regular function. These internal reference or housekeeping and the target gene are then simultaneously amplified, with the assumptions that their copy number is equal or their expression is similar between all samples and resistant to experimental conditions, since it undergoes all steps of the qPCR with the same kinetics as the target gene[35,38]. The choice of the reference gene must be made with regard to the specificities of each experiment, since it significantly influences the results reliability. However, there has not been identified an unequivocal reference gene yet and the use of several simultaneous reference genes has been suggested, with their mean quantitation used for normalization[38]. This approach is useful when the housekeeping gene is affected by the experimental design and treatments or when comparing different tissue types and provides a more robust, reliable and accurate quantitation[35,42]. Some of the commonly used reference genes include *β-actin*, *GAPDH*[35,38,49,50], *ZNF80*[41,51–55] or *GPR15*[41,52,54].

### *4) Amplicon Detection*

During the qPCR experiment, the product accumulation may be followed by one of several different chemistries, including double stranded DNA specific binding dyes, i.e. dsDNA-intercalating agents, or fluorescent probes, such as hydrolysis probes, dual hybridization probes and molecular beacons[38,41]. It is important to select the adequate method for a specific PCR assay, according to its objective.

The dsDNA-intercalating agents are nonsequence-specific fluorogenic DNA-binding dyes that intercalates into dsDNA. It is the simplest method, relatively cheap, and includes SYBR® Green I and EvaGreen®. When bounded to double-stranded DNA, the dye gives a fluorescent signal, measuring the total amount of double-stranded DNA at the end of the elongation step of each PCR cycle. However, it might overestimate the true amount of product because it binds to all dsDNA, such as non-specific PCR products and primer-dimers that sometimes are formed. In order to increase the reaction specificity, melting

curves should be generated and compared. The melting peak at the melting temperature of the amplicon differentiates it from the artifacts with lower melting temperatures and broader smaller peaks[10,38]. EvaGreen® outperforms SYBR® Green, achieving high reaction efficiencies and giving higher and sharper peaks in melt curve analysis, due to the fact that it can be used at relatively high concentrations[56,57]. Also, EvaGreen® is less inhibitory to PCR than SYBR® Green and has a relatively low tendency to cause nonspecific amplification. Moreover, EvaGreen® is very stable and spectrally compatible with existing PCR instruments[57].

The hydrolysis probes require a forward and a reverse primer, and their efficiency is dependent on the 5' to 3' exonuclease activity from *Taq* polymerase, for example. The fluorogenic non-extendable "TaqMan" probe has a covalently bonded fluorescent reporter dye attached at its 5' end and a quencher dye attached to the 3' end. When intact, the quencher is close to the reporter dye, decreasing the fluorescent signal, through fluorescence resonance energy transfer (FRET). However, during the real-time quantitative TaqMan® assay, the fluorogenic probe hybridizes with the target sequence and, during the extension phase, is cleaved by the 5' nuclease activity of *Taq* polymerase. The cleavage removes the probe from the target strand and separates the dyes, increasing the fluorescence of the reporter dye. At each cycle, more and more probes are cleaved, increasing the fluorescence intensity in a way that is directly proportional to the amount of generated fragments[38].

The method of dual hybridization probes requires two different oligonucleotides specific for the PCR product probes, in a tandem-like manner. The downstream probe carries a donor fluorophore at its 3' end and the upstream probe has an acceptor fluorophore attached to its 5' end. Each of the probes anneals with its respective target sequence in a head-to-tail manner, after denaturation. This way, the two fluorophores are brought closer in spatial proximity, allowing FRET. The donor dye transfers energy to the acceptor, that dissipates fluorescence at a different wavelength. As a consequence, this method is highly specific because the fluorescence only occurs when the two independent probe hybridize correctly to their target sequences[38,58].

In the same way as the hydrolysis probes, the molecular beacons have a fluorescent reporter dye attached at its 5' end and a quenching compound attached to the 3' end.

However, when near or below the annealing temperature and free in solution, the intact probe has the two ends base pairing to one another due to short complementary sequences, forming a hairpin or a stem-loop structure. As a result, the quencher is next to the reporter dye, decreasing the fluorescent signal, based on the FRET principle. During the annealing step, the loop portion, complementary to the target sequence hybridizes, causing a conformational change, and the stem is cleaved. Then, the dyes are separated and the quencher no longer decreases the reporter dye fluorescence, generating a fluorescent signal. At each cycle of the PCR, the molecular beacons rehybridize to the target sequence for fluorescence emission, indicating the amount of accumulated amplicon[10,38,58].

### *5) Quantitation Methods*

One of the methods to quantify real-time PCR results is the standard-curve or absolute quantitation, often used to measure a small number of genes in a few or many samples[47]. It is based on the plotting of a standard curve, constructed with the logs of copy numbers from standards with known quantities of initial DNA versus their corresponding Cts, which generates a straight line. Then, the Ct of the test samples is measured and compared with the standard curve, in order to determine the initial target amount[38]. In other words, the amount of product amplified from the test sample is compared with the amount of product amplified from standards with known quantities of initial DNA[10].

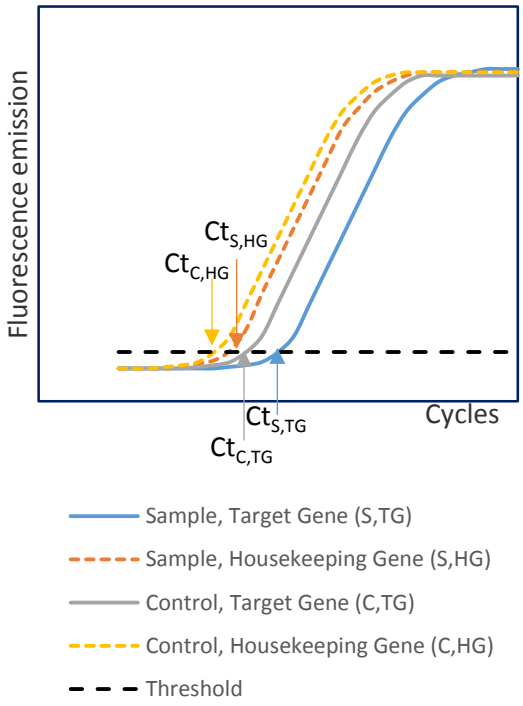


*Figure 5 – Model of an amplification plot illustrating the mathematical basis of the $2^{-\Delta\Delta Ct}$ method.*
*This method allows relative quantitation through the measurement of cycle thresholds from the target and housekeeping genes of the test and control samples. Adapted from Vanguilder, H. D., Vrana, K. E. & Freeman, W. M. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 44, 619–626 (2008).*

Another method of quantitation in qPCR is the relative quantitation or Delta Delta Ct method ($2^{-\Delta\Delta Ct}$ method), which requires a reference control and a

housekeeping gene (Figure 5). First of all, if this method is applied, the results only are considered valid and reliable if the amplification efficiencies of the housekeeping and target genes are very similar and at or above 90%[38,40]. To guarantee that, a 10-fold serial dilution of a positive control template can be performed, followed by the plotting of the Ct as a function of $\log_{10}$ (concentration of template). The trend line slope (M) gives the PCR efficiency, through the function $Efficiency = 10^{1/-M} - 1$, with –3,32 indicating an efficiency of 100%[47]. The relative expression levels of a target are calculated relative to a reference control (calibrator) through mathematical equations. The target amount of the sample is normalized to an endogenous housekeeping gene and relative to the normalized control. It is calculated by the formula[38]:

$$2^{-\Delta\Delta Ct},$$

where $\Delta\Delta Ct = \Delta Ct_{sample} - \Delta Ct_{control}$,

and $\Delta Ct_{sample/control} = Ct_{\ target\ gene} - Ct_{housekeeping\ gene}$.

## Objectives

The present work describes the development and validation of a qPCR method for CNVs detection. A qPCR-based methodology was developed to search for large deletions and/or duplications in genes in two situations: a) sequences which are not covered by aCGH or MLPA kits and b) confirmation or exclusion of previous MLPA and aCGH results.

To accomplish that, initially the methodology was validated using samples with known results for the genes *MECP2*, *NF1* and *PARK2*. Then, the method was developed in order to search for CNVs upon request. At the moment, 33 genes were assessed. The work flow begins with the selection of regions and primers design, followed by the primers validation. When the primers are validated, the samples are subjected to qPCR and the results analysed.

## Methods

### Samples

For validation of the study, 3 samples with normal copy number, 1 sample with a known deletion and 1 sample with a known duplication, all regarding the *MECP2* gene, were used. Also, 5 samples with normal copy number for the tested fragment, 1 sample with a known deletion and 1 unknown sample were used for each *NF1* and *PARK2* genes. All samples were extracted from blood (EDTA) and studied previously by another method, such as MLPA, aCGH or other.

For the development of the study 36 samples, extracted from blood (EDTA), were used as calibrators and/or normal samples. The calibrator samples were selected from genomic DNA samples with prior diagnosis of another disease without relation to CNVs of the tested fragment, which were beforehand studied by aCGH or MLPA. At first, in some cases, calibrator samples were commercial normal genomic DNAs, used in NGS or aCGH, and pools of 3 samples used for MLPA. After several tests it was decided that calibrator samples should be a pool of 3 normal samples. When the tested fragment was from a gene of the X chromosome, the calibrators were male pools if the test sample was male or female pools if the test sample was female. Also, 32 unknown samples, extracted from blood (EDTA) or, in some cases, purified genomic DNA, were tested for several genes.

The negative control consisted of water, without template DNA, in order to check for primer-dimer and contaminations. All samples (Table S1, ANNEX I) were diluted with TE buffer to a concentration of 20ng/µL ±1,5ng/µL, and quantified in EPOCH (BioTek).

### Selection of Regions and Primers Design

One region of the the *ZNF80* gene and another of the *GPR15* gene, having no CNVs described in the literature, were used as reference genes for normalization, during the relative quantitation[51–53].

The regions to be amplified in the test samples were selected based on the described gross deletions and gross insertions of the gene, recurring to The Human Gene Mutation Database (HGMD; http://www.hgmd.cf.ac.uk/), in order to ensure the detection of small deletions and duplications. The genes of interest were searched on the Ensembl Genome

Browser (http://www.ensembl.org/), with the organism "human" selected. Each gene sequence was downloaded in FASTA format and converted into Microsoft Office Word format for primer design.

Primers were designed with the online program Primer3Plus (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi), according to the following criteria: primer size of 18/20/22bp (minimum/optimal/maximum); GC content between 40 and 60%; Tm of 58/60/62ºC (minimum/optimal/maximum); amplicon size of 100/150/200bp (minimum/optimal/maximum). When not possible, primers were designed manually through the software Alamut Visual v. 2.8 rev. 1 (©2015 Interactive Biosoftware), following the same specifications as for Primer3Plus. The selected primers were analysed on the online program SNPCheck v3 (http://secure.ngrl.org.uk/SNPCheck/snpcheck.htm), to verify that these did not target SNPs. The primers structure analysis, such as primers complementarity, hairpin formation and dimers, was made through the online program Oligo Analyzer 3.1 (http://eu.idtdna.com/calc/analyzer), using default settings for qPCR, with the exclusion criterion of $\Delta G \leq -9,0$ kcal/mol and the software Oligo Analyzer 1.0.2 (Copyright© 2000-2002 Teemu Kuulasmaa), with the exclusion criterion of $\Delta G \leq -6,0$ kcal/mol. To ensure primer specificity, primer sequences were run in the online program Primer search 2.52 (http://bisearch.enzim.hu/) and in the UCSC *In Silico* PCR online tool (http://genome.ucsc.edu/cgi-bin/hgPcr), which also gave primers melting temperatures. The secondary structure of the amplicon sequences was predicted with the online program DNA Folding Form from The mfold Web Server (http://unafold.rna.albany.edu), using default settings and a temperature of 60ºC, with the exclusion criterion of $\Delta G \leq -7,0$ kcal/mol. In order to predict the melting curve for each fragment, its sequence was inserted in the web-based application uMelt$^{SM}$ – DNA Melting Curve Prediction (https://www.dna.utah.edu/umelt/um.php).

The designed primers are summarized in Table S2 (ANNEX II); primers pairs for the *ZNF80*, *MECP2*, *PLCB1*, *PARK2* and *NF1* genes were supplied by Sigma-Aldrich and the remaining were supplied by Invitrogene, due to financial constraints.

**Empirical Validation of Primers**

After *in silico* testing of primers, an empirical validation was performed, similar to the one described by D'haene and colleagues, in order to optimize primers[41]. To estimate qPCR efficiency, the standard curve is the most reliable and robust approach, broadly accepted by the scientific community[48]. A human normal genomic DNA dilution series (1:2 with 5 dilution points of 26ng/µL, 13ng/µL, 6,5ng/µL, 3,25ng/µL and 1,625ng/µL and a no-template control) was prepared and amplified. The concentrations of the dilution series were set in order to extend over the anticipated range of interest (18,5 – 21,5ng/µL) by at least 20%[48]. Assays were performed in duplicate, in a 20µL reaction volume, containing 4µL of 5x HOT FIREPol® EvaGreen® qPCR Mix Plus, 1µL of gDNA, 0,1-1,0µL of forward and reverse primers at an initial concentration of 10nM (final concentration of 200nM) and DEPC Treated Water (Ambion, Life Technologies) up to the reaction volume. Reactions were performed in 0,1mL Strip Tubes in a 72-well rotor using the Rotor-Gene Q (Qiagen, Germany) detection system. Cycling conditions were 95ºC for 15min, for initial activation step, followed by 40 cycles of 95ºC or 99ºC for 15sec, 60ºC for 20sec and 72ºC for 20sec. A standard curve was generated, with the evaluation of the amplification efficiency, coefficient of determination ($R^2$) and slope. The PCR efficiency was considered satisfactory when between 90% and 110%, although in some cases an efficiency above 80% was accepted, and the $R^2$ was considered suitable when not lower than 0,98, preferably over 0,99[41]. Afterwards, a melt curve analysis was performed to check the specificity of the PCR reactions. The Ct values were extracted and analysed with the Rotor-Gene Q software, version 1.7.94.

The PCR programme with a melting temperature of 95ºC was used as standard. When the fragment melting curve prediction by the uMelt software was near or over 95ºC, the PCR program with a melting temperature of 99ºC was used for validation. Also, if a melting curve of a fragment first tested at 95ºC was highly close to 95ºC or the melting was incomplete, the same fragment was validated with a melting temperature of 99ºC. When the efficiency was lower than 80% or higher than 110%, the primers volume was increased or reduced, respectively, in the following validation run, until it was validated.

**qPCR Assays for CNVs Detection**

All real-time qPCR assays were performed in a 20µL reaction volume, containing 1µL of genomic DNA at an initial concentration of 20ng/mL, 4µL of 5x HOT FIREPol® EvaGreen® qPCR Mix Plus, 0,1-1,0µL of forward and reverse primers at an initial concentration of 10nM (final concentration 200nM) and DEPC Treated Water (Ambion, Life Technologies) up to the reaction volume. Reactions were performed in 0,1mL Strip Tubes in a 72-well rotor using the Rotor-Gene Q (Qiagen, Germany) detection system. Cycling conditions were 95ºC for 15min, for initial activation step, followed by 40 cycles of 95ºC or 99ºC for 15sec, 60ºC for 20sec and 72ºC for 20sec. Primers volume and cycling conditions were used according to the ones used in the primers validation, when optimized.

During validation of the method, the *MECP2* experiment contained 2 calibrator samples (in duplicate), a sample with normal copy number (in duplicate), a sample with a known deletion (in duplicate), a sample with a known duplication (in duplicate), and a NTC (in duplicate). The experiments regarding the *NF1* and *PARK2* genes contained 2 calibrator samples (in triplicate), a sample with normal copy number (in triplicate), a sample with a known deletion (in triplicate), an unknown sample (in triplicate), and a NTC. Throughout the method development, each experiment contained 3 calibrator samples (in triplicate), a sample with normal copy number (in triplicate), at least, a test sample (in triplicate), and a NTC. In order to test genes located at the X chromosome, the 2 calibrator samples (in triplicate) and the sample with normal copy number (in triplicate) were pools of male samples only, since the test sample was male; also, we used a pool of female samples as a positive control (in triplicate), as well as a test sample (in triplicate), and a NTC.

The copy number level of target genes was normalized to the *ZNF80* or the *GPR15* gene. After PCR amplification, a melting curve was generated for each PCR product and analysed in order to ensure amplicons specificity. Data collection and analysis was performed through Rotor-Gene Q software, version 1.7.94, with the Ct values exported to Microsoft Office Excel.

**Data Analysis and Statistics**

The data analysis was performed using the Microsoft Office Excel software, based on the method described by Hoebeeck and colleagues[52]. Reaction wells with obvious PCR reaction failure or clear outlier values (difference between Ct and mean Ct greater than 0,5) were excluded from data analysis. Copy number alterations were calculated according to the Delta Delta Ct method ($2^{-\Delta\Delta Ct}$) and the samples were scored as normal, deleted or duplicated based on the scoring limits established. A haploid copy number of 1 is expected for a normal sample, a value of 0,5 for a sample with a deletion and a value of 1,5 for a sample with a duplication[22,59,60].

The expected cut-off values used were 0,7 for deletion upper limit[51–53], 0,8 for normal lower limit[51–53], 1,2 for normal upper limit[53] and 1,25 for duplication lower limit[50]. As such, fragments with a haploid copy number below the deletion upper limit were scored as deleted, while fragments with a haploid copy number between the normal lower and upper limits were scored as normal, and finally fragments with a haploid copy number above the duplication lower limit were scored as duplicated. A fragment with an ambiguous haploid copy number was reanalysed.

To determine the accuracy and to evaluate the precision of the qPCR approach, the haploid copy number values and their standard deviations for the deleted vs. control samples and duplicated vs. control samples were compared. In order to evaluate the equality of the two variances, a F test was performed (p-value < 0,05). If the two variances were not significantly different, then the two-sample t test for independent samples with equal variances was used; otherwise, a two-sample t test for independent samples with unequal variances was selected. The two-sample *t-test* was performed to evaluate if the copy numbers are significantly different between deleted and control samples as well as between duplicated and control samples (p-value < 0,05).

At the end of all tests, based on the observed haploid copy number and the standard deviation, the limits for scoring a deletion, a normal copy number or a duplication were assessed. The mean value of the haploid copy number in deleted fragments plus two times the standard deviation was used as upper limit for scoring a deletion. For scoring a normal fragment, the mean value of the normal haploid copy number minus two times the standard deviation was used as lower limit and the mean value of the normal haploid copy number

plus two times the standard deviation was used as higher limit. Similarly, the mean value of the haploid copy number in duplicated fragments minus two times the standard deviation was used as lower limit for scoring a duplication.

The qPCR results from the method validation were compared to previous results of MLPA, aCGH, Sanger sequencing and NGS, in order to evaluate the sensitivity (false-negative results), specificity (false-positive results) and accuracy (true-value determination) of the qPCR approach.

## Results and Discussion

### Method Validation

In order to validate the planned methodology, qPCR for the *MECP2*, *NF1* and *PARK2* genes were performed. The primers were designed according to the CNVs detected in the samples previously studied by another method and validated through the generation of a standard curve and melt curve analysis. The standard curve allowed the efficiency assessment of each primers' pair (Figure 6).

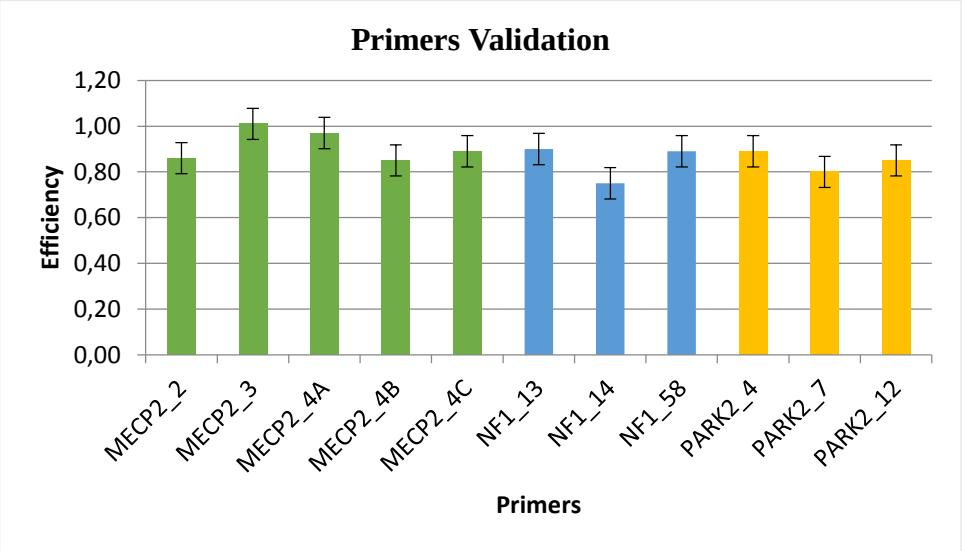


*Figure 6 - Efficiency of the primers pairs for the MECP2, NF1 and PARK2 genes, used in the methodology validation.*
*The mean efficiency of the primers pairs was 0,88±0,07, with the primers pair for exon 14 of the NF1 gene being the one with the lowest accepted efficiency (0,75). The primers pair MECP2_4B was discarded from the relative quantitation analysis, since it was considered not specific through the melt curve analysis.*

### *1) The MECP2 gene*

Regarding the *MECP2* gene, 3 samples with normal copy number (calibrators and samples A1), 1 sample with a known deletion (sample A4) and 1 sample with a known duplication (sample A5), were used. The *MECP2* gene encodes a methyl-CpG binding protein, which is likely to be involved in transcriptional regulation. The most frequent mutation in the *MECP2* are point mutations, small insertions or deletions[55]. Mutations in this gene are associated with Rett syndrome (MIM 312750), as well as with a severe neonatal form of encephalopathy (MIM 300673) and with X-linked syndromic mental retardation (MIM 300055, MIM 300260). The efficiency and specificity of each primers pair was evaluated through the generation of standard curves from a dilution series of DNA from two of the samples with normal copy number and a melting curve, respectively. Although primers MECP2_2 and MECP2_4C had efficiencies slightly below 90%, the initial cut-off value for acceptance, their melting curves showed specificity, and they followed to quantitative assay (Figure 6). The primer pair MECP2_4B was discarded from the subsequent quantitative assay, since it had a low efficiency in the standard curve from primers validation and the melt curve analysis showed that it was not specific.

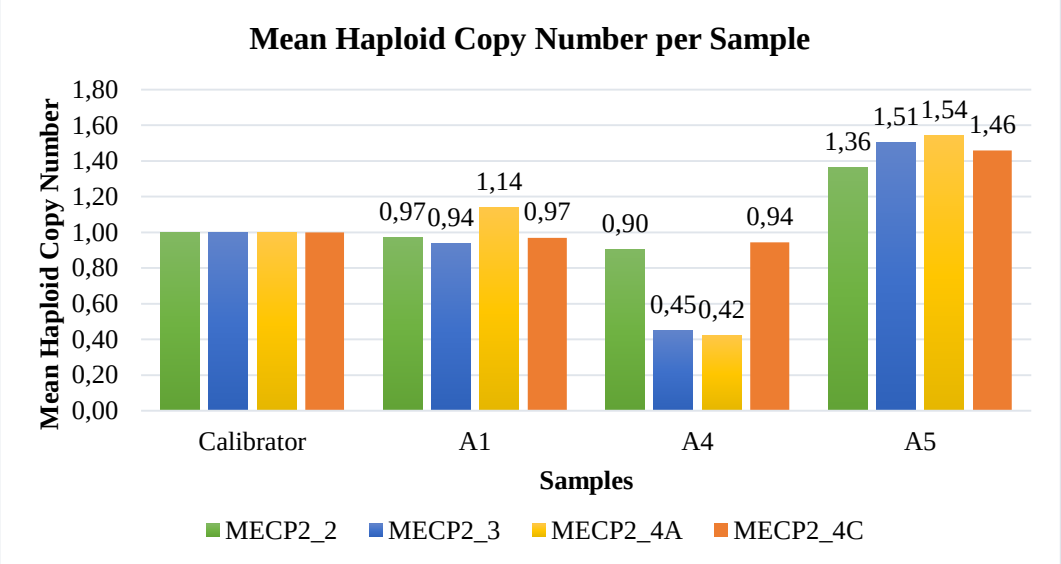


*Figure 7 - Relative quantitation of the MECP2 gene, during method validation.*
*Sample A1 had a normal haploid copy number (1,01±0,08) for every primers pair. Sample A4 presented a normal copy number (0,92±0,02) for exon 2 and the end of exon 4, as well as a heterozygous deletion (0,44±0,01) between exon 3 and part of exon 4. Sample A5 had a whole gene heterozygous duplication (1,47±0,07).*

The results from the quantitative PCR (Figure 7) were concordant with the previous findings by aCGH and MLPA. Sample A1 had a normal haploid copy number (1,01±0,08) for every primers pair as expected, and MLPA was negative for deletions or duplications. Sample A4 showed a heterozygous deletion of exon 3 in the aCGH study and a heterozygous deletion of exon 3 and part of exon 4 in the MLPA. qPCR of this sample presented a normal copy number (0,92±0,02) for exon 2 and the end of exon 4, as well as a deletion (0,44±0,01) between exon 3 and part of exon 4. Also, sample A5 had a whole gene heterozygous duplication (1,47±0,07), in agreement with the heterozygous duplication detected by MLPA. Since the relative quantitation results were in agreement with previous results, the efficiency cut-off value for primers acceptance was redefined to 80%. However, primers efficiency was still preferred between 90 and 110%.

### *2) The NF1 gene*

Subsequently, the *NF1* gene was tested, by analysing 3 samples with normal copy number, 1 sample with a known deletion and 1 unknown sample. The unknown sample, blinded for qPCR testing, was previously tested at another lab by NGS and qPCR, which reported a duplication of exon 13; however, the results of MLPA at this lab were negative for deletions or duplications of the NF1 gene. The Neurofibromin 1 (*NF1*) gene (OMIM* 613113) is a tumour suppressor, located at 17q11.2 locus, which contains 57 constitutive coding exons and 3 alternative spliced exons over 280kb[61,62]. *NF1* is a negative regulator of the oncogene *Ras*, being a Ras-GTPase activating protein and its loss is associated with increased formation of Ras-GTP leading to increased downstream signalling[63,64]. Heterozygous mutations in this gene cause neurofibromatosis type I (MIM 162200). Even though exon 14 primer pair had a low efficiency of 75% (Figure 6), all primers were considered validated, since the melt curve analysis indicated amplification of specific fragments.

Sample A10 had a normal haploid copy number (0,99±0,04) for every pair of primers, as predicted, since it was a sample with a positive result for DMD deletion, and therefore it was not expected to have another deletion/duplication. Sample A11 showed a heterozygous deletion of the entire gene in the MLPA study. qPCR of this sample presented a heterozygous deletion (0,65±0,03) of all tested fragments. Furthermore, the

blinded sample A12 showed a normal copy number (0,99±0,07), discordant with the results from another lab, however in agreement with the MLPA results. As such, the duplication of exon 13 was ruled out as a false-positive result.

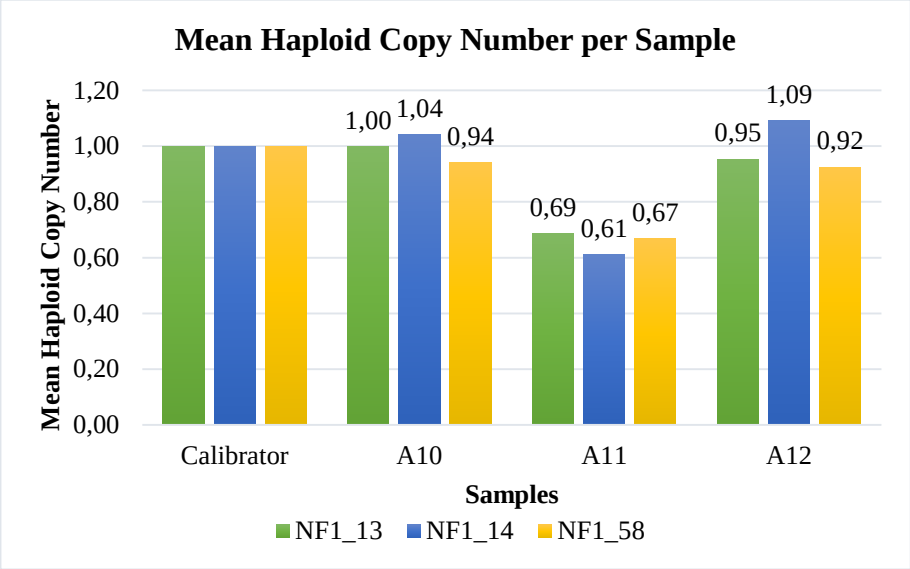


*Figure 8 - Relative quantitation of the NF1 gene, during method validation. Sample A10 had a normal haploid copy number (0,99±0,04) for every primers pair. Sample A11 presented a heterozygous deletion (0,65±0,03) of all tested fragments. Sample A12 showed a normal copy number (0,99±0,07).*

### *3) The PARK2 gene*

In order to test for the *PARK2* gene, 3 samples with normal copy number, 1 sample with a known deletion and 1 unknown sample, were used. The unknown sample was previously tested by MLPA, which showed a heterozygous deletion of exons 4 to 7, but it was blinded for qPCR testing. The *PARK2* gene (OMIM* 602544) encodes an E3 ubiquitin ligase, being located on chromosome 6, locus 6q25.2-q27[65,66]. Autosomal recessive juvenile Parkinson disease-2 (MIM 600116) is caused by homozygous or compound heterozygous mutation in this gene. *PARK2* encompasses a genomic region of 1.3Mbp, consisting of 12 exons, and its 4.5kbp transcript contains a 1,395 bp open reading frame[65]. *PARK2* is located within FRA6E, one of the most unstable common fragile sites of the human genome[65]. All primers were considered validated, since their efficiency was above 80% (Figure 6) and the melt curve analysis indicated specific fragments amplification.

Sample A10 had a normal haploid copy number (1,02±0,01) for every primers pair as expected, since it was a sample with a positive result for DMD deletion. Sample A13

showed a compound heterozygous deletion of exons 3 to 4 and 3 to 6 in the MLPA study and a sequencing failure of exons 3 and 4 after Sanger sequencing. qPCR of this sample presented a normal copy number (1.03±0,05) for exons 7 and 12, as well as a false homozygous deletion (0,00±0,00) of exon 4, since it cannot differentiate between a homozygous and a compound heterozygous deletion. Moreover, the blinded sample A14 had a heterozygous deletion (0,55±0,09) of exons 4 and 7, but it had a normal copy number for exon 12, in agreement with the posteriorly known MLPA results. Therefore, the qPCR results were concordant with the previous findings for this gene either from MLPA or Sanger Sequencing.

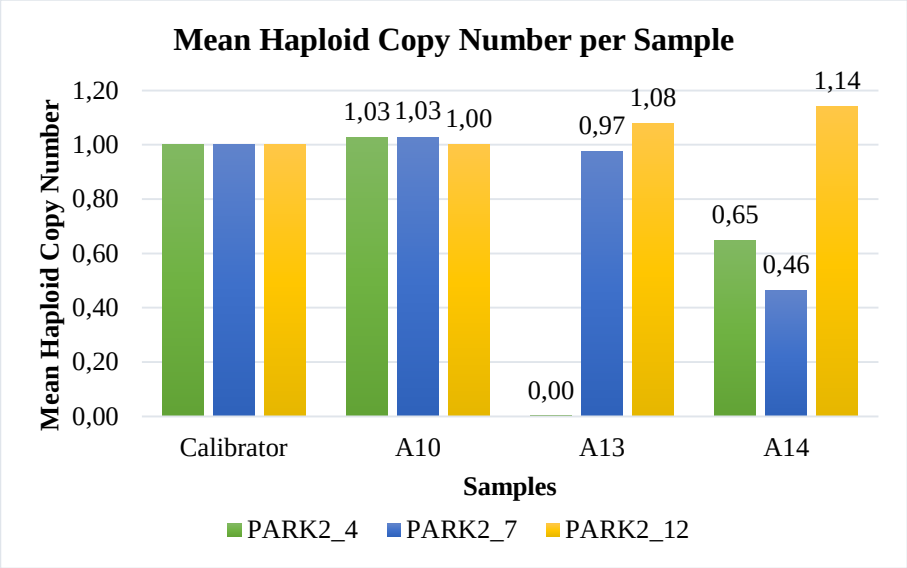


*Figure 9 - Relative quantitation of the PARK2 gene, during method validation. Sample A10 had a normal haploid copy number (1,02±0,01) for every primers pair. Sample A13 presented a normal copy number (1.03±0,05) for exons 7 and 12, as well as a false homozygous deletion (0,00±0,00) of exon 4. Sample A14 had a heterozygous deletion (0,55±0,09) of exons 4 and 7, but it had a normal copy number for exon 12.*

### *4) Results comparison*

The qPCR results were compared to the previous results of aCGH, MLPA or other tests, in order to evaluate the sensitivity (false-negative results), specificity (false-positive results) and accuracy (true-value determination) of the qPCR approach. The total rate of concordant results was of 95%, the total rate of false-positive results was of 2,5%, and the rate of total false-negative results was of 2,5% (Figure 10).

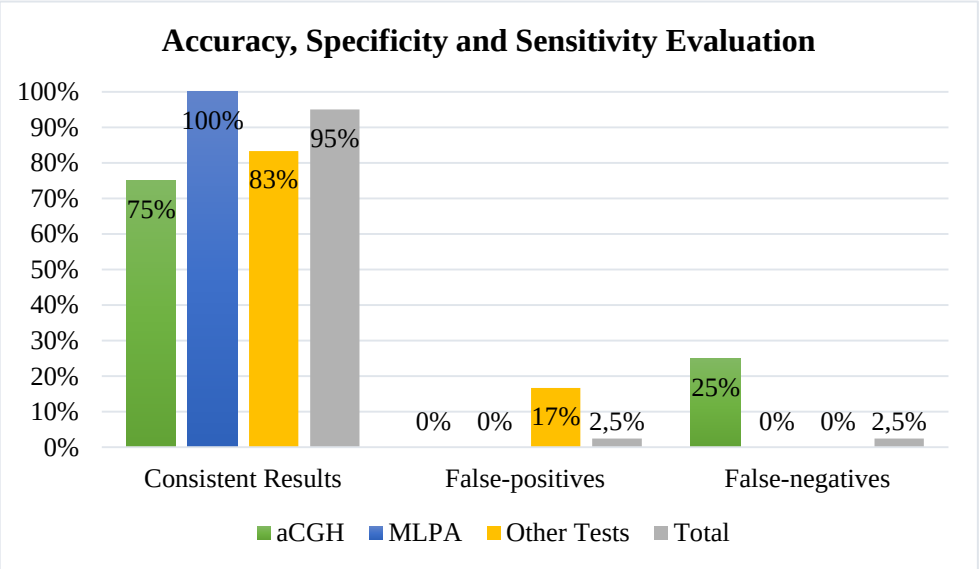


*Figure 10 - Evaluation of the accuracy, specificity and sensitivity of the developed methodology.*
*qPCR and aCGH had 75% of their results in agreement, with the remaining 25% being false-negative aCGH results. MLPA and qPCR had all results consistent with each other. Other tests were 83% concordant with qPCR, however with a 17% rate of false positive results. Overall, qPCR had a total of 95% consistent results with other methods, as well as 2,5% of false-positives and 2,5% of false-negative results*

## Reference Genes

The relative quantitation method for CNVs detection requires the use of at least one reference gene for normalization[38]. The *ZNF80* and *GPR15* genes were selected to be used as reference genes for the developed method, since they possess two copies in the human genome and deletions or duplications are not described in the Human Gene Mutation Database®. Both genes are located on the long arm of chromosome 3[52]. The zinc finger protein 80 (*ZNF80*) (MIM 194553) gene is located at the 3q13.31 locus, and the G protein-coupled receptor 15 (*GPR15*) (MIM 601166) gene is located at the 3q11.2 locus[51,54,67,68].

Both *ZNF80* and *GPR15* genes were validated at the melting temperatures of 95ºC and 99º C, with a mean efficiency of 0,89(±0,07) and a $R^2$ of 0,9911 to 0,99708 (Figure 11). During the relative quantitation for CNVs detection, *ZNF80* was first used as the reference gene. When the mean haploid copy numbers calculated using the *ZNF80* gene were in close proximity with the cut-off values or between the cut-off values for deletion/normal or normal/duplication, the tested fragments were re-tested using the *GPR15* gene as reference gene. Both genes proved to be good candidate genes for normalization in the developed methodology.

## Primers validation

In order to validate the designed primers, standard curves were retrieved for each primers pair, from qPCR assays of a dilution series of genomic DNA with normal copy number for the tested fragments. The efficiencies were considered acceptable when between 90% and 110%, however, efficiencies over 80% were accepted in several cases. The $R^2$, which indicates the reproducibility of the results, was accepted when superior to 0,98, even though it should be preferably over 0,99.

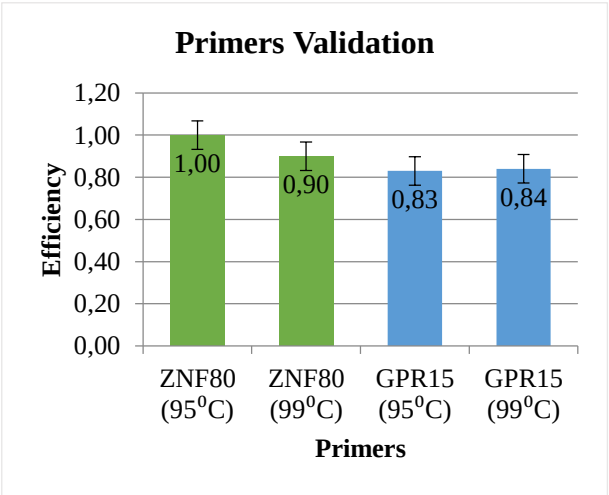


*Figure 11 - Efficiency of the primers for the reference genes, using the cycling programs with melting temperatures of 95ºC and 99ºC.*
*The mean efficiency of the primers pairs was 0,89±0,07, with the primers pair the GPR15 gene with lower efficiencies than the ones for ZNF80.*

Nonetheless, in a few cases an inferior value had to be accepted.

A melt curve was performed to evaluate the specificity of the primers pairs. The primers pairs with a single sharp peak in the melt curve were considered acceptable, since they show the amplification of only one fragment, and do not form primer dimers. Also, primers pairs with a melting curve according to the prediction by uMelt were accepted, since small sequences with a high CG ou AT content might cause small deviations of the peak.

However, some primer pairs could not be validated since the efficiency was much higher than 110% or much lower than 80%, the $R^2$ were too low and/or the melting curves showed a non-specific fragment or failure of amplification (Table 1). In these cases, primers were redesigned in the proximities of the first sequence.

*Table 1 - Efficiency, $R^2$ and PCR conditions of the not validated primers.*

| Primer | Efficiency | $R^2$ | Primer Volume (µL) | Melting Temperature ($^0$C) |
|---|---|---|---|---|
| **CC2D1A_26** | 1,80 | 0,905 | 0,4 | 95 |
| **COL6A1_12** | 2,66 | 0,781 | 0,4 | 99 |
| **COL6A1_12v2** | 2,48 | 0,724 | 0,4 | 99 |
| **COL6A1_35** | 0,36 | 0,890 | 0,4 | 95 |
| **COL6A2_6** | 8,56 | 0,755 | 0,4 | 99 |
| **COL6A2_29** | 2,20 | 0,943 | 0,2 | 95 |
| **COL6A3_44** | 0,60 | 0,947 | 0,4 | 95 |
| **SBF1_2** | 5,61 | 0,889 | 0,4 | 95 |
| **SLC5A2_3** | 1,50 | 0,986 | 0,2 | 95 |

In the end, 128 primer pairs were validated for the 33 genes that were assessed. The validated primers had a mean efficiency of 0,91±0,08 and a mean $R^2$ of 0,99±0,01 (Figure 12, Table S4).

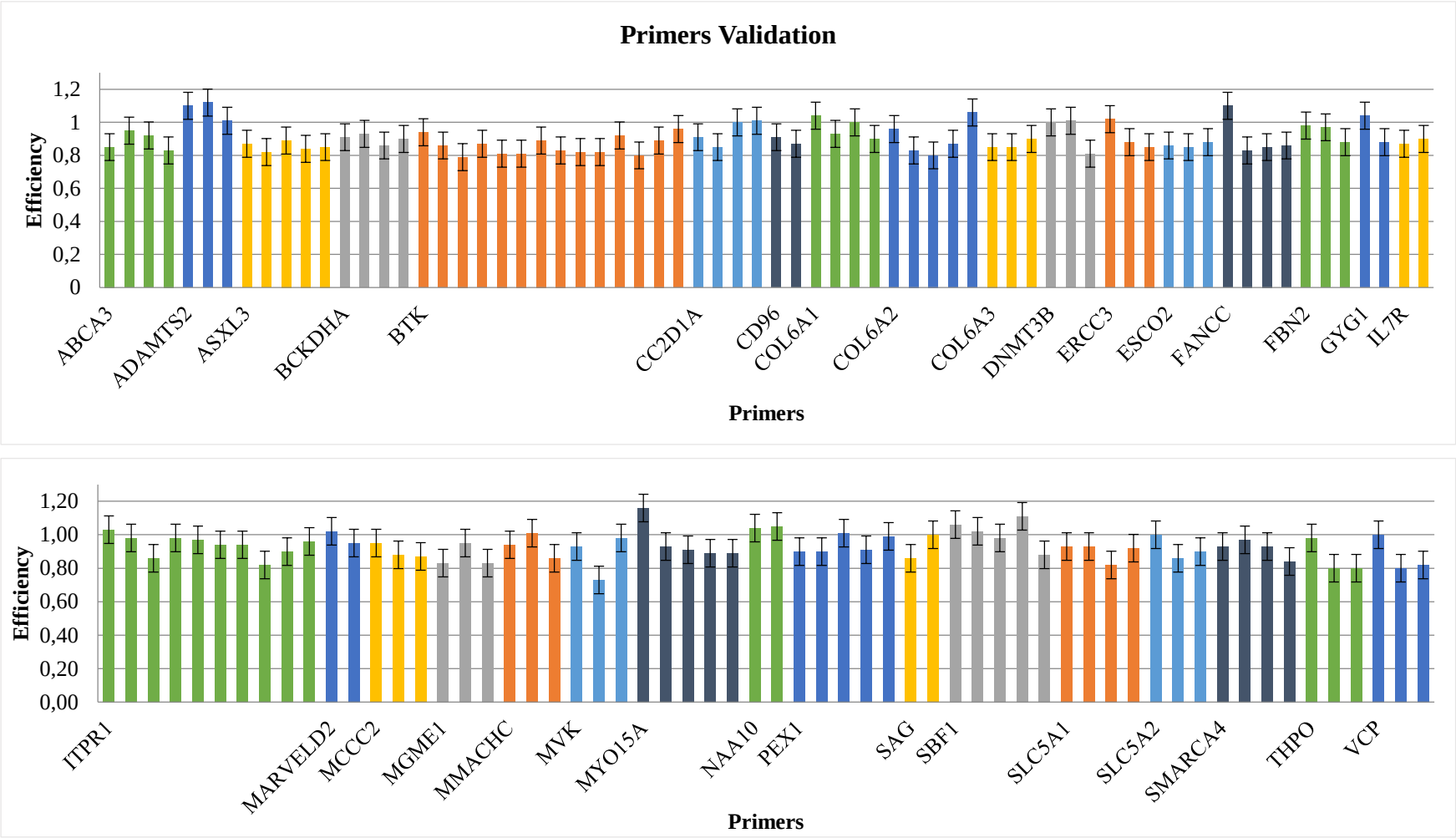


*Figure 12 – Efficiencies of the primers pairs of the tested genes.*

*The mean efficiency of the primers pairs was 0,91±0,08, for the 33 tested genes. Primers were validated with a DNA dilution series of samples with normal copy number for the tested fragments. The efficiencies were acceptable when higher than 0,80, although in some primers pairs an inferior value had to be accepted.*

**Scoring limits**

For scoring a sample as having a deletion, a duplication or a normal copy number for the tested fragment, theoretical cut-off values were set. For the tested fragments, the cut-off values were set based on the values of previous studies, such as 0,7 for deletion upper limit[51–53], 0,8 for normal lower limit[51–53], 1,2 for normal upper limit[53] and 1,25 for duplication lower limit[50]. At the time this methodology was initiated, no hemizygous duplication cut-off values were set theoretically. However, it was expected that the mean haploid copy number for a hemizygous duplication was approximately 2, since one copy would be represented as 1. Similarly, a homozygous deletion was expected to present a mean haploid copy number close to 0 (zero), since replication should not occur.

In the future, for the next tested fragments, the mean values of the haploid copy number of fragments previously scored as deleted, duplicated and normal was used for the calculation of the cut-off values (Table S5; Table S6; Table S7). The mean value of the haploid copy number in deleted fragments plus two times the standard deviation is set as the upper limit for scoring a deletion. For scoring a normal fragment, the mean value of the normal haploid copy number minus two times the standard deviation is set as the lower limit and the mean value of the normal haploid copy number plus two times the standard deviation is set as the higher limit. However, since at this time no heterozygous duplications other than the ones used in the methodology validation were detected, the lower limit for scoring a duplication could not be readjusted. The mean haploid copy numbers of the Pool F (female pool), scored as duplicated for the fragments from the X chromosome genes *BTK* and *NAA10* can be used to calculate the lower limit for a hemizygous duplication.

At the moment, it is set that the values are: 0,67 for heterozygous deletion upper limit; 0,85 for normal lower limit and 1,15 for normal upper limit; 1,25 for heterozygous duplication lower limit; 1,84 for hemizygous duplication lower limit and 2,33 for hemizygous duplication upper limit (Table 2).

*Table 2 - Scoring limits for scoring a deletion, duplication or normal copy number sample, based on the mean haploid copy numbers and standard-deviation of the test and control samples.*

| Sample Type | Mean Haploid Copy Number | Standard-deviation | Lower Limit | Upper Limit |
|---|---|---|---|---|
| **Normal** | 1,00 | 0,07 | 0,85 | 1,15 |
| **Heterozygous Deletion** | 0,53 | 0,07 | 0,38 | 0,67 |
| **Heterozygous Duplication** | - | - | 1,25 | - |
| **Hemizygous Duplication** | 2,08 | 0,12 | 1,84 | 2,33 |

**Relative Quantitation for CNVs Detection**

Data analysis of the qPCR results from RotorGene-Q was performed using the Microsoft Office Excel software and copy number alterations were calculated according to the Delta Delta Ct method ($2^{-\Delta\Delta Ct}$). This method allows the normalization of the test sample results through a normal copy number control (calibrator) and a housekeeping gene.

During the implementation of the methodology, the majority of the analysed samples had a normal copy number for the tested gene fragments. One homozygous deletion was detected for the *BCKDHA* gene and heterozygous deletions were found in the *COL6A1* and *ABCA3* genes. Moreover, two genes from the X chromosome, *BTK* and *NAA10*, were assessed. Significant results are presented in more detail below.

The comparison of the haploid copy number values for the deleted vs. control samples and duplicated vs. control samples, through the F test and two-sample *t-test*, allowed the determination of the statistical strength of the qPCR approach. The fragments analysed were considered statistically significant when p-value<0,05 and confidently assessed as deleted or duplicated.

The results from the quantitative PCR showed that test fragments scored as normal, had a mean haploid copy number of 0,98±0,08, fragments scored as deleted (in heterozygosis) had a haploid copy number of 0,53±0,07 and fragments scored as duplicated (in hemizygosis) had a haploid copy number of 2,08±0,12. Moreover, the negative control samples used had a mean haploid copy number of 1,01±0,06.

### *1) Heterozygous deletion in COL6A1 gene*

The *COL6A1*, *COL6A2* and *COL6A3* genes encode three of the collagen VI chains, which is an important component of the extracellular matrix[69,70]. Mutations in each of the previous genes cause two main types of muscle disorders, being Ullrich congenital muscular dystrophy (UCMD) or Bethlem myopathy (BM)[70]. Although not very frequent, large deletions (deletions of exons 9 to 10, 9 to 13 and 9 to 14 and deletion of the exon 12) have been described in the *COL6A1* gene, namely in association with Bethlem Myopathy[71]. Also, the *COL6A2* gene can present some large deletions (deletion of intron 1, exons 1 to 18, exon 5, exons 5 to 8, intron 5 to exon 7, exons 6 to 10, exons 9 to 28, exon 10, exon 11, exons 26 to 28, entire gene), mainly associated with UCMD[72]. Until present, no CNVs have been described for the *COL6A3* gene[73].

The patient analyzed by us had a clinical suspicious of Bethlem myopathy or Ulrich congenital muscular dystrophy. Previous molecular studies for *MYOT*, Pompe, *LMNA*, *CAV3* and *SEP1* were negative. Also, immunohistochemical analysis was negative for sarcoglycans and dysferline. Moreover, the NGS panel for congenital muscular dystrophies did not detect any mutations in the analysed genes. Therefore, a deletion/duplication search through qPCR was requested for the *COL6A1*, *COL6A2* and *COL6A3* genes.

A heterozygous deletion c.(1002+1_1003-1)_(1056+1_1057-1)del was detected in the *COL6A1* gene. The detected variant is a deletion in heterozygosis which encompasses at least exon 14 of the *COL6A1* gene. The result was first detected using the *ZNF80* gene as the reference gene (Figure 13 - a) and the sample was re-tested for fragment using the *GPR15* as reference gene (Figure 13 - b), which confirmed the deletion. This deletion is not described in the literature neither in databases of polymorphisms. Therefore, it was considered a variant of unknown clinical significance (VOUS) that could not confirm or exclude the clinical diagnosis of Bethlem myopathy. Nevertheless, the patient's father is also affected and has a similar clinical presentation; being so, a recommendation was made in order to study the father to assess if the deletion segregates with the phenotype.

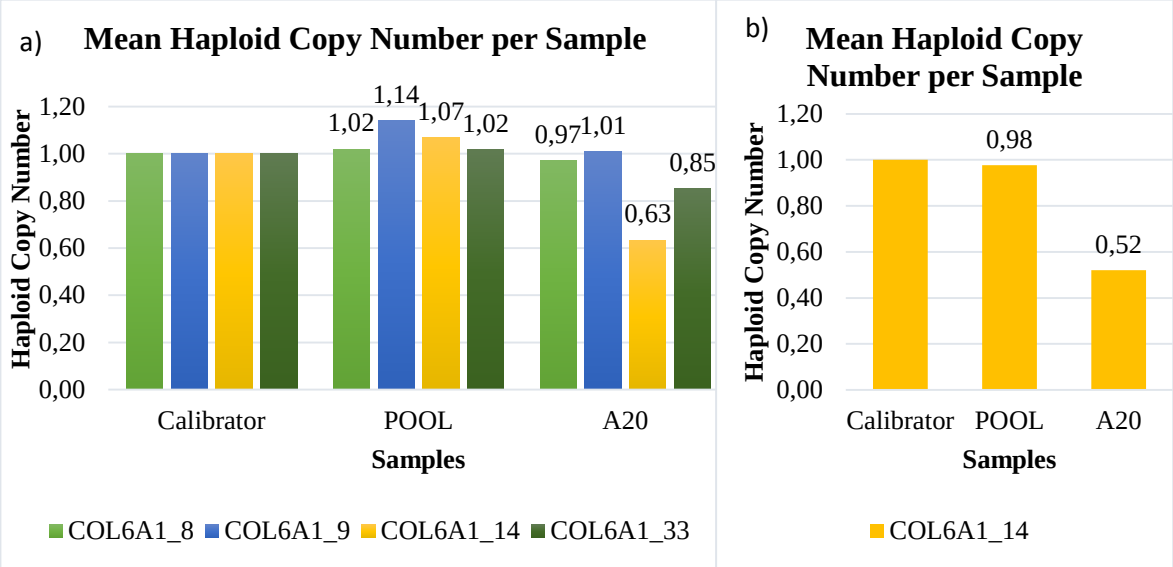


*Figure 13 - Relative quantitation for the COL6A1 gene.*
*The pool of samples used as normal control showed a normal copy number (1,04±0,06). Sample A20 had a normal haploid copy number (0,94±0,07) for the primers pairs of exon 8, 9 and 33, as well as a heterozygous deletion of exon 14, using the ZNF80 as reference gene (left). The heterozygous deletion of exon 14 was confirmed using the GPR15 as reference gene (right).*

No deletions or duplications were detected in the remaining genes studied, *COL6A2* (Figure 14) and *COL6A3* (Figure 15). Therefore, it was considered that these results did not allow us to attribute a genetic etiology to the patient clinical presentation.

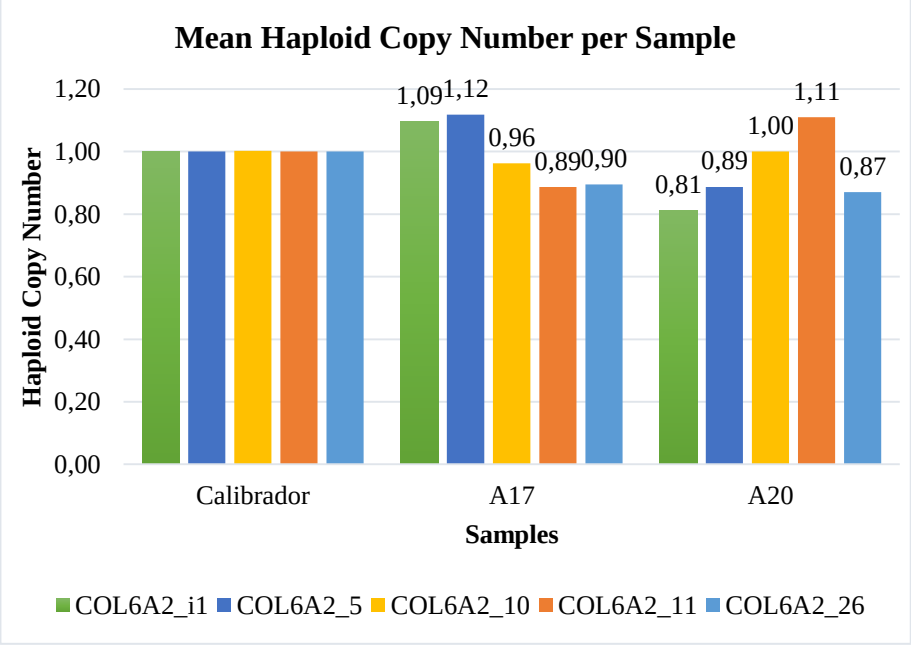


*Figure 14 - Relative quantitation for the COL6A2 gene*
*Sample A17 used as normal control showed a normal copy number (1,00±0,10). Sample A20 had a normal haploid copy number (0,93±0,10) for all primers pairs.*

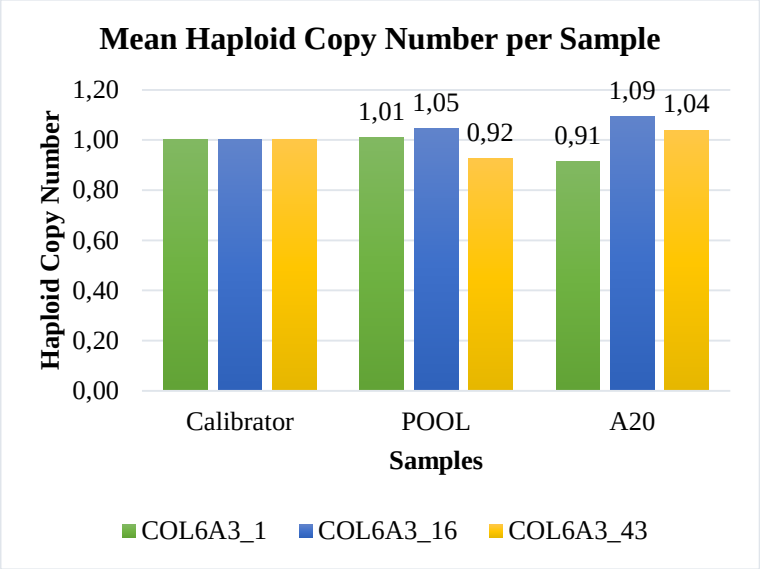


*Figure 15 - Relative quantitation of the COL6A3 gene.*
*The pool of samples used as normal control showed a normal copy number (0,99±0,05). Sample A20 had a normal haploid copy number (1,02±0,08) for all primers pairs.*

### *2) Homozygous deletion in the BCKDHA gene*

The *BCKDHA* gene encodes the alpha subunit of the decarboxylase component of the branched-chain alpha-keto acid dehydrogenase (BCKD) complex. This complex catalyses the oxidative decarboxylation of the branched-chain α-ketoacids derived from the branched-chain amino acids leucine, isoleucine, and valine[74]. Maple Syrup Urine Disease (MSUD) (MIM #248600) is a rare autosomal recessive disorder of metabolism caused by the defective function of the mitochondrial BCKD. When one of the four catalytic subunits, encoded by nuclear genes (*BCKDHA*, *BCKDHB*, *DBT* and *DLD*), is functionally impaired the complex's activity is reduced and the respective branched-chain α-ketoacids accumulate in body fluids to toxic levels[75]. HGMD reports 4 gross deletions (deletion of exons 2 to 4, deletion of exon 4, deletion of exon 6 and deletion of exons 5 to 8) and one gross duplication (duplication of exons 2 to 4) of the *BCKDHA* gene, all related to MSUD[76].

The patient under study had a clinical suspicion of MSUD. First, Sanger sequencing of the *BCKDHB* and *BCKDHA* genes were performed, which did not detect any mutations in the analysed regions. However, no amplification of exons 2 and 3 of the *BCKDHA* gene could be obtained. Subsequently, CNVs detection through qPCR was requested, which confirmed a homozygous deletion c.(108+1_109-1)_(375+1_376-1)del in the *BCKDHA* gene (Figure 16). This mutation is a novel deletion, not described in the literature, which includes at least exons 2 to 3 of the *BCKDHA* gene. Nonetheless, a similar deletion has been described in patients with maple syrup urine disease[77].

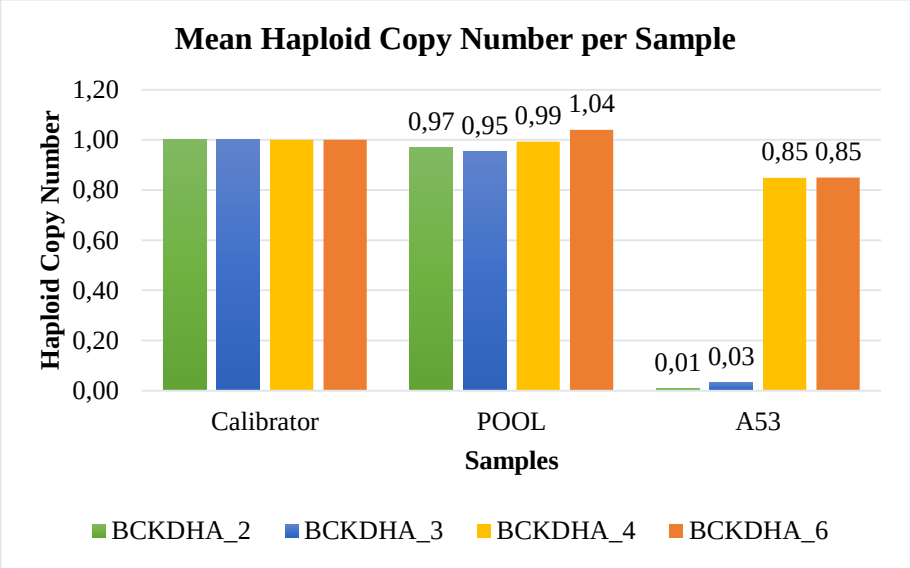


*Figure 16 - Relative quantitation for the BCKDHA gene.*
*The pool of samples used as normal control showed a normal copy number (0,99±0,03). Sample A53 had a normal haploid copy number (0,85±0,00) for primers pairs of exons 4 and 6, as well as a false homozygous deletion (0,02±0,01) of exons 2 and 3.*

### *3) Heterozygous deletion in the ABCA3 gene*

The ATP-binding cassette A3 (*ABCA3*) gene is an 80kB gene mapped to human chromosome 16p13.3[78]. It encodes a 1704-amino-acid protein highly expressed in the lung, which has been localized to the limiting membrane of lamellar bodies, implicating *ABCA3* as possibly important in the maturation of lamellar bodies and surfactant production[79]. The concept that *ABCA3* mutations disrupt surfactant homeostasis is supported by clinical and pathologic findings in patients with *ABCA3* mutations, which present respiratory distress after birth associated with abnormally small lamellar body inclusions[78]. Bi-allelic

mutations of *ABCA3* are the most frequent cause of surfactant metabolism dysfunction, pulmonary type 3 (MIM#610921)[80]. HGMD reports 2 gross deletions (deletion of exons 2 to 5 and deletion of exon 12) of the *ABCA3* gene, related to Lung disease and Fatal surfactant deficiency[81].

The patient had a clinical suspicion of pulmonary surfactant metabolism dysfunction, and a Surfactant Metabolism Dysfunction Panel (NGS panel) was requested. This panel detected the variant c.4442C>T (p.Ala1481Val), in apparent homozygosis, in the *ABCA3* gene. However, no mutations were detected in the remaining genes analysed in this panel. The detected variant is not described in the literature, being located in a highly conserved residue and bioinformatic analysis suggests that it is probably a pathogenic variant. Therefore, it was considered a VOUS. Also, the study of the parents revealed that this variant is of paternal origin.

Subsequently, a CNVs analysis through qPCR was performed. A heterozygous deletion c.(?_4910-1)_(4983+1_?)del was detected in the *ABCA3* gene. The detected variant is a deletion in heterozygosis which encompasses at least exon 32 of the *ABCA3* gene and most likely will include exon 29, where the variant was detected. The result was first detected using the *ZNF80* gene as the reference gene (Figure 17 - a) and the sample was re-tested for fragment using the *GPR15* as reference gene (Figure 17 - b), which confirmed the deletion. This deletion is not described in the literature neither in databases of polymorphisms. Therefore, it was considered a VOUS.

The qPCR result, together with the previous NGS panel result, might support the clinical diagnosis of pulmonary surfactant metabolism dysfunction type 3. Also, the apparent homozygous variant c.4442C>T (p.Ala1481Val) detected in the NGS panel corresponds, in fact, to a compound heterozygosity. Nevertheless, the study of the mother to confirm if she carries the detected deletion was recommended, as well as an additional qPCR to delimit the extent of the detected deletion.

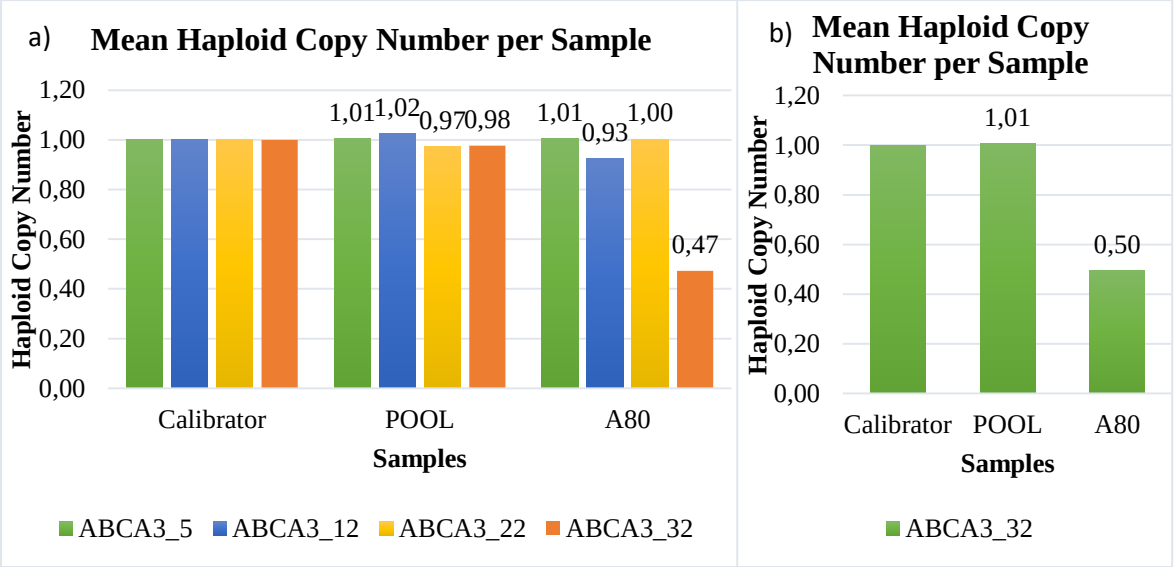


*Figure 17 - Relative quantitation for the ABCA3 gene.*

*The pool of samples used as normal control showed a normal copy number (1,00±0,02). Sample A80 had a normal haploid copy number (0,98±0,04) for the primers pairs of exon 5, 12 and 22, as well as a heterozygous deletion of exon 32, using the ZNF80 as reference gene (left). The heterozygous deletion of exon 32 was confirmed using the GPR15 as reference gene (right).*

### *4) CNVs search in genes from the X chromosome*

The Bruton's tyrosine kinase (*BTK*) gene is localized on Xq21.3–Xq22, being a member of the Tec family of kinases[82]. X-linked agammaglobulinemia (XLA; MIM #300300) is a primary immunodeficiency caused by deleterious variations in *BTK*. XLA presents a drastic decrease in the number of B lymphocytes, a severe reduction of all immunoglobulin isotypes and recurrent bacterial and enteroviral infections. Therefore, the defect blocks B-lymphocyte development between the pro- and pre-B-cell stage[83]. HGMD reports several gross deletions and duplications of the BTK gene, mostly related to XLA, with a few related to *BTK* deficiency or Mohr-Tranebjaerg syndrome. Furthermore, the described deletions and duplications encompass almost every single exon of the gene[84].

The patient studied was a male with a clinical suspicion of XLA, for which it was requested both sequencing and deletion/duplication analysis of *BTK* gene. Sanger sequencing of the *BTK* gene was performed, which detected a hemizygous c.82C>T (p.Arg28Cys) variant, described in the literature and already reported in other patients and families with XLA[85], confirming the clinical diagnosis of X-linked agammaglobulinemia type 1 caused by mutation in the *BTK* gene. On the other hand, CNVs detection through qPCR did not detect any deletions or duplications in the analysed regions (Figure 18).

Since it was a gene from the X chromosome and the test sample was male, the pools of samples used as calibrators, as well as the pool of samples used as normal control, only contained male samples. As expected, the later showed a normal copy number, so did the test sample, for all pairs of primers. Nevertheless, the positive control (pool of female samples) had a copy number of 2,05±0,17, accordingly with the predicted results, because female samples possess two X chromosomes, therefore the double of the *BTK* sequence. However, exon 19 showed normal copy number for the normal control and test samples, but a slightly lower relative quantitation for the female pool. This might be explained by germline mosaicism of at least one of the samples from the pool, or a side effect of the X-chromosome inactivation[86,87].

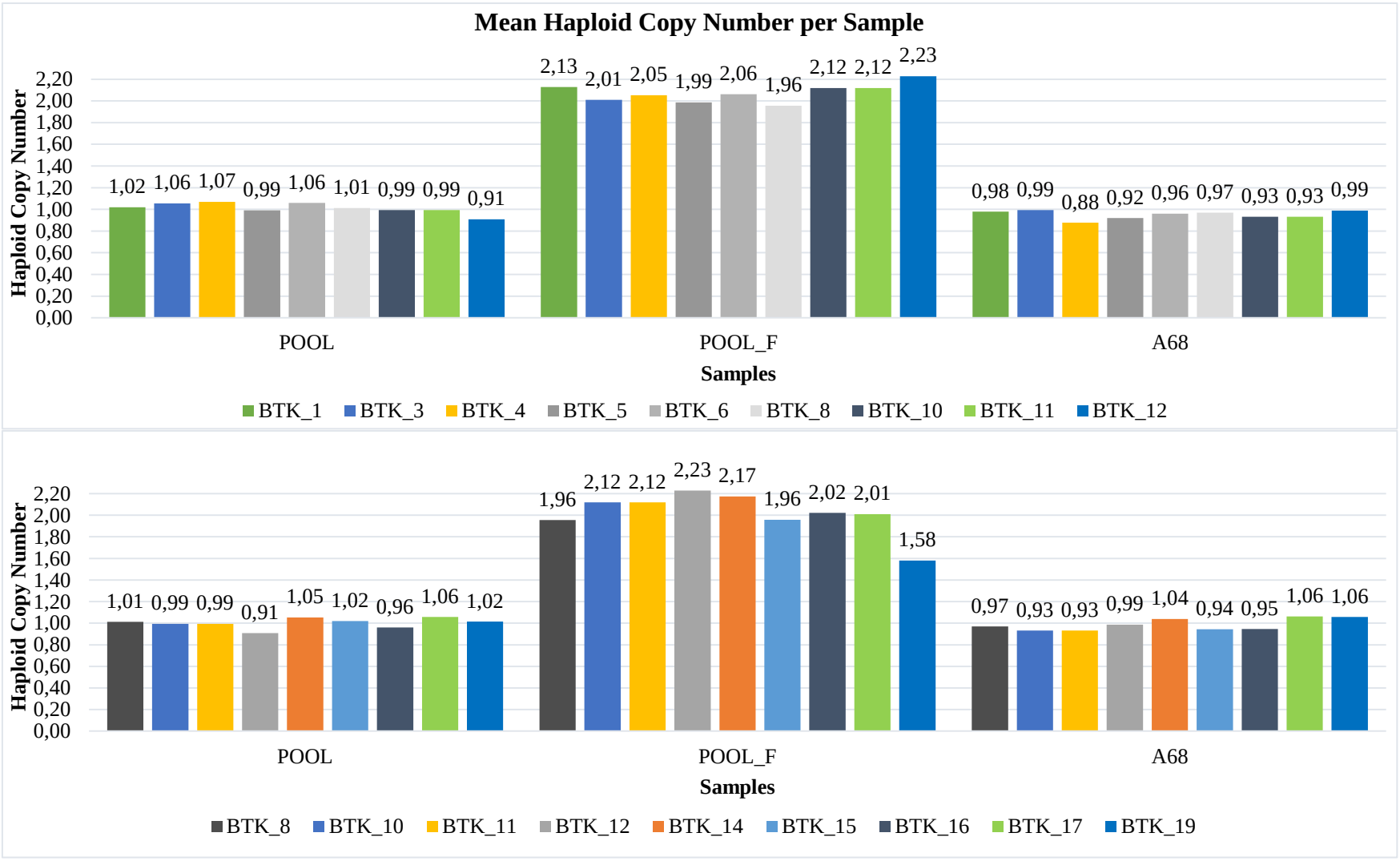


*Figure 18 - Relative quantitation of the BTK gene.*

*The pool of samples used as normal control showed a normal copy number (1,01±0,04), while POOL_F had a copy number of 2,05±0,17, confirming the duplication of the gene sequence. Sample A68 had a normal haploid copy number (0,97±0,05) for all primers pairs.*

The *NAA10* gene is localized at Xq28, and codes for the catalytic subunit of the hNaTA complex[88]. Pathogenic NAA10 mutations are related to Lenz microphthalmia syndrome (LMS, MIM 309800), intellectual disability, arrhythmia and developmental delays[89]. LMS is an X-linked, genetically heterogeneous disorder, characterized by unilateral or bilateral microphthalmia or anophthalmia and defects in the skeletal and genitourinary systems[88]. HGMD does not have any described deletions or duplications of the *NAA10* gene[90].

The patient studied was a male with a clinical diagnosis of Lenz microphthalmia, with multiple studies, including a whole exome sequencing (WES), performed in another laboratory, with normal results. Therefore, deletion/duplication analysis of *NAA10* gene was requested, which did not detect any deletions or duplications in the analysed regions (Figure 19). This result could not confirm the clinical diagnosis of syndromic type 1 (or Lenz) microphthalmia caused by deletions/duplications in the analysed regions of the *NAA10* gene.

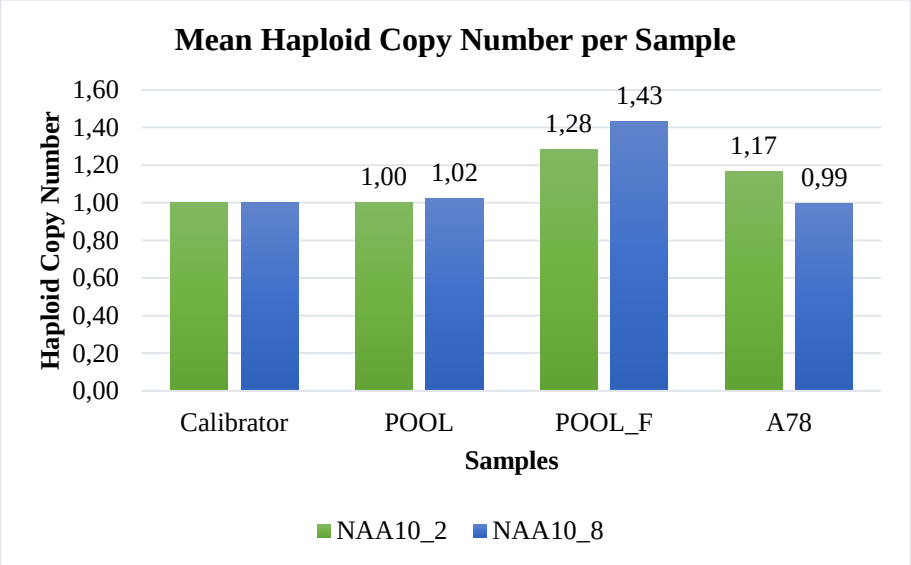


*Figure 19 - Relative quantitation of the NAA10 gene.*
*The pool of samples used as normal control showed a normal copy number (1,01±0,01), while POOL_F had a copy number of 1,36±0,08. Sample A78 had a normal haploid copy number (1,08±0,09) for all primers pairs.*

In the same manner as the *BTK* gene, *NAA10* is a gene from the X chromosome with a male test sample. Therefore, the calibrators and normal control pools only contained male samples. The normal control and the test sample showed a normal copy number.

Again, the positive control had a mean haploid copy number slightly lower than expected, which also might be explained by germline mosaicism or a side effect of the X-chromosome inactivation[86,87].

## Conclusion

This work allowed the development and validation of a methodology for CNVs detection in clinical diagnostics, through qPCR, using the relative quantitation method. The developed and optimized methodology proved to be accurate and specific, allowing the detection of autosomal homozygous and heterozygous deletions, heterozygous duplications, as well as deletions and duplications on fragments from the X chromosome.

However, this methodology cannot differentiate compound heterozygous deletions from homozygous deletions. Also, it is sensitive to SNPs which may impair the primers annealing and subsequent PCR, as well as cause allele dropout. Pseudogenes or other highly homologous sequences might interfere with the technical ability of variant identification in this analysis. Moreover, secondary structures and sequences very rich in CGs might interfere with the efficiency of PCR.

Therefore, the qPCR methodology requires a well-though planning, from selection of regions and primers design up to data analysis. The presented method is easily adaptable for the screening of CNVs in several genetic diseases. So, this approach may be implemented as the method of choice in routine DNA diagnosis for deletion/duplication analysis in genes for which there is no MLPA kit, as well as to confirmf CNVs in a particular gene, previously detected by another method (MLPA, aCGH, etc.).

## References


1. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat. Rev. Genet.* **7,** 85–97 (2006).
2. Pirooznia, M. Whole-genome CNV analysis: advances in computational approaches. *Front. Genet.* **6,** 1–9 (2015).
3. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444,** (2006).
4. Clancy, S. Copy Number Variation. *Nat. Educ.* **1,** 95 (2008).
5. Knouse, K. A., Wu, J. & Amon, A. Assessment of megabase-scale somatic copy number variation using single-cell sequencing. *Genome Res.* **26,** 376–384 (2016).
6. Riggs, E. R., Ledbetter, D. H. & Martin, C. L. Genomic Variation: Lessons Learned from Whole-Genome CNV Analysis. *Curr. Genet. Med. Rep.* **2,** 146–150 (2014).
7. Ceulemans, S., Ven, K. Van Der & Del-favero, J. in *Genomic Structural Variants: Methods and Protocols* (ed. Feuck, L.) **838,** 311–328 (Springer Science & Business Media, 2012).
8. Whale, A. S. *et al.* Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. *Nucleic Acids Res.* **40,** (2012).
9. Zhang, X. *et al.* A modified multiplex ligation-dependent probe amplification method for the detection of 22q11.2 copy number variations in patients with congenital heart disease. *BMC Genomics* 1–11 (2015). doi:10.1186/s12864-015-1590-5
10. Brown, T. A. *Gene Cloning & DNA Analysis*. *Gene Cloning & DNA Analysis: An Introduction* (Wiley-Blackwell, 2010).
11. O'Connor, C. Fluorescence In Situ Hibridization (FISH). *Nat. Educ.* **1,** 171 (2008).
12. Kallioniemi, a, Visakorpi, T., Karhu, R., Pinkel, D. & Kallioniemi, O. Gene Copy Number Analysis by Fluorescence in Situ Hybridization and Comparative Genomic Hybridization. *Methods* **9,** 113–21 (1996).
13. Jansen, F. A. R., Blumenfeld, Y. J., Fisher, A., Cobben, J. M. & Odibo, A. O. Array comparative genomic hybridization and fetal congenital heart defects: a systematic review and meta-analysis. *Ultrasound Obs. Gynecol* **45,** 27–35 (2015).
14. Theisen, A. Microarray-based Comparative Genomic Hybridization (aCGH). *Nat.*

*Educ.* **1,** (2008).

15. McNamee, K., Dawood, F. & Farquharson, R. Evaluation of array comparative genomic hybridization in recurrent miscarriage. *Br. J. Hosp. Med.* **74,** 36–40 (2013).
16. Hillman, S. C., Mcmullan, D. J., Williams, D., Maher, E. R. & Kilby, M. D. Microarray comparative genomic hybridization in prenatal diagnosis: a review. *Ultrasound Obs. Gynecol* **40,** 385–391 (2012).
17. Miller, D. T. *et al.* Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *Am. J. Hum. Genet.* **86,** 749–764 (2010).
18. Szuhai, K. & Vermeer, M. Microarray Techniques to Analyze Copy-Number Alterations in Genomic DNA: Array Comparative Genomic Hybridization and Single-Nucleotide Polymorphism Array. *J. Invest. Dermatol.* **135,** 1–5 (2015).
19. McDonnell, S. K. *et al.* Experimental Designs for Array Comparative Genomic Hybridization Technology. *Cytogenet. Genome Res.* **139,** 250–257 (2013).
20. Pinkel, D. *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20,** 207–211 (1998).
21. Pyatt, R. E. & Astbury, C. Interpretation of Copy Number Alterations Identified Through Clinical Microarray-Comparative Genomic Hybridization. *Clin Lab Med* **31,** 565–580 (2011).
22. Wang, R., Carter, J. & Lench, N. Evaluation of Real-Time Quantitative PCR as a Standard Cytogenetic Diagnostic Tool for Confirmation of Microarray (aCGH) Results and Determination of Inheritance. *Genet. Test. Mol. Biomarkers* **17,** 821–825 (2013).
23. Yau, C. & Holmes, C. C. CNV discovery using SNP genotyping arrays. *Cytogenet. Genome Res.* **123,** 307–312 (2009).
24. Lin, C. F., Naj, A. C. & Wang, L. S. Analyzing copy number variation using SNP array data: Protocols for calling CNV and association tests. *Curr. Protoc. Hum. Genet.* 1–17 (2013). doi:10.1002/0471142905.hg0127s79
25. Tattini, L., D'Aurizio, R. & Magi, A. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Front. Bioeng. Biotechnol.* **3,** 92 (2015).
26. Wang, H. *et al.* Copy number variation detection using next generation sequencing read counts. *BMC Bioinformatics* **15,** 109 (2014).

27. Zhao, M., Wang, Q., Wang, Q., Jia, P. & Zhao, Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives - Springer. *BMC Bioinformatics* **14 Suppl 1,** S1 (2013).
28. Schouten, J. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30,** (2002).
29. Marcinkowska-swojak, M., Klonowska, K., Figlerowicz, M. & Kozlowski, P. An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs. *Gene* (2014). doi:10.1016/j.gene.2014.05.072
30. Konialis, C. *et al.* Routine application of a novel MLPA-based first-line screening test uncovers clinically relevant copy number aberrations in haematological malignancies undetectable by conventional cytogenetics. *Hematology* **19,** 217–224 (2014).
31. González, J. R. *et al.* Probe-specific mixed-model approach to detect copy number differences using multiplex ligation-dependent probe amplification (MLPA). *BMC Bioinformatics* **9,** (2008).
32. Saxena, S., Gowdhaman, K., Kkani, P. & Vennapusa, B. Improved Multiplex Ligation-dependent Probe Amplification (i-MLPA) for rapid copy number variant (CNV) detection. *Clin. Chim. Acta* **450,** 19–24 (2015).
33. Willis, A. S., Veyver, I. Van Den & Eng, C. M. Multiplex ligation-dependent probe amplification (MLPA) and prenatal diagnosis. *Prenat. Diagn.* **32,** 315–320 (2012).
34. Stuppia, L., Antonucci, I., Palka, G. & Gatta, V. Use of the MLPA Assay in the Molecular Diagnosis of Gene Copy Number Alterations in Human Genetic Diseases. *Int. J. Mol. Sci.* **13,** 3245–3276 (2012).
35. Vanguilder, H. D., Vrana, K. E. & Freeman, W. M. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **44,** 619–626 (2008).
36. Mullis, K. B. & Faloona, F. A. Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction. *Methods Enzymol.* **155,** 335–350 (1987).
37. Griffiths, A. J. F., Wessler, S. R., Carroll, S. B. & Doebley, J. *Introduction to Genetic Analysis*. (W. H. Freeman & Company, 2012).
38. Arya, M., Shergill, I. S., Williamson, M. & Gommersall, L. Basic principles of real-time quantitative PCR. *Expert Rev Mol Diagn* **5,** 209–219 (2005).
39. Ferre, F. Quantitative or Semi-Quantitative PCR: Reality Versus Myth. *PCR*

*Methods Appl.* **2,** 1–9 (1992).

40. Bar, T., Kubista, M. & Tichopad, A. Validation of kinetics similarity in qPCR. *Nucleic Acids Res.* **40,** 1395–1406 (2012).
41. D'haene, B., Vandesompele, J. & Hellemans, J. Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* **50,** 262–270 (2010).
42. Hoebeeck, J., Speleman, F. & Vandesompele, J. in *Methods in Molecular Biology: Protocols for Nucleic Acid Analysis by Nonradioactive Probes* (eds. Hilario, E. & Mackay, J.) **353,** 205–226 (Humana Press Inc, 2007).
43. Li, Y. *et al.* Identification of genome-wide copy number variations among diverse pig breeds by array CGH. *BMC Genomics* **13,** 1 (2012).
44. Hasstedt, S. J. *et al.* A Copy Number Variant on Chromosome 20q13.3 Implicated in Thinness and Severe Obesity. *J. Obes.* **2015,** (2015).
45. Wang, J. *et al.* Analyses of Copy Number Variation Reveal Putative Susceptibility Loci in MTX-Induced Mouse Neural Tube Defects. *Dev Neurobiol.* (2014). doi:10.1002/dneu.22170
46. Wang, J. *et al.* Enhancing Genome-Wide Copy Number Variation Identification by High Density Array CGH Using Diverse Resources of Pig Breeds. *PLoS One* **9,** (2014).
47. Ginzinger, D. G. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Exp. Hematol.* **30,** 503–512 (2002).
48. Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W. & Kubista, M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomol. Detect. Quantif.* **3,** 9–16 (2015).
49. Orlando, C., Pinzani, P. & Pazzagli, M. Developments in Quantitative PCR. *Clin. Chem. Lab. Med.* **36,** 255–269 (1998).
50. Yin, C. *et al.* Genome-wide analysis of copy number variations identifies PARK2 as a candidate gene for autism spectrum disorder. *Mol. Autism* **7,** (2016).
51. Ebenazer, A. & Rajaratnam, S. Detection of large deletions in the VHL gene using a Real-Time PCR with SYBR Green. *Fam. Cancer* **12,** 519–524 (2013).
52. Hoebeeck, J. *et al.* Rapid detection of VHL exon deletions using real-time quantitative PCR. *Lab. Investig.* **85,** 24–33 (2005).
53. Muscarella, L. A. *et al.* Small Deletion at the 7q21.2 Locus in a CCM Family

Detected by Real-Time Quantitative PCR. *J. Biomed. Biotechnol.* **2010,** 1–8 (2010).

54. Hattori, K. *et al.* Detection of germline deletions using real-time quantitative polymerase chain reaction in Japanese patients with von Hippel-Lindau disease. *Cancer Sci.* **97,** 400–405 (2006).
55. Vieira, J. P. *et al.* Variant Rett syndrome in a girl with a pericentric X-chromosome inversion leading to epigenetic changes and overexpression of the. *Int. J. Dev. Neurosci.* **46,** 82–87 (2015).
56. Eischeid, A. C. SYTO dyes and EvaGreen outperform SYBR Green in real-time PCR. *BMC Res. Notes* **4,** 263 (2011).
57. Mao, F., Leung, W.-Y. & Xin, X. Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol.* **7,** 76 (2007).
58. Jung, R. & Soondrum, K. Quantitative PCR. *Clin. Chem. Lab. Med.* **38,** 833–836 (2000).
59. Truong, H. T. *et al.* Diagnosing Smith-Magenis Syndrome and Duplication 17p11.2 Syndrome by RAI1 Gene Copy Number Variation Using Quantitative Real-Time PCR. *Genet. Test.* **12,** 67–74 (2008).
60. Kim, S. *et al.* Rapid Detection of Duplication/Deletion of the PMP22 Gene in Patients with Charcot-Marie-Tooth Disease Type 1A and Hereditary Neuropathy with Liability to Pressure Palsy by Real-time Quantitative PCR using SYBR Green I Dye. *J. Korean Acad. Med. Sci.* **18,** 727–732 (2003).
61. Imbard, A. *et al.* NF1 single and multi-exons copy number variations in neurofibromatosis type 1. *J. Hum. Genet.* 1–4 (2015). doi:10.1038/jhg.2015.6
62. Hsiao, M. *et al.* Decoding NF1 Intragenic Copy-Number Variations. *Am. J. Hum. Genet.* **97,** 238–249 (2015).
63. Singh, H. *et al.* Whole exome sequencing of urachal adenocarcinoma reveals recurrent NF1 mutations. *Oncotarget* **7,** 29211–29215 (2016).
64. Hutter, S. *et al.* No correlation between NF1 mutation position and risk of optic pathway glioma in 77 unrelated NF1 patients. *Hum. Genet.* (2016). doi:10.1007/s00439-016-1646-x
65. Ambroziak, W., Koziorowski, D. & Duszyc, K. Genomic instability in the PARK2 locus is associated with Parkinson's disease. *J. Appl. Genet.* **56,** 451–461 (2015).

66. Huttenlocher, J., Stefansson, H., Steinberg, S. & Helgadottir, H. T. Heterozygote carriers for CNVs in PARK2 are at risk of PD. *Hum. Mol. Genet.* 1–27 (2015).

67. Cristofano, A. Di, Strazzullo, M., Longol, L. & Mantia, G. La. Characterization and genomic mapping of the ZNF80 locus: expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucleic Acids Res.* **23,** 2823–2830 (1995).

68. Heiber, M. *et al.* A Novel Human Gene Encoding a G-Protein-Coupled Receptor (GPR15) Is Located on Chromosome 3. *Genomics* **32,** 462–465 (1996).

69. Deconinck, N. *et al.* Bethlem myopathy: long-term follow-up identifies COL6 mutations predicting severe clinical evolution. *J. Neurol. Neurosurg. Psychiatry* **86,** 1337–1346 (2015).

70. Bushby, K. M. D., Collins, J. & Hicks, D. in *Advances in Experimental Medicine and Biology* (ed. Halper, J.) **802,** 77–94 (Springer Netherlands, 2014).

71. Cardiff University. HGMD COL6A1 Gene Result. *HGMD® Professional 2017.1* (2016). at <https://portal.biobase-international.com/hgmd/pro/gene.php?gene=col6a1>

72. Cardiff University. HGMD COL6A2 Gene Result. *HGMD® Professional 2017.1* (2016). at <https://portal.biobase-international.com/hgmd/pro/gene.php?gene=col6a2>

73. Cardiff University. HGMD COL6A3 Gene Result. *HGMD® Professional 2017.1* (2016). at <https://portal.biobase-international.com/hgmd/pro/gene.php?gene=col6a3>

74. Wynn, R. M., Davie, J. R., Chuang, J. L., Cote, C. D. & Chuang, D. T. Impaired Assembly of E1 Decarboxylase of the Branched-chain a-Ketoacid Dehydrogenase Complex in Type IA Maple Syrup Urine Disease. *J. Biol. Chem.* **273,** 13110–13118 (1998).

75. Quental, S. *et al.* Revisiting MSUD in Portuguese Gypsies: Evidence for a Founder Mutation and for a Mutational Hotspot within the BCKDHA Gene. *Ann. Hum. Genet.* **73,** 298–303 (2009).

76. Cardiff University. HGMD BCKDHA Gene Result. *HGMD® Professional 2017.1* (2016). at <https://portal.biobase-international.com/hgmd/pro/gene.php?gene=bckdha>

77. Quental, S., Martins, E., Vilarinho, L., Amorim, A. & Prata, M. J. Maple syrup urine disease due to a new large deletion at BCKDHA caused by non-homologous recombination. *J Inherit Metab Dis* **31,** 457–460 (2008).
78. Garmany, T. H. *et al.* Surfactant Composition and Function in Patients with ABCA3 Mutations. *Pediatr. Res.* **59,** 801–805 (2006).
79. Shulenin, S. *et al.* ABCA3 Gene Mutations in Newborns with Fatal Surfactant Deficiency. *N. Engl. J. Med.* **350,** 1296–1303 (2004).
80. Piersigilli, F. *et al.* New ATP-binding cassette A3 mutation causing surfactant metabolism dysfunction pulmonary type 3. *Pediatr. Int.* **57,** 970–974 (2015).
81. Cardiff University. HGMD ABCA3 Gene Result. *HGMD® Professional 2017.1* (2007). at <https://portal.biobase-international.com/hgmd/pro/gene.php?gene=abca3>
82. Aadam, Z. *et al.* X-Linked Agammagobulinemia in a Large Series of North African Patients: Frequency, Clinical Features and Novel BTK Mutations. *J. Clin. Immunol.* **36,** 187–194 (2016).
83. Väliaho, J., Faisal, I., Ortutay, C., Smith, C. I. E. & Vihinen, M. Characterization of All Possible Single-Nucleotide Change Caused Amino Acid Substitutions in the Kinase Domain of Bruton Tyrosine Kinase. *Hum. Mutat.* **36,** 638–647 (2015).
84. Cardiff University. HGMD BTK Gene Result. *HGMD® Professional 2017.1* (2017). at <https://portal.biobase-international.com/hgmd/pro/gene.php?gene=BTK>
85. Conley, M. E., Mathias, D., Treadaway, J., Minegishi, Y. & Rohrer, J. Mutations in Btk in Patients with Presumed X-Linked Agammaglobulinemia. *Am. J. Hum. Genet.* **62,** 1034–1043 (1998).
86. Sakamoto, M. *et al.* Maternal germinal mosaicism of X- linked agammaglobulinemia. *Am. J. Med. Genet.* **99,** 234–237 (2001).
87. Chear, C. T., Ripen, A. M., Mohamed, S. A. S. & Dhaliwal, J. S. A novel BTK gene mutation creates a de-novo splice site in an X-linked agammaglobulinemia patient. *Gene* **560,** 245–248 (2015).
88. Esmailpour, T. *et al.* A splice donor mutation in NAA10 results in the dysregulation of the retinoic acid signalling pathway and causes Lenz microphthalmia syndrome. *J. Med. Genet.* **51,** 185–196 (2014).

89. Myklebust, L. M., Støve, S. I. & Arnesen, T. Naa10 in development and disease. *Oncotarget* **6,** 34041–34042 (2015).
90. Cardiff University. HGMD NAA10 Gene Result. *HGMD® Professional 2017.1* (2017). at <https://portal.biobase-international.com/hgmd/pro/gene.php?gene=NAA10>

## ANNEX I – Samples

*Table S1 - Samples identification, matrix of extraction, CNV status and studied genes for the validation and development of the methodology.*

| Sample | Matrix | CNV Status | Studied Gene | Observations |
|---|---|---|---|---|
| **A1** | DNA | Normal | *MECP2* | |
| **A2** | Blood (EDTA) | Normal | *MECP2* | |
| **A3** | Blood (EDTA) | Normal | *MECP2* | |
| **A4** | Blood (EDTA) | Deleted | *MECP2* | |
| **A5** | Blood (EDTA) | Duplicated | *MECP2* | |
| **A6** | Blood (EDTA) | Normal | *PARK2; FANCC; COL6A2; VCP; DNMT3B; SBF1* | Del SMN1 |
| **A7** | Blood (EDTA) | Normal | *NF1; FANCC; COL6A1; CC2D1A; COL6A2; BCKDHA; SLC5A1; MMACHC; CD96; ASXL3; MYO15A; SLC5A2* | Del SMN1 |
| **A8** | Blood (EDTA) | Normal | *NF1; PARK2; FANCC; COL6A2; MYO15A; SLC5A2; CC2D1A; SBF1* | Del SMN1 |
| **A9** | Blood (EDTA) | Normal | *NF1; PARK2; FANCC; RBM8A; SLC5A1* | Del SMN1 |
| **A10** | Blood (EDTA) | Normal | *NF1; PARK2; FANCC; VCP; DNMT3B* | Del DMD |
| **A11** | Blood | Deleted | *NF1* | |
| **A12** | Blood (EDTA) | Test | *NF1* | |
| **A13** | Blood (EDTA) | Deleted | *PARK2* | |
| **A14** | Blood (EDTA) | Test | *PARK2* | |
| **A15** | Blood (EDTA) | Normal | *COL6A1; COL6A3; COL6A2; MYO15A; ESCO2; ASXL3* | Del SMN1 |
| **A16** | DNA | Test | *FANCC* | |
| **A17** | Blood (EDTA) | Normal | *FANCC; COL6A2; THPO* | Del DMD |
| **A18** | Blood (EDTA) | Normal | *FANCC; COL6A2; CC2D1A; SLC5A2; SBF1; PEX1* | Del NF1 |
| **A19** | Blood (EDTA) | Normal | *FANCC; COL6A2; THPO; ESCO2* | Del DMD |
| **A20** | Blood (EDTA) | Test | *COL6A1; COL6A2; COL6A3* | |
| **A21** | Blood (EDTA) | Normal | *IL7R; MVK; COL6A2; THPO; MGME1* | Del DMD |
| **A22** | Blood (EDTA) | Normal | *MVK; THPO; MGME1; PEX1; COL6A1; COL6A3* | Del DMD |

*(cont.) Table S1 - Samples identification, matrix of extraction, CNV status and studied genes for the validation and development of the methodology.*

| Sample | Matrix | CNV Status | Studied Gene | Observations |
|---|---|---|---|---|
| **A23** | N/A | Normal | *COL6A2; THPO; COL6A3; IL7R; MGME1; ESCO2; MYO15A; ASXL3* | CT_NGS |
| **A24** | N/A | Normal | *COL6A2; THPO; COL6A3; IL7R; MGME1; ESCO2; MYO15A; ASXL3* | CT_aCGH |
| **A25** | DNA | Test | *THPO* | |
| **A26** | Blood (EDTA) | Normal | *COL6A3; IL7R; MGME1; ESCO2; MYO15A* | POOL_M_MLPA |
| **A27** | Blood (EDTA) | Normal | *COL6A3; IL7R; MGME1; ESCO2; MYO15A* | POOL_M_MLPA |
| **A28** | Blood (EDTA) | Normal | *COL6A3; IL7R; MGME1; ESCO2; MYO15A* | POOL_M_MLPA |
| **A29** | Blood (EDTA) | Normal | *COL6A3; IL7R; MGME1; ESCO2; MYO15A* | POOL_F_MLPA |
| **A30** | Blood (EDTA) | Normal | *COL6A3; IL7R; MGME1; ESCO2; MYO15A* | POOL_F_MLPA |
| **A31** | Blood (EDTA) | Normal | *COL6A3; IL7R; MGME1; ESCO2; MYO15A* | POOL_F_MLPA |
| **A32** | Blood (EDTA) | Test | *IL7R* | |
| **A33** | Blood (EDTA) | Test | *MGME1* | |
| **A34** | DNA | Test | *ESCO2* | |
| **A35** | Blood (EDTA) | Test | *MYO15A* | |
| **A36** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; BTK; MARVELD2; GYG1; SMARCA4; NAA10; ESCO2* | POOL_M |
| **A37** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; BTK; MARVELD2; GYG1; SMARCA4; NAA10; ESCO2* | POOL_M |
| **A38** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; BTK; MARVELD2; GYG1; SMARCA4; NAA10; ESCO2* | POOL_M |

*(cont.) Table S1 - Samples identification, matrix of extraction, CNV status and studied genes for the validation and development of the methodology.*

| Sample | Matrix | CNV Status | Studied Gene | Observations |
|---|---|---|---|---|
| **A39** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; BTK; MARVELD2; GYG1; SMARCA4; NAA10; ESCO2* | POOL_F |
| **A40** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; BTK; MARVELD2; GYG1; SMARCA4; NAA10; ESCO2* | POOL_F |
| **A41** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; BTK; MARVELD2; GYG1; SMARCA4; NAA10; ESCO2* | POOL_F |
| **A42** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; BTK; MARVELD2; GYG1; SMARCA4; ESCO2* | POOL_3 |
| **A43** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; MARVELD2; GYG1; SMARCA4; ESCO2* | POOL_3 |
| **A44** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; MARVELD2; GYG1; SMARCA4; ESCO2* | POOL_3 |

*(cont.) Table S1 - Samples identification, matrix of extraction, CNV status and studied genes for the validation and development of the methodology.*

| Sample | Matrix | CNV Status | Studied Gene | Observations |
|---|---|---|---|---|
| **A45** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; MARVELD2; GYG1; SMARCA4; ESCO2* | POOL_4 |
| **A46** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; MARVELD2; GYG1; SMARCA4; ESCO2* | POOL_4 |
| **A47** | Blood (EDTA) | Normal | *ASXL3; PEX1; MVK; FANCC; COL6A1; SLC5A1; BCKDHA; MYO15A; COL6A2; MMACHC; SAG; ADAMTS2; CC2D1A; ERCC3; FBN2; SLC5A2; VCP; CD96; SBF1; ITPR1; DNMT3B; IL7R; THPO; MARVELD2; GYG1; SMARCA4; ESCO2* | POOL_4 |
| **A48** | Blood (EDTA) | Test | *ASXL3* | |
| **A49** | Blood (EDTA) | Test | *ASXL3* | |
| **A50** | DNA | Test | *PEX1; MVK* | |
| **A51** | Blood (EDTA) | Normal | *FBN2; MMACHC; ASXL3; BCKDHA; ADAMTS2; SAG; COL6A2; SLC5A2; COL6A1; ERCC3; FANCC* | |
| **A52** | DNA | Test | *SLC5A1* | |
| **A53** | Blood (EDTA) | Test | *BCKDHA* | |
| **A54** | Blood (EDTA) | Test | *MMACHC* | |
| **A55** | Blood (EDTA) | Test | *SAG* | |
| **A56** | Blood (EDTA) | Test | *ADAMTS2* | |
| **A57** | Blood (EDTA) | Test | *CC2D1A* | |
| **A58** | Blood (EDTA) | Test | *ERCC3* | |
| **A59** | Blood (EDTA) | Test | *FBN2* | |
| **A60** | Blood (EDTA) | Test | *SLC5A2* | |

*(cont.) Table S1 - Samples identification, matrix of extraction, CNV status and studied genes for the validation and development of the methodology.*

| Sample | Matrix | CNV Status | Studied Gene | Observations |
|---|---|---|---|---|
| **A61** | Blood (EDTA) | Test | *VCP* | |
| **A62** | Blood (EDTA) | Test | *CD96* | |
| **A63** | Blood (EDTA) | Test | *SBF1* | |
| **A64** | Blood (EDTA) | Test | *ITPR1* | |
| **A65** | Blood (EDTA) | Test | *DNMT3B* | |
| **A66** | Blood (EDTA) | Test | *IL7R* | |
| **A67** | DNA | Test | *THPO* | |
| **A68** | DNA | Test | *BTK* | |
| **A69** | Blood (EDTA) | Normal | *BTK; NAA10* | POOL_M1_MLPA |
| **A70** | Blood (EDTA) | Normal | *BTK; NAA10* | POOL_M1_MLPA |
| **A71** | Blood (EDTA) | Normal | *BTK; NAA10* | POOL_M1_MLPA |
| **A72** | Blood (EDTA) | Normal | *BTK; NAA10* | POOL_M2_MLPA |
| **A73** | Blood (EDTA) | Normal | *BTK; NAA10* | POOL_M2_MLPA |
| **A74** | Blood (EDTA) | Normal | *BTK; NAA10* | POOL_M2_MLPA |
| **A75** | Blood (EDTA) | Test | *MARVELD2* | |
| **A76** | DNA | Test | *GYG1* | |
| **A77** | Blood (EDTA) | Test | *SMARCA4* | |
| **A78** | Blood (EDTA) | Test | *NAA10* | |
| **A79** | Blood (EDTA) | Test | *MCCC2* | |
| **A80** | Blood (EDTA) | Test | *ABCA3* | |

## ANNEX II -Primers

*Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **ZNF80_1-F** | GTGGCAGTAGGAAAGCTGTG | 20 | 58,5 | 133 | NM_007136.3 |
| **ZNF80_1-R** | AGTTCTCTGACGTTGACTGATG | 22 | 57,0 | 133 | NM_007136.3 |
| **MECP2_2-F** | CCCAGAATACACCTTGCTTCT | 22 | 58,3 | 119 | NM_004992 |
| **MECP2_2-R** | GGCCAAACCAGGACATATACTA | 22 | 58,0 | 119 | NM_004992 |
| **MECP2_3-F** | GGTCCAAGCCTGCCTCTG | 18 | 62,0 | 186 | NM_004992 |
| **MECP2_3-R** | CTCATGCTTGCCCTCTTTC | 19 | 58,5 | 186 | NM_004992 |
| **MECP2_4A-F** | CATCCGCTCTGCCCTATCT | 19 | 60,3 | 150 | NM_004992 |
| **MECP2_4A-R** | AGGGATGTGTCGCCTACCT | 19 | 59,5 | 150 | NM_004992 |
| **MECP2_4B-F** | GAGAAGAGCGGGAAAGGAC | 19 | 59,0 | 118 | NM_004992 |
| **MECP2_4B-R** | GGTGGTGGTGCTCCTTCT | 18 | 58,6 | 118 | NM_004992 |
| **MECP2_4C-F** | ACATTGTTTCATCCTCCATGC | 21 | 59,8 | 164 | NM_004992 |
| **MECP2_4C-R** | TCAGAGCCCTACCCATAAGG | 20 | 59,1 | 164 | NM_004992 |
| **PARK2_4-F** | TTCTCCAGCAGGTAGATCAATC | 22 | 58,4 | 152 | NM_004562.2 |
| **PARK2_4-R** | TGCTGACACTGCATTTCCTT | 20 | 59,4 | 152 | NM_004562.2 |
| **PARK2_7-Fv2** | CCTCCAGGATTACAGAAATTGG | 22 | 59,8 | 115 | NM_004562.2 |
| **PARK2_7-Rv2** | TGCACTGGAAAACCAGGA | 18 | 58,7 | 115 | NM_004562.2 |
| **PARK2_12-F** | GCTCAGAAAGTGATGTCTAGGC | 22 | 58,2 | 163 | NM_004562.2 |
| **PARK2_12-R** | ACAGTTCCAGCACCACTCG | 19 | 59,9 | 163 | NM_004562.2 |
| **NF1_13-F** | ACAGACCTGGAGACAAGAAGC | 21 | 58,5 | 151 | NM_000267.3 |
| **NF1_13-R** | GGCGTTTCAGCTAAACCC | 18 | 58,3 | 151 | NM_000267.3 |
| **NF1_14-F** | CCTTGCAGAATCCAAGAAAAC | 21 | 58,8 | 148 | NM_000267.3 |
| **NF1_14-R** | GAACATGGAATTCATTTTCCC | 21 | 58,2 | 148 | NM_000267.3 |
| **NF1_58-F** | CGACACATGACTGCAATGAA | 20 | 59,2 | 173 | NM_000267.3 |
| **NF1_58-R** | CTTCTTTGCTTCTGCACTTGG | 21 | 60,2 | 173 | NM_000267.3 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **FANCC_1-F** | CCAGAATGCACTGCTGACAC | 20 | 60,5 | 106 | NM_000136.2 |
| **FANCC_1-R** | AATTTGGCTTTGCCTCGTATT | 21 | 60,0 | 106 | NM_000136.2 |
| **FANCC_2-F** | CACTTTGGAAACCCAGCAA | 19 | 59,7 | 166 | NM_000136.2 |
| **FANCC_2-R** | GTGGCATTCTGTCTTGGTGA | 20 | 59,7 | 166 | NM_000136.2 |
| **FANCC_4-F** | CCTTTTTGTAGATGAAAGCCAA | 22 | 58,4 | 140 | NM_000136.2 |
| **FANCC_4-R** | TCAAAGAAGTGCAGAGCAAGA | 21 | 58,9 | 140 | NM_000136.2 |
| **FANCC_15-F** | GTCCCTGGACAAAGGACAAA | 20 | 59,9 | 120 | NM_000136.2 |
| **FANCC_15-R** | CTGGTCAAGAAAGCCAATGA | 20 | 58,8 | 120 | NM_000136.2 |
| **COL6A2_i1-F** | GCCGCTGCCATCAGTAAT | 18 | 59,8 | 173 | NM_001849.3 |
| **COL6A2_i1-R** | CAAGGACAAGCACTTCATGC | 20 | 59,4 | 173 | NM_001849.3 |
| **COL6A2_5-F** | ACCCATGCAACCTTCTGTC | 19 | 58,5 | 123 | NM_001849.3 |
| **COL6A2_5-R** | AGGTAATCTGGGCATCTTACCT | 22 | 58,2 | 123 | NM_001849.3 |
| **COL6A2_7-F** | GCTCACACTGCTGCGTTGT | 19 | 61,3 | 199 | NM_001849.3 |
| **COL6A2_7-R** | TGCTTCTGTTCCTTGGTAACTG | 22 | 59,4 | 199 | NM_001849.3 |
| **COL6A2_10-F** | TTCCTTCTCTCTTCAGGGGG | 20 | 60,7 | 113 | NM_001849.3 |
| **COL6A2_10-R** | CTGGGATGCCTCTGTGAGAC | 20 | 60,8 | 113 | NM_001849.3 |
| **COL6A2_11-F** | GGCTGTGTCTTGGTCGTTG | 19 | 60,3 | 127 | NM_001849.3 |
| **COL6A2_11-R** | GTTTCCAGGGTCTCCCTTG | 19 | 59,5 | 127 | NM_001849.3 |
| **COL6A2_29-F** | GAATGGAAGGGCACAGGTG | 19 | 61,5 | 101 | NM_001849.3 |
| **COL6A2_29-R** | TGCCACGGACAGCTCTGT | 18 | 61,7 | 101 | NM_001849.3 |
| **COL6A1_8-F** | AACCTGAGTCTGGGGTCCT | 19 | 58,5 | 108 | NM_001848.2 |
| **COL6A1_8-R** | GCAGGAGTCACCACTCACC | 19 | 59,2 | 108 | NM_001848.2 |
| **COL6A1_9-F** | CCCAACCTTGACCTGTTTTG | 20 | 60,4 | 185 | NM_001848.2 |
| **COL6A1_9-R** | ACTTTGGCACTCAGAACCCA | 20 | 60,7 | 185 | NM_001848.2 |
| **COL6A1_12-F** | CACTGTCAGTCCCCATGATTC | 21 | 60,4 | 178 | NM_001848.2 |
| **COL6A1_12-R** | CTCACCTTGTAGCCCTTGGG | 20 | 61,9 | 178 | NM_001848.2 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **COL6A1_35-F** | CAGACAGTCTCCAGGAAGGTG | 21 | 59,9 | 119 | NM_001848.2 |
| **COL6A1_35-R** | GGAAAATTCGCATCACATTTT | 21 | 58,9 | 119 | NM_001848.2 |
| **COL6A3_1-F** | ACGCAGTGAGTGGGAAAAGT | 20 | 59,8 | 124 | NM_004369.3 |
| **COL6A3_1-R** | CCCAAACCTGGAAAAACAGA | 20 | 59,9 | 124 | NM_004369.3 |
| **COL6A3_16-F** | GTGCTGCTCTGCCAAATG | 18 | 59,1 | 113 | NM_004369.3 |
| **COL6A3_16-R** | CTTCTCCAGGAATACCCTGAA | 21 | 58,3 | 113 | NM_004369.3 |
| **COL6A3_44-F** | GTGCCCTTTTGGGTCAGTAA | 20 | 60,0 | 177 | NM_004369.3 |
| **COL6A3_44-R** | ATGGCTGACTCCTTCTTCTTCA | 22 | 60,4 | 177 | NM_004369.3 |
| **IL7R_3-F** | GAACATGCCTCCACTCACC | 19 | 59,0 | 154 | NM_002185.3 |
| **IL7R_3-R** | GCCCCACTGTAGGAATAAAAAG | 22 | 59,1 | 154 | NM_002185.3 |
| **IL7R_8-F** | CCTGACATTGAACCCAGTTG | 20 | 59,0 | 191 | NM_002185.3 |
| **IL7R_8-R** | GCTGTGAGGGAGACTAGGAACT | 22 | 59,0 | 191 | NM_002185.3 |
| **THPO_3-F** | AACTGCAAGGCTAACGCTGT | 20 | 60,1 | 123 | NM_000460.3 |
| **THPO_3-R** | GGGATAATGTTGGGAGTTCTCA | 22 | 60,2 | 123 | NM_000460.3 |
| **THPO_4-F** | CAGAGGTTCACCCTTTGCCTA | 21 | 61,5 | 101 | NM_000460.3 |
| **THPO_4-R** | CAAGGTTAGGGATGGCTTTCTT | 22 | 60,8 | 101 | NM_000460.3 |
| **THPO_6-F** | CCCCTACCAGCCCTCTTCTA | 20 | 60,6 | 179 | NM_000460.3 |
| **THPO_6-R** | GGCTTTGGGTTTCAGGAGA | 19 | 60,2 | 179 | NM_000460.3 |
| **MGME1_2-F** | ACAGCGGATGATTCTGGAAC | 20 | 60,1 | 126 | NM_052865.3 |
| **MGME1_2-R** | CATTCTTTTGACACCGTTGC | 20 | 59,2 | 126 | NM_052865.3 |
| **MGME1_3-F** | AAGTGCTGTTCAACATGAAACC | 22 | 59,1 | 174 | NM_052865.3 |
| **MGME1_3-R** | GCACAAAGGTCTGGACATTAAA | 22 | 59,1 | 174 | NM_052865.3 |
| **MGME1_5-F** | TGCCCTGGAGTAAGAAGGAA | 20 | 59,8 | 186 | NM_052865.3 |
| **MGME1_5-R** | GAAGAAGCCACTTGGTCCAG | 20 | 59,8 | 186 | NM_052865.3 |
| **PEX1_6-F** | TCAGGATAACTCCAGTGGAAGTT | 23 | 59,2 | 140 | NM_000466.2 |
| **PEX1_6-R** | GCTAAAGCAACTAGAAATCTTACAAAA | 27 | 58,1 | 140 | NM_000466.2 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **PEX1_9-F** | GCAACTGTAGGTCCTTCTAGATCC | 24 | 59,7 | 165 | NM_000466.2 |
| **PEX1_9-R** | TTACATTGAAAACTCTGCCAGAT | 23 | 58,0 | 165 | NM_000466.2 |
| **PEX1_12-F** | GTGCTTTCAAAAGGACAACCTG | 22 | 61,0 | 149 | NM_000466.2 |
| **PEX1_12-R** | ACAGATGGCTGCATCCACA | 19 | 61,3 | 149 | NM_000466.2 |
| **PEX1_14-F** | TTTACTTAAGGAGGACGAATCCA | 23 | 59,2 | 123 | NM_000466.2 |
| **PEX1_14-R** | CAGTCATCTTCACTAATGGATGG | 23 | 58,6 | 123 | NM_000466.2 |
| **PEX1_24-F** | GCTTTCAAAATCCAAAGAGGA | 21 | 58,4 | 139 | NM_000466.2 |
| **PEX1_24-R** | GTTACAACATATGGAAAAGCCATC | 24 | 58,9 | 139 | NM_000466.2 |
| **MVK_3-F** | CAATGGGAAAGTGGACCTCA | 20 | 60,9 | 144 | NM_000431.3 |
| **MVK_3-R** | CTGCCCTCTCTGTAGGCTCTT | 21 | 60,2 | 144 | NM_000431.3 |
| **MVK_4-F** | CATCCATGTTCCAATTCCAGT | 21 | 59,7 | 113 | NM_000431.3 |
| **MVK_4-R** | TTGCTCTGTTCCACAGGCA | 19 | 61,6 | 113 | NM_000431.3 |
| **MVK_6-F** | CAGGTGGACCAAGGAGGA | 18 | 59,6 | 188 | NM_000431.3 |
| **MVK_6-R** | AGCACAGACTCTTGGGCA | 18 | 58,0 | 188 | NM_000431.3 |
| **MYO15A_2-F** | AAGCTGGCGGAACAAGGTAT | 20 | 61,0 | 121 | NM_016239.3 |
| **MYO15A_2-R** | AAGCAGCCTCTCTCCAAGGT | 20 | 60,5 | 121 | NM_016239.3 |
| **MYO15A_16-F** | CAGGACCTGAGCTTCAACAG | 20 | 58,6 | 101 | NM_016239.3 |
| **MYO15A_16-R** | ACCTGCTCCTCCTGGAAGA | 19 | 59,9 | 101 | NM_016239.3 |
| **MYO15A_30-F** | CACAATGCCCACAATGCT | 18 | 59,0 | 107 | NM_016239.3 |
| **MYO15A_30-R** | AGTCCCACTCACTTGAGAAGG | 21 | 58,4 | 107 | NM_016239.3 |
| **MYO15A_45-F** | CCTGCATCACAGCCTGTTC | 19 | 60,4 | 152 | NM_016239.3 |
| **MYO15A_45-R** | TGTGCCTCTGACTGGTTGC | 19 | 61,1 | 152 | NM_016239.3 |
| **MYO15A_66-F** | CTCCACCTATCTCCCAACCA | 20 | 59,9 | 124 | NM_016239.3 |
| **MYO15A_66-R** | ACCACACGACACAGTTCCAG | 20 | 59,6 | 124 | NM_016239.3 |
| **ESCO2_2-F** | CCTCCCAGAAAATAGGGTTTT | 21 | 58,5 | 201 | NM_001017420 |
| **ESCO2_2-R** | CACAGGCTGAGATTCACCTG | 20 | 59,4 | 201 | NM_001017420 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **ESCO2_7-F** | TGCAGCATGTACAGCATCAC | 20 | 59,4 | 129 | NM_001017420 |
| **ESCO2_7-R** | GGTGGTAGAGGTGGTGGAAA | 20 | 59,8 | 129 | NM_001017420 |
| **ESCO2_11-F** | TTGCTTCATGTTTGGCTGTT | 20 | 59,3 | 107 | NM_001017420 |
| **ESCO2_11-R** | GGAAATTAGGGGTGTTGCAG | 20 | 59,4 | 107 | NM_001017420 |
| **ASXL3_2-F** | TAAAAGCCCCTTCCCCTCTA | 20 | 60,0 | 152 | NM_030632.2 |
| **ASXL3_2-R** | CTGCTTTGCTGTCATTGGTG | 20 | 60,4 | 152 | NM_030632.2 |
| **ASXL3_11-F** | GAAAAGCCCAGCTTCTCCA | 19 | 60,5 | 179 | NM_030632.2 |
| **ASXL3_11-R** | CATCCTGGCATTCACAGATACT | 22 | 59,1 | 179 | NM_030632.2 |
| **CC2D1A_2-F** | AGCCATGTGGTGAACCAAAG | 20 | 61,0 | 189 | NM_017721.4 |
| **CC2D1A_2-R** | TCACCTTTGCCTTTGAGCTT | 20 | 60,0 | 189 | NM_017721.4 |
| **CC2D1A_14-F** | ACTGCGAGCCAAGCAGAA | 18 | 60,9 | 138 | NM_017721.4 |
| **CC2D1A_14-R** | CACAAGCCCAGAAGGTTCAC | 20 | 60,7 | 138 | NM_017721.4 |
| **CC2D1A_16-F** | GACTCACTGCCTTCTGTTTCC | 21 | 58,9 | 118 | NM_017721.4 |
| **CC2D1A_16-R** | GGTGCAAGTGTCTGGCAC | 18 | 58,7 | 118 | NM_017721.4 |
| **CC2D1A_28-F** | CTACCTGCTGAATGCCCATC | 20 | 60,6 | 101 | NM_017721.4 |
| **CC2D1A_28-R** | CCACCCCATCTCCTAAGCT | 19 | 59,1 | 101 | NM_017721.4 |
| **COL6A1_12v2-F** | CTCCCCACCCCAAATACC | 18 | 60,0 | 148 | NM_001848.2 |
| **COL6A1_12v2-R** | AAGGCAGGAGTCAGATGCA | 19 | 59,5 | 148 | NM_001848.2 |
| **COL6A3_43-F** | TGGATCACAGAAAGAATGTGAA | 22 | 58,2 | 192 | NM_004369.3 |
| **COL6A3_43-R** | GGGTTGTATTTGAACGTCTTCC | 22 | 59,7 | 192 | NM_004369.3 |
| **CD96_2-F** | CTCATGGCTCATGTTCTTTCTT | 22 | 58,5 | 184 | NM_198196.2 |
| **CD96_2-R** | GGATGATAGACAGCAATCAGGT | 22 | 58,1 | 184 | NM_198196.2 |
| **CD96_15-F** | GTGTGTGCATAAGACTCCCTCA | 22 | 60,2 | 124 | NM_198196.2 |
| **CD96_15-R** | GGCTCTTGAATGCAAGTGTACT | 22 | 58,5 | 124 | NM_198196.2 |
| **COL6A2_6-F** | TCAGAGGAGCTGTGGCAG | 18 | 58,7 | 135 | NM_001849.3 |
| **COL6A2_6-R** | CCTGTCTTCCCTTCTGGC | 18 | 58,3 | 135 | NM_001849.3 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **SLC5A2_3-F** | AGCCCTGCTCACTCCCTC | 18 | 60,5 | 123 | NM_003041.3 |
| **SLC5A2_3-R** | CAGCAACAGCCAAGCCAC | 18 | 61,6 | 123 | NM_003041.3 |
| **SLC5A2_14-F** | CCAGTGCCTGCTCTGGTT | 18 | 60,0 | 116 | NM_003041.3 |
| **SLC5A2_14-R** | GGTCCTCGCTGATGTCCT | 18 | 58,7 | 116 | NM_003041.3 |
| **FBN2_2-F** | CGCTTGTCCATCACAGGC | 18 | 61,4 | 118 | NM_001999.3 |
| **FBN2_2-R** | ATCCCTGCCAAGCACTCA | 18 | 60,4 | 118 | NM_001999.3 |
| **FBN2_23-F** | GGAAGTAGGACCACACCTGCT | 21 | 60,6 | 182 | NM_001999.3 |
| **FBN2_23-R** | CAGTCCCATCCAACGTAAGG | 20 | 60,4 | 182 | NM_001999.3 |
| **FBN2_65-F** | CCCAAATCATTTCCTGTTCTTC | 22 | 59,8 | 131 | NM_001999.3 |
| **FBN2_65-R** | GCCTTAGTTCCAGGATGTGCT | 21 | 60,6 | 131 | NM_001999.3 |
| **ADAMTS2_3-F** | CCGTGTGCATGTGGTGTATC | 20 | 60,9 | 108 | NM_014244.4 |
| **ADAMTS2_3-R** | GCTCTCCACACAGCCTGC | 18 | 60,7 | 108 | NM_014244.4 |
| **ADAMTS2_13-F** | GATGCCACCTGTACTGCG | 18 | 58,8 | 111 | NM_014244.4 |
| **ADAMTS2_13-R** | GCACACAGAGGCTGAAGG | 18 | 58,0 | 111 | NM_014244.4 |
| **ADAMTS2_22-F** | CCAAGGCGACAAGTCAATATT | 21 | 59,1 | 113 | NM_014244.4 |
| **ADAMTS2_22-R** | TGGTGAGGTTGTTGTACAGGTT | 22 | 59,4 | 113 | NM_014244.4 |
| **MMACHC_1-F** | GCTCAGCGTGTAACGTGC | 18 | 59,1 | 110 | NM_015506.2 |
| **MMACHC_1-R** | AAACTAACCTGGAAGGGGTAAA | 22 | 58,1 | 110 | NM_015506.2 |
| **MMACHC_2-F** | CTCAAGCCCTTCTTGCAGA | 19 | 59,3 | 147 | NM_015506.2 |
| **MMACHC_2-R** | GAGGAACTGGAGGCAGCT | 18 | 58,0 | 147 | NM_015506.2 |
| **MMACHC_4-F** | GGACCTCCATGACCTTGCT | 19 | 60,1 | 113 | NM_015506.2 |
| **MMACHC_4-R** | CCTGGCAGCAGCACTACC | 18 | 60,6 | 113 | NM_015506.2 |
| **SLC5A1_2-F** | AGGTATGTCCTTTTGGCTGG | 20 | 59,1 | 120 | NM_000343.3 |
| **SLC5A1_2-R** | CTGCCAGGAAGAAGCCTC | 18 | 58,6 | 120 | NM_000343.3 |
| **SLC5A1_15-F** | CCTATGACCTATTTTGTGGGCT | 22 | 59,4 | 148 | NM_000343.3 |
| **SLC5A1_15-R** | GTCACCAGGATGATGCCA | 18 | 58,9 | 148 | NM_000343.3 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **COL6A2_26-F** | GGCGTGGTGCAGTACAGC | 18 | 61,7 | 153 | NM_001849.3 |
| **COL6A2_26-R** | GTAGGCAAACTTGAGGGCTGA | 21 | 61,5 | 153 | NM_001849.3 |
| **BCKDHA_2-F** | CTGGGTGCTGCTTCTGATG | 19 | 60,6 | 178 | NM_000709.3 |
| **BCKDHA_2-R** | GGTAGATGGGGATTCCAGAGA | 21 | 60,3 | 178 | NM_000709.3 |
| **BCKDHA_3-F** | AGGAGAAGGTGCTGAAGCTC | 20 | 58,8 | 138 | NM_000709.3 |
| **BCKDHA_3-R** | ACAGGTAGGGGACCTCAGGT | 20 | 59,8 | 138 | NM_000709.3 |
| **BCKDHA_4-F** | GCAGGATTTGGATGGCTC | 18 | 59,1 | 100 | NM_000709.3 |
| **BCKDHA_4-R** | TTGGTCATGTAGAAGGAGATCC | 22 | 58,1 | 100 | NM_000709.3 |
| **BCKDHA_6-F** | ATGCCAACAGGGTCGTCA | 18 | 61,1 | 101 | NM_000709.3 |
| **BCKDHA_6-R** | GATGGGGCACTCAAGTGTG | 19 | 60,1 | 101 | NM_000709.3 |
| **COL6A1_14-F** | AGGGCACCAAGTCTGACAG | 19 | 58,8 | 211 | NM_001848.2 |
| **COL6A1_14-R** | ACAGAAGTCAAAACGGTCCAC | 21 | 59,1 | 211 | NM_001848.2 |
| **COL6A1_33-F** | CATCCCAGGCTCAGACCA | 18 | 60,8 | 131 | NM_001848.2 |
| **COL6A1_33-R** | TGACATTCTTCAGGAAGGCAT | 21 | 59,7 | 131 | NM_001848.2 |
| **SAG_2-F** | GCACAGGATCTCGTGAGTAGG | 21 | 59,9 | 136 | NM_000541.4 |
| **SAG_2-R** | CATGTTATCTGTCCCTGGCA | 20 | 59,5 | 136 | NM_000541.4 |
| **SAG_16-F** | GAGTTTGCTCGCCATAATCTG | 21 | 59,9 | 128 | NM_000541.4 |
| **SAG_16-R** | CGTTGCACTGGTAACTACAGGT | 22 | 59,2 | 128 | NM_000541.4 |
| **MYO15A_3-F** | CTCTCACACAGATGCACTCCA | 21 | 60,0 | 105 | NM_016239.3 |
| **MYO15A_3-R** | TGCCCACTCACTCCAGCT | 18 | 60,6 | 105 | NM_016239.3 |
| **ASXL3_3-F** | AGAAGCTGTACCTCAGACTGCTT | 23 | 58,9 | 141 | NM_030632.2 |
| **ASXL3_3-R** | CTGACTTTCCAGGGATTTTGA | 21 | 59,2 | 141 | NM_030632.2 |
| **ASXL3_8-F** | ACCAAAGCTGAGGACATTGAC | 21 | 59,2 | 149 | NM_030632.2 |
| **ASXL3_8-R** | TGCCTATCCACTTCTGGGAG | 20 | 60,2 | 149 | NM_030632.2 |
| **ASXL3_12-F** | GAAGAGAGTCCCTGGTGCAG | 20 | 60,0 | 161 | NM_030632.2 |
| **ASXL3_12-R** | GCAATGGAAGCCTGTTGATT | 20 | 60,1 | 161 | NM_030632.2 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **FANCC_14-F** | ACCTCCTGGCAATGTCCAG | 19 | 61,1 | 126 | NM_000136.2 |
| **FANCC_14-R** | AGTTGAGGAGAAGGTGCCTGA | 21 | 61,3 | 126 | NM_000136.2 |
| **ERCC3_2-F** | AGAAATCCAGGAAGCGGC | 18 | 60,3 | 129 | NM_000122.1 |
| **ERCC3_2-R** | CATCCACTTTGGTGCCTGA | 19 | 60,7 | 129 | NM_000122.1 |
| **ERCC3_8-F** | CCAGTTCAAGATGTGGTCCAC | 21 | 60,4 | 113 | NM_000122.1 |
| **ERCC3_8-R** | CCAGCATGGAGTAGGTGCTAA | 21 | 60,3 | 113 | NM_000122.1 |
| **ERCC3_14-F** | AGAGCAACAGCAGCTCTTACAG | 22 | 59,1 | 104 | NM_000122.1 |
| **ERCC3_14-R** | TACTCACCTGGCTGGATCTG | 20 | 58,8 | 104 | NM_000122.1 |
| **SLC5A1_7-F** | TAGACAGGTCACCCTCCACA | 20 | 59,1 | 122 | NM_000343.3 |
| **SLC5A1_7-R** | GATTAAAGACCCCACCAGCA | 20 | 59,9 | 122 | NM_000343.3 |
| **SLC5A1_i14-F** | AGGAGCTGTTCAGACCAGTTCT | 22 | 59,6 | 101 | NM_000343.3 |
| **SLC5A1_i14-R** | GTCTGGCTGGCATGATTCA | 19 | 60,8 | 101 | NM_000343.3 |
| **GPR15_1-F** | TTCCTGACTGGAGTGCTGG | 19 | 60 | 169 | NM_005290.3 |
| **GPR15_1-R** | TCCACAGTCCTAGAGATGCTTC | 22 | 58,6 | 169 | NM_005290.3 |
| **VCP_2-F** | AGGTGATGACCTATCAACAGCC | 22 | 60,4 | 112 | NM_007126.3 |
| **VCP_2-R** | GCTTACCTGGGACAAGGACAC | 21 | 60,9 | 112 | NM_007126.3 |
| **VCP_8-F** | GCTGGTGAGTCTGAGAGCAA | 20 | 59,3 | 134 | NM_007126.3 |
| **VCP_8-R** | TATCCCCTCAGGTAAGCTCCT | 21 | 59,2 | 134 | NM_007126.3 |
| **VCP_16-F** | GGAGGTAGAAGAGGATGATCCA | 22 | 59,5 | 137 | NM_007126.3 |
| **VCP_16-R** | GACTCTGCTGAAGGGTCTGG | 20 | 60 | 137 | NM_007126.3 |
| **DNMT3B_3-F** | ACTCTCCAAGAGGGAGGTGTC | 21 | 59,7 | 103 | NM_006892.3 |
| **DNMT3B_3-R** | CGTGATGAAAGCCAAAGACA | 20 | 59,8 | 103 | NM_006892.3 |
| **DNMT3B_11-F** | ACGACTCAGCCACCTCTGAC | 20 | 60,5 | 103 | NM_006892.3 |
| **DNMT3B_11-R** | TCACCTCGGCTCTGATCTTC | 20 | 60,5 | 103 | NM_006892.3 |
| **DNMT3B_22-F** | ACAGACAATAACCACCAAGTCG | 22 | 59 | 118 | NM_006892.3 |
| **DNMT3B_22-R** | CCTTGCTCACCTTTCGAGC | 19 | 61,1 | 118 | NM_006892.3 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **SBF1_2-F** | GATCCTCAAGCCTCCTGGT | 19 | 59,2 | 123 | NM_002972.3 |
| **SBF1_2-R** | CCTCCCAGTCCTTCTCTGG | 19 | 59,8 | 123 | NM_002972.3 |
| **SBF1_12-F** | CGTGGGCTGGTAGAGGAC | 18 | 59,2 | 115 | NM_002972.3 |
| **SBF1_12-R** | GCACCTCATCGAACAGGTC | 19 | 59,2 | 115 | NM_002972.3 |
| **SBF1_26-F** | ATGACCATGAGCAGCCTG | 18 | 58,2 | 105 | NM_002972.3 |
| **SBF1_26-R** | GAAGGGCTCAGACTTGGC | 18 | 58,5 | 105 | NM_002972.3 |
| **SBF1_35-F** | GTGCAGGGCTGCTGTATG | 18 | 58,9 | 145 | NM_002972.3 |
| **SBF1_35-R** | CAGCTTACCTCTGCGTCCT | 19 | 58,2 | 145 | NM_002972.3 |
| **SBF1_40-F** | TGTGGACACAGAGTGCAAGG | 20 | 60,9 | 123 | NM_002972.3 |
| **SBF1_40-R** | CCAAGGCTCACGTCAAAGA | 19 | 60 | 123 | NM_002972.3 |
| **ITPR1_4-F** | GGAAAAGCCCTGGGATATCT | 20 | 59,4 | 127 | NM_002222.5 |
| **ITPR1_4-R** | CGGTTTCTGGCTGTACAACAC | 21 | 60,6 | 127 | NM_002222.5 |
| **ITPR1_6-F** | GAAAACAGGAAATTGCTGGG | 20 | 59,5 | 137 | NM_002222.5 |
| **ITPR1_6-R** | TCCAAACCCTGTCTTAGGTGTT | 22 | 59,9 | 137 | NM_002222.5 |
| **ITPR1_10-F** | AGGAAGAAGCAGCACGTCTT | 20 | 59,2 | 100 | NM_002222.5 |
| **ITPR1_10-R** | ACCCTACCCTTACCTCCACCT | 21 | 60,1 | 100 | NM_002222.5 |
| **ITPR1_20-F** | GCTGGATGTCGATCTCATTCT | 21 | 59,3 | 123 | NM_002222.5 |
| **ITPR1_20-R** | GTGACTTGTTCCTGGGGATC | 20 | 59,4 | 123 | NM_002222.5 |
| **ITPR1_38-F** | CCTGAGGGAAATGATGACCA | 20 | 60,9 | 118 | NM_002222.5 |
| **ITPR1_38-R** | AGGTGGGTTTTGAAAGGTCC | 20 | 60,2 | 118 | NM_002222.5 |
| **ITPR1_40-F** | AGCTCTATGAGCAGGGGTGA | 20 | 60 | 123 | NM_002222.5 |
| **ITPR1_40-R** | TTCATGGAACACTCGGTCAC | 20 | 59,5 | 123 | NM_002222.5 |
| **ITPR1_44-F** | ACTGCATAGCCACCCATGA | 19 | 60,1 | 113 | NM_002222.5 |
| **ITPR1_44-R** | TGCCTTCAGTTCTAACACAAGG | 22 | 59,4 | 113 | NM_002222.5 |
| **ITPR1_48-F** | GAACAGATAGTCTTTCCCGTGC | 22 | 60,1 | 136 | NM_002222.5 |
| **ITPR1_48-R** | AGAGGTCTTCAGACCGCAGA | 20 | 60,1 | 136 | NM_002222.5 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **ITPR1_52-F** | TGGACGGTCCATCATCCT | 18 | 59,8 | 119 | NM_002222.5 |
| **ITPR1_52-R** | CATTGGGCAGCCTATCTACTTC | 22 | 60,1 | 119 | NM_002222.5 |
| **ITPR1_57-F** | CCAGGATGAGAGCCATGTC | 19 | 59,1 | 141 | NM_002222.5 |
| **ITPR1_57-R** | CCTGATCCTTTAATTCCGACAG | 22 | 60 | 141 | NM_002222.5 |
| **SMARCA4_3-F** | TTCCACATGCTGACCCTG | 18 | 59,1 | 112 | NM_001128849.1 |
| **SMARCA4_3-R** | ATCCCCATTCCTTTCATCTG | 20 | 58,8 | 112 | NM_001128849.1 |
| **SMARCA4_16-F** | ATCAAAGGTTTGGAGTGGCT | 20 | 58,6 | 159 | NM_001128849.1 |
| **SMARCA4_16-R** | AGGCACGATGATGAGGAAG | 19 | 58,8 | 159 | NM_001128849.1 |
| **SMARCA4_30-F** | CATGACACAGCCAGCAGTG | 19 | 60 | 131 | NM_001128849.1 |
| **SMARCA4_30-R** | CACGTCCCCAGTGTTGTG | 18 | 59,5 | 131 | NM_001128849.1 |
| **SMARCA4_34-F** | GAAGACTCCATCGTCTTGCAG | 21 | 60 | 112 | NM_001128849.1 |
| **SMARCA4_34-R** | CTTCCTCCTCGCCCTCTT | 18 | 59,5 | 112 | NM_001128849.1 |
| **SLC5A2_1-F** | AGAATGGAGGAGCACACAGAG | 21 | 59,5 | 105 | NM_003041.3 |
| **SLC5A2_1-R** | CAGGAAATATGCAGCAATGAC | 21 | 58,3 | 105 | NM_003041.3 |
| **SLC5A2_4-F** | GGTCATCACGATGCCACA | 18 | 60,1 | 114 | NM_003041.3 |
| **SLC5A2_4-R** | GCACTCACTGAGATCTTGGTGA | 22 | 60,5 | 114 | NM_003041.3 |
| **MARVELD2_2-F** | ACCAGTGAATGCAGTGTTCTG | 21 | 58,8 | 137 | NM_001038603.2 |
| **MARVELD2_2-R** | GTCTCCGAGCTGCCTCAT | 18 | 59 | 137 | NM_001038603.2 |
| **MARVELD2_5-F** | CGCTATAAAGCTGTGTTCCAAG | 22 | 59,1 | 105 | NM_001038603.2 |
| **MARVELD2_5-R** | GCTCATCACTGCATCCAGC | 19 | 60,5 | 105 | NM_001038603.2 |
| **NAA10_2-F** | CCAGAGGACCTAATGAACATGC | 22 | 60,9 | 115 | NM_003491.3 |
| **NAA10_2-R** | TTCAGAAGCTGCCCACCT | 18 | 59,5 | 115 | NM_003491.3 |
| **NAA10_8-F** | GAGCTGAAAGAGAAGGGCAG | 20 | 59,3 | 132 | NM_003491.3 |
| **NAA10_8-R** | ACCACTATCCTCGGCAGC | 18 | 58,8 | 132 | NM_003491.3 |
| **CC2D1A_26-F** | CCAGCAGTACCAGGACATCA | 20 | 59,7 | 109 | NM_017721.4 |
| **CC2D1A_26-R** | CCACCAGATGCCCATCTC | 18 | 60 | 109 | NM_017721.4 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **SBF1_3-F** | ACTGCTGAGCCTTCCCTG | 18 | 59,1 | 141 | NM_002972.3 |
| **SBF1_3-R** | CTCCCAGAAGGTCAAGCAG | 19 | 58,5 | 141 | NM_002972.3 |
| **SBF1_24-F** | ACGAGGAGGTGGGGTCTG | 18 | 61,1 | 118 | NM_002972.3 |
| **SBF1_24-R** | CCAGGTGTGTGGGCAGAG | 18 | 61,4 | 118 | NM_002972.3 |
| **GYG1_2-F** | CATCTCTGAAACAGCACAGGA | 21 | 59 | 132 | NM_004130.3 |
| **GYG1_2-R** | GCAACAGGGAGAAGGATGTC | 20 | 59,7 | 132 | NM_004130.3 |
| **GYG1_8-F** | GGAAGGAACGATGGGAACA | 19 | 60,9 | 136 | NM_004130.3 |
| **GYG1_8-R** | CAAGGCTTGTGAAGTGGATGT | 21 | 60,2 | 136 | NM_004130.3 |
| **BTK_1-F** | AGCAGGGAACCAGATAGCAT | 20 | 58,8 | 121 | NM_000061.2 |
| **BTK_1-R** | AGAGGAAGGACAGTCTGAGCA | 21 | 59,2 | 121 | NM_000061.2 |
| **BTK_3-F** | CATCTGGTTGCTTAATCCCTCT | 22 | 59,6 | 158 | NM_000061.2 |
| **BTK_3-R** | GGAATCTGTCTTTCTGGAGGAG | 22 | 59,3 | 158 | NM_000061.2 |
| **BTK_4-F** | GGTGAAGAGTCCAGTGAAATGG | 22 | 60,9 | 160 | NM_000061.2 |
| **BTK_4-R** | CTGGGTCCTCGTAGCCTTC | 19 | 59,8 | 160 | NM_000061.2 |
| **BTK_5-F** | CTGAACACCATTGCTGACTGA | 21 | 59,9 | 194 | NM_000061.2 |
| **BTK_5-R** | GTTCTTCAGTTGGGGAGAAGAC | 22 | 59,2 | 194 | NM_000061.2 |
| **BTK_6-F** | TGCTTGAAGTTCACCACTAAGC | 22 | 59,6 | 149 | NM_000061.2 |
| **BTK_6-R** | GATCCAGAAGCAAGGGTGA | 19 | 58,7 | 149 | NM_000061.2 |
| **BTK_8-F** | ACCAGTCTCCACAAGTGAGCT | 21 | 59 | 124 | NM_000061.2 |
| **BTK_8-R** | TGGTAAGTTGCTTTCCTCCAAG | 22 | 60,6 | 124 | NM_000061.2 |
| **BTK_10-F** | CATGGACAAGCCCTGGAG | 18 | 60,2 | 126 | NM_000061.2 |
| **BTK_10-R** | AGCAGTTGCTCAGCCTGAC | 19 | 59,3 | 126 | NM_000061.2 |
| **BTK_11-F** | CCACTTCCTCCTACAGACAGC | 21 | 59 | 127 | NM_000061.2 |
| **BTK_11-R** | CACTCACCCTGTGGATTTAGC | 21 | 59,6 | 127 | NM_000061.2 |
| **BTK_12-F** | GTTGACCTTTGTGCCCAAG | 19 | 59,1 | 103 | NM_000061.2 |
| **BTK_12-R** | TACTGGCTCTGAGGTGTGGA | 20 | 59,4 | 103 | NM_000061.2 |

*(cont.) Table S2 – qPCR primers sequence, base pairs, melting temperatures, amplicon base pairs and RefSeq used during design.*

| Primer name | Primer sequence (5'-3') | Primer bp | Primer Tm | Amplicon bp | RefSeq |
|---|---|---|---|---|---|
| **BTK_14-F** | GATCATGGGAAATTGATCCAA | 21 | 59,6 | 145 | NM_000061.2 |
| **BTK_14-R** | TCATCTTCAGACATGGAGCCT | 21 | 59,8 | 145 | NM_000061.2 |
| **BTK_15-F** | TGAGAAGCTGGTGCAGTTGT | 20 | 59,6 | 149 | NM_000061.2 |
| **BTK_15-R** | CCTTGCACATCTCTAGCAGC | 20 | 58,8 | 149 | NM_000061.2 |
| **BTK_16-F** | GTTTGCACTACAGGCAGCTC | 20 | 58,7 | 143 | NM_000061.2 |
| **BTK_16-R** | CCTACCCATGTTTCATACTGTGC | 23 | 60,7 | 143 | NM_000061.2 |
| **BTK_17-F** | ATGTCCTGGATGATGAATACACA | 23 | 59,2 | 101 | NM_000061.2 |
| **BTK_17-R** | AGATTTGCTGCTGAACTTGCT | 21 | 59,3 | 101 | NM_000061.2 |
| **BTK_19-F** | ATGTCATGGATGAAGAATCCTG | 22 | 58,9 | 116 | NM_000061.2 |
| **BTK_19-R** | GCTCCAGGGCTCCTAAGC | 18 | 60,1 | 116 | NM_000061.2 |
| **MCCC2_1-F** | CCTCTACCAGGTAGGCTGAGC | 21 | 60,4 | 102 | NM_022132.4 |
| **MCCC2_1-R** | GACAAGAGAGCAGCAGTTGTTG | 22 | 60,2 | 102 | NM_022132.4 |
| **MCCC2_10-F** | TGCAGGTCACCATTGAACC | 19 | 60,5 | 137 | NM_022132.4 |
| **MCCC2_10-R** | GCTGGTCATAAAGCTCACAGTTC | 23 | 60,3 | 137 | NM_022132.4 |
| **MCCC2_17-F** | TTCCTTGGCATTTCTGCAC | 19 | 59,8 | 141 | NM_022132.4 |
| **MCCC2_17-R** | ACTGAGACCCAAGACCAGTCTG | 22 | 60,7 | 141 | NM_022132.4 |
| **ABCA3_5-F** | GCTTGAAGATTCAGTCGGAAA | 21 | 59,4 | 133 | NM_001089.2 |
| **ABCA3_5-R** | CTGTGAGAAGGGATGTAGGCA | 21 | 60,3 | 133 | NM_001089.2 |
| **ABCA3_12-F** | CATCCAGTGGCGAGACCT | 18 | 59,8 | 121 | NM_001089.2 |
| **ABCA3_12-R** | CTCCATGTACCAGGTCACCAG | 21 | 60,4 | 121 | NM_001089.2 |
| **ABCA3_22-F** | TCTTCAGGGCTTCTGTGGAG | 20 | 60,5 | 131 | NM_001089.2 |
| **ABCA3_22-R** | AGTGGCTGGAGAGTGGTACG | 20 | 60,3 | 131 | NM_001089.2 |
| **ABCA3_32-F** | GAGGACTCCAGTGCTGAGC | 19 | 58,6 | 116 | NM_001089.2 |
| **ABCA3_32-R** | CATGCCTTGGTGCTCATCT | 19 | 59,8 | 116 | NM_001089.2 |

## ANNEX III - Method Validation – Results Comparison

*Table S3- Results comparison between qPCR and previous results for methodology validation*

| Sample | Gene | Fragment | MLPA Result | aCGH Result | Other Tests | qPCR Result | Prediction Value | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | MLPA | aCGH | Other |
| **A1** | *MECP2* | MECP2_2 | Normal | - | - | Normal | 0 | - | - |
| **A1** | *MECP2* | MECP2_3 | Normal | - | - | Normal | 0 | - | - |
| **A1** | *MECP2* | MECP2_4A | Normal | - | - | Normal | 0 | - | - |
| **A1** | *MECP2* | MECP2_4C | Normal | - | - | Normal | 0 | - | - |
| **A4** | *MECP2* | MECP2_2 | Normal | Normal | - | Normal | 0 | 0 | - |
| **A4** | *MECP2* | MECP2_3 | c.27-?_704+?del; Heterozygous deletion of exon 3 and part of exon 4 | arr Xq28(153,297,503-153,314,668)x1; Heterozygous deletion of exon 3 | - | Heterozygous Deletion | 0 | 0 | - |
| **A4** | *MECP2* | MECP2_4A | c.27-?_704+?del; Heterozygous deletion of exon 3 and part of exon 4 | Normal | - | Heterozygous Deletion | 0 | 1 | - |
| **A4** | *MECP2* | MECP2_4C | Normal | Normal | - | Normal | 0 | 0 | - |
| **A5** | *MECP2* | MECP2_2 | c.1-?_c.1461+?dup; Heterozygous duplication | - | - | Heterozygous Duplication | 0 | - | - |

*(cont.) Table S3- Results comparison between qPCR and previous results for methodology validation.*

| Sample | Gene | Fragment | MLPA Result | aCGH Result | Other Tests | qPCR Result | Prediction Value | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | MLPA | aCGH | Other |
| **A5** | *MECP2* | MECP2_3 | c.1-?_c.1461+?dup; Heterozygous duplication | - | - | Heterozygous Duplication | 0 | - | - |
| **A5** | *MECP2* | MECP2_4A | c.1-?_c.1461+?dup; Heterozygous duplication | - | - | Heterozygous Duplication | 0 | - | - |
| **A5** | *MECP2* | MECP2_4C | c.1-?_c.1461+?dup; Heterozygous duplication | - | - | Heterozygous Duplication | 0 | - | - |
| **A10** | *NF1* | NF1_13 | Hemizygous deletion in *DMD* gene | - | - | Normal | 0 | - | - |
| **A10** | *NF1* | NF1_14 | Hemizygous deletion in *DMD* gene | - | - | Normal | 0 | - | - |
| **A10** | *NF1* | NF1_58 | Hemizygous deletion in *DMD* gene | - | - | Normal | 0 | - | - |
| **A11** | *NF1* | NF1_13 | c.1-?_8457+?del; Heterozygous deletion | - | - | Heterozygous Deletion | 0 | - | - |

*(cont.) Table S3- Results comparison between qPCR and previous results for methodology validation.*

| Sample | Gene | Fragment | MLPA Result | aCGH Result | Other Tests | qPCR Result | Prediction Value | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | MLPA | aCGH | Other |
| **A11** | *NF1* | NF1_14 | c.1-?_8457+?del; Heterozygous deletion | - | - | Heterozygous Deletion | 0 | - | - |
| **A11** | *NF1* | NF1_58 | c.1-?_8457+?del; Heterozygous deletion | - | - | Heterozygous Deletion | 0 | - | - |
| **A12** | *NF1* | NF1_13 | Normal | - | Duplication of exon 13 detected by NGS and qPCR at another lab | Normal | 0 | - | -1 |
| **A12** | *NF1* | NF1_14 | Normal | - | Normal | Normal | 0 | - | 0 |
| **A12** | *NF1* | NF1_58 | Normal | - | Normal | Normal | 0 | - | 0 |
| **A10** | *PARK2* | PARK2_4 | Hemizygous deletion in *DMD* gene | - | - | Normal | 0 | - | - |
| **A10** | *PARK2* | PARK2_7 | Hemizygous deletion in *DMD* gene | - | - | Normal | 0 | - | - |
| **A10** | *PARK2* | PARK2_12 | Hemizygous deletion in *DMD* gene | - | - | Normal | 0 | - | - |

*(cont.) Table S3- Results comparison between qPCR and previous results for methodology validation.*

| Sample | Gene | Fragment | MLPA Result | aCGH Result | Other Tests | qPCR Result | Prediction Value | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | MLPA | aCGH | Other |
| **A13** | *PARK2* | PARK2_4 | c.172-?_534+?del; c.172-?_734+?del; Compound heterozygous deletion of exons 3 to 4 and 3 to 6 | - | Sequencing failure of exons 3 and 4 (Sanger) | Homozygous/ Compound Heterozygous Deletion | 0 | - | 0 |
| **A13** | *PARK2* | PARK2_7 | Normal | - | Normal | Normal | 0 | - | 0 |
| **A13** | *PARK2* | PARK2_12 | Normal | - | Normal | Normal | 0 | - | 0 |
| **A14** | *PARK2* | PARK2_4 | c.413-?_871+?del; Heterozygous deletion of exons 4 to 7 | - | - | Heterozygous Deletion | 0 | - | - |
| **A14** | *PARK2* | PARK2_7 | c.413-?_871+?del; Heterozygous deletion of exons 4 to 7 | - | - | Heterozygous Deletion | 0 | - | - |
| **A14** | *PARK2* | PARK2_12 | Normal | - | - | Normal | 0 | - | - |

# ANNEX IV – Primers Validation

*Table S4 - Primers efficiency, $R^2$ and PCR characteristics for the tested genes.*

| Primer | Efficiency | $R^2$ | Primer Volume (µL) | Melting Temperature ($^0$C) |
|---|---|---|---|---|
| **ABCA3_5** | 0,85 | 0,996 | 0,4 | 99 |
| **ABCA3_12** | 0,95 | 0,996 | 0,4 | 99 |
| **ABCA3_22** | 0,92 | 0,993 | 0,4 | 99 |
| **ABCA3_32** | 0,83 | 0,992 | 0,4 | 99 |
| **ADAMTS2_3** | 1,10 | 0,988 | 0,4 | 95 |
| **ADAMTS2_13** | 1,12 | 0,997 | 0,4 | 95 |
| **ADAMTS2_22** | 1,01 | 0,996 | 0,4 | 95 |
| **ASXL3_2** | 0,87 | 0,957 | 0,4 | 95 |
| **ASXL3_3** | 0,82 | 0,994 | 0,9 | 95 |
| **ASXL3_8** | 0,89 | 0,990 | 0,4 | 95 |
| **ASXL3_11** | 0,84 | 0,983 | 0,4 | 95 |
| **ASXL3_12** | 0,85 | 0,997 | 0,4 | 95 |
| **BCKDHA_2** | 0,91 | 0,995 | 0,9 | 95 |
| **BCKDHA_3** | 0,93 | 0,987 | 0,4 | 95 |
| **BCKDHA_4** | 0,86 | 0,992 | 0,4 | 95 |
| **BCKDHA_6** | 0,90 | 0,994 | 0,4 | 95 |
| **BTK_1** | 0,94 | 0,974 | 0,4 | 95 |
| **BTK_3** | 0,86 | 0,993 | 0,4 | 95 |
| **BTK_4** | 0,79 | 0,978 | 1,0 | 95 |
| **BTK_5** | 0,87 | 0,991 | 1,0 | 95 |
| **BTK_6** | 0,81 | 0,981 | 1,0 | 95 |
| **BTK_8** | 0,81 | 0,977 | 0,6 | 95 |
| **BTK_10** | 0,89 | 0,991 | 0,4 | 95 |
| **BTK_11** | 0,83 | 0,979 | 0,8 | 95 |
| **BTK_12** | 0,82 | 0,983 | 0,8 | 95 |
| **BTK_14** | 0,82 | 0,983 | 0,4 | 95 |
| **BTK_15** | 0,92 | 0,993 | 1,0 | 95 |
| **BTK_16** | 0,80 | 0,991 | 0,4 | 95 |
| **BTK_17** | 0,89 | 0,988 | 0,8 | 95 |
| **BTK_19** | 0,96 | 0,996 | 0,8 | 95 |
| **CC2D1A_2** | 0,91 | 0,994 | 0,4 | 95 |
| **CC2D1A_14** | 0,85 | 0,987 | 0,4 | 95 |
| **CC2D1A_16** | 1,00 | 0,978 | 0,4 | 95 |
| **CC2D1A_28** | 1,01 | 0,992 | 0,4 | 99 |
| **CD96_2** | 0,91 | 0,996 | 0,4 | 95 |
| **CD96_15** | 0,87 | 0,995 | 0,4 | 95 |
| **COL6A1_8** | 1,04 | 0,996 | 0,4 | 99 |
| **COL6A1_9** | 0,93 | 0,995 | 0,3 | 99 |
| **COL6A1_14** | 1,00 | 0,989 | 0,4 | 99 |
| **COL6A1_33** | 0,90 | 0,992 | 0,4 | 99 |
| **COL6A2_i1** | 0,96 | 0,981 | 0,4 | 95 |

*(cont.) Table S4 - Primers efficiency, R2 and PCR characteristics for the tested genes.*

| Primer | Efficiency | $R^2$ | Primer Volume (µL) | Melting Temperature ($^0$C) |
|---|---|---|---|---|
| **COL6A2_5** | 0,83 | 0,981 | 0,4 | 95 |
| **COL6A2_10** | 0,80 | 0,990 | 0,4 | 95 |
| **COL6A2_11** | 0,87 | 0,991 | 0,4 | 95 |
| **COL6A2_26** | 1,06 | 0,957 | 0,4 | 95 |
| **DNMT3B_3** | 1,00 | 0,996 | 0,6 | 95 |
| **DNMT3B_11** | 1,01 | 0,997 | 0,6 | 95 |
| **DNMT3B_22** | 0,81 | 0,992 | 0,4 | 95 |
| **ERCC3_2** | 1,02 | 0,991 | 0,4 | 95 |
| **ERCC3_8** | 0,88 | 0,990 | 0,4 | 95 |
| **ERCC3_14** | 0,85 | 0,992 | 0,4 | 95 |
| **ESCO2_2** | 0,86 | 0,968 | 0,4 | 95 |
| **ESCO2_7** | 0,85 | 0,983 | 0,4 | 95 |
| **ESCO2_11** | 0,88 | 0,938 | 0,4 | 95 |
| **FANCC_1** | 1,10 | 0,972 | 0,4 | 99 |
| **FANCC_2** | 0,83 | 0,973 | 0,4 | 95 |
| **FANCC_4** | 0,85 | 0,986 | 0,8 | 95 |
| **FANCC_14** | 0,86 | 0,991 | 0,8 | 95 |
| **FBN2_2** | 0,98 | 0,989 | 0,4 | 95 |
| **FBN2_23** | 0,97 | 0,998 | 0,4 | 95 |
| **FBN2_65** | 0,88 | 0,995 | 0,4 | 95 |
| **GYG1_2** | 1,04 | 0,992 | 0,6 | 99 |
| **GYG1_8** | 0,88 | 0,992 | 0,4 | 99 |
| **IL7R_3** | 0,87 | 0,973 | 0,4 | 95 |
| **IL7R_8** | 0,90 | 0,974 | 0,4 | 95 |
| **ITPR1_4** | 1,03 | 0,994 | 0,4 | 95 |
| **ITPR1_6** | 0,98 | 0,993 | 0,4 | 95 |
| **ITPR1_10** | 0,86 | 0,989 | 0,4 | 95 |
| **ITPR1_20** | 0,98 | 0,994 | 0,4 | 95 |
| **ITPR1_38** | 0,97 | 0,992 | 0,4 | 95 |
| **ITPR1_40** | 0,94 | 0,985 | 0,4 | 95 |
| **ITPR1_44** | 0,94 | 0,984 | 0,4 | 95 |
| **ITPR1_48** | 0,82 | 0,992 | 0,6 | 95 |
| **ITPR1_52** | 0,90 | 0,981 | 0,4 | 95 |
| **ITPR1_57** | 0,96 | 0,992 | 0,4 | 95 |
| **MARVELD2_2** | 1,02 | 0,997 | 0,4 | 95 |
| **MARVELD2_5** | 0,95 | 0,997 | 0,4 | 95 |
| **MCCC2_1** | 0,95 | 0,993 | 0,4 | 99 |
| **MCCC2_10** | 0,88 | 0,993 | 0,4 | 95 |
| **MCCC2_17** | 0,87 | 0,990 | 0,4 | 95 |
| **MGME1_2** | 0,83 | 0,947 | 0,4 | 95 |
| **MGME1_3** | 0,95 | 0,996 | 0,6 | 95 |
| **MGME1_5** | 0,83 | 0,979 | 0,4 | 95 |

*(cont.) Table S4 - Primers efficiency, R2 and PCR characteristics for the tested genes.*

| Primer | Efficiency | $R^2$ | Primer Volume (µL) | Melting Temperature ($^0$C) |
|---|---|---|---|---|
| **MMACHC_1** | 0,94 | 0,990 | 0,6 | 95 |
| **MMACHC_2** | 1,01 | 0,993 | 0,4 | 95 |
| **MMACHC_4** | 0,86 | 0,996 | 0,4 | 95 |
| **MVK_3** | 0,93 | 0,980 | 0,4 | 95 |
| **MVK_4** | 0,73 | 0,965 | 0,9 | 95 |
| **MVK_6** | 0,98 | 0,972 | 0,4 | 95 |
| **MYO15A_3** | 1,16 | 0,989 | 0,4 | 95 |
| **MYO15A_16** | 0,93 | 0,991 | 0,4 | 95 |
| **MYO15A_30** | 0,91 | 0,991 | 0,4 | 95 |
| **MYO15A_45** | 0,89 | 0,985 | 0,4 | 95 |
| **MYO15A_66** | 0,89 | 0,995 | 0,4 | 95 |
| **NAA10_2** | 1,04 | 0,983 | 0,4 | 95 |
| **NAA10_8** | 1,05 | 0,989 | 0,4 | 95 |
| **PEX1_6** | 0,90 | 0,988 | 0,6 | 95 |
| **PEX1_9** | 0,90 | 0,990 | 0,4 | 95 |
| **PEX1_12** | 1,01 | 0,994 | 0,4 | 95 |
| **PEX1_14** | 0,91 | 0,986 | 0,4 | 95 |
| **PEX1_24** | 0,99 | 0,996 | 0,4 | 95 |
| **SAG_2** | 0,86 | 0,993 | 0,4 | 95 |
| **SAG_16** | 1,00 | 0,989 | 0,4 | 95 |
| **SBF1_3** | 1,06 | 0,993 | 0,4 | 95 |
| **SBF1_12** | 1,02 | 0,992 | 0,4 | 95 |
| **SBF1_26** | 0,98 | 0,985 | 0,2 | 99 |
| **SBF1_35** | 1,11 | 0,987 | 0,45 | 95 |
| **SBF1_40** | 0,88 | 0,983 | 0,4 | 95 |
| **SLC5A1_2** | 0,93 | 0,992 | 0,4 | 95 |
| **SLC5A1_7** | 0,93 | 0,995 | 0,4 | 95 |
| **SLC5A1_i14** | 0,82 | 0,997 | 0,4 | 95 |
| **SLC5A1_15** | 0,92 | 0,964 | 0,4 | 95 |
| **SLC5A2_1** | 1,00 | 0,992 | 0,6 | 95 |
| **SLC5A2_4** | 0,86 | 0,994 | 0,6 | 99 |
| **SLC5A2_14** | 0,90 | 0,995 | 0,4 | 95 |
| **SMARCA4_3** | 0,93 | 0,969 | 0,4 | 95 |
| **SMARCA4_16** | 0,97 | 0,994 | 0,4 | 95 |
| **SMARCA4_30** | 0,93 | 0,990 | 0,4 | 95 |
| **SMARCA4_34** | 0,84 | 0,995 | 0,6 | 99 |
| **THPO_3** | 0,98 | 0,983 | 0,6 | 95 |
| **THPO_4** | 0,80 | 0,978 | 0,4 | 95 |
| **THPO_6** | 0,80 | 0,949 | 0,4 | 95 |
| **VCP_2** | 1,00 | 1,000 | 0,6 | 95 |
| **VCP_8** | 0,80 | 0,992 | 0,4 | 95 |
| **VCP_16** | 0,82 | 0,992 | 0,4 | 95 |

## ANNEX V – Relative Quantitation Results

*Table S5 - Relative quantitation results for the samples used as normal controls.*

| Control Samples | | | Control Samples | | |
|---|---|---|---|---|---|
| **Primer** | **Sample** | **Copy Number** | **Primer** | **Sample** | **Copy Number** |
| **ABCA3_5** | POOL | 1,01 | **COL6A1_33** | POOL | 1,02 |
| **ABCA3_12** | POOL | 1,02 | **COL6A1_8** | POOL | 1,02 |
| **ABCA3_22** | POOL | 0,97 | **COL6A1_9** | POOL | 1,14 |
| **ABCA3_32** | POOL | 1,01 | **COL6A2_10** | A8 | 1,00 |
| **ADAMTS2_13** | POOL | 0,93 | **COL6A2_11** | A17 | 0,89 |
| **ADAMTS2_22** | POOL | 1,10 | **COL6A2_26** | POOL | 0,90 |
| **ADAMTS2_3** | POOL | 1,09 | **COL6A2_5** | A8 | 1,12 |
| **ASXL3_11** | POOL | 0,99 | **COL6A2_i1** | A17 | 1,09 |
| **ASXL3_12** | POOL | 0,88 | **COL6A3_1** | POOL | 1,01 |
| **ASXL3_2** | POOL | 0,90 | **COL6A3_16** | POOL | 1,05 |
| **ASXL3_3** | POOL | 0,93 | **COL6A3_43** | POOL | 0,92 |
| **ASXL3_8** | POOL | 1,00 | **DNMT3B_11** | POOL | 0,97 |
| **BCKDHA_2** | POOL | 0,97 | **DNMT3B_22** | POOL | 0,96 |
| **BCKDHA_3** | POOL | 0,95 | **DNMT3B_3** | POOL | 0,99 |
| **BCKDHA_4** | POOL | 0,99 | **ERCC3_14** | POOL | 1,01 |
| **BCKDHA_6** | POOL | 1,04 | **ERCC3_2** | POOL | 0,99 |
| **BTK_1** | POOL | 1,02 | **ERCC3_8** | POOL | 0,99 |
| **BTK_10** | POOL | 0,99 | **ESCO2_11** | POOL | 1,01 |
| **BTK_11** | POOL | 0,99 | **ESCO2_2** | POOL | 0,99 |
| **BTK_12** | POOL | 0,91 | **ESCO2_7** | POOL | 1,04 |
| **BTK_14** | POOL | 1,05 | **FANCC_1** | POOL | 1,10 |
| **BTK_15** | POOL | 1,02 | **FANCC_14** | POOL | 1,07 |
| **BTK_16** | POOL | 0,96 | **FANCC_2** | POOL | 1,02 |
| **BTK_17** | POOL | 1,06 | **FANCC_4** | POOL | 0,97 |
| **BTK_19** | POOL | 1,02 | **FBN2_2** | POOL | 1,08 |
| **BTK_3** | POOL | 1,06 | **FBN2_23** | POOL | 1,09 |
| **BTK_4** | POOL | 1,07 | **FBN2_65** | POOL | 0,98 |
| **BTK_5** | POOL | 0,99 | **GYG1_2** | POOL | 1,05 |
| **BTK_6** | POOL | 1,06 | **GYG1_8** | POOL | 1,07 |
| **BTK_8** | POOL | 1,01 | **IL7R_3** | POOL | 1,07 |
| **CC2D1A_14** | POOL | 0,98 | **IL7R_3** | POOL | 0,96 |
| **CC2D1A_16** | POOL | 1,02 | **IL7R_8** | POOL | 1,10 |
| **CC2D1A_2** | POOL | 0,96 | **IL7R_8** | POOL | 1,05 |
| **CC2D1A_28** | POOL | 1,03 | **ITPR1_10** | POOL | 1,09 |
| **CD96_15** | POOL | 1,03 | **ITPR1_20** | POOL | 0,96 |
| **CD96_2** | POOL | 0,95 | **ITPR1_38** | POOL | 0,92 |
| **COL6A1_14** | POOL | 1,07 | **ITPR1_4** | POOL | 0,95 |
| **COL6A1_14** | POOL | 0,98 | **ITPR1_40** | POOL | 1,00 |

*(cont.) Table S5 - Relative quantitation results for the samples used as normal controls.*

| Control Samples | | | Control Samples | | |
|---|---|---|---|---|---|
| **Primer** | **Sample** | **Copy Number** | **Primer** | **Sample** | **Copy Number** |
| **ITPR1_44** | POOL | 1,06 | **PEX1_6** | POOL | 0,98 |
| **ITPR1_48** | POOL | 0,97 | **PEX1_9** | POOL | 1,03 |
| **ITPR1_52** | POOL | 1,13 | **SAG_16** | POOL | 1,01 |
| **ITPR1_57** | POOL | 1,07 | **SAG_2** | POOL | 0,99 |
| **ITPR1_6** | POOL | 0,92 | **SBF1_12** | POOL | 1,00 |
| **MARVELD2_2** | POOL | 0,98 | **SBF1_26** | POOL | 1,03 |
| **MARVELD2_5** | POOL | 1,05 | **SBF1_3** | POOL | 1,06 |
| **MCCC2_10** | POOL | 0,98 | **SBF1_35** | POOL | 1,11 |
| **MCCC2_17** | POOL | 0,99 | **SBF1_40** | POOL | 1,01 |
| **MCCC2_2** | POOL | 1,07 | **SLC5A1_15** | POOL | 1,06 |
| **MGME1_2** | POOL | 1,08 | **SLC5A1_2** | POOL | 1,05 |
| **MGME1_3** | POOL | 1,13 | **SLC5A1_7** | POOL | 1,03 |
| **MGME1_5** | POOL | 1,09 | **SLC5A1_i14** | POOL | 1,03 |
| **MMACHC_1** | POOL | 1,09 | **SLC5A2_1** | POOL | 1,01 |
| **MMACHC_2** | POOL | 1,07 | **SLC5A2_14** | POOL | 0,99 |
| **MMACHC_4** | POOL | 1,07 | **SLC5A2_4** | POOL | 1,08 |
| **MVK_3** | POOL | 0,95 | **SMARCA4_16** | POOL | 1,11 |
| **MVK_4** | POOL | 1,00 | **SMARCA4_3** | POOL | 1,05 |
| **MVK_6** | POOL | 1,07 | **SMARCA4_30** | POOL | 0,99 |
| **MYO15A_16** | POOL | 0,96 | **SMARCA4_34** | POOL | 1,02 |
| **MYO15A_3** | POOL | 1,00 | **THPO_3** | POOL | 1,02 |
| **MYO15A_30** | POOL | 1,11 | **THPO_3** | CT_aCGH | 1,08 |
| **MYO15A_45** | POOL | 0,94 | **THPO_4** | POOL | 1,07 |
| **MYO15A_66** | POOL | 0,95 | **THPO_4** | CT_aCGH | 1,02 |
| **NAA10_2** | POOL | 1,00 | **THPO_6** | POOL | 1,00 |
| **NAA10_8** | POOL | 1,02 | **THPO_6** | CT_aCGH | 1,15 |
| **PEX1_12** | POOL | 1,05 | **VCP_16** | POOL | 0,91 |
| **PEX1_14** | POOL | 1,00 | **VCP_2** | POOL | 0,98 |
| **PEX1_24** | POOL | 1,05 | **VCP_8** | POOL | 0,95 |

*Table S6 - Relative quantitation results for the test samples scored as normal.*

| Normal Samples | | |
|---|---|---|
| **Primer** | **Sample** | **Copy Number** |
| **ABCA3_5** | A80 | 1,01 |
| **ABCA3_12** | A80 | 0,93 |
| **ABCA3_22** | A80 | 1,00 |
| **ADAMTS2_13** | A56 | 0,95 |
| **ADAMTS2_22** | A56 | 0,91 |
| **ADAMTS2_3** | A56 | 1,08 |
| **ASXL3_11** | A49 | 1,09 |
| **ASXL3_11** | A48 | 0,99 |
| **ASXL3_12** | A49 | 1,03 |
| **ASXL3_12** | A48 | 0,94 |
| **ASXL3_2** | A49 | 1,19 |
| **ASXL3_2** | A48 | 0,99 |
| **ASXL3_3** | A49 | 1,15 |
| **ASXL3_3** | A48 | 1,03 |
| **ASXL3_8** | A49 | 1,10 |
| **ASXL3_8** | A48 | 1,12 |
| **BCKDHA_4** | A53 | 0,85 |
| **BCKDHA_6** | A53 | 0,85 |
| **BTK_1** | A68 | 0,98 |
| **BTK_10** | A68 | 0,93 |
| **BTK_11** | A68 | 0,93 |
| **BTK_12** | A68 | 0,99 |
| **BTK_14** | A68 | 1,04 |
| **BTK_15** | A68 | 0,94 |
| **BTK_16** | A68 | 0,95 |
| **BTK_17** | A68 | 1,06 |
| **BTK_19** | A68 | 1,06 |
| **BTK_3** | A68 | 0,99 |
| **BTK_4** | A68 | 0,88 |
| **BTK_5** | A68 | 0,92 |
| **BTK_6** | A68 | 0,96 |
| **BTK_8** | A68 | 0,97 |
| **CC2D1A_14** | A57 | 1,05 |
| **CC2D1A_16** | A57 | 1,15 |
| **CC2D1A_2** | A57 | 1,09 |
| **CC2D1A_28** | A57 | 1,08 |
| **CD96_15** | A62 | 0,98 |
| **CD96_2** | A62 | 1,09 |
| **COL6A1_33** | A20 | 0,85 |
| **COL6A1_8** | A20 | 0,97 |

| Normal Samples | | |
|---|---|---|
| **Primer** | **Sample** | **Copy Number** |
| **COL6A1_9** | A20 | 1,01 |
| **COL6A2_i1** | A17 | 0,81 |
| **COL6A2_10** | A20 | 0,96 |
| **COL6A2_11** | A20 | 1,11 |
| **COL6A2_26** | A20 | 0,87 |
| **COL6A2_5** | A20 | 0,89 |
| **COL6A3_1** | A20 | 0,91 |
| **COL6A3_16** | A20 | 1,09 |
| **COL6A3_43** | A20 | 1,04 |
| **DNMT3B_11** | A65 | 1,06 |
| **DNMT3B_22** | A65 | 0,99 |
| **DNMT3B_3** | A65 | 0,97 |
| **ERCC3_14** | A58 | 0,92 |
| **ERCC3_2** | A58 | 0,90 |
| **ERCC3_8** | A58 | 0,96 |
| **ESCO2_11** | A34 | 0,93 |
| **ESCO2_2** | A34 | 1,12 |
| **ESCO2_7** | A34 | 0,97 |
| **FANCC_1** | A16 | 0,91 |
| **FANCC_14** | A16 | 0,94 |
| **FANCC_2** | A16 | 1,12 |
| **FANCC_4** | A16 | 1,14 |
| **FBN2_2** | A59 | 0,93 |
| **FBN2_23** | A59 | 1,06 |
| **FBN2_65** | A59 | 0,97 |
| **GYG1_2** | A76 | 0,98 |
| **GYG1_8** | A76 | 0,90 |
| **IL7R_3** | A66 | 1,11 |
| **IL7R_3** | A32 | 0,97 |
| **IL7R_8** | A66 | 1,07 |
| **IL7R_8** | A32 | 0,90 |
| **ITPR1_10** | A64 | 0,97 |
| **ITPR1_20** | A64 | 0,93 |
| **ITPR1_38** | A64 | 0,99 |
| **ITPR1_4** | A64 | 1,06 |
| **ITPR1_40** | A64 | 0,93 |
| **ITPR1_44** | A64 | 1,00 |
| **ITPR1_48** | A64 | 0,90 |
| **ITPR1_52** | A64 | 0,90 |
| **ITPR1_57** | A64 | 0,88 |

*(cont.) Table S6 - Relative quantitation results for the test samples scored as normal.*

| Normal Samples | | |
|---|---|---|
| **Primer** | **Sample** | **Copy Number** |
| **ITPR1_6** | A64 | 0,93 |
| **MARVELD2_2** | A75 | 1,13 |
| **MARVELD2_5** | A75 | 1,16 |
| **MCCC2_10** | A79 | 0,98 |
| **MCCC2_17** | A79 | 0,90 |
| **MCCC2_2** | A79 | 0,96 |
| **MGME1_2** | A33 | 0,88 |
| **MGME1_3** | A33 | 0,94 |
| **MGME1_5** | A33 | 1,11 |
| **MMACHC_1** | A54 | 0,93 |
| **MMACHC_2** | A54 | 0,88 |
| **MMACHC_4** | A54 | 1,02 |
| **MVK_3** | A50 | 1,07 |
| **MVK_4** | A50 | 1,14 |
| **MVK_6** | A50 | 0,91 |
| **MYO15A_16** | A35 | 1,04 |
| **MYO15A_3** | A35 | 1,08 |
| **MYO15A_30** | A35 | 1,08 |
| **MYO15A_45** | A35 | 1,08 |
| **MYO15A_66** | A35 | 1,08 |
| **NAA10_2** | A78 | 1,17 |
| **NAA10_8** | A78 | 0,99 |
| **PEX1_12** | A50 | 0,89 |
| **PEX1_14** | A50 | 0,95 |
| **PEX1_24** | A50 | 0,89 |
| **PEX1_6** | A50 | 0,93 |
| **PEX1_9** | A50 | 0,92 |

| Normal Samples | | |
|---|---|---|
| **Primer** | **Sample** | **Copy Number** |
| **SAG_16** | A55 | 1,09 |
| **SAG_2** | A55 | 1,00 |
| **SBF1_12** | A63 | 0,94 |
| **SBF1_26** | A63 | 0,97 |
| **SBF1_3** | A63 | 1,10 |
| **SBF1_35** | A63 | 0,95 |
| **SBF1_40** | A63 | 0,93 |
| **SLC5A1_15** | A52 | 0,85 |
| **SLC5A1_2** | A52 | 0,91 |
| **SLC5A1_7** | A52 | 1,14 |
| **SLC5A1_i14** | A52 | 0,98 |
| **SLC5A2_1** | A60 | 1,06 |
| **SLC5A2_14** | A60 | 0,93 |
| **SLC5A2_4** | A60 | 0,91 |
| **SMARCA4_16** | A77 | 0,93 |
| **SMARCA4_3** | A77 | 1,13 |
| **SMARCA4_30** | A77 | 0,94 |
| **SMARCA4_34** | A77 | 1,07 |
| **THPO_3** | A67 | 1,02 |
| **THPO_3** | A25 | 0,92 |
| **THPO_4** | A67 | 0,89 |
| **THPO_4** | A25 | 0,91 |
| **THPO_6** | A67 | 0,93 |
| **THPO_6** | A25 | 0,97 |
| **VCP_16** | A61 | 0,92 |
| **VCP_2** | A61 | 0,92 |
| **VCP_8** | A61 | 0,91 |

*Table S7 - Relative quantitation results for the test samples scored as deleted and duplicated.*

| Deleted Samples | | | Duplicated Samples | | |
|---|---|---|---|---|---|
| **Primer** | **Sample** | **Copy Number** | **Primer** | **Sample** | **Copy Number** |
| **COL6A1_14** | A20 | 0,63 | **BTK_1** | POOL_F | 2,13 |
| **COL6A1_14** | A20 | 0,52 | **BTK_3** | POOL_F | 2,01 |
| **ABCA3_32** | A80 | 0,47 | **BTK_5** | POOL_F | 1,99 |
| **ABCA3_32** | A80 | 0,50 | **BTK_6** | POOL_F | 2,06 |
| **BCKDHA_2** | A53 | 0,01 | **BTK_8** | POOL_F | 1,96 |
| **BCKDHA_3** | A53 | 0,03 | **BTK_10** | POOL_F | 2,12 |
| | | | **BTK_11** | POOL_F | 2,36 |
| | | | **BTK_12** | POOL_F | 2,23 |
| | | | **BTK_14** | POOL_F | 2,17 |
| | | | **BTK_15** | POOL_F | 1,96 |
| | | | **BTK_16** | POOL_F | 2,02 |
| | | | **BTK_17** | POOL_F | 2,01 |
| | | | **BTK_19** | POOL_F | 1,58 |
| | | | **NAA10_2** | POOL_F | 1,28 |
| | | | **NAA10_8** | POOL_F | 1,43 |