UNRAVELING RARE AND INHERITED DISEASES WITH GENETIC TECHNOLOGIES

- **Page 2**
  Rare and Inherited Diseases: Uncommon Yet Widespread
- **Page 3**
  Infographic: Exploring Rare Diseases on a Global and Molecular Scale
- **Page 4**
  Somatic CAG Repeats in Huntington's Disease Expand with Age
- **Page 5**
  Smad8 Dysregulates microRNAs in Duchenne Muscular Dystrophy
- **Page 6**
  Technology Overview

Sponsored by

ThermoFisher SCIENTIFIC





RARE AND INHERITED DISEASES: UNCOMMON YET WIDESPREAD

Rare diseases are disorders or conditions that afflict a small number of the population.[1] There is no worldwide consensus on the official definition, as different countries have distinct thoughts as to what they consider rare. For instance, Europe and Canada define rare diseases as affecting fewer than one in 2,000 people, while the US defines rare diseases as afflicting fewer than 200,000 Americans, which is equivalent to one in 1,630 people.[2,3] Although each disease is rare individually, collectively scientists estimate that they affect 473 million people worldwide.[3] Based on data extracted from the rare disease database Orphanet by the website Orphadata, as of 2023 there are over 10,000 known rare diseases.[4] Additionally, of the known rare diseases, 39 percent of them have a confirmed genetic origin, and this number is increasing as researchers discover new disease-associated genes.[3] These rare and inherited diseases, which are also known as rare genetic diseases, result from mutations within the genome affecting one or more genes.[5] Generally, rare genetic diseases are chronic and lead to decreased quality of life and premature death, making diagnosis and therapeutic development of the highest priority for these patients.

Tools Used to Diagnose and Study Rare Genetic Diseases

Patients with rare and inherited diseases need to receive a correct diagnosis as quickly as possible. This information is vital for clinicians to assess a patient's treatment options and find ways to manage their disease symptoms.[6] However, the genetic and phenotypic heterogeneity of these diseases, as well as the small number of patients, makes diagnosing rare diseases difficult.[7] The mutations underlying rare genetic diseases result from genomic alterations, such as single nucleotide polymorphisms, insertions, deletions, inversions, translocations, or microsatellite repeat expansions affecting one or many genes.[8] Additionally, the same phenotype can result from different mutations in the same gene or in separate genes.[7] This further complicates diagnosis and the prescription of therapies, as not all patients with the same phenotype will respond to the same treatment.[9]

Historically, clinicians used Sanger sequencing and PCR for diagnosis, but these techniques limited the number of genes that they could examine simultaneously.[7] Clinicians now frequently use next-generation sequencing (NGS) to identify mutations and make a diagnosis, but they also confirm their findings using established molecular biology techniques including quantitative real-time PCR (qPCR), digital PCR (dPCR), microarrays, and Sanger sequencing.[7] Moreover, researchers use these approaches and NGS to determine which genomic changes are associated with a disease to establish diagnostic biomarkers, the disease etiology, or affected cellular pathways. This information can inform therapeutic development or the repurposing of existing drugs.[9]

Most Researched Rare Genetic Diseases

Although there are numerous rare and inherited diseases to examine, the diseases most well-funded by the National Institutes of Health (NIH) and, thus, most researched include lupus, cystic fibrosis, muscular dystrophy, and Huntington's disease.[10]

Lupus is a group of chronic autoimmune diseases, where systemic lupus erythematosus (SLE) is the most common.[11] The disorder is characterized by autoantibody production resulting in organ damage. Although the etiology of SLE is unknown, researchers have shown that some patients are genetically predisposed to develop the disorder.[11]

Cystic fibrosis is an autosomal recessive disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator gene (*CFTR*), which encodes a chloride channel that regulates the water concentration in body secretions including mucus and sweat.[12] Consequently, the disease is characterized by viscous mucus leading to chronic respiratory infections.

Muscular dystrophies are a group of inherited diseases characterized by progressive muscle weakening and atrophy.[13] Duchenne muscular dystrophy, an X-linked recessive disorder caused by mutations in the muscle-expressed dystrophin gene (*DMD*), is the most common and severe of these neuromuscular diseases.[14]

Huntington's disease is an autosomal dominant disorder resulting from mutations in the huntingtin gene (*HTT*). The encoded protein is critical for various cellular processes, such as transcriptional regulation, cell signaling, and axonal trafficking.[15] The mutant huntingtin protein aggregates within cells, leading to their death and the associated neurodegenerative symptoms including motor impairment and cognitive decline.

See References on Page 13.

EXPLORING RARE DISEASES ON A GLOBAL AND MOLECULAR SCALE

Rare Diseases Collectively

Estimated global prevalence is **1 in 2,500** people[1]

In 2019, estimated economic burden reached **$1 trillion** annually in the US alone[2]

510% increase in the number of publications in 2023 compared to 2009

Estimated **$7.1 billion** budgeted by NIH for 2024[3]

Systemic Lupus Erythematosus

Global prevalence ranges between **13 to 7,700 in 100,000** people[4]

Estimated **$144 million** budgeted by NIH for 2024[3]

180% increase in the number of publications in 2023 compared to 2009

Cystic Fibrosis

Estimated global prevalence is **2 in 100,000** people[5]

Estimated **$96 million** budgeted by NIH for 2024[3]

170% increase in the number of publications in 2023 compared to 2009

Duchenne Muscular Dystrophy

Estimated global prevalence is **4.8 in 100,000** people[6]

Estimated **$34 million** budgeted by NIH for 2024[3]

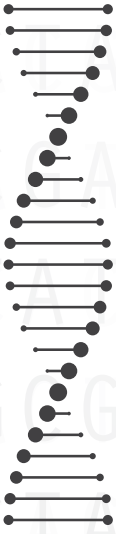
210% increase in the number of publications in 2023 compared to 2009

Huntington's Disease

Estimated global prevalence is **4.88 in 100,000** people[7]

Estimated **$54 million** budgeted by NIH for 2024[3]

170% increase in the number of publications in 2023 compared to 2009



Comparison of Techniques Employed to Study Rare Genetic Diseases

	Sanger Sequencing	Quantitative Real-Time PCR (qPCR)	Digital PCR (dPCR)	Microarray	Next-Generation Sequencing
Prior Knowledge of Gene(s) of Interest Required	No	Yes	Yes	Yes	No
Number of Genes Researchers Can Analyze	1 to 100s	1 to 100s	1 to 100s	1 to 1,000s	1 to 1,000s

See References on Page 13.

SOMATIC CAG REPEATS IN HUNTINGTON'S DISEASE EXPAND WITH AGE

Alexandra Durr, MD, PhD
Professor, Pitié-Salpêtrière University Hospital and Sorbonne University
Team Leader, Paris Brain Institute



Above Image: Courtesy of the Paris Brain Institute

Huntington's disease is a progressive, autosomal dominant disease caused by an unstable, expanded CAG trinucleotide repeat in the huntingtin gene (*HTT*), which leads to cognitive decline in functions including attention, planning, and thought. Alexandra Durr, a neurologist and clinical researcher at the Pitié-Salpêtrière University Hospital and Paris Brain Institute, studies rare inherited diseases including Huntington's disease. In an *eLife* paper, she showed that CAG repeats within somatic cells increase with age, where both the expansion rate and repeat length affect the age of onset.[1]

How did you become involved in Huntington's disease research?

It started with the families of patients with neurogenetic disorders that we see here at the Salpêtrière Hospital. We wanted to learn why there was variability in disease symptoms from one generation to the other. Before the huntingtin gene was discovered in 1993, I had a lot of questions from families about predictive testing. So, I started the first clinic in the Salpêtrière Hospital together with Josué Feingold and Marcela Gargiulo. We became interested in Huntington's disease because patients came in for predictive testing for this disease more frequently than others, it was the first rare disease-associated gene identified, and it was very easy to diagnose.

What motivated you to perform this study?

The onset of Huntington's disease was first linked to the number of CAG repeats, but we realized that there was variation in its onset and symptoms that could not explained by the repeat alone. Also, when performing prenatal testing for Huntington's disease, the CAG repeat in the fetus is very stable, but there is somatic instability, or mosaicism in CAG repeat lengths, within the parents. Clinicians use this difference to distinguish fetal blood from parental blood. This made me question what is happening in patients' cells throughout their lifetime to acquire somatic instability. Additionally, there was a computational model published by Shai Kaplan and his colleagues at the Weizmann Institute of Science that proposed that the disease's onset is driven by somatic instability.[2] We had already collected multiple blood samples from our patients with Huntington's disease over a 20 year period, so we tested them to see if somatic instability changed over a patient's lifetime.

How did you determine the CAG repeat length?

We extracted DNA from patient blood samples and performed nested PCR amplification of the CAG repeat. We then ran fragment analysis of the various PCR products and plotted their abundance relative to the number of CAG repeats. Interestingly, the profile is quite symmetrical. The highest peak in the middle is known as the main peak, while the peaks on the left and right sides are considered the minus and plus peaks. The number and height of these additional plus peaks indicates the level of somatic instability, while we do not examine the minus peaks because polymerase slippage during PCR could have produced them. Through analysis of the longitudinal blood samples, we observed that somatic instability increased over a patient's lifetime.

If I was publishing this paper today, I would also sequence the CAG repeat because it is not only the size that matters but also its structure. The CAG repeat is not completely pure in Huntington's disease. There is a CAA interruption before the end of the CAG repeat and its loss leads to an earlier onset. Researchers can see the structure through sequencing rather than PCR.

What effect did you hope this study would have on your field?

Until this study, no one had monitored lifetime somatic instability in the real world. We showed that instability exists, and it is likely driving the onset of the disease, as Kaplan originally hypothesized. We know that both having the mutation and the aging process causes a patient to get Huntington's disease, and what comes with aging is a decline in DNA repair capacity and efficiency. If somatic cells are unable to properly repair their DNA, this could lead to greater instability and make the cells more vulnerable to dying. I think it will be important to determine what exactly is driving this somatic instability because this could give us a therapeutic target.

This interview has been condensed and edited for clarity.

See References on Page 13.

Smad8 DYSREGULATES MICRORNAS IN DUCHENNE MUSCULAR DYSTROPHY

Michael Lopez, MD, PhD
Assistant Professor of Pediatrics
Children's of Alabama
University of Alabama at Birmingham



Above Image: Courtesy of Michael Lopez

After working with a muscular dystrophy researcher through an undergraduate research program, Michael Lopez became fascinated with muscle biology and the disorders that could affect its normal function. Since then, he has become a pediatric neuromuscular physician-scientist at Children's of Alabama and the University of Alabama at Birmingham focusing particularly on muscular dystrophies. In his *International Journal of Molecular Sciences* paper, Lopez along with his mentors, Peter King and Matthew Alexander, examined the role of receptor-regulated Smads (R-Smads) in Duchenne muscular dystrophy (DMD).[1]

What is DMD?

DMD is an X-linked recessive disorder of progressive skeletal muscle wasting, which affects one in 3000 to 5000 newborn males. It results from mutations that disrupt the production of the dystrophin protein, which is required at the sarcolemmal membrane and destabilizes the muscle fiber. This leads to chronic instability and susceptibility to injury, which eventually culminates in the replacement of healthy muscle tissue with fibrotic and adipocytic tissue.

What motivated you to perform this study?

We already knew that increased TGFβ signaling is a driver of DMD pathology, where R-Smads are known transducers of this receptor superfamily. Although the Smad1/5/8 pathway was not established as a disease-driving pathway in DMD, Peter King identified Smad8 as a biomarker of amyotrophic lateral sclerosis (ALS) muscle tissue.[2] Because of the similarities between ALS and DMD, including the aberrant TGFβ signaling and that both disorders lead to muscle atrophy, we thought that the R-Smads, particularly Smad8, would be important in DMD as well. Additionally, Smads regulate the processing of muscle-enriched microRNAs, called myomiRs. In ALS, the increased Smad8 levels corresponded to decreased myomiR levels.[3] As we already knew that myomiRs are aberrantly regulated in DMD, we decided to determine if R-Smads were causing this dysregulation.

How did you measure the R-Smads and myomiRs?

We took muscle tissue from patients with DMD and a DMD mouse model and looked at R-Smad protein and mRNA using standard molecular biology techniques. It is important to use both human and mouse samples because we wanted to ensure that our findings are highly relevant to the disease in humans. Through employing western blotting, we examined the ratio of active phosphorylated forms of R-Smads to their total levels within the tissues. We also used quantitative real-time PCR to measure the mRNA levels of the R-Smads and the myomiR transcripts in those same muscle samples.

What are your next steps for this project?

Through this paper, we observed that increased Smad8 expression in DMD muscle tissue corresponded to a strong suppression of the myomiRs. Now, we are developing the tools to be able to answer how Smad8 represses the myomiRs and other mechanistic questions. For instance, we are developing a conditional knockout of Smad8 in the DMD mouse model. After isolating myoblasts from these mice and comparing them to cells from wildtype mice, we will be able to understand the role of Smad8 in healthy and diseased myoblasts. While we are actively looking at the role of Smad8 in the disease model, I think examining the role of R-Smads in wound repair and regeneration is also going to be important.

Why are studies like this one important for the future treatment of DMD?

Gene therapies are now available for many rare diseases, but one of the challenges for DMD is that the dystrophin gene is the largest in the whole body, so replacing it is a technical challenge that is not easily solved. I believe that treating DMD is going to require researchers to perform an in-depth interrogation of the affected pathways beyond just thinking about the TGFβ receptor superfamily independently. We need to learn more about the drivers of TGFβ signaling and understand the complexities of the downstream effectors including R-Smads. Through this knowledge, we potentially could turn this neuromuscular disorder into more of a chronic disease, where we can strengthen the muscle in these patients and improve their quality of life.

This interview has been condensed and edited for clarity.

See References on Page 13.

TECHNOLOGY OVERVIEW

Written by Thermo Fisher Scientific

No matter what the disease is, the human genome is likely to be a factor in some aspect of that disease. In some cases, a disease is heritable—variations in the germline nucleic acid sequence and structure get passed from one generation to the next. However, the link between genomic variation and dysfunction and the etiology of these pathologies is still an intense field of investigation. The Human Genome Project was instrumental in introducing tools that both generated sequence information and used that information to understand human health and disease. The techniques and data produced by these tools have critically contributed to the knowledge of these inherited diseases. The Applied Biosystems™ brand has been at the forefront of developing these tools and has an extensive portfolio of solutions used for understanding rare and inherited genetic syndromes.

The Applied Biosystems portfolio can be thought of as a continuum that gives researchers options based on the types of questions being asked (Figure 1). For example, some investigators might be interested in **discovery**-based experimentation, using unbiased queries spanning the entire genome to uncover new sequences or relationships between genes. These types of experiments are facilitated by Ion Torrent™ AmpliSeq™ next-generation sequencing (NGS) workflows or by Applied Biosystems™ microarray solutions. Once a set of sequences of interest is decided, investigators might move to a **focus**-based experimental approach, where these selected sequences are analyzed in medium-throughput analyses, such as Sanger sequencing to focus on a defined region across many samples or Applied Biosystems™ TaqMan™ panels on array cards to focus on sets containing many genes. Finally, if a very specific set of mutations or genes needs to be analyzed in a large number of samples with minimal effort, individual Applied Biosystems™ TaqMan™ Assays may be used for the **detection** of these sequences.

Importantly, these approaches and technologies are complementary to each other. For example, a researcher might start with a discovery-based approach to



Figure 1. The genetic analysis continuum. In inherited disease research, the tools and techniques used depend on the question being asked, the scale of the data that needs to be analyzed, and the logistical needs of the investigator. Applied Biosystems™ genetic analysis systems fall into a continuum, from very large-scale **discovery**-based research to medium-scale focused research to very targeted detection of specific genes and mutations.

catalog the mutations commonly found in an affected individual, then focus on those mutations by confirmatory Sanger sequencing in more samples, and finally design a TaqMan probe to the mutation(s) that can be used to track and understand the dysfunction in model organisms. The continuum of Applied Biosystems solutions is meant to streamline the research path from discovery to gaining valuable insight.

In this chapter, we give an overview of the technologies in the Applied Biosystems continuum and how they can be used for research on rare and inherited diseases. However, this is not a complete or comprehensive list. The flexibility of these solutions as applied to inherited disease research depends only on the researcher's experimental creativity.

Microarray Methods and Tools

When a normal cell goes down a path that ultimately ends in a dysfunctional cell, it can acquire mutations that may range from single-nucleotide changes and small indels to copy number changes and large chromosomal rearrangements. One way to characterize these genomic changes is to use high-density DNA microarrays. For these analyses, hundreds of thousands of probes tiled across the genome are arrayed onto a single chip. Hybridization of samples to these arrays can determine which sequences are present and in how many copies (Figure 2). The advantage of microarrays is that hundreds of thousands to millions of sequences can be interrogated in a single experiment. If the appropriate probes that detect specific mutations are present on the chip, they can also detect common single-nucleotide polymorphism (SNP) variants. Microarrays are also useful for detecting copy number changes. The resolution of the copy number sequences is determined by the number of probes available; when probes are directly designed to target predefined common deletions or duplications, these can detect anomalies as small as 1,000 nucleotides. Whole-genome microarrays that cover both polymorphic (e.g., SNPs) and nonpolymorphic regions of the genome can be used to assess DNA copy number alterations at a much higher resolution than conventional cytogenetic analyses.

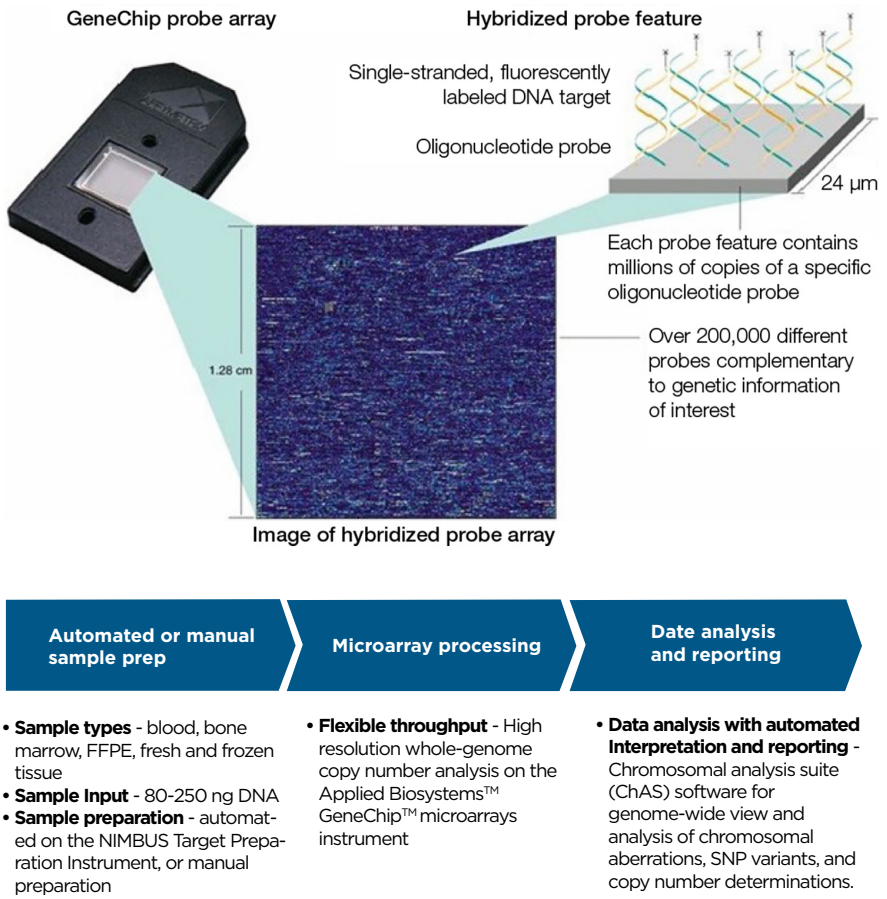


Figure 2. Basics of a microarray assay.

With modern microarray technologies, various types of causative genetic aberrations associated with inherited disorders can be studied by chromosomal microarray analysis (CMA). For genome-wide association studies (GWAS), several population-based microarrays have been designed for the discovery of alleles associated with a disorder. For example, the Applied Biosystems™ UK Biobank Axiom™ Array and similar arrays for other populations facilitate high-throughput, high-value genotyping of large sample cohorts with a single comprehensive low-cost solution. These assays contain one 96-array plate that allows for genome-wide analysis of large sample collections such as those screened at biobanks, genome centers, and core labs. The Applied Biosystems™ Axiom™ Precision Medicine Diversity Research Array includes content that was selected to emphasize variants that are commonly seen in clinical research on inherited diseases. All these arrays include variants associated with known developmental anomalies, clinically actionable variants, pharmacogenomic variants, and more. The markers were chosen from the list of published variants associated with phenotypes identified via GWAS, as per the NHGRI-EBI GWAS catalog, as well as some recently published and unpublished cancer-associated SNPs. The clinically actionable variants include relevant variants from the ClinVAR database, which enable the assessment of actionable genetic risks across a wide range of populations.

Several microarrays have been designed specifically for reproductive health research. The Applied Biosystems™ CarrierScan 1S Assay Kit includes over 6,000 sequences and structural variants informed by the American College of Medical Genetics (ACMG) and the American College of Obstetricians and Gynecologists (ACOG) for 600 inherited diseases.[1-6] The Applied Biosystems™ CytoScan™ HD Suite—comprising microarrays, reagents,

and analysis software—is a comprehensive high-resolution, whole-genome solution designed to assist in the understanding and characterization of biomarkers in hematological malignancies. The Applied Biosystems™ OncoScan™ CNV Plus Assay is a microarray-based assay that can analyze copy number changes over the whole genome. It is designed to query degraded genomic DNA, such as DNA extracted from formalin-fixed paraffin-embedded (FFPE) samples. Similarly, the Applied Biosystems™ OncoScan™ CNV Assay has the same copy number coverage as the OncoScan CNV Plus Assay but does not include somatic SNP mutation probes. These arrays may be useful for discovering copy number variations (CNVs), aneuploidies, or other structural anomalies associated with inherited diseases.[7]

miRNAs and other noncoding RNAs are important regulators of gene activity. It is estimated that more than 30% of protein translation of coding genes is regulated by miRNAs. These can be important biomarkers in inherited diseases (for reviews, see[8,9]). Importantly, miRNAs can be found in circulating fluids, such as blood or cerebrospinal fluid (CSF), and can be analyzed in liquid biopsies.[10] Because of their small size, stability, potential for regulating several different pathways, and accessibility in CSF or blood samples, miRNAs are ideal biomarkers. To keep pace with the discovery of new and novel miRNAs, the Applied Biosystems portfolio offers the GeneChip™ miRNA 4.0 Array. This array is designed to interrogate all mature miRNA sequences in miRBase, including human, mouse, and rat miRNAs. The analysis files contain host gene ID, predicted and validated miRNA target genes, and clustered miRNA information, facilitating downstream analysis.

The Applied Biosystems™ Clariom™ microarrays provide whole-transcriptome analysis solutions. The Human Clariom S Assay provides extensive coverage of all known well-annotated genes, is compatible with most research sample types, and is the tool of choice for finding gene expression patterns with known functions as quickly, easily, and cost-effectively as possible. The more advanced Clariom D Assays are based on coding and noncoding sequences culled from 16 different databases,* representing over 540,000 transcripts in over 6.7 million probes. Included in these sequences are probes for different splice isoforms, noncoding RNAs (pre-miRNAs, lincRNAs, Piwi-interacting RNAs or piRNAs), and circular RNAs, as well as annotated and speculative sequences. Clariom D Assays allow translational research scientists to generate high-fidelity biomarker signatures quickly and easily.[11] Querying the multiple transcript models on the Clariom D chip in a single experiment helps ensure that important pathways and potential biomarkers are not missed.

Transcriptome analysis is facilitated by data generated by Applied Biosystems™ Transcriptome Analysis Console (TAC) Software. A report of all transcripts and the associated relative expression is provided using statistical analysis and visualization tools. Additionally, information is supplied regarding the role of these transcripts in biological pathways. Together, the combination of microarrays and analysis software provides rapid, easy, and economical tools for identifying meaningful gene expression signature biomarkers.

Fragment Analysis Methods and Tools

Fragment analysis is a highly flexible method that can be applied to a wide variety of research fields. The flexibility afforded by the choice of PCR primers (PCR being a necessary step in any fragment analysis experiment) means that a specifically sized fragment corresponding to a PCR target sequence is straightforward to generate (Figure 3). Coupled with the ability to label fragments with up to four different fluorophores, researchers have great flexibility in their fragment analysis experimental design. Below, we describe some examples of how fragment analysis can be used in biomarker research problems.

Mutation detection

Biomarker researchers frequently need to investigate or authenticate a limited num-

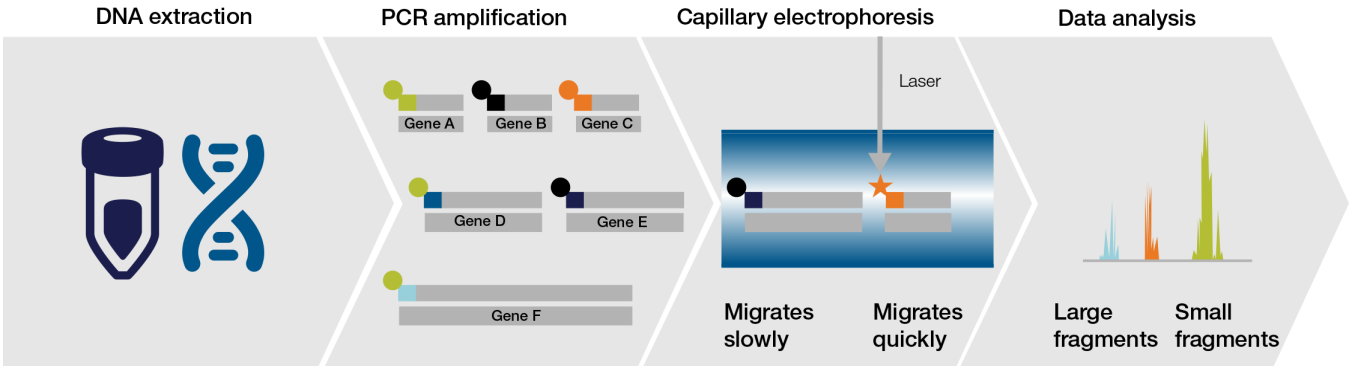


Figure 3. Basics of a fragment analysis assay. PCR primers are designed such that different target sequences are amplified with primers with different fluorophore labels, which will generate amplicons of different sizes. Following PCR, the amplicons are electrophoretically separated by size using capillary electrophoresis. A laser will excite the fluorophores as the fragments migrate through the capillary. The size and color of the resulting fragments reflect the abundance of the target sequence in the sample.

* Public database sources for the Clariom D Human Assays: Refseq, ccdsGENE, mgcGenes, UCSC KnownGene, lincRNATranscripts, VEGA, GENCODE, NONCODE, MiTranscriptome, RNACentral, ENSEMBL, circBase, Guo CircRNAs, AceView, lncRNAWiki, lncRNADB.

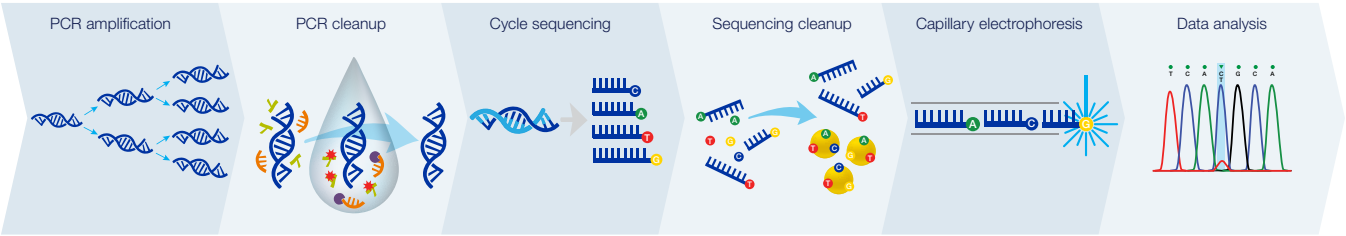


Figure 4. Basics of a Sanger sequencing workflow. The target region that will be sequenced is amplified by PCR. The primers and PCR reagents are removed before a second, linear amplification is performed. This step generates fragments that are chain-terminated with a fluorescent dideoxynucleotide. The cycle sequencing reaction is purified, and the resulting fragments are separated by capillary electrophoresis and detected using a laser. The sequence can be read from the lengths of the fragments and the colors of the dideoxynucleotide terminators.

ber of variations in one or more genes of interest. The Applied Biosystems™ SNaPshot™ Multiplex System is a versatile and economical method for performing SNP genotyping. Up to ten SNP markers (on different genes) can be investigated simultaneously by PCR amplification, followed by dideoxy single-base extension (SBE) using an unlabeled primer, and then capillary electrophoresis of the resulting fragments. After electrophoresis and fluorescent detection, the alleles of a single marker appear as differently colored peaks at roughly the same size in the electropherogram plot.[12]

Microsatellite instability assay

Microsatellite instability (MSI) is a hallmark of several cancers and is characterized by changes in the length of microsatellites due to defective mismatch repair (MMR) mechanisms. Microsatellites are genetic motifs consisting of 1-6 base pair repeats. These sequences are susceptible to replication errors that can result in deletions and insertions. Normally, these errors are corrected by DNA MMR; however, when DNA MMR systems are defective, microsatellite replication errors accumulate in the genome.[13] Although MSI is commonly seen in sporadic colorectal cancers,[14] it is also associated with Lynch syndrome, which can produce inherited colon cancer and other non-colon cancers.[15]

The Applied Biosystems™ TrueMark™ MSI Assay from Thermo Fisher Scientific interrogates 13 mononucleotide MSI markers.[16] Eight of these markers are derived from the literature and guidelines from the National Cancer Institute,[17] and the other five markers were internally identified for monomorphism and high sensitivity in multiple cancer types. The assay also contains two sample identification markers to determine sample mix-ups or contamination. To facilitate analysis of MSI data, we developed the Applied Biosystems™ TrueMark™ MSI Analysis Software. This is a desktop-based application that takes the fragment analysis files of the TrueMark MSI Assay and provides easy-to-interpret analysis of the results. A groundbreaking feature of the software is a proprietary algorithm that can analyze samples without the need to run normal sample controls concurrently, saving time and expense.

Neuroscience researchers have long relied on fragment analysis for understanding a class of inherited neurodegenerative diseases known as triplet repeat diseases. These diseases are linked to expansions of microsatellite repeats found within translated or untranslated regions of certain genes.[18] The pathology often arises when the repeat sequence, which in normal alleles can range from 15 to 40 repeats, expands to exceed a disease-specific threshold, usually greater than 45 repeats. The disease severity may worsen from generation to generation due to *de novo* germline expansion of repeats. It is therefore critical to monitor the repeat length of these microsatellite sequences, especially in individuals who are at risk. Fragment analysis by capillary electrophoresis has been recommended by the European Molecular Genetics Quality Network (EMQN) as the default analytical method for analyzing triplet repeat length in spinocerebellar ataxias (SCAs)[19] and Huntington's disease.[20] Kits that use capillary electrophoresis to analyze repeat lengths in Fragile X, Huntington's, and other neurodegenerative diseases are commercially available.[21]

Copy number variability

Multiplex Ligation-Dependent Probe Amplification™ (MLPA™) technology, developed and commercialized by MRC Holland BV, is a flexible technique (www.mlpa.com) that is commonly used to detect aberrations in gene copy number, such as loss of heterozygosity. It is based on the ligation and PCR amplification of up to 50 multiplexed pairs of probe oligonucleotides, which hybridize to the loci of interest. Each oligonucleotide pair is designed to give an amplification product of a specific length; by using sequence-tagged ends, all ligated probes can be amplified with a single primer pair in a PCR reaction. The forward PCR primer carries a fluorescent label, allowing for the detection and quantification of size-separated probes on an automated capillary electrophoresis system.[22] MLPA technology has been used to understand the role of CNV regions in inherited cancers[23] and other inherited diseases.[24]

Epigenetic modification

Changes in DNA methylation patterns can be important in syndromes that are inherited.[25] Genome-wide hypomethylation or hypermethylation, particularly of cytosine 5´ to guanosine (CpG) within the promoter regions of tumor repressor genes, has been shown to be associated with changes in gene expression that could lead to Rett syndrome.[26] The SNaPshot assay can also be used to investigate DNA methylation patterns following bisulfite treatment of genomic DNA. Sodium bisulfite converts unmethylated cytosines into uracils, which are subsequently replaced as thymines in PCR; in contrast, methylcytosines remain unchanged. This feature can be used to distinguish base differences by PCR amplification and T/C genotyping using SNaPshot single-base extension (SBE) followed by fragment analysis.

MLPA technology can also be used to detect epigenetic modifications involving DNA methylation.[27] The methylation specific-MLPA (MS-MLPA) assay has probes that bind and ligate over a GCGC sequence, which is also a cleavage site for a methylation-sensitive restriction enzyme, HhaI. The enzyme cleaves the probes that ligate and hybridize to unmethylated DNA, while the probes bound to methylated DNA remain intact and are subsequently amplified by PCR. An MS-MLPA assay is described as a sensitive test for detecting methylation changes in genetic imprinted diseases, such as Angelman and Prader-Willi syndromes.[28]

Cell line authentication and human sample matching

Human cell lines are heavily relied upon for biomedical research. To ensure the reproducibility of scientific research, cell line authentication (CLA) is of paramount importance to confirm the cells' origin as well as check for contamination and genomic instability. CLA is performed by generating a profile of highly variable short tandem repeat (STR) markers from microsatellite loci with a varying number of repeats for a particular cell line/type and then comparing it against the allelic profiles present at these loci of known standards. A study aimed at authenticating 278 human tumor cell lines used in China found that nearly 46% of the samples were either cross-contaminated or misidentified.[29] Such findings can have massive implications for studies that utilize such cell lines. The Applied Biosystems™ CLA GlobalFiler™ kit generates a molecular fingerprint for 24 different STR loci, while the Applied Biosystems™ CLA IdentiFiler™ Plus kit analyzes 16 STR loci.

Sanger Sequencing Methods and Tools

Sanger sequencing is the trusted standard for obtaining DNA sequence information. It powered the Human Genome Project, and investigators continue to rely on this method to generate highly accurate and reliable sequencing results. Sanger sequencing is a specialized form of fragment analysis: it relies on chain-terminating fluorescent nucleotides to generate a series of fragments that differ by one nucleotide (Figure 4). The fast and straightforward Applied Biosystems™ Sanger sequencing workflows enable a high degree of accuracy, long-read capabilities, and simple data analysis. Applied Biosystems™ BigDye™ Terminator v1.1 and v3.1 cycle sequencing chemistries are the gold standard for Sanger sequencing by capillary electrophoresis. After cycle sequencing, various options exist for cleanup before electrophoresis, including Applied Biosystems™ ExoSAP-IT™ enzyme mix and Applied Biosystems™ BigDye XTerminator™ kits. An entire sequencing workflow can be completed in less than one workday, from sample to answer, providing the flexibility to support a diverse range of applications in many research areas.

The simplicity of Sanger sequencing facilitates the analysis of methylated DNA sequences (for example, see[30]). Primers that can be used for sequencing can be designed using the freely available Applied Biosystems™ Methyl Primer Express™ Software.[31] After bisulfite conversion, the DNA can be sequenced using the standard Sanger sequencing workflow. Data analysis is also straightforward since the converted sequence can be analyzed directly without a need for comparison to a specialized reference genome.

Discovery-based genomic research, such as NGS, often uncovers novel or unexpected variants or other sequence anomalies. Investigators look for ways to verify these new discoveries using orthogonal methods. Sanger sequencing is the method of choice for confirming NGS results because of its workflow simplicity and accuracy. For these confirmatory studies, short amplicons, usually covering only the region to be confirmed, need to be sequenced. Moreover, minor allelic variants present in a heterogeneous sample can be identified and confirmed by Sanger sequencing. Applied Biosystems™ Minor Variant Finder Software is simple, easy-to-use desktop software designed for accurate detection and reporting of minor variants in Sanger sequencing traces, with a detection level of minor alleles as low as 5%. On a test set of 632,452 base positions, it exhibited a 5% limit of detection with 95.3% sensitivity and 99.83% specificity.[32] Minor Variant Finder Software can also readily align sequences with the human reference genome and VCF files from NGS experiments, providing a smooth workflow for NGS confirmation with annotations in the SNP database (dbSNP).

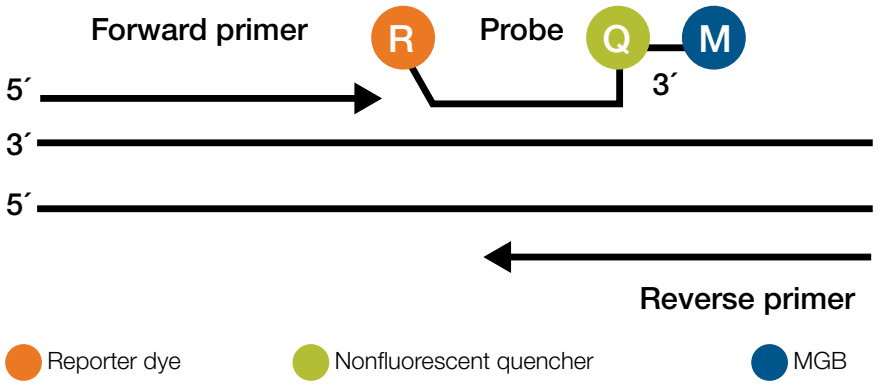


Figure 5. Anatomy of a TaqMan assay. Unlabeled forward and reverse oligonucleotide primers define the region that will be queried. A third oligonucleotide binds to the region between the primers. The probe has an end-labeled fluorophore reporter dye, a nonfluorescent quencher, and a minor groove binding (MGB) moiety that stabilizes the binding of the short probe sequence. When the probe is hybridized to its target sequence, the quencher greatly reduces the fluorescence of the reporter. When the sequence is copied during the PCR reaction, the nuclease activity of the polymerase cleaves the reporter from the rest of the probe, releasing it so it can be detected by the instrument. The amount of fluorescence is therefore directly related to the amount of target amplified.

PCR Methods and Tools

The TaqMan Assay family is among the most comprehensive set of real-time PCR (qPCR) products available for analyzing gene expression, miRNA levels, protein abundance, copy number variation, SNP genotyping, and rare-allele mutation detection. The typical

TaqMan Assay consists of forward and reverse primers to amplify a target by PCR (Figure 5). TaqMan Assays also contain a sequence-specific fluorescent probe, designed with a minor groove binder (MGB) moiety at the 3´ end that increases the melting temperature (T_m) of the probe and stabilizes probe-target hybrids. This means that TaqMan MGB probes can be significantly shorter than traditional probes, allowing better sequence discrimination and flexibility to accommodate more targets. In addition, TaqMan probes have a nonfluorescent quencher (NFQ) that binds to the fluorophore until cleaved by the polymerase and minimizes background. All TaqMan Assays come with a performance guarantee such that if they do not perform as promised, they can be replaced or refunded.[33]

One strategy that helps ensure high-quality PCR data is that investigators follow standardized practices for performing and reporting qPCR results. The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) recommendations are a set of guidelines for qPCR experimental design and data reporting practices, as well as standards for sharing experimental information with colleagues.[34] The standards are designed to ensure that published real-time PCR research is meaningful and accurate and provides researchers with the information necessary to faithfully reproduce results. The Applied Biosystems product portfolio supports these guidelines and provides all information necessary to enable MIQE compliance when publishing the results of experiments with TaqMan Assays. Information requested for the qPCR target, qPCR oligonucleotide, and qPCR protocol sections of the guidelines is readily available in the Applied Biosystems reagent protocol or from the product website.

The TaqMan portfolio is made up of over 20 million predesigned assays. These assays can be ordered as individual assays in tubes. Alternatively, Applied Biosystems™ TaqMan™ Arrays contain TaqMan Assays dried down in three array formats: 384-well TaqMan Array microfluidic cards, 96- and 384-well TaqMan Array plates, and Applied Biosystems™ TaqMan™ OpenArray™ formats. The TaqMan Assay collection includes genotyping and gene expression assays. In addition, the highly specific TaqMan Assay chemistry can also be used to analyze miRNAs for neural pathways, pathology, and biomarker research.

To help with choosing the right TaqMan Assays for an experiment, an assay search wizard was developed.[35] Entering a gene name, keyword, pathway, or disease can return a list of all the assays that meet the search criterion. The results page highlights the "best coverage" assay for each gene and publications that describe the use of the highlighted assay. Additionally, it provides an opportunity to select and design a combination of assays for an array format. Furthermore, predesigned arrays are available that have been prespotted with gene expression assays targeting common pathways, diseases, and gene families, including several important in neuroscience research.

dPCR Methods and Tools

Digital PCR (dPCR) is a method that quantifies sequences present in a sample by counting them. The basis of dPCR is that a nucleic acid sample is segregated into thousands of parallel PCR reactions, such that each reaction well contains one target molecule on average (Figure 6). In this scenario, some reactions will not contain the target molecule at all, while others will contain one or more copies. The collection of segregated reactions is subjected to endpoint PCR, and the number of wells containing a positive signal and no signal are tallied. The number of target copies can be calculated from the fraction of negative reactions, based on the assumption that the segregation follows a Poisson distribution (thus accounting for the possibility that multiple target molecules occupy the same reaction). The number of individual reactions influences the sensitivity of the assay—the more reactions there are, the lower the limit of detection and the higher the accuracy.

dPCR is often used to detect rare mutant alleles in cancer samples and to analyze circulating tumor DNA (ctDNA)

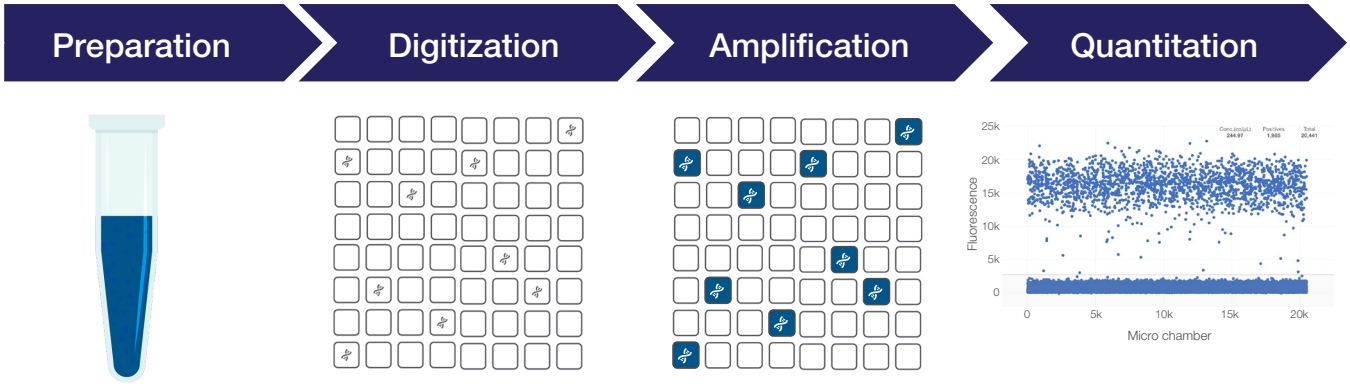


Figure 6. Basics of a digital PCR (dPCR) assay. Absolute Q dPCR reactions make use of TaqMan Assays to detect a target sequence. In dPCR assays, the sample is diluted and loaded into a matrix containing thousands of individual reaction chambers. If the dilution is correct, some of the chambers will have the target and some will not. Subjecting the chambers to PCR will cause a positive signal in the chambers where there is a target, but no signal where there is no target. The wells that are positive and negative are counted, and the starting concentration can be calculated based on dilution and other factors. Because the sample is loaded with one molecule per partition, an overabundance of one sequence will not necessarily overwhelm the detection of rare sequences in the wells.

in liquid biopsy research. Because the entire sample is segregated into individual wells, the detection of rare alleles is not masked by an overabundance of normal alleles. The sensitivity is dependent on the amount of DNA; to achieve a sensitivity of 0.1% (1 in 1,000 copies), 1,000 copies are needed. For diploid genomic DNA (gDNA), at least 6 ng of input gDNA is required for this sensitivity. Thus, the amount of recoverable DNA limits the sensitivity of the assay.

The Applied Biosystems™ QuantStudio™ Absolute Q™ Digital PCR System consists of just one instrument, consolidating all steps required into a single plate and transforming a multistep, multi-instrument workflow into a one-step qPCR-like workflow. Absolute Q Liquid Biopsy dPCR Assays detect and quantify the most common glioma-related mutations (e.g., *IDH1, TP53, EGFR, PTEN*[36]) as well as therapy-resistant mutations (e.g., *EGFR* T790M). They have been verified to detect allele frequencies as low as 0.1% and are guaranteed to perform on the QuantStudio Absolute Q dPCR System.[37] The Absolute Q Liquid Biopsy dPCR Assays run on the QuantStudio Absolute Q Digital PCR System provide a precise, cost-effective, and rapid method for monitoring response and resistance to treatment by testing for biomarkers that may indicate tumor-driver and therapy-resistant mutations.

© 2024 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. MLPA is a trademark of De Luwe Hoek Octrooien B.V. TaqMan is a trademark of Roche Molecular Systems, Inc., used under permission and license. TaqMan Assay qPCR Guarantee: Terms and conditions apply. For complete details, go to www.thermofisher.com/taqmanguarantee

See References on Page 13.

Article 1: Rare and Inherited Diseases: Uncommon Yet Widespread

1) Pogue RE, et al. Rare genetic diseases: Update on diagnosis, treatment and online resources. *Drug Discov Today*. 2018;23(1):187-195.
2) Divino V, et al. Pharmaceutical expenditure on drugs for rare diseases in Canada: A historical (2007–13) and prospective (2014–18) MIDAS sales data analysis. *Orphanet J Rare Dis*. 2016;11(1):68.
3) Ferreira CR. The burden of rare diseases. *Am J Med Genet A*. 2019;179(6):885-892.
4) Orphadata. Orphadata in numbers. Accessed January 31, 2024.
5) Braga LAM, et al. Future of genetic therapies for rare genetic diseases: What to expect for the next 15 years? *Ther Adv Rare Dis*. 2022;3:26330040221100840.
6) Kruse J, et al. Genetic testing for rare diseases: A systematic review of ethical aspects. *Front Genet*. 2022;12.
7) Vinkšel M, et al. Improving diagnostics of rare genetic diseases with NGS approaches. *J Community Genet*. 2021;12(2):247-256.
8) Manolio TA, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747-753.
9) Sun W, et al. Drug discovery and development for rare genetic disorders. *Am J Med Genet A*. 2017;173(9):2307-2322.
10) National Institutes of Health (NIH). Estimates of funding for various research, condition, and disease categories (RCDC). Research Portfolio Online Reporting Tools (RePORT). Table published on March 31, 2023. Accessed January 31, 2024.
11) Maidhof W, Hilas O. Lupus: An overview of the disease and management options. *Pharm Ther*. 2012;37(4):240-249.
12) Ong T, Ramsey BW. Cystic fibrosis: A review. *JAMA*. 2023;329(21):1859-1871.
13) Theadom A, et al. Prevalence of muscular dystrophies: A systematic literature review. *Neuroepidemiology*. 2014;43(3-4):259-268.
14) Ryder S, et al. The burden, epidemiology, costs and treatment for Duchenne muscular dystrophy: An evidence review. *Orphanet J Rare Dis*. 2017;12:79.
15) D'Egidio F, et al. Therapeutic advances in neural regeneration for Huntington's disease. *Neural Regen Res*. 2024;19(9):1991.

Infographic: Exploring Rare Diseases on a Global and Molecular Scale

1) Ferreira CR. The burden of rare diseases. *Am J Med Genet A*. 2019;179(6):885-892.
2) Tisdale A, et al. The IDeaS initiative: Pilot study to assess the impact of rare diseases on patients and healthcare systems. *Orphanet J Rare Dis*. 2021;16(1):429.
3) National Institutes of Health (NIH). Estimates of funding for various research, condition, and disease categories (RCDC). Research Portfolio Online Reporting Tools (RePORT). Table published on March 31, 2023. Accessed January 31, 2024.
4) Barber MRW, et al. Global epidemiology of systemic lupus erythematosus. *Nat Rev Rheumatol*. 2021;17(9):515-532.
5) Guo J, et al. Worldwide rates of diagnosis and effective treatment for cystic fibrosis. *J Cyst Fibros*. 2022;21(3):456-462.
6) Salari N, et al. Global prevalence of Duchenne and Becker muscular dystrophy: A systematic review and meta-analysis. *J Orthop Surg*. 2022;17(1):96.
7) Medina A, et al. Prevalence and incidence of Huntington's disease: An updated systematic review and meta-analysis. *Mov Disord*. 2022;37(12):2327-2335.

Article 2: Somatic CAG Repeats in Huntington's Disease Expand with Age

1) Kacher R, et al. Propensity for somatic expansion increases over the course of life in Huntington disease. *eLife*. 2021;10:e64674.
2) Kaplan S, et al. A universal mechanism ties genotype to phenotype in trinucleotide diseases. *PLOS Comput Biol*. 2007;3(11):e235.

Article 3: Smad8 Dysregulates microRNAs in Duchenne Muscular Dystrophy

1) Lopez MA, et al. Smad8 is increased in Duchenne Muscular Dystrophy and suppresses miR-1, miR-133a, and miR-133b. *Int J Mol Sci*. 2022;23(14):7515.
2) Si Y, et al. Smads as muscle biomarkers in amyotrophic lateral sclerosis. *Ann Clin Transl Neurol*. 2014;1(10):778-787.
3) Si Y, et al. Muscle microRNA signatures as biomarkers of disease progression in amyotrophic lateral sclerosis. *Neurobiol Dis*. 2018;114:85-94.

Article 4: Technology Overview

1) Grody WW, et al. ACMG position statement on prenatal/preconception expanded carrier screening. *Genet Med*. 2013;15(6):482-483.
2) Landrum MJ, et al. ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862-868.
3) Stenson PD, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*. 2003;21:577-581.
4) Zlotogora J, et al. The Israeli national population program of genetic carrier screening for reproductive purposes. *Genet Med*. 2015;18(2):203-206.
5) Langfelder-Schwind E, et al. Molecular testing for cystic fibrosis carrier status practice guidelines: Recommendations of the National Society of Genetic Counselors. *J Genet Couns*. 2014;23(1):5-15.
6) Sosnay PR, et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet*. 2013;45:1160-1167.
7) Wen J, et al. Detection of cytogenomic abnormalities by OncoScan microarray assay for products of conception from formalin-fixed paraffin-embedded and fresh fetal tissues. *Mol Cytogenet*. 2021;14(1):21.
8) Wang F, et al. MicroRNAs in β-thalassemia. *Am J Med Sci*. 2021;362(1):5-12.
9) Fortunato F and Ferlini A. Biomarkers in Duchenne muscular dystrophy: Current status and future directions. *J Neuromuscul Dis*. 2023;10(6):987-1002.
10) Doroszkiewicz, et al. Molecular biomarkers and their implications in the early diagnosis of selected neurological diseases. *Int J Mol Sci*. 2022;23(9):4610.
11) Applied Biosystems. Clariom D microarrays provide a deep view of our transcriptome. Thermo Fisher Scientific Technical Note. 2019.
12) Applied Biosystems. Using the SNaPshot® Multiplex System with the POP-7™ Polymer on Applied Biosystems 3730/3730xl DNA Analyzers and 3130/3130xl Genetic Analyzers. Applied Biosystems User Bulletin. 2005.
13) Reyes GX, et al. New insights into the mechanism of DNA mismatch repair. *Chromosoma*. 2015;124:443-462.
14) Kang S, et al. The significance of microsatellite instability in colorectal cancer after controlling for clinicopathological factors. *Medicine (Baltimore)*. 2018;97(9):e0019.
15) Biller LH, et al. Lynch syndrome-associated cancers beyond colorectal cancer. *Gastrointest Endosc Clin N Am*. 2022;32(1):75-93.
16) Thermo Fisher Scientific. TrueMark MSI Assay for microsatellite instability analysis. 2024.
17) Boland CR, et al. A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: Development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res*. 1998;58(22):5248-5257.
18) Paulson H. Repeat expansion diseases. *Handb Clin Neurol*. 2018;147:105-123.
19) Sequieros J, et al. EMQN best practice guidelines for molecular genetic testing of SCAs. *Eur J Hum Genet*. 2010;18:1173-1176.
20) Losekoot M, et al. EMQN/CMGS best

practice guidelines for the molecular genetic testing of Huntington disease. *European Journal of Human Genetics*. 2013;21:480-486.
21) Asuragen. Kitted Assays for Inherited Genetic Disorders. 2024.
22) Applied Biosystems. MLPA assays on the SeqStudio Genetic Analyzer. Thermo Fisher Scientific Application Note. 2019.
23) Leite Rocha D, et al. Reviewing the occurrence of large genomic rearrangements in patients with inherited cancer predisposing syndromes: Importance of a comprehensive molecular diagnosis. *Expert Rev Mol Diagn*. 2022;22(3):319-346.
24) Ishida C, et al. Molecular genetics testing. In: *StatPearls*. StatPearls Publishing; 2023.
25) Cerrato F, et al. DNA Methylation in the diagnosis of monogenic diseases. *Genes (Basel)*. 2020;26:11(4):355.
26) Boxer LD, et al. MeCP2 Represses the rate of transcriptional initiation of highly methylated long genes. *Mol Cell*. 2020;77(2):294-309.e9.
27) Nygren AO, et al. Methylation-specific MLPA (MS-MLPA): Simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res*. 2005;33(14):e128.
28) Ma VK, et al. Prader-Willi and Angelman syndromes: Mechanisms and management. *Appl Clin Genet*. 2023;16:41-52.
29) Huang Y, et al. Investigation of cross-contamination and misidentification of 278 widely used tumor cell lines. *PLoS One*. 2017;12(1):e0170384.
30) Salcedo-Tacuma D, et al. Differential methylation levels in CpGs of the BIN1 gene in individuals with Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2019;33(4):321-326.
31) Thermo Fisher Scientific. Applied Biosystems™ Methyl Primer Express™ Software v1.0. 2024
32) Applied Biosystems. Low-level somatic variant detection in tumor FFPE samples by Sanger sequencing. Thermo Fisher Scientific Application Note. 2016.
33) Thermo Fisher Scientific. TaqMan real-time PCR assays. 2024
34) Bustin SA, et al. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. 2009;55(4):611-22.
35) Thermo Fisher Scientific. TaqMan assay search wizard. 2024.
36) Liu A, et al. Genetics and epigenetics of glioblastoma: Applications and overall incidence of *IDH1* mutation. *Front Oncol*. 2016;6:16.
37) Thermo Fisher Scientific. Applied Biosystems™ TaqMan™ Liquid Biopsy dPCR Assay. 2024

For Research Use Only.
Not for use in diagnostic procedures.