Sapienza University of Rome
Department of Experimental Medicine

Ph.D. program
"Human Biology and Medical Genetics"
XXXI cycle

Intragenic modulators of CFTR gene expression involved in Cystic Fibrosis clinical variability

Candidate: Dr. Manuela Sterrantino

Scientific Tutor:
Prof. Marco Lucarelli (Sapienza University of Rome)

Director of Doctoral Program:
Prof. Antonio Pizzuti (Sapienza University of Rome)

Board of Examiners:
Prof. Federica Sangiuolo (Tor Vergata University, Rome)
Prof. Anna La Teana (Marche Polytechnic University, Ancona)
Prof. Annalisa Botta (Tor Vergata University, Rome)

Academic year 2017-2018

INDEX

ABSTRACT

Alterations in Cystic Fibrosis Transmembrane conductance Regulator (CFTR) gene cause Cystic Fibrosis (CF), a disease whose complications lead to death due to respiratory failure. CFTR produces a primary transcript of 6132 bases, with a complex transcriptional regulation that depends on the entire locus of the gene. The correlation between genotype and phenotype in this pathology is still difficult to understand.

We evaluated the possible mechanisms involved in the intragenic modulation of CFTR expression in CF and CF-like populations, already extensively characterized at the mutational level. Our study is based on three experimental approaches, applied to the CFTR gene of these populations: analysis of intragenic haplotypes; analysis of variant tracts (TG)mTn and effects on mRNA splicing; analysis of 3'-UTR for potential miRNAs recognition sites. We performed the structural analysis on DNA (by primer extension and Sanger sequencing) and the functional analysis on RNA (by reverse transcriptase PCR, densitometric analysis and Real-time PCR).

Among the haplotypes characterized, one showed a significantly higher frequency in the populations of interest

(CF and CF-like) compared to control populations (general population and normospermic men), with a trend that varies according to the severity of the phenotypic manifestations. It has been highlighted that no anomalous splicing is produced by the presence of the individual sequence variations of the haplotype. Future studies will evaluate whether the simultaneous presence in *cis* of all the variations that compose the haplotype may influence the levels of CFTR mRNA. Some variant tracts (TG)mTn resulted more frequent in CF-like populations than in classical CF or in controls. Functional analysis at RNA level evidenced a higher percentage of anomalous splicing produced by some specific tracts. This correlates with the clinical phenotype of these patients. We evidenced some variants within the CFTR 3'-UTR. We did not highlight 3'-UTR sequence variations that could influence the binding of miRNAs currently known in the literature. Future studies will evaluate what effect the variations we found in this area can have.

INTRODUCTION

Cystic Fibrosis (CF), also called mucoviscidosis or fibrocystic disease of the pancreas, is the most common lethal genetic disease of the Caucasian population, with a very variable prevalence, from 1/1350 to 1/25000 depending on the geographical region, and with a carrier frequency between 1/20 and 1/80 *[Farrell PM, 2008]*. In the report of the Italian CF Registry (RIFC 2014) there is an incidence of 1 CF affected on 4052 born alive. On this basis, the presumed number of carriers is calculable equal to about 1/32 in the Italian general population. These values are also contained in the range reported in studies on selected populations and of a more limited number (incidence: from 1/4356 to 1/2438; carrier frequency: from 1/33 to 1/25). Taking into consideration a probable underestimation of RIFC data, and also a principle of caution, the prevalence value of the carrier, that would seem representative of the Italian population, is 1/30.

It is defined as a "multi-systemic disease" because it involves numerous organs and systems, including airways, pancreas, liver, small intestine and male reproductive system. The hallmarks of CF are salty sweat, poor growth and poor weight

gain (despite a normal food intake), pancreatic insufficiency, steatorrhea, malnutrition, cirrhosis of the liver, accumulation of thick and sticky mucus at the pulmonary level, frequent infections and persistent cough or shortness of breath. Near all affected males are not fertile due to the bilateral congenital absence of the vas deferens (CBAVD). Symptoms often appear during childhood, such as intestinal obstruction due to pathologic meconium ileus in newborns. On the basis of the symptoms, from the diagnostic point of view, differentiated clinical forms can be distinguished: on one side there are the severe polysymptomatic forms (the so-called "typical" forms which in this thesis will be called Cystic Fibrosis - CF), on the other side the lighter oligosymomatic and monosymptomatic forms ("atypical" forms; CFTR-pathies; CFTR related diseases or CFTR-RD. In this thesis will be called CF-like forms). Our group had also evidence that the idiopathic hyperviscosity of the seminal fluid (ISHV) may be, in some cases, a CF-like form *[Rossi et al., 2004; Elia et al., 2009]*.

The disease is monogenic with mendelian inheritance and it is caused by mutations in the CFTR (Cystic Fibrosis Transmembrane conductance Regulator) gene. This gene encodes a protein of 1480 amino acids, which performs the main function of cAMP-dependent chloride channel and

regulator of transmembrane ion exchange of epithelial cells. The functionality of the protein influences, therefore, the composition in water and electrolytes in glandular secretions. An alteration at CFTR level leads to the secretion of very thick, viscous and sluggish mucus, with consequent obstruction of the main ducts. 2031 sequence variations (database CFTR1, December 2018) were identified in the responsible gene, of which about 374 were functionally characterized and recognized as: CF-causing (312), variants of varying clinical consequence (36), non CF-causing (13), variants of unknown significance (13) (database CFTR2, December 2018). It is an autosomal recessive disease, which means that the full manifestation of the pathology occurs only in homozygous individuals (same mutation on both alleles) and in compound heterozygote ones (two different mutations on alleles). Eliminating the single dosage of the gene is not enough to show any manifestation of the disease, but it is necessary that the residual functionality of the protein decreases below 50% to originate the symptomatology. Heterozygotes with only one mutation on an allele are carriers and do not have any clinical symptoms. Laboratory identification of these carriers results difficult because the biochemical tests are not enough sensitive. The only way to identify carriers is the genetic investigation,

which, however, shows a variable ability to detect CFTR mutations depending on the number and type of mutations included in the test. To date there is no decisive cure, but the use of some therapeutic approaches is useful to ensure a high standard of living to patients. In particular, the prognosis for CF has improved considerably: the integrated treatment programs have allowed, in fact, to increase the median survival of patients from the 25 years of the 80s to the 40 years of today *[Wilschanski, 2013]*, and are supported both by the development of new diagnostic methods, to identify precociously the disease, and by the institution of specialized centers of reference, able to guarantee an adequate assistance to the patient. The research points to further improve the knowledge of this pathology and to find new therapies for a more effective treatment. They may be, in the near future, the solution to the inconveniences that CF causes.

AIM OF THE THESIS

The main problems in the comprehension of CFTR-dependent mechanisms in CF are:

- the great variety of the clinical forms of the disease, although it is monogenic;
- the recent identification of a complex set of new related pathologies (CF-like forms);
- the still uncertain genotype-phenotype correlation in this pathology;
- an incomplete genetic characterization, often for strictly technical reasons related to the mutational search methods;
- the limited knowledge of the mechanisms of transcriptional regulation of the gene, which may affect the protein levels and, consequently, the severity of the pathology.

For these reasons, we conceived a research project that, in addition to the classical mutations of CFTR, takes under consideration some players of CFTR transcriptional control, possibly modulating CFTR mRNA levels. This may help to better understand CFTR-dependent mechanisms originating and modulating clinical severity of CF and also to provide new

insights about the control of CFTR expression. In particular, we evaluated the possible mechanisms involved in the transcriptional modulation through a study based on three experimental approaches: 1) analysis of intragenic haplotypes in CF and CF-like populations (CFTR-RD, CBAVD and ISHV); 2) analysis of variant tracts (TG)mTn and effects on the splicing of mRNA; 3) analysis of CFTR 3'-UTR for potential variants in miRNAs recognition sites. We performed the structural analysis on DNA by sequencing and primer extension and the functional analysis on RNA by reverse transcriptase PCR, densitometric analysis and Real-time PCR. The results obtained with these experiments show that there are some intragenic actors actively involved in the process of regulation of messenger RNA levels and, consequently, able to influence the quantity of functional protein that reaches the cellular membrane. In turn, this may modulate the final phenotype of each patient.

A better knowledge of these aspects is very useful not only to understand the transcriptional activity of CFTR and the contribution of intragenic factors to the modulation of gene expression, but also to obtain an optimal definition of CF and CF-like clinical categories. In addition, an improvement of the understanding of the genotype-phenotype relationship in this

disease, an optimization of current screening programs and a development of more focused treatment strategies (customized for the individual patient's better-defined genotype) are expected.

1. CFTR AND CYSTIC FIBROSIS

1.1 CFTR: gene, protein and mutations

The CFTR is a multifunctional membrane protein that in vertebrates is encoded by the homonymous gene. This gene was first cloned in 1989 *[Kerem et al., 2006]* by Lap-Chee Tsui and his team of researchers in Toronto and it is a member of the superfamily of membrane transporters or ABC carriers ("nucleotide adenine-binding cassette"). It is located on chromosome 7 (position 7q31.2) *[Riordan JR et al., 1989]* and consists of 27 exons (about 250 Kb). The most common transcript is 6128 bases long and is translated into a protein of 1480 amino acids (**Figure 1**). The CFTR is under the control of a housekeeping promoter with a regulated time-specific and tissue-specific expression, established by alternative start transcription sites and/or by alternative splicing. The gene is highly expressed in various epithelial tissues of organs such as lung, pancreas, liver, intestine, testis, sweat glands and salivary glands.

The protein is an ATP- and cAMP- dependent chloride channel that performs its major function on the apical membrane of

epithelial cells: this function is the secretion of Cl- ions in the colon and in the airways or its reabsorption in the sweat glands. The transport direction depends, in fact, on the epithelial tissue involved: in the respiratory epithelium and in the intestinal epithelium the flow goes from the inside towards the outside of the cells, while in the epithelium of the sweat glands it happens from the lumen of the gland to the inside of the cell. The product of the gene is, therefore, an important channel in the production of sweat, of digestive juices and of airway surface fluid.

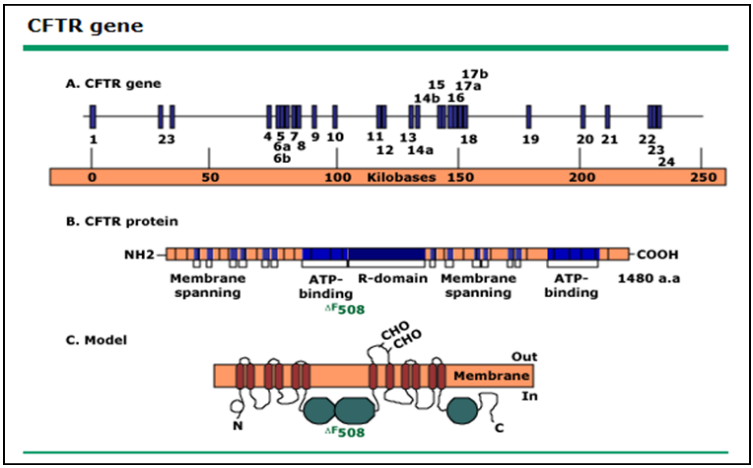


Figure 1. CFTR gene. The figure shows the structure of the CFTR gene with the 27 exons (A), the polypeptide with the various domains (B) and the protein model (C). *[Human Genetics: Concepts and Applications by Ricki Lewis (1994)]*

The structure of this glycoprotein is characterized by five distinct domains (**Figure 2**): two homologous MSD domains, that physically constitute the channel (each consisting of six transmembrane segments of hydrophobic α-helices); two nucleotide binding domains (NBD) for ATP, located in the cytoplasm; a central regulating domain (R-domain), also cytoplasmic, which joins the two structural units of the protein and which contains various phosphorylation sites *[Morales et al., 1999]*. The ion channel opens only when the regulatory domain is in the phosphorylated state (phosphorylation occurs by PKA, ie a cAMP-dependent kinase) *[Morales et al., 1999]* and the ATP is bound to the NBD domains. The terminal -NH2 portion interacts with the 1A syntaxin, whereas the terminal -COOH is involved in the interaction with the proteins containing the PDZ domain.

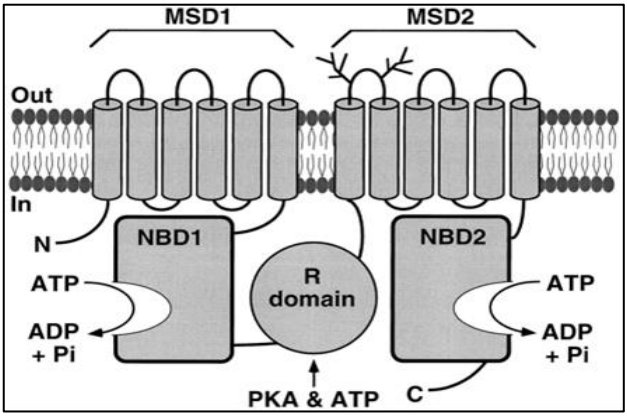


Figure 2. Structure of CFTR with its domains and the fundamental reactions to the activity of the channel. *[Sheppard et al, 1999]*

However, CFTR does not only perform the chloride ion channel function. Below there are the additional features of this protein:

a. regulation of the function of other ion channels, such as the EnaC (the epithelial sodium channel), the ORCC (Outwardly Rectifying Chloride Channel), the CaCC (calcium-dependent chloride channel), the ROMK (renal potassium channel), the sodium/proton exchanger NHE3 and the aquaporin channel *[Lubamba B et al., 2012]*. These regulations allow, directly or indirectly, the transport of sodium, potassium, water, glutathione and bicarbonate;

b. pH regulation of cytoplasmic organelles (Golgi reticulum, phagosomes, endosomes and prelisosomes). In CF cells, in fact, an alkalization of these compartments has been observed and the glycoproteins produced show an altered glycosylation pattern. It has been hypothesized that these alterations may, at least in part, contribute to the increased colonization of the airways by pathogenic bacteria *[Vankeerberghen et al., 2002]*;
c. receptor for *Pseudomonas aeruginosa* and its consequent endocytosis *[Goldberg et al., 2000; Pier, 2000]*;
d. regulation of innate immunity *[Pier, 2000]*;
e. regulation of inflammatory processes, by action on chemokines and neutrophils *[Vankeerberghen A et al., 2002]*;
f. intercellular communication (gap junctions) and activation of enzymes *[Goldberg et al., 2000]*.

The interaction of CFTR with ENaC is the most characterized. ENaC is located on the apical surface of the exocrine epithelial cells and it is responsible for the sodium intracytoplasmic transport from luminal secretions. The regulation of the transport of Cl^- and Na^+ influences the homeostasis of the superficial liquid of the airways. It is formed by mucus and

periciliary fluid and performs various tasks: it guarantees clearance of the airways; it contains numerous antibacterial and anti-toxic enzymes (neutrophil proteases and protease inhibitors), lysozyme (disorganizing of peptidoglycans of the bacterial wall of GRAM+), lactoferrin and transferrin (antimicrobial proteins with bacteriostatic and bactericidal action), immunoglobulins (especially secretory IgA) and cytokines (IL-1, IL-8 and TNF); it protects the underlying epithelium from dehydration.

The pathway of production of CFTR protein starts at the nucleus level, where the pre-mRNA is transcribed and undergoes the process of splicing of the introns; from here, it passes into the cytoplasmic compartment: at the endoplasmic reticulum level the mRNA is translated into protein which undergoes processing, then in the Golgi apparatus it undergoes further modifications and it is glycosylated at the level of the transmembrane domains. Finally, the mature protein reaches its definitive site on the apical membrane of epithelial cells.

Mutations in CFTR gene are responsible for CF, since they cause an anomalous or absent production and/or function of the protein. The most common mutation, F508del (or ΔF508 in old notation), is a deletion of three nucleotides located in exon 10, which results in the loss of a phenylalanine in the position 508

of the protein. This mutation is found in two thirds (66-70%) of all worldwide patients of CF [*Bobadilla et al., 2002*]. CF develops when there is no allele capable of producing a functional CFTR protein and, therefore, it is considered an autosomal recessive disease. Mutations and polymorphisms are distributed throughout the entire coding sequence of the gene, in the promoter regions, in the portions of exon-intron junction and, in a minor part, in the introns. Modifications at the exonic level have consequences of variable impact on the amino acid sequence, while those at the intronic level may cause changes in the splicing process as consequence, for example, of the elimination of a splicing site or the creation of alternative splicing sites (and, therefore, of new exons).

CFTR mutations were grouped into 6 mutational classes *[Estvill et al., 1996; Fanen et al., 2014; Amaral, 2015]*, depending on their effects on transcription, cellular processing, final localization and quantitative level of functional protein (**Figure 3**).

- ✓ Class I. These are production defects that determine the total lack of protein synthesis. These mutations (for example G542X, 3950delT, R553X, W1282X) are usually nonsense mutations, severe splicing mutations (which produce only

aberrant mRNAs), small or large deletions or insertions. They act by generating premature inframe or frameshift stop signals, unstable transcripts and/or aberrant proteins, destined to have no functionality or to be rapidly degraded. This degradation is mediated, at RNA level, by a control mechanism known as NMD (Nonsense-Mediated Decay). The class I mutations cause, therefore, the loss of the chloride channel activity in the affected epithelia.

- ✓ Class II. These are the defects of maturation of the protein and of decreased stability and they determine an altered tertiary structure. They are, therefore, alterations of trafficking (membrane transport), based on the recognition of the anomaly by the cell systems, ubiquitination and increased degradation of the misfolded protein within the endoplasmic reticulum. These defects of processing/maturation severely decrease the amount of protein in the apical cell membrane, although often with a tissue-specificity. They can be missense mutations (N1303K) or deletions (F508del). The class II mutations cause the production of an abnormal protein, which is rapidly degraded and does not reach the membrane, or reaches it only in part often with a reduced functionality. The clinical phenotype will usually be severe.

- ✓ Class III. These are regulation defects that lead to an impairment in channel opening and in chloride transport. Although the CFTR protein is complete and mature and able to reach the apical membrane, it is not properly activated/inactivated by ATP or cAMP. The mutations of this class interfere, in fact, with the phosphorylation of the R domain, with the binding and hydrolysis of the ATP at the level of both the NBD domains and with the channel stimulation by the ATP. These are missense mutations (R553G and G551D), whose final effect is a decrease in functional protein activity, even though the membrane quantity is normal.
- ✓ Class IV. These are the channel defects that cause a reduction in the transport of chloride through the CFTR. In this case, the protein is present on the apical membrane, but it is not able to adequately support the ionic flow due to a defect in the physical structure of the channel. This low conductance of chloride ions is caused by missense mutations (for example R117H, R334W, R234P) involving the domains that pass through the cell membrane and form the pore. This class of mutations is mostly associated with a moderate phenotype (alleles presenting such mutations are usually associated with mild pancreatic insufficiencies).

- ✓ Class V. These are defects of reduced synthesis or slowed maturation. The lower production of protein is due to mutations on the promoter or that cause splice defects, but, unlike the class I mutations, the alterations belonging to the fifth group do not completely abolish the correctly folded form, since only a reduction of wild-type mRNA is realized. In fact, since a small percentage of the transcripts may be normal and that, therefore, a small amount of functioning CFTR channels can be generated, the resulting phenotype will be mild or even mono-symptomatic (as, for example, in CF-like forms of CF). An example of variants able to influence the splicing of the CFTR pre-mRNA is the repetition (TG)mTn, located in intron 8, at the splicing acceptor site of exon 9 of the CFTR gene *[Chu et al., 1993; Chillon et al., 1995; Dork et al., 1997]*. In this tract there are repeats of the TG dinucleotide followed by a variable number of the nucleotide T: in general, it has been seen that when there are a high number of TG repeats coupled with a low number of T, there is a greater exclusion of exon 9 from the messenger RNA of CFTR *[Chu et al., 1993; Teng et al., 1997]*.
- ✓ Class VI. These are the defects that cause a destabilization of CFTR or that affect the regulation of other channels.

These alterations affect the terminal domains of the protein, especially the last 70-98 amino acid residues of the C-terminal region *[Haardt M et al., 1999]*, and this influences the relationships with other proteins and their regulation. The mutations are nonsense or frameshift (for example 4326delTC, 4279insA, 4271delC). Despite the biosynthesis, the processing and the capacity to transport the chloride by the truncated protein are essentially normal or only partially compromised, various anomalies in the regulation of many other proteins have been highlighted, for example of the epithelial sodium channel EnaC, due to lower metabolic stability of CFTR.

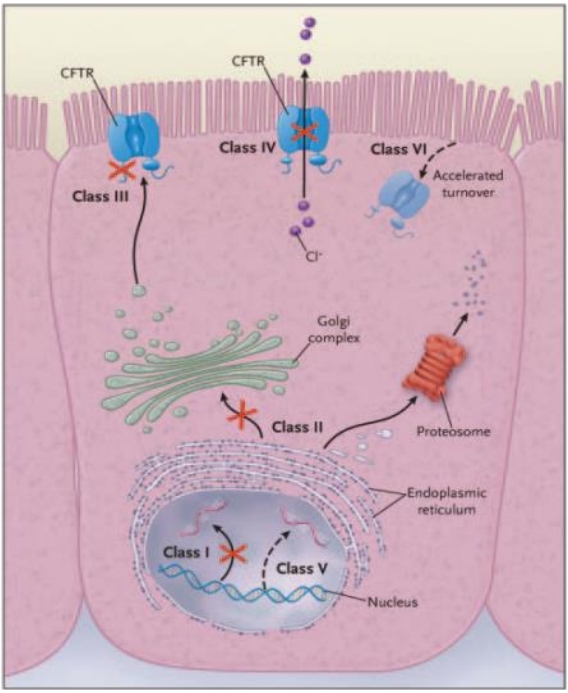


Figure 3. Mutational classes of CFTR. *[Ratjen et al. 2009]*

Only a partial correlation between variations in the gene sequence (genotype) and clinical manifestations of the disease (phenotype) exists. This means that the great variability and severity of the disease are related to multiple mutations in the CFTR gene and even more to their possible combinations. The determination of the relationship between genotype and phenotype is important, but still difficult to establish. Furthermore, the presence of intragenic modifiers, as specific

haplotypes, and of complex alleles, that consist of two or more mutations in *cis* position, contribute to the lack of clarity of the genotype-phenotype relationship. Additional variability is introduced by the intervention of modifying genes, which can influence the severity of the CF phenotype through, for example, an alternative chloride conduction, the regulation of splicing, the modulation of CFTR gene expression and the modulation of susceptibility to infections/inflammatory responses *[Slieker et al., 2005]*. Penetrance is usually 100% in homozygotes or compound heterozygotes with severe mutations, but the gravity of the disease is variable, with even mild and clinically negligible clinical forms, without pancreatic insufficiency or with moderate respiratory failure. It should be also taken into account that phenotypic variability is also determined by numerous other factors besides those listed, such as random phenomena and environmental factors *[Salvatore et al., 2011]*.

From the molecular point of view, this broad spectrum of clinical manifestations can be related to the amount of residual functional protein at the cell membrane level *[Chillon et al., 1995; Zhang et al., 2009]*: there is an inverse correlation between the amount of residual functional protein and the severity of the clinical phenotype. In general, 100% of channel

activity is present in an individual with no mutation, as both alleles encode for a functional protein. In an individual carrying a severe mutation, which completely deactivates an allele, only 50% of functional protein will be present on the membrane: this residual function is sufficient to ensure an overall normal CFTR activity. The first clinical symptoms of the disease are shown when the percentage of residual function of the channel is less than about 10-15% *[Zhang et al., 2009]*, because at this level it is possible that a mild CF-like disease, the CBAVD, takes place. If the amount of residual functionality is less than 5%, in addition to CBAVD, there will be a further reduced chloride transport, with an amount of NaCl in the sweat higher than normal, then a positive and pathological sweat test will occur. Coming down to 4.5%, lung infections will begin to appear; around 1% of residual function, pancreatic insufficiency will also happen. At this level the complete and most severe form of CF will be manifested, with pancreatic insufficiency, pulmonary involvement, positive sweat test and male infertility for CBAVD.

Some mutations may cause the impairment of different aspects of CFTR function and many epidemiological studies show that different populations have different mutations *[Bobadilla et al, 2002]*. The disease is more common, in fact, among the nations

of the Western world, but the distribution of CF alleles varies among individuals. The frequency of F508del carriers, for example, was estimated at 1:200 in the North of Sweden, in 1:143 in Lithuania and 1:38 in Denmark. Because of this variability, it is very important to pay attention to the geographical origin of the individuals analyzed at the time of diagnosis, in order to choose the best panel of mutations for each case. The F508del affects about 10 million people in Europe *[Bobadilla et al., 2002]*. Some other mutations may have a frequency between 2% and 5% or are characteristic of some ethnic groups, such as W1282X in Jews from Central Europe and 3659delC in Swedes. In Italy the characteristic mutations are the T338I in the Sardinians, the 2183AA>G and R1162X in Northern Italy and a group of five specific mutations in Southern Italy.

1.2 Physiopathology of Cystic Fibrosis

The symptomatology of CF is certainly complex and variable and depends on the percentage of residual function of the CFTR channel and on the consequent involvement of various

organs capable of secretion, such as lung, salivary and sweat glands, pancreas, liver and biliary tract, deferent ducts *[Knowles MR et al., 2002]*. It is an extremely heterogeneous disease at the phenotypic level for what concern the age of onset, the clinical course and the impairment of the different organs. Moreover, some elements that influence the production or the process of the protein and, therefore, the clinical manifestation in each individual, such as the presence of various environmental factors and the interaction of CFTR with the so-called "modifier" genes, should be emphasized. Often, the symptoms are already evident in the first weeks or months of life, while in other cases they occur later, in adolescence or even adult age.

The organs mostly involved are the lung (more than 90% of cases) and the pancreas (85% of cases). As consequences of the first aspect, airway obstruction leads to continuous bacterial infections, tissue damage and inflammation, up to respiratory failure and death; relatively to the second aspect, malabsorption occurs and this can lead to delays in growth in younger patients.

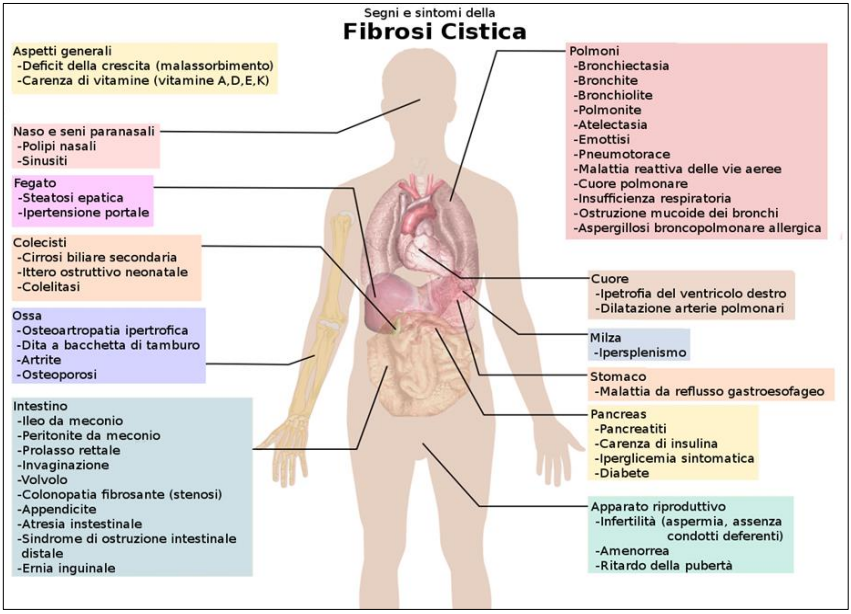


Figure 4. Clinical manifestations of Cystic Fibrosis. *[Kliegman, Robert; Richard M Kliegman (2006) Nelson essentials of pediatrics, St. Louis, Mo: Elsevier Saunders]*

The main manifestations of CF concern:

- Lungs and paranasal sinuses. Pulmonary disease is the consequence of airway obstruction, caused by mucus accumulation, reduction of mucociliary clearance and chronic inflammation. Inflammation and infection cause injury and structural changes in the lungs, leading to a variety of symptoms: in the initial stages, incessant cough,

abundant production of sputum and reduced lung capacity are common conditions (many of these symptoms occur when bacteria, which normally inhabit dense sputum, grow out of control and cause pneumonia); in the following phases, the pathologies of the main airways (bronchiectasis) further aggravate the difficulties in breathing. The respiratory system presents evident histological alterations, such as hyperplasia and hypertrophy, at the level of the tracheobronchial glands, and this inevitably leads to bronchial obstruction. Other symptoms include: cough with blood (hemoptysis), high blood pressure in the lungs (pulmonary hypertension), heart failure, difficulty obtaining oxygen (hypoxia) and respiratory failure. *Staphylococcus aureus*, *Haemophilus influenzae*, *Pseudomonas aeruginosa*, *Burkholderia cepacia*, *Candida albicans* and *Aspergillus flavia* are the most common organisms that cause severe and frequent pulmonary infections in patients with CF, because they find a fertile place to develop and grow in the wet and rich of mucus environment of the sick lungs. In addition, this thick and viscous mucus is not easily eliminated from the airways and stagnates, creating a beneficial situation for opportunistic bacteria, which are not eradicated despite a strong inflammatory reaction. This inflammatory process

then leads, with the passage of time, to the progressive destruction of the lung. In addition to typical bacterial infections, patients usually develop other types of lung diseases: among these, the allergic bronchopulmonary aspergillosis, in which the response to the common fungus *Aspergillus fumigatus* causes a worsening of respiratory problems, or the infection with *Mycobacterium avium complex* (MAC), a group of bacteria related to tuberculosis, which can cause lung damage and do not respond to common antibiotics. The mucus present in the paranasal sinuses is equally thick and can also cause obstruction, resulting in infection. This can cause facial pain, fever, nasal discharge, headache and increase breathing difficulties. Infections of the paranasal sinuses are usually treated with prolonged cycles of antibiotics. Subjects with CF may develop excessive growth of nasal tissue (nasal sinus polyposis), as well as chronic rhinitis and pansinusitis. The development of nasal polyps can seriously limit the flow of air through the nose, and over time it can even reduce the sense of smell. The impairment of the respiratory system is present in almost all patients with the severe classical disease and cardiorespiratory complications are the most common cause of death (about 80%).

- Exocrine pancreas. The pancreas is physiologically responsible for providing digestive juices that help break down food. In the affected subjects, however, the exocrine secretions of this organ are made of a thicker and more viscous mucus, whereby there is a stagnation of pancreatic juices and the formation of cysts, with a fibrosis around the ducts (from here the name of the pathology). As a result, the latter are eventually obstructed and the movement of digestive enzymes in the duodenum is blocked, causing irreversible damage to the pancreas and often resulting in painful inflammation (pancreatitis) with destruction of the acinar tissue and parenchymal fibroadipose infarction. In the most severe and advanced cases, the pancreatic ducts appear even atrophic. The lack of digestive enzymes leads to difficulties in the absorption of nutrients, in particular of fats, with their subsequent excretion in the stool (steatorrhea). This disorder is known as "malabsorption" and leads to protein-caloric malnutrition and poor growth. Individuals with CF also have difficulty in absorbing the fat-soluble vitamins A, D, E, K and this can result in night blindness, xerophthalmia, neuropathies, asthenia, osteopenia and osteoporosis, anemia and coagulopathy. The exocrine pancreatic insufficiency occurs in the majority (about 85%)

of patients and it is mainly associated with severe mutations of the CFTR gene, in which both alleles are completely non-functional. The replacement therapy with pancreatic enzymes allows to correct it and to avoid malnutrition.

- Intestine. The meconium ileus is an intestinal obstruction of the newborn, which is unable to excrete feces, and it occurs in 5-10% of newborns with CF. Meconium can completely block the intestine and cause a serious condition. Another very common condition (about 10% of affected children) is the protrusion of the rectal internal membrane (rectal prolapse) that is caused by an increase in fecal volume, malnutrition and increased intra-abdominal pressure caused by coughing. Pictures of intestinal sub-occlusion (Distal Intestinal Obstruction Syndrome or DIOS), due to the presence of fecal masses, can arise in adolescence and adult age. In addition, people with CF often complain of heartburn, intestinal blockage from intussusception and constipation. Adult patients may also develop distal bowel obstruction due to thickened feces.
- Liver. Dense secretions also involve the liver, because the bile, secreted to aid digestion, results more viscous for the reduced functionality of CFTR expressed in cholangiocytes and can block the ducts and cause liver damage, as well as

impair digestion. Over time, this can lead to the development of gallstones in the gallbladder and in the bile ducts and to scars and nodularities in the liver (cirrhosis). Among the most important consequences there is also the fact that the liver cannot clear the blood from toxins and does not synthesize important proteins, such as those responsible for blood coagulation.

- Endocrine system. The pancreas contains the Islets of Langerhans, which are responsible for the production of insulin, and damages to this organ can lead to the loss of insular cells, resulting in a form of insulin-dependent diabetes, Cystic Fibrosis Related Diabetes (CFRD). This is one of the most important non-pulmonary complications of the disease, with late onset, and affects 15-20% of the affected adult population. CFRD presents characteristics mixed with type I and II diabetes and the main risk factors for its onset are female sex, longevity, pancreatic insufficiency, homozygosity for F508del and corticosteroid therapy. Its most common complications are the microvasculopathy: retinopathy, neuropathy and nephropathy. The only recommended treatment consists of insulin injections. Furthermore, the low absorption in the diet of vitamin D, due to malabsorption, can lead to

osteoporosis, since it is involved in the regulation of calcium and phosphate. People affected often develop the so-called Hippocratic fingers (also called clubbing), due to the low oxygen content in their tissues.

- Reproductive system. Infertility mainly affects men, although female infertility cases have been reported in CF patients. At least 97-98% of affected men are infertile, but not sterile (they can have children with assisted procreation techniques). The main cause of infertility in men is the CBAVD (vas deferens normally connect the testicles to the penile ejaculatory ducts). In practice an obstruction of the lumen of the vas deferens is carried out that is a mechanical obstruction of the spermatic transport system, capable of rapidly provoking bilateral atresia, but potentially there may also be other problems that can cause azoospermia, teratospermia and oligoastenospermia. The reduced female fertility is less documented. Some women have difficulty in procreation due to thickening of the cervical mucus *[Kredentser et al., 1986]* or because of malnutrition and, in severe cases, the latter stops ovulation and causes amenorrhea.

1.3 Diagnosis and treatment

Because of the wide range of signs and symptoms, the diagnosis of CF is particularly difficult, especially during childhood. On the other hand, however, an early diagnosis, able to detect pancreatic insufficiency, to prevent complications such as malnutrition and to allow timely treatment of lung infections, improves both duration and quality of life of patients. Therefore, for a correct diagnosis of CF, it is very important to carry out combined biochemical, genetic and clinical diagnostic investigations. A faster diagnosis also allows the early selection of couples with high genetic risk associated with subsequent pregnancies. For these reasons, neonatal screening programs have been activated around the world.

In neonatal screening programs, the first test to be performed is the evaluation of the blood concentration of immunoreactive trypsin (IRT) in all newborns: in children with CF, the blood trypsin is higher than a threshold (depending on the screening protocol) and this fact is due to pancreatic suffering, even in absence of secretive insufficiency. Trypsin levels may be increased, however, even in individuals who have only one

mutated copy of CFTR (that is in carriers) and, in rare cases, in those ones with two normal copies of the gene (the unavoidable false positives of screening programs). Moreover, it has been shown that, in addition to newborns with CF, these screening programs also select CF-like neonates. So, this test is certainly simple to perform but, as every screening programs, shows optimal diagnostic sensitivity but reduced diagnostic specificity. Furthermore, most countries do not perform routine screenings at the time of birth and, as a result, most people are diagnosed only after the manifestation of symptoms.

Infants with an abnormal trypsin dosage (hypertripsinemia) are submitted to the sweat test, for the evaluation of the concentration of chloride ions, to assess the diagnosis of CF. The sweat test (TS) *[Gibson and Cooke, 1995]* involves the application of a drug (pilocarpine) that stimulates sweating (ionophoretic stimulation) in the sweat glands of the forearm. The resulting sweat is collected on a paper sheet or in a capillary tube or on previously weighed gauze and analyzed for values of chloride. To date, the TS is the diagnostic test of reference, in terms of sensitivity and specificity, for the diagnosis of CF, since it is able to measure, in a reproducible and rapid way, the defect of functioning of the CFTR membrane protein: the test measures the event closest to the

basic defect of the disease, which, in the sweat glands, results in high concentrations of chloride in the sweat. A chloride ion concentration higher than 60 mEq/L leads to the diagnosis of CF; an intermediate (or "borderline") concentration, between 40 and 60 mEq/L, is indicative but not diagnostic for CF; a concentration lower than 40 mEq/L indicates a negative test. Up to 6 months of age, the normal chloride cut-off should be lowered to 30 mEq/L. At least two positive tests are needed to confirm the diagnosis of the disease. Unfortunately, around 10% of sweat tests can be false positive. In fact, TS is also positive in several other diseases. In addition, some CF subjects and many CF-like ones have a borderline or even negative test. Pancreatic functionality, often impaired in CF patients, is assessed by determination of steatorrhea or dosage in the feces of pancreatic digestive enzymes such as chymotrypsin and elastase. In some cases the measurement of the nasal transepithelial electrical potential difference and/or of the intestinal chloride flow appears to be useful additional diagnostic procedures.

Taking into account the high variability of the manifestations of this disease, it is clear that no laboratory practice alone is able to provide a complete and secure diagnosis of CF. The screening programs, therefore, have been made more reliable

thanks to the genetic analysis of CFTR and to the identification of significant mutations in this gene. The various techniques, in this case, allow not only the identification of the affected patients, but also of the carriers. An example of the application of this type of analysis can be found in prenatal diagnosis. In fact, couples who are expecting a baby or who are planning a pregnancy can investigate the mutations of CFTR gene, in order to determine the risk that their child may have CF. The test is typically performed first on parents and, if the risk is high (both carriers), investigations are carried out on the fetus. In some protocols, the survey is initially performed on only one parent, both because the test is expensive and because the negativity of one parent is able to decrease the risk for an affected baby. Thus, if a mutation is detected in the first parent, the tests are carried out on the other one to calculate the risk of the couple. The reduction of the costs of genetic tests and the need to reduce reproductive risk, however, suggest an evolution towards the contemporary investigation of both parents. CF can derive from many different mutations and the genetic tests are not able to investigate all of them; therefore, a negative genetic test does not guarantee that the unborn child will not be affected, but it reduces the risk proportionally to the frequency, in the regions of origin of the family, of the

mutations investigated. Despite the numerous problems related to mutational search in CFTR, namely the fact that the mutations are various and heterogeneous, that the analysis differs according to the geographical area and that some sequence variations are not included in the commonly used panels, the difficulty of distinguishing between variants causing and non-causing disease, several efficient approach for genetic assessment exist. For this purpose, a widely accepted approach is the multistep one, with various levels of molecular analysis, which are distinguished for different execution times, use of different technologies, costs, sensitivity and specificity. The various levels and their applications are listed below, while the table (**Table 1**) shows the most common techniques related to each level of analysis.

- First level analysis: it uses targeted methods to search for known specific mutations. The first level genetic analysis consists in the search of the most frequent mutations in the population and generally it is the most used method. Mainly commercial kits are used, expensive but also quick and reproducible. They allow to analyze a uniform mutation panel in different laboratories. Therefore, the first level tests show a rather limited diagnostic sensitivity since they allow

the identification of only known and frequent mutations; they are not very suitable if the mutations sought are specific for a geographical area. Moreover, they have little diagnostic utility in the analysis of subject presenting CF-like forms, with borderline clinical and/or biochemical values and specific mutations usually not included in commercial panels.

- Second level analysis: it is characterized by methods that allow to examine large regions of the gene, to search for any type of mutations, even previously unknown. These systems usually allow the recognition of sequence variations in all the coding regions, in the adjacent intronic regions (those that may influence splicing) and in the proximal 5'-flanking of the CFTR gene. Thanks to technological improvements and to cost reduction, the Sanger sequencing and the "New Generation Sequencing" (NGS) are the most used methods, in place of other obsolete scanning approaches. However, no mutational scanning technique currently used for diagnostic purposes is able to detect all existing CFTR mutations. About 3% of the mutated alleles, in fact, present mutations that are not easily identifiable even by extensive sequencing protocols. The second level tests are characterized, therefore, by a greater, though not complete, sensitivity,

identifying known or even not previously described mutations. Moreover, their use raises the problem of the interpretation of the new sequence variations found: while mutational panels are usually designed to include only CFTR mutations that cause disease, with the scanning procedures we also include changes in the sequence not still characterized from the functional point of view and this fact complicates the genetic counseling.

- Third level analysis: it allows to study the mutations that escape the first and second level, that are rare gene rearrangements (for example large deletions), fully intronic mutations not adjacent to exons and those ones in the distal 5'-flanking and in the 3'-UTR. As already mentioned, in 3% of affected subjects the mutations responsible for the disease can't be identified, despite the analysis of all coding portions and of the splicing regions of the CFTR gene. Since the rearrangements involving deletions/duplications of single or multiple exons usually escapes to this type of analysis, it is reasonable to suppose that the presence of insertions, duplications or deletions in CFTR is underestimated and can account for at least a part of those alleles not characterized. These investigations are commonly performed using MLPA (Multiplex Ligation-

dependent Probe Amplification), which is a procedure based on the use of probes that are amplified, in a PCR reaction, in a target-specific manner: each probe, in fact, is constituted by two oligonucleotides, which hybridize adjacent sequences on DNA and which are ligated only in presence of the target. There is also the possibility to study mRNA usually obtained from nasal epithelial cells with the brushing technique. This analysis is able to identify even the fully intronic mutations that alter the correct splicing of the CFTR gene; furthermore, it allows to highlight the great genetic rearrangements, provided that the mutation does not cause the missing expression of the transcript. The use of genetic analysis up to the third level allows to reach a diagnostic sensitivity (detection rate) of genetic tests in CF of about 98%.

Table 1. List of the most common techniques related to each level of CFTR genetic analysis.

Genetic analysis of CFTR gene	Techniques
First level analysis	
1	Multiplex analysis RDB or OLA or ARMS of the most frequent mutations (commercial kits)
2	Specific analysis for the mutation of the family (in case the mutation of the family is known and not included among those detected in the kit)
3	Mini-Sequencing for the analysis of single mutation sites
Second level analysis	
4	D-HPLC or DGGE analysis (almost no longer used)
5	Sanger Sequencing or NGS for the entire coding portion and of the exon-intron junctions of CFTR gene
Third level analysis	
6	MLPA or QMPSF analysis for the identification of genetic rearrangements
7	Study of the mRNA of CFTR gene extracted from nasal epithelium

The practical application of this multistep approach changes according to its use in subjects with suspected disease for diagnostic purposes or in subjects of the general population to reduce the genetic risk. In the first case it is reasonable to move progressively through the levels until you find two mutations

of CFTR on different alleles. If no mutations are found (or even if only one mutation is not found) even at third level, the genetic test contributes to a sensible exclusion of diagnosis of CF. On the contrary, in the second case, since it can be difficult to apply all levels of mutational search to each subject to check its carrier status, an appropriate residual genetic risk is usually chosen and the mutation search is performed with a suitable panel of mutations. Following the above considerations, it can be said that the often incomplete genetic characterization of CF and CF-like patients may often be due to technical limitations and constitutes the major obstacle to the understanding of the genotype-phenotype relationship. Although the techniques have progressively evolved, becoming increasingly more sensitive and effective, the situation still remains complex for CF-like forms, characterized by mutational patterns that are different from those of the classical forms of CF and with a definition of the diagnosis often not clear.

Although there is not yet a definitive cure for CF, currently there are different types of treatment, which concern all the clinical aspects and the complications related to this disease. These approaches, in addition to ensure that most patients will reach adulthood, allow them to spend an existence less and less burdened by the damage of the pathology. However, the

management of the condition must be protracted throughout the life of the individual and is aimed at maximizing the functionality of the organs and, consequently, the quality of life. Among the most used treatments we have to remember:

1. Encouragement to a good nutrition. The diet must be high calorie and high in fats, to compensate for malabsorption, and in proteins, to guarantee normal growth. In the event of a deficiency, supplementation of fat-soluble vitamins, essential fatty acids, B12, salts and iron is also required and complications (for example diabetes) should be treated in a specific way.
2. Additional therapy for pancreatic enzymes. In case of pancreatic insufficiency, it is also important to administer replacement enzymes appropriately, based on digestion and on absorption of ingested fats.
3. Active lifestyle. Respiratory physiotherapy is one of the most widely used practices. It is already started during the early stages of the disease because it allows to improve lung ventilation and to effectively eliminate secretions.
4. Hydration of the mucus. Some compounds with osmotic effect are able to recall water from inside to outside of cells.

In this way the mucus is more hydrated and slides more easily along the lungs, freeing them.

5. Treatment of airway infections. The most consistent aspect of CF therapy is the limitation of lung injury caused by dense sputum and subsequent infections. First, the therapy tries to prevent infections by eliminating the thickened mucus that obstructs the airways and by following appropriate hygiene rules; instead, to treat chronic and acute infections in progress, antibiotics administered intravenously, inhaled and oral, accompanied by anti-inflammatory drugs, are prescribed.
6. Psychological support. Both the patient and the family must be followed carefully, especially since the disease often manifests itself at an early age, it leads to serious problems in everyday life and aggressive therapies are needed to counter it.
7. Pregnancy support. Also in the particular case of a pregnancy planning, couples must be helped, whether they have already had a sick child or if one of the parents is affected (in particular the father): in the first case we use molecular biology techniques to calculate the genetic risk, in the second one we can intervene, if possible, with assisted reproduction practices.

8. Lung transplantation. This practice is recommended only when pulmonary function is expected to decline to the point where assistance with mechanical devices, due to respiratory failure, is required or when patient survival is threatened. Contrary to the past, the most modern protocols are anyway oriented to include patients in good condition in the transplant list without waiting for an excessive decline in lung function. In fact, it has been shown that, in this way, the short and long term outcomes of the transplant are much better. It is a bilateral lung transplant. Although single lung transplantation is possible for other conditions, individuals with CF require replacement of both lungs: the remaining lung may, in fact, contain bacteria that will infect the new transplanted one. A pancreas or liver transplant can be performed at the same time in order to relieve liver deficits and/or diabetes.
9. New therapies. They consist essentially of mutation-specific therapeutic approaches and gene therapy. The first ones employ particular small molecules that correct the functional defect induced by the specific mutations of the CF, targeting directly the defective protein, and are designed according to the belonging mutational class of the mutation itself *[Wilschanski M, 2013; Wilschanski M et al., 2011]*.

Mutation-specific pharmacological therapies are now a therapeutic reality, already in use on patients, for a limited number of mutations. In fact, we already have two mutational class-specific treatments: the Ivacaftor, a potentiator that facilitates the opening of the Cl- channel in the cell (effective on the so-called mutations of gating), and the Orkambi, a combined therapy of a corrector (Lumacaftor) and a potentiator (Ivacaftor), effective also on those mutations with a defective maturation as, for example, the F508del (although with a limited efficacy in respect to the treatment with potentiator); the extension to an increasing number of mutations is now in place and this fact will allow the mutation-specific therapy to evolve in a patient-specific therapy in the future. Instead, gene therapy aims either to introduce the entire normal gene into the airway cells (gene augmentation) or to perform the specific correction of the mutation at the genomic DNA level (gene editing). At moment, this last approach is the most advanced genetic medicine approach for the disease, because the correction could be more stable over time and could reach physiological level of protein functionality; however, the quantitative level of correction is very low and this solution

is still under development in laboratory, requiring a longer period of time to be effectively transferred to patients.

1.4 Cystic Fibrosis and CF-like clinical forms

CF is a multi-system pathology with a chronic and progressive trend. It is usually associated with chronic bronchopneumopathy, exocrine pancreatic insufficiency (PI), high concentrations of chloride and sodium in the sweat, familiarity with the disease, male infertility. These events are typical of what is commonly called the "classical" form of CF and, in this case, the diagnosis is clear (thanks to both the clinical manifestations and the various laboratory tests already mentioned).

The different level of functional impairment of the CFTR protein and the different deficit of the various affected apparatuses, however, originates a wide phenotypic variability, with clinical pictures that differ depending on the organs involved, the severity of the induced lesions and the age of onset of the symptomatology. The so-called "non-classical" forms may originate, characterized by a more complex and

often late diagnosis, since the symptoms are different and can only be detected in the juvenile-adult age. They can manifest themselves as mono-symptomatic or oligo-symptomatic forms, characterized by a normal function of the exocrine pancreas (PS), even in the presence of the typical respiratory impairment, and this incompleteness makes their identification more difficult. The common clinical features of these CF-like forms are the following: they are oligo or mono-symptomatic; the symptoms may start from infancy, but the disease becomes clinically significant only after several years (usually after 10 years of age); survival reaches adulthood; there are chronic diseases of the respiratory system, but with pancreatic sufficiency; the sweat test is less than 60 mmol/l. In recent years, new and in-depth knowledges in the genetic/biomolecular field have allowed us to classify as CF-like forms atypical clinical pictures characterized by this nuanced respiratory symptomatology or clinical situations in which there is even only one of the typical manifestations of CF: some examples are chronic pansinusitis, nasal polyposis, biliary cirrhosis and recurrent pancreatitis of undefined etiology, portal hypertension, hypochloraemic alkalosis, glucose intolerance or non-type I diabetes in adolescents, cholelithiasis at a young age *[Kerem et al., 2006]*. Among the

most common CF-like forms there are: minor pulmonary diseases (bronchiectasis and chronic bronchitis); idiopathic chronic pancreatitis (ICP); male infertility due to CBAVD; male reduced fertility due to ISHV. All these categories of patients are associated to various mutations of the CFTR gene.

Patients with mild lung diseases, such as bronchiectasis, have milder airway symptoms than those of the classical form, although *Pseudomonas Aeruginosa* infections are common also in these individuals. The disease manifests itself, in this case, in old age and there is no pancreatic insufficiency, with sweat test in the norm or borderline. The lower severity of this form could be explained by the fact that often such subjects present a "severe" mutation on a gene allele and a "mild" mutation on the other one. CF is often associated also with ICP: a lot of patients with pancreatitis have mutations on both alleles of the CFTR gene *[Noone and Knowles, 2001]*. The difference with those affected by the classical form of CF is that the manifestations at the level of the pancreas arise, in this case, only in adulthood, while in the previous ones the functionality of this organ is altered since birth. At the level of the reproductive system the most studied monosymptomatic form in men is the CBAVD. These subjects do not show any classic clinical symptoms and the sweat test is generally borderline. Patients

with absence of vas deferens, without other CF symptoms, often present also other abnormalities of the reproductive tract: obstructions in the epididymis, unilateral/bilateral absence of the seminal vesicles, cystic dilation of the rete testis *[Jarvi et al., 1998]*. The phenotype is less severe than the classical form because, also in this case, often a CFTR allele has a "severe" mutation, while the other allele a "mild" one, and these mutations frequently are at intronic level, altering the maturation process of messenger RNA. Another pathology related to reproductive disorders in CF is the ISHV. It is a form of subfertility due to an excessive viscosity of the seminal fluid and to a consequent reduction of sperm motility. The causes of seminal hyperviscosity are not yet completely clear, but they seem to be the result of various factors, such as the presence of an acute or chronic infection or inflammation of the genital tract or the establishment of autoimmune processes, but could be traced back also to sources of different types, such as biochemical, enzymatic and genetic alterations, like in CF. Thanks to our previous studies, in some cases ISHV can be now considered a "minimal clinical form" of CF. There are several prerequisites to support this statement, for example the fact that the secretions in CF are thick and viscous and that the epididymis, in which the CFTR protein is expressed for the

hydroelectrolytic balance, is the main responsible for the seminal viscosity, but most of all that many variations of CFTR sequence are present in individuals with ISHV *[Rossi et al., 2004; Lucarelli et al., 2007; Elia et al., 2009]*.

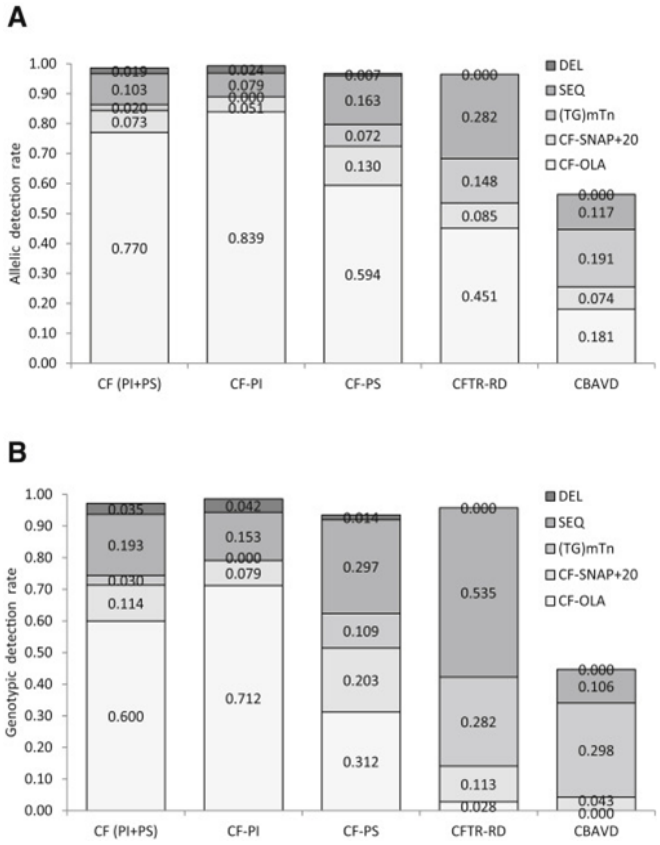


Figure 5. Detection rates at each mutational step in the clinical CF and CF-like macro-categories. Both the allelic (**A**) and genotypic (**B**) detection rates at each mutational step are shown. *[Lucarelli et al., J Mol Diagn., 2016]*

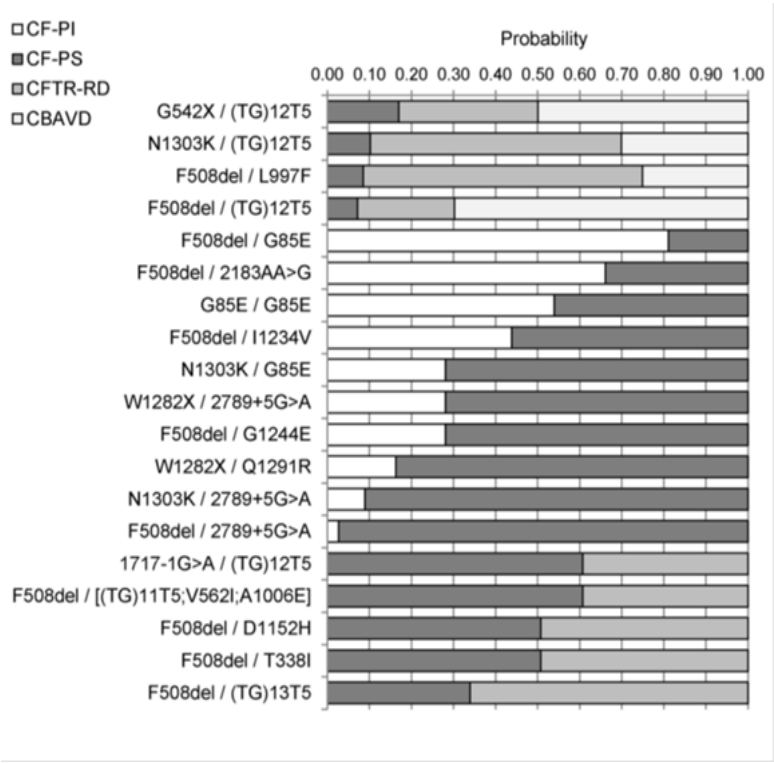


Figure 6. On a total of 225 different CFTR mutated genotypes identified, 19 (8.4%) were found in at least two different CF and CF-like population, whereas the remaining 206 (91.6%) were exclusive to a single population. *[Lucarelli et al., Mol Med., 2015]*

The great variability of clinical manifestations in CF and CF-like populations has been examined in two previous studies of my research group *[Lucarelli et al., 2015; Lucarelli et al., 2016]*. Different mutational patterns and different detection rates (the frequency of alleles or genotypes fully characterized

from genetic point of view) were evidenced between CF and CF-like populations. Each mutational search step is influenced by the genetic heterogeneity between the different populations. Both the allelic and the genotypic detection rates were found to vary greatly between the CF and CF-like populations and at each step of the mutational search (**Figure 5**). Moreover, we found that this genetic heterogeneity between populations is amplified at the phenotypical level; in fact, the same genotype was found in populations with different clinical profiles (**Figure 6**). We interpreted at least part of this variability with the possible presence of intragenic modifiers with a possible effect of CFTR transcriptional regulation. We assume that they can act as modulators when two different mutations are already identified (as for example in most of CF patients) or as causing-disease variants when only one mutation is present (as for example in most CF-like patients).

1.5 Transcriptional control of CFTR expression

As already seen, the transcription of CFTR gene generally produces a primary transcript of 6.1 Kb. However, on this gene

a various and complex transcriptional regulation is performed at different levels on the development of the messenger RNA. Here we discuss some examples of processes that influence the final production of the protein, both in terms of quantity and quality.

I. On this gene different transcription start sites are found. Since the CFTR promoter has the characteristics of a house-keeping promoter, the use of different transcription start sites can support a regulation on gene expression that explains the time- and tissue-specific expression patterns seen for CFTR. Alternatively, the production of a functional CFTR protein can be influenced by tissue- and time-specific alternative splicing *[Vankeerberghen et al., 2002]*. In general, this gene is more expressed in airways and in secretory organs than in other tissues and in a variable quantity with the increasing of the age (CFTR is more expressed at birth and then it decreased during life); the regulation of the expression, in this case, is guaranteed by lung development-specific transcription factors (FOXA, C/EBP) and some microRNAs (miR-101, miR-145,

miR-384), that regulate the switch from a strong fetal to a very low adult CFTR amount *[Viart et al., 2014]*.

II. Different spliced forms of CFTR have been detected *[Vankeerberghen et al., 2002]*. Some examples of this alternative splicing of the gene are: exon 4^{-}; exon 5^{-} (a splice variant that is exclusively expressed in heart, with a still unclear functional relevance); exon 9^{-} (the percentage of non-functional CFTR may vary between tissues: it is higher in the vas deferens than in nasal epithelium; moreover, the overexpression of splicing factors can modulate exon 9 skipping, as well as the presence of alleles carrying specific (TG)mTn polymorphism on intron 8); insertion of 119 bp of intron 10; exon 12^{-}; TNR-CFTR (a kidney-specific alternatively spliced form of CFTR that, as a result of the introduction of a premature stop-codon, gives rise to a protein that only contains the N-terminal part of CFTR).

III. The transcriptional modulation of a gene can be also associated to genetic haplotypes, that consist of a combination of allelic variants along a chromosomal segment containing closely associated loci. They are usually inherited together and can be often typical of

certain geographical areas. Recent studies underlined that the presence of specific intragenic haplotypes of CFTR can be also linked to particular mutations of this gene: this happens, for example, to the rare CF mutation E1104X in a study on a population of Tunisian patients *[Oueslati et al., 2015],* and can be useful both as a molecular method in clinical linkage analysis and in cases where one or both mutations of the disease remain unidentified. Moreover, these intragenic haplotypes can be associated to different aspects of this complicated pathology, because the presence of common CFTR variants often results in aberrantly spliced transcripts and, consequently, in susceptibility, for example, to chronic pancreatitis and CBAVD *[Steiner et al., 2011]*; so the determination of these haplotypes helps to stratify patients into high- and low-risk subjects, providing helpful information for genetic counseling.

IV. In the last 10 years the role of miRNAs on the regulation of gene expression has been partially described. These small endogenous non-coding RNA molecules are able to act on RNA silencing and post-transcriptional regulation via base-pairing with

complementary sequences on 3'-UTR of target mRNAs. miRNAs have a dual role, both as regulators of development and in the maintenance of homeostasis; however, their expression is altered in pathological states, such as lung inflammation and diseases like CF. The cause of this alteration can be an increased/reduced production of a certain miRNA or a mutation on the binding site of miRNAs on 3'-UTR (these mutations may modify the affinity and thus the binding of a miRNA to the transcript, or may cause the appearance of novel target sites for miRNAs thus impairing the levels of gene expression). In CF, in particular, it has been highlighted that various aspects of this disease are influenced by the presence of miRNAs, in fact they intervene in: innate immunity, both in epithelium (miR-126 regulates TOM1) and in bone marrow derived cells (miR-9 is decreased in CF neutrophils and miR-126 is increased in CF mononuclear cells); inflammation (a decrease in miR-31 expression is correlated with an increase in IRF-1 and cathepsin S levels, while a decrease in miR-93 expression is correlated with an increase in IL-8 levels); ion conductance (miR-145, miR-223 and miR-494 target directly CFTR mRNA;

miR-433 and miR-509-3p are increased in CF; miR-145, miR-150 and miR-451 are upregulated in adult lung); ER stress (ex. miR-145, miR-221 and miR-494 target ATF6) *[McKiernan and Greene, 2015]*. Often, the intervention of miRNAs, that directly bind CFTR, results in a negative modulation of the amount of messenger in the cell; so many current therapies are aimed at creating anti-miRNA agents able to protect the CFTR mRNA from their action *[Amato et al., 2014; Zarrilli et al., 2017*].

V. Obviously the presence of DNAse I hypersensitive sites (DHS) on promoter (5'UTR) of CFTR allows the binding on this gene of various transcription factors, that perform local changes in chromatin structure and histone modifications. The area at the 5´UTR of the gene play a role in its expression because it has a high GC content, several potential protein binding sites and many transcriptional start sites. Mutations in this region, especially on DHS, may change the interaction between the gene and its regulatory factors, thus acting as disease-causing mutations. Moreover, these mutations can modify the gene expression and, consequently, the clinical expression of the disease at

different organs levels. To date, a few disease-causing mutations in the proximal promoter region have been described in CF patients *[Castaldo and Tomaiuolo, 2013]*.

VI. The elements already studied in the basal promoter of the gene do not fully explain all CFTR expression patterns, suggesting that other regulatory elements are located elsewhere, outside the basal promoter, either within the locus or in the adjacent chromatin. We talk about the so-called "long range interactions". In particular, they are insulator elements, that maintain distinct the expression domains, and may be involved in the mechanisms that demarcate the transition between open and closed chromatin states, so between activation and inactivation of the transcription of the gene. When positioned between an enhancer and promoter, insulators block their interaction ("enhancer-blocking" or "barrier" activity). The CTCF (CCCTC-binding factor) protein plays a crucial role in maintaining the enhancer-blocking activity and several CTCF-binding sites have been characterized in many vertebrates. However, the precise mechanism by which insulators work is already unknown: a looping mechanism, with

CTCF sites being bound to the nucleolus or the nuclear matrix has been suggested; alternatively, CTCF may form part of a protein complex directly involved in epigenetic remodelling and in histone modification *[Blackledge et al., 2007]*. The CFTR gene is flanked by two genes, ASZ1 (Ankyrin repeat, SAM and basic leucine zipper domains) on the 5' side and CTTNBP2 (cortactin-binding protein 2) on the 3' side, and they have very different expression profiles: CFTR is expressed primarily in specialized epithelial cells, while ASZ1 is transcribed exclusively in testis and ovary and CTTNBP2 is highly expressed in brain, kidney and pancreas. In CFTR locus there are two DHS of particular interest, that flank the CFTR gene upstream at −20.9 kb (to the 5' of CFTR gene) with respect to the translational start site, and downstream at +15.6 kb (to the 3' of CFTR gene). These DHS possess enhancer-blocking activity and bind proteins that are characteristic of known insulator elements, but the first one is associated with a classical CTCF-dependent insulator element, while the second one is independent of CTCF binding. In addition to the prominent site at +15.6 kb, other DHS are evident 3′ to the coding region

of the gene: in particular, there is a complex cluster of sites at +5.4, +6.8, +7.0 and +7.4 kb from the CFTR translation end point [*Ott et al., 2009]*. The DHS at +5.4 kb and +7.0 kb were observed in a variety of cell types, irrespective of CFTR expression; however, the DHS at +6.8 kb and +7.4 kb were only found in a restricted number of CFTR-expressing cell-types, including, for example, primary epididymis cells, suggesting that they may contain tissue-specific regulatory elements that participate in controlling CFTR-expression *[Blackledge et al., 2009]*. So these chromatin insulators that flank CFTR prevent interference between the regulatory elements of neighbour loci and maintain independent the expression domains, participating in a correct and regulated transcription of the gene.

2. CASE SERIES, MATERIALS AND METHODS

2.1 Analyzed populations and experimental approach

The populations used in this study and their composition are listed below:

- A total of 424 patients (848 alleles), already diagnosed and followed by the CF Regional Reference Center of Lazio (Institute of Pediatric Clinic, Sapienza, University of Rome). These patients were mainly from Central Italy, classified according to recent guidelines in the following 4 clinical macro-categories: affected CF with pancreas insufficiency (CF-PI: 222 patients, 444 alleles), affected CF with pancreas sufficiency (CF-PS: 106 patients, 212 alleles), mono- or oligo-symptomatic CF forms (CFTR-RD: 50 patients, 100 alleles), congenital bilateral absence of the vas deferens (CBAVD: 46 patients, 92 alleles). These patients had already been examined and characterized from a clinical, biochemical, microbiological and genetic point of view.

- 60 sub-fertile subjects (120 alleles) with hyperviscosity *sine causa* of seminal fluid (ISHV), selected by the Andrology Unit of the Department of Clinical and Molecular Medicine, Faculty of Medicine and Psychology of Sapienza University of Rome, analyzed in a previous research work of my group and considered as a minimal form of CF. *[Rossi et al., 2004; Elia et al., 2009]*
- 119 normospermic (NORMO) subjects (238 alleles), selected by the Andrology Unit of the Department of Clinical and Molecular Medicine, Faculty of Medicine and Psychology of Sapienza University of Rome, used as a negative control and also already characterized by the genetic point of view.
- 102 male and female subjects (204 alleles) belonging to the general population (GEN), consisting of both blood donors and healthy partners of couples received at the CF Regional Reference Center of Lazio, Department of Pediatrics and Infantile Neuropsychiatry, Faculty of Medicine and Dentistry of Sapienza University of Rome, used as a further negative control and already analyzed in this laboratory.

The experimental approach we have devised for this study is based on three kinds of analysis and explores three

corresponding possible intragenic CFTR modulators, already previously mentioned: intragenic haplotypes and the individual functional effect of each of the comprising variant; variant tracts (TG)mTn and their effect on the splicing of CFTR mRNA; 3'-UTR for potential recognition sites of miRNAs.

a) The intragenic haplotypes analysis derive from a previous study, carried out by my research group, in which 8 markers of CFTR gene were selected, specifically 6 SNPs and 2 repetitions: 125G/C (*rs1800501*), 875+40A/G (*rs1800502*), (GATT)n, 1001+11C/T (*rs1800503*), (TG)mTn, 1540A/G or M470V (*rs213950*), 2694T/G (*rs1042077*), 4521G/A (*rs1800136*). In this study their experimental study was completed. They were examined in all the populations, both in targets and in controls, on DNA, through MiniSequencing for the SNPs and Cycle Sequencing for the repetitions. With this analysis we obtained all the possible combinations of variants of each marker and evaluated the frequency of the haplotypes in each population. Then, these sequence variations were also studied individually from a functional point of view, through RT-PCR, to evidence any unusual forms of splicing.

b) The polymorphic tract (TG)mTn varies from one subject to another: in fact, at the level of intron 8, we can have a variable number of TG dinucleotide (from 9 to 13) followed by an equally variable repetition of T (T3, T5, T7, T9; we also analyzed a rare T6 repetition). It regulates the splicing of exon 9: in general, a high number of TG repeats followed by a low number of T repetitions has been seen to promote the exclusion of exon 9 from the transcript. So we studied this variant tract in our patients through Cycle Sequencing on DNA to evaluate the frequency of each (TG)mTn in each population. Then we decided to study these tracts also from a functional point of view on RNA, using RT-PCR, densitometric analysis and Real-time PCR; in this way it was possible to identify the abnormal splicing forms and to quantify them by the calculation of the percentage of both wild-type and mutated mRNA.

c) The CFTR 3-'UTR was included in this study to evaluate the sequence variations that could influence the action of miRNAs. In fact, if the binding site of a miRNA on the 3'-UTR of the CFTR gene is altered by a sequence variation, it is possible that the miRNA binding and, consequently, the amount of mRNA, will be changed. So, initially, we carried out a bibliographic, bioinformatic and database research,

whose purpose was to highlight which are the so far known miRNAs able to bind the 3'-UTR of CFTR; then we examined all the sequence of 3'-UTR (1557 bp) through its partition in five overlapping zones (≃400 bp for each one) (**Figure 7**), design of specific primers and Cycle Sequencing on DNA. This analysis did not concern all the subjects previously listed, but only patients with specific characteristics.

As regards the basic genetic analysis, all populations had already been analyzed for the main mutations and genetically characterized before this work. My experimental contribution of this thesis was, therefore:

- the completion of the structural analysis of intragenic haplotypes and variant tracts (TG)mTn, for a total of 250 patients on 484. The 221 control subjects (NORMO and GEN populations) were already analyzed;
- the complete bibliographic and structural analysis of CFTR 3'-UTR;
- the complete functional analysis of the individual variations included in the haplotype;
- the complete functional analysis of the splicing of variant tracts (TG)mTn;

❖ the complete setting up and all the analysis with Real-time PCR.

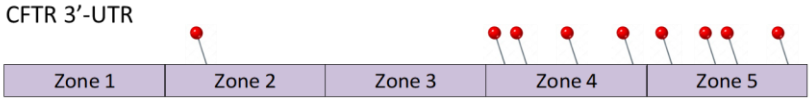


Figure 7. Schematic representation of the five overlapping zones of CFTR 3'-UTR. In red we indicated the points on which known miRNAs binding sites were found.

2.2 Extraction and quantification of DNA and RNA

Genomic DNA was extracted from whole blood (only in some cases from sperm pellets) collected in tubes with EDTA, using the kit QIAamp DNA Blood Midi Kit (Qiagen), which exploits a chromatographic system on single-use columns. The quantification and purity of the extracted DNA was evaluated through the spectrophotometric reading of the absorbance at 260/280

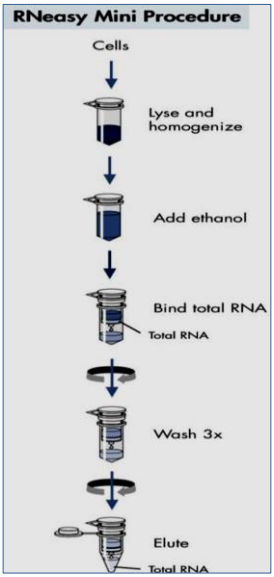


Figure 8. Phases of the RNA extraction procedure.

nm or the use of relative fluorescence given by Qubit Fluorometer (Invitrogen). All the extracted DNA were then brought to the concentration of 50 ng/µl and a 10 ng/µl diluted aliquot was made; it corresponds to the concentration usually used in the amplification protocol.

The RNA was extracted from nasal epithelial cells, obtained through the "brushing" technique *[Ramalho As et al., 2004]*, or from stem-like epithelial cell cultures of CRC (conditionally reprogrammed cells) *[Liu et al., 2012; Suprynowicz et al., 2012]* derived from nasal epithelium of CF patients, in collaboration with the Higher Institute of Health. The sampling consists in the introduction of a stick, whose end is endowed with bristles, in the nasal cavities for a depth of about 3 cm. Once inserted, it is rotated 360° in both directions, in order to obtain a quantity of cells suitable for the analysis; it is necessary to brush both nostrils. The brush used is sterile and has a thickness of 2,5 mm. After collection, it is placed inside a 2 ml eppendorf containing 990 µl of RLT buffer (Qiagen), that is a lysis buffer and is the first one used for RNA extraction procedure; to this, 10 µl of β-mercaptoethanol are added, and subsequently the cells are frozen at -80°C. The next day the extraction is performed using RNeasy mini kit (Qiagen), which allows to extract the total RNA from cell pellets. RNeasy

technology simplifies the isolation of total RNA, combining the use of guanidine thiocyanate for cell lysis with purification through the use of single-use columns with silica membrane, ideal for obtaining a high level of RNA purity. The system used is therefore a chromatographic system. The quantification and the purity of the extract are evaluated by spectrophotometry, analyzing the absorbance at 260/280 nm, or reading the relative fluorescence given by Qubit Fluorometer, while the quality of the extracted RNA is evaluated by agarose denaturing gel electrophoresis, which makes it possible to distinguish the characteristic rRNA bands.

2.3 MiniSequencing (primer extension) and sequencing of the CFTR gene

The second level investigation of CFTR gene has the function to investigate all the rare or unknown mutations and to identify possible sequence variations of pathological interest by means of direct sequencing (Cycle Sequencing) of genomic DNA; it can be used also to study variant tracts and repetitions that are present in the sequence. Instead, when a sequence variation is

already known, it is preferred to use a faster methodology, called MiniSequencing, that is a primer-extension reaction in which each primer binds in a complementary manner to the target, in presence of a DNA polymerase and fluorescently labeled ddNTPs (dideoxy-terminators): the polymerase extends the primer of a single nucleotide, adding a single ddNTP to its 3' end *[Lucarelli et al.,2015; Lucarelli et al.,2016]*. In this thesis, I used the results of a previously executed first level analysis and completed the second level analysis, together with the MiniSequencing method, to study at the structural level the intragenic modifiers of CFTR. In particular, I analyzed the six single nucleotide polymorphisms (SNPs) of the intragenic haplotype (125G/C; 875+40A/G; 1001+11C/T; 1540A/G or M470V; 2694T/G; 4521G/A) using MiniSequencing and the two repetitions (that are (GATT)n and (TG)mTn) using a classical Sanger Sequencing (**Figure 9**).

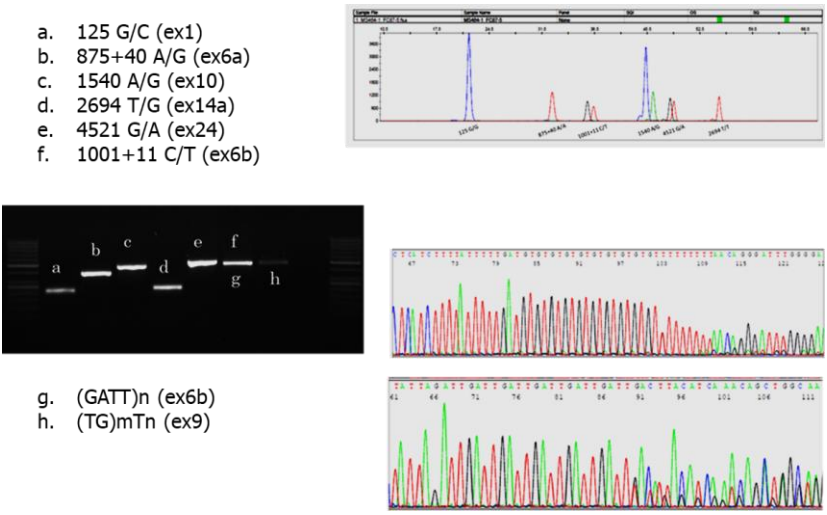


Figure 9. Schematic representation of the approach used for the study of the intragenic haplotype. On the left there is an example of amplification of the DNA zones of interest; on the right there are an example of peaks of the six SNPs included in the intragenic haplotype and obtained through MiniSequencing (on the top) and an example of electropherograms of the two repetitions included in the intragenic haplotype and obtained through CyCle Sequencing (on bottom).

The sequencing of the exons including the 8 markers that compose the selected haplotype, including the (TG)mTn tract, and of the 3'-UTR (divided in 5 zones) was performed, as well as the sequencing of the entire coding region and adjacent intronic zones. The extended analysis (amplification, purification and DNA sequencing) was performed in 96-well

format: in this way it was possible to simultaneously amplify all the zones of interest and analyze up to 12 samples.

For sequencing, it is first necessary to proceed with the amplification of the DNA zone of interest, then to the PCR product purification and subsequently to the Cycle Sequencing reaction, from which the final products will be obtained and will be separated by the genetic analyzer *[Lucarelli et al., 2006]*. The PCR reagents are initially assembled into a single master mix, not containing the pairs of primers, and this is subsequently divided into aliquots. Later, PCR primer pairs are added, so that each aliquot contains the specific primer pair for each zone. These aliquots are dispensed into the plate by a multi-channel pipettor. The DNA samples are divided into 12-well strips and subsequently dispensed into the plate. This disposition makes sure that the samples and the pairs of primers will be crossed in such a way that in each well there is a single combination of primers amplifying a given zone of the CFTR of a specific sample. This multiwell organization was maintained for subsequent purification and sequencing. For the amplification, "external" primers are used, while for the subsequent sequencing reactions "internal" primers are employed; they are designed internally compared to the previous ones, in order to increase the specificity of the

reaction. Both external and internal primers are designed in such a way that all known small variations in the exons and in the adjacent intronic zones are identifiable. The internal primers used in the sequencing reactions are purified by HPLC, in order to eliminate the background of n-1. The description of the single steps of this approach follows.

2.3.1 DNA amplification by PCR

The amplification of the samples was carried out by means of a PCR of the DNA segment which is subsequently intended to be sequenced; the reaction mix (for a single sample) is prepared as follows:

MM PCR:	**x 1**
1. H_2O	2.75 µl
2. Buffer 5X	3.0 µl
3. $MgCl_2$ (25 mM)	0.9 µl
4. dNTPs (1,25 mM)	2.1 µl
5. Go Taq Promega (5 U/µl)	0.05 µl
6. External Primer Forward (10 pmol/µl)	0.6 µl
7. External Primer Reverse (10 pmol/µl)	0.6 µl
8. Sample (10 ng/µl)	5.0 µl
	15.0 µl

First the reaction mix is prepared; after dispensing it, 5 µl of each sample are subsequently added. DNA amplification is performed using the PTC 100™ (MJ Reasearch) thermocycler and the program used is as follows:

- 95°C for 2 minutes (step of activation of Taq Polimerase)
- 94°C for 45 seconds (denaturation of DNA)
- Ta° from 62°C to 54°C in TD for 1,30 minutes (primer-specific temperature for pairing)
- 72°C for 4 minutes (fragment extension)

40 cycles

- 72°C for 7 minutes (final extension)

2.3.2 Electrophoresis on agarose gel

The amplicons are electrophoretically controlled to evaluate their size in number of bases (bp); 3.3 µl of amplicon together with 3.3 µl of Loading 2x are mixed, migrated and compared with a molecular weight standard (Ladder 6.6 µl). The verification by electrophoresis is necessary not only to check the amplicons, but also to rule out any contamination of the blank reactions. The bands are visualized on an ultraviolet light

transilluminator, which excites the DNA to which an intercalating dye (GelRed) is bound.

2.3.3 Enzymatic purification of the amplicons

Before MiniSequencing: The amplicons are joined in a single tube (2 µl from each PCR product) and are brought to a final volume of 15 µl with water. To these mix of amplicons, 5 µl of FastAp (Thermosensitive Alkaline Phosphatase, Thermo Fisher) [1 U/µl] and 0.1 µl of Exo1 (Exonuclease 1, Thermo Fisher) [20 U/µl] for each sample are added. The solution is incubated at 37°C for 60 minutes to remove the excess of primers and the non-incorporated dNTPs, and then placed at 80°C for 15 minutes to deactivate the enzymes.

Before Cycle Sequencing: The amplicons are purified by adding to the 11.7 µl left of the amplification reaction, a reaction mix composed of: 1.5 µl of Buffer, 1.2 µl of the enzyme FastAp [1 U/µl] and 0.6 µl of the enzyme Exo1 [20 U/µl] for each sample. The solution is incubated at 37°C for 45 minutes to remove the excess of primers and the non-incorporated dNTPs, and then placed at 80°C for 15 minutes to deactivate the enzymes. This step is performed in a 96-well format.

2.3.4 MiniSequencing (primer extension)

MiniSequencing technique allows the study of single nucleotide polymorphisms (SNPs) through the primer extension of a single fluorescent nucleotide. In this method, much importance is assumed by the MiniSequencing primers; in fact, primers with certain characteristics have been planned and designed for this job:

- these primers are designed close to the mutation site and they must differ in length to avoid any overlap between the final products (a difference in length of at least 4-6 nucleotides between successive primers is necessary): we reached a common temperature of annealing of 50°C;
- in some cases, the length of the primer was modified by the addition at its 5' end of a polynucleotide tail (poly-A tail), not homologous to the target sequence;
- the mobility of an oligonucleotide in capillary electrophoresis is determined by its length, by its nucleotide composition and by the type of fluorochrome used: all these variables were taken into consideration both in the design phase and, experimentally, in the test phase (the effect of the nucleotide composition on the mobility can be not negligible when the primer is short);

- finally, the MiniSequencing primers are purified HPLC to avoid the appearance of unwanted products that could interfere with the final analysis.

Table 2. List of the six primers designed and used for MiniSequencing. In the first column the primers names are indicated; in the second column there are the relative sequences of the primers, with the poly-A tail in bold; in the third column the number of bp for each primer is evidenced.

Oligo Name	Sequence	bp
H125G>C	**ATAAT**GGGACCCCAGCGCCC	20
H875+40A>Gback	GGTTTTACTAAAGTGGGCTTTTTGAAAACA	30
H1001+11C>T	GAAAAAATGATTGAAAACTTAAGACAGTAAGT TGTT	36
H1540A>G	**ATAAATTATTAATA**GGGTTTTATTTCCAGACTT CACTTCTAATG	44
H4521G>Aback	**TAAAAAAATAAAAAATAAAAAAAAAA**TGTC TCCTCTTTCAGAGCAGCAAT	50
H2694T>G	**TATAAATATAATATAATAATAAATA**TACATGG AACACATACCTTCGATATATTAC	55

These primers, before use, are mixed together to give a mix of CF-SNaP primers. This mix is composed by 15 µl of primers (2.5 µl x 6), from the stocks at 100 pmol/µl, and 35 µl of H2O, for a total volume of 50 µl; so, each primer is 5 pmol/µl concentrated.

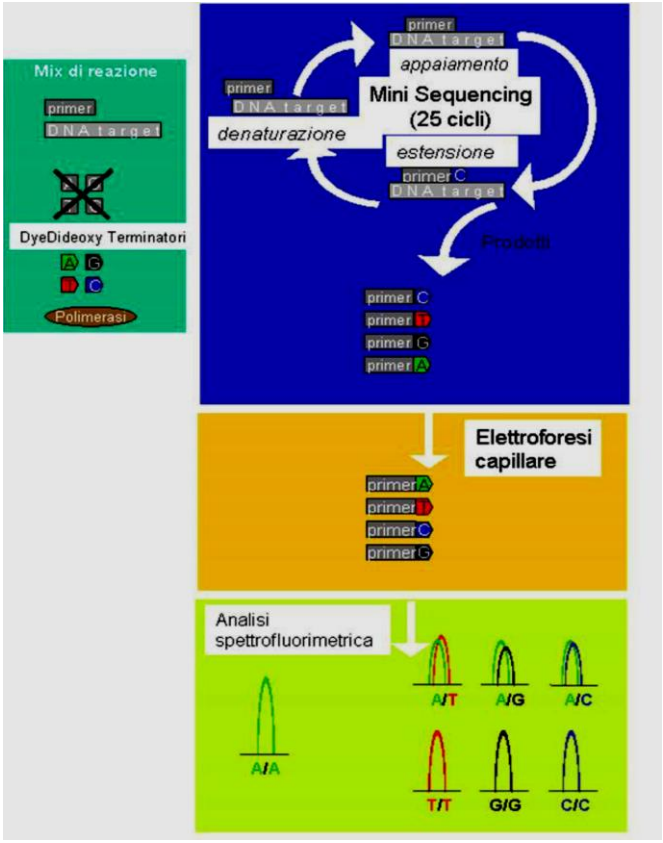


Figure 10. Schematic representation of the MiniSequencing technique.

The reaction mixture is prepared using the SNaPshot Multiplex Ready Reaction Mix (MRRM, Applied Biosystems), containing the DNA polymerase and fluorescently labeled ddNTPs. The reaction mix for a single sample is prepared with

MRRM and the CF-SNaP primers mix; finally, 2 µl of purified PCR product are added.

MM MiniSequencing:	**x 1**
1. MRRM	2.5 µl
2. CF-SNaP primers mix	0.5 µl
3. Sample (purified PCR product)	2.0 µl
	5.0 µl

The solution obtained is placed in the thermocycler with the following program:

- 96°C for 10 seconds
- 50°C for 5 seconds
- 60°C for 30 seconds

} 30 cycles

At the end of the MiniSequencing reaction, each sample undergoes a further purification treatment to remove the non-incorporated ddNTPs, that could interfere with the migration of the fragments of interest: for each sample an alkaline phosphatase treatment is carried out, in order to remove the phosphate groups at the 5' end of each ddNTPs (37°C for 60

minutes followed from 80°C for 15 minutes as inactivation step).

2.3.5 Cycle Sequencing

Sanger Sequencing is based on the principle that the DNA polymerase is unable to extend a nucleotide sequence after the insertion of a dideoxynucleotide (ddNTP); the latter, an analog of the deoxynucleotide (dNTP), is characterized by the absence of the hydroxyl group (OH) also in 3', necessary to promote the formation of the subsequent phosphodiesteric bond.

Each amplicon obtained by PCR reaction represents the template for two different sequencing reactions, performed with the internal primers, forward and reverse. Thus, with a single primer at a time, two sequencing reactions are performed, in which both the standard deoxynucleotides (dATP, dTTP, dCTP, dGTP) and the analogues fluorescents ddNTP (dideoxynucleotide) are mixed. The latter, once inserted into the polynucleotide chain in elongation, do not allow a further extension.

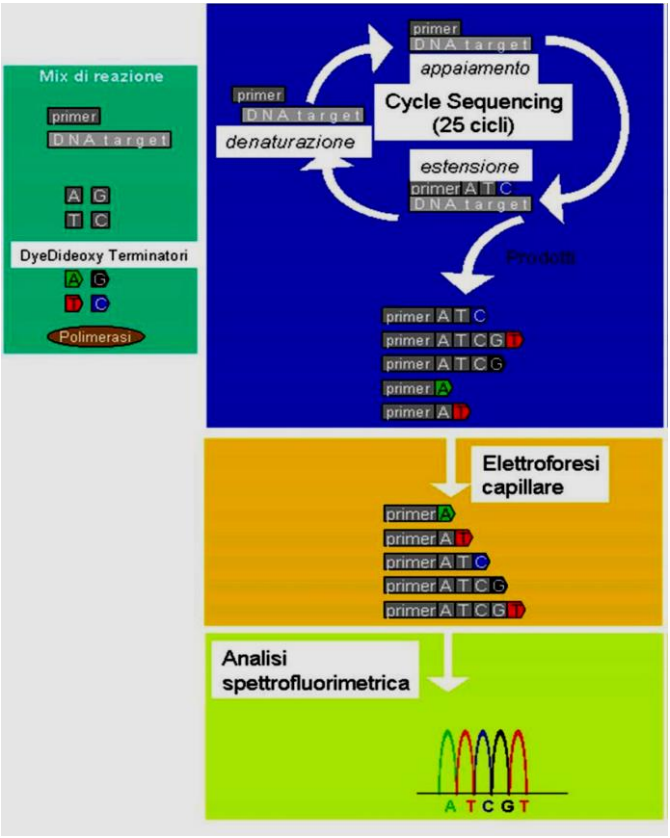


Figure 11. Schematic representation of the Cycle Sequencing technique.

The sequencing reaction mix is prepared using a kit provided by the Applied Biosystem (ABI PRISM Big Dye Terminator v.1.1 Cycle Sequencing Kit), consisting of the TRRM (Terminator Ready Reaction Mix) containing, among other

components, a mixture of dNTPs and ddNTPs labeled with four different fluorescent molecules; to this mix only the primer, purified by HPLC and corresponding to the DNA strand that has to be sequenced, must be added.

MM Cycle Sequencing:	**x 1**
1. TRRM (v.1.1)	0.5 µl
2. Buffer 5X	1.75 µl
3. Internal Primer Forward or Reverse (10 pmol/µl)	0.5 µl
4. H_2O	3.25 µl
5. Sample (purified PCR product)	4.0 µl
	10.0 µl

The reaction is placed in the thermocycler with the following program:

- 96°C for 1 minute
- 96°C for 10 seconds
- 50°C for 5 seconds
- 60°C for 4 minutes

} 30 cycles

To work in 96-well format, a single reaction mix is prepared, primers free, and divided into aliquots, to which the specific

primers for sequencing are added. These aliquots are dispensed into the plate and, subsequently, the amplicons are added. Each well contains, therefore, a single combination of amplicon and specific primer for sequencing. Starting from an amplification plate, two sequencing plates can be assembled, one for the forward strand and the other one for the reverse. At the end of the sequencing cycles, the samples undergo a further purification process with a QIAGEN (DyeEx 2.0 Spin Kit) chromatographic system, to remove the labeled dideoxynucleotides not incorporated: in this way, the ddNTPs are prevented from migrating with the fragments of interest and creating interference in the detection.

2.3.6 Capillary electrophoresis

Sequencing is evaluated with electrophoretic runs through a 36 cm long capillary. The apparatus used is the ABI PRISM 3130*xl* Genetic Analyzer (Applied Biosystems) with sixteen capillaries and POP6 as matrix.

After MiniSequencing: Before proceeding with the electrophoresis of the samples, a molecular weight standard labeled with a fifth fluorochrome (GeneScan-120 LIZ size standard) must be added; it is different from those with which

the ddNTPs are marked and it is specific for small fragments and necessary for the final analysis of the products. Capillary electrophoresis separates the different fragments that a specific software show as a series of peaks of different color; the data analysis is carried out using the ABI PRISM® GeneMapper software (Applied Biosystems): this program allows to verify the size of a fragment based on the migration of a molecular weight standard and to view which fluorescently labeled ddNTP was inserted. A single peak indicates that the variation of interest is found in homozygosis (wild-type or mutated), while a double peak indicates that it is found in heterozygosis.

After Cycle Sequencing: During the electrophoretic run, the fluorescence signal of the different fragments, detected by the CCD camera present in the instrument, is converted into a series of peaks of different colors, generally called an electropherogram. The obtained electropherogram is analyzed with the aid of the ABI PRISM® programs Sequencing Analysis and SeqScape (Applied Biosystems); every single peak is checked and the correct nucleotide call is evaluated, correcting where necessary. In this way any anomalies in the path are identified, such as the presence of overlapping peaks (possible mutation in heterozygosis), irregular peaks (possibly due to current variations), background disturbances

(attributable to poor quantity of the amplicons) or insufficient resolution of the peaks (attributable to possible polymer defects). The sequence obtained is compared to the wild-type sequence; this allows to highlight any areas of non-homology within experimental sequence.

Figure 12. Example of a project template created with the SeqScape software.

2.4 Expression analysis of the CFTR gene by RT-PCR

Retrotranscription is an *in vitro* synthesis of a cDNA conducted by the use of a reverse transcriptase enzyme. The total RNA extracted was retrotranscribed by means of the iScript cDNA Syntesis kit (Biorad), which consists of the reverse transcriptase already provided with RNase inhibitors, as well as random hexamers and a mixture of dNTPs. RT-PCR was performed using 1 µg of total RNA (for each sample) in a retrotranscription volume of 20 µl. As a negative control, a reaction without reverse transcriptase (RT) is prepared, with the same amount of RNA and of Reaction Mix.

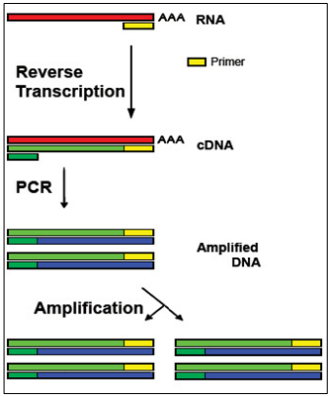


Figure 13. Schematic representation of the RT-PCR technique.

MM RT-PCR:	**x 1**
1. iScript Reaction Mix 5X	4.0 µl
2. RT enzyme	1.0 µl
3. H_2O	10.0 µl
4. Sample (200 ng/ µl)	5.0 µl
	20.0 µl

The reactions are then incubated in a thermocycler, according to a program that allows the synthesis of the double stranded cDNA:

- 25°C for 5 minutes
- 42°C for 30 minutes
- 85°C for 5 minutes

In order to amplify the cDNA obtained, specific pairs of primers are available, designed by my research group to cover the entire length of CFTR messenger (six overlapping fragments) and able to highlight the splicing of each exon *[Auriche et al., 2010]*. The fragments of CFTR cDNA analyzed are the following: 5'UTR-EX7 (1118 bp); EX6a-EX10 (865 bp); EX7-EX13 (780 bp); EX11-EX15 (1142 bp); EX13-EX19 (1141 bp); EX18-3'-UTR (1060 bp). The RT-PCR products are subsequently analyzed by electrophoretic run on a 1.5% agarose gel, to verify possible deletions/insertions and alternative splicing forms, observing the electrophoretic mobility of the amplicons and knowing exactly the expected dimension of each exon. In presence of an abnormal amplicon, we proceeded with the recovery of the amplicon from the gel and with the subsequent sequencing, to characterize the anomalies found. We also used the UVP's bioimaging and

analysis system “VisionWorks LS” (Life Science Software) to study these alterations by a semiquantitative densitometric analysis. This program allows to establish the optical density of each band. After the amplification and the electrophoretic run, we can visualize the bands on the gel, verify their dimension and calculate the percentage of both wild-type and mutated retrotranscribed mRNA. Bands with higher or lower molecular weight, in respect to wild-type mRNA, may highlight anomalous splicing.

2.5 Expression analysis of the CFTR gene by Real-time PCR

Real-time PCR is a method that simultaneously amplifies and quantifies DNA. Common methods of quantification include the use of fluorescent stains intercalating with double-stranded DNA (like in the SYBR Green method) and modified DNA oligonucleotides, called probes, that are fluorescent once hybridized with DNA and then degraded by the DNA polymerase during the elongation step of PCR amplification (like in the TaqMan method). It is therefore a quantitative

assay to estimate the expression levels of specific RNAs: in fact, the retro-transcription produces complementary cDNA, while maintaining unchanged the relative concentration ratios of the different RNA species present in the sample. In this way it is possible, for example, to measure the relative expression of a gene at a particular time, either in a cell or in a particular type of tissue.

The technology we decided to apply is that of fluorescent probes, that are synthetic oligonucleotide constructs (20/30 bp long). The use of a detector probe significantly increases the specificity, and allows regular quantification even in the presence of non-specific amplification products (on the contrary fluorescent dyes, such as SYBR Green, reveal the amplification of any DNA fragment). We use 3 TaqMan probes (Single Tube TaqMan® Gene Expression Assays) (**Figure 14**), provided by the Applied Biosystems (ThermoFisher Scientific), to study the variant tract (TG)mTn and the related splicing levels of CFTR exon 9:

- Exon8 – Exon9: probe used to study the percentage of CFTR mRNA with exon 9. It works when no splicing is present.

- Exon8 – Exon10: probe used to study the percentage of CFTR mRNA without exon 9. It works when splicing is detected.
- Exon18 – Exon19: probe used to study the total quantity of CFTR mRNA. It is an internal control, not related to exon 9.

These probes are designed at the edges of exons of interest and they are FAM dye-labeled.

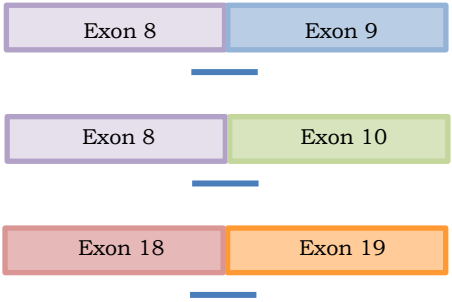


Figure 14. Schematic representation of the approach used for the study of the variant tract (TG)mTn with Real-time PCR and TaqMan probes. The CFTR exons of interest are represented with different colors; the blue lines at the esonic junctions represent the TaqMan probes and their relative positions.

An absolute quantification of the concentrations of specific RNAs can be carried out by producing a standard calibration curve; alternatively, a relative quantification can be performed

by comparing their quantity with respect to that of a control gene. Absolute quantification can use standard samples (for example plasmid DNA), whose absolute concentration is known. It must be certain, however, that the efficiency of the PCR is the same for all the samples. The relative quantification method, instead, requires the quantification of control genes or constitutive genes to normalize the expression of the studied ones. However, selecting the appropriate control genes can cause problems because they may not be expressed equally through all the unknown samples. This can be solved by normalizing the measurements with a set of housekeeping genes to avoid the possible variability. For our work we used an absolute reference, through the use of plasmids pCR®2.1 (Invitrogen), containing exon 9 (PL 9+) and not containing it (PL 9-); therefore, we created plasmid mixes with known concentrations of both PL 9+ and PL 9- and we constructed a calibration curve against which we compared the results of splicing of our samples. We evaluated the expression of total CFTR in the samples with the probe relative to Exon18 – Exon19, insensitive to exon 9 splicing.

For each probe a unique mix is initially prepared and, in addition to the samples, a reaction blank is also used to evaluate that no contamination is present. All Real-time PCR

reactions are performed in triplicate, both for the samples and for the reaction blanks. The mix for each probe contains: FluoCycle II™ Master Mix for Probe (EuroClone), that is a specific Real-time PCR mix in which reaction buffer, dNTPs, $MgCl_2$ and Taq DNA polymerase are already present; TaqMan® Gene Expression Assay (FAM dye-labeled probe), that contains both the specific probe and its relative primers; RNase-Free Water; cDNA of interest. The final volume for each reaction is 20 µl.

MM Real-Time PCR:	**x 1**
1. Master Mix FluoCycle II 2X	10.0 µl
2. TaqMan probe 20X	1.0 µl
3. H_2O	8.0 µl
4. Sample (50 ng/ µl)	1.0 µl
	20.0 µl

The Real-time PCR instrument is a Bio-Rad Opticon 2 and the program used is the following:

- 95°C for 5 minutes
- 95°C for 15 seconds
- 60°C for 1 minute

} 45 cycles

2.6 Statistical analysis

Statistical analysis were performed on all the data reported, to evidence statistically significant differences between target and control populations.

All data were analyzed using the ANOVA variance analysis test in the comparison of all the populations under investigation, with a $p<0.05$ considered statistically significant.

For the selection of intragenic haplotypes, we use the ANOVA variance analysis test with Bonferroni's correction for multiple tests in the comparison of all the populations under investigation, with a $p<0.001$ considered statistically significant. The specific analysis program used for this selection is "Haploview", that is a commonly used bioinformatics software which is designed to analyze and visualize patterns of linkage disequilibrium (LD) in genetic data. It can also perform association studies, choosing tagSNPs and estimating haplotype frequencies.

In the scatterplots, the regression curve was extrapolated and a correlation analysis (Pearson R correlation index) was performed, considered statistically significant with a $p <0.05$.

3. RESULTS

3.1 Analysis of intragenic haplotypes

The analysis of intragenic haplotypes was performed on all the seven populations under examination (CF-PI, CF-PS, CFTR-RD, CBAVD and ISHV as populations of interest, NORMO and GEN as controls). Following preliminary analyses starting from a higher number of markers, we selected 8 markers (6 SNPs and 2 repetitions): 125G/C, 875+40A/G, (GATT)n, 1001+11C/T, (TG)mTn, 1540A/G (M470V), 2694T/G, 4521G/A. Then, we obtained all the possible combinations, for a total of 48 haplotypes and, by means of statistical analysis, we highlighted those that show a significant difference in frequency among all the populations analyzed, for a total of 5 haplotypes. The most common haplotype in our populations of interest, that we considered as "target" for subsequent studies, is as follows: 125G, 875+40A, (GATT)6, 1001+11T, (TG)10T9, 1540A, 2694T, 4521G. The same analysis allowed us also to highlight a control haplotype, which results, unlike the first, significantly more frequent in the NORMO and GEN populations than in the populations of interest. The control

haplotype is as follows: 125G, 875+40A, (GATT)7, 1001+11C, (TG)11T7, 1540G, 2694T, 4521G. The first graph (**Figure 15**) shows the compared trend of these two haplotypes, based on their relative frequency in the seven populations. It is particularly important to underline that the target haplotype is present in 63% of the CF-PI alleles, in 33% of the CF-PS alleles, in 36% of the CFTR-RD alleles, in 20% of the CBAVD alleles and in 13% of ISHV alleles as compared to 6% in NORMO alleles and 3% in GEN alleles and, therefore, it has a trend that varies with the clinical condition of the population (in a statistically significant manner). In the second graph (**Figure 16**), the other 6 most frequent haplotypes are indicated, but they do not show a particular trend through the various populations, with frequency differences statistically significant only in 3 of them.

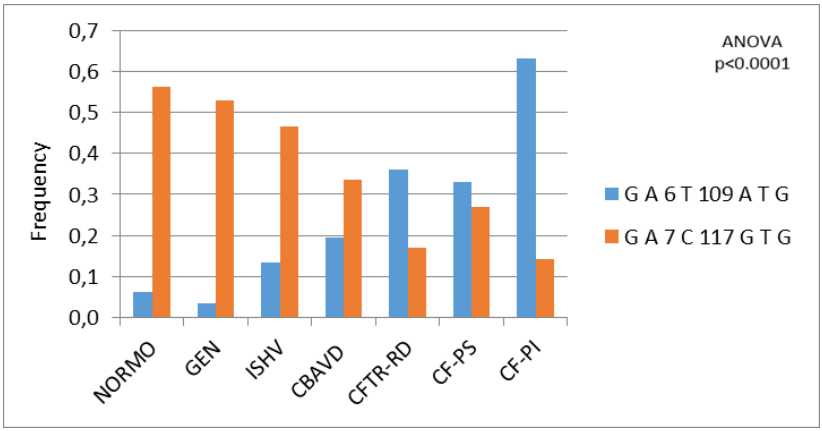


Figure 15. Frequencies of the target (in blue) and control (in orange) haplotypes in the populations under examination.

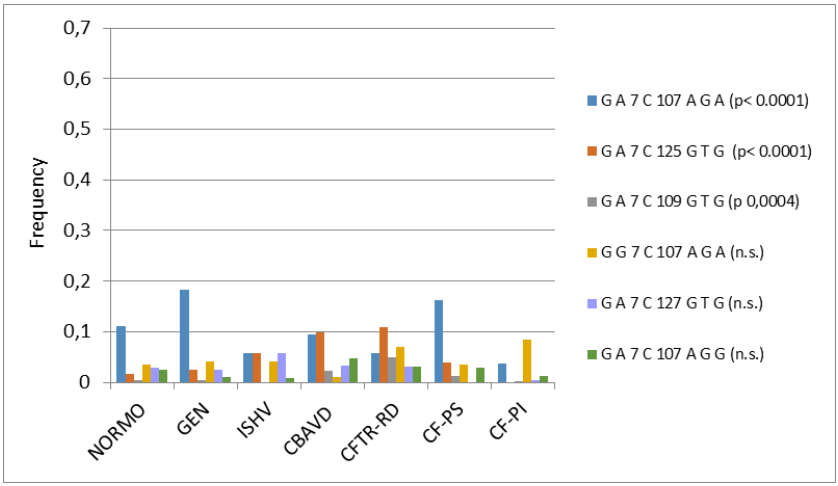
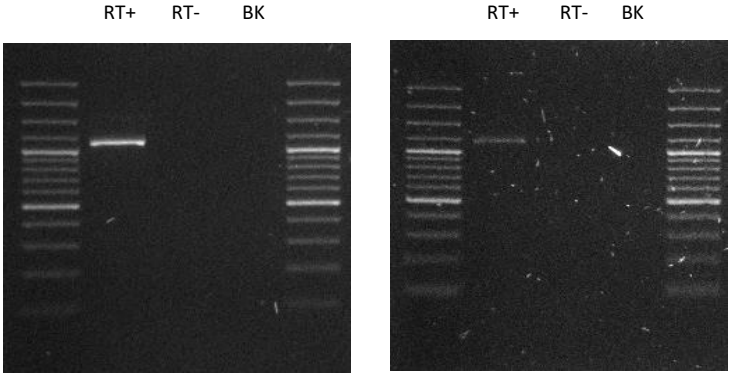


Figure 16. Frequencies of the remaining six haplotypes in the populations under examination.

After the structural analysis, the single sequence variations that compose the haplotypes were studied also from a functional point of view. One of them is located in the 5'-flanking region of the CFTR gene, 3 are intronic variants and 3 are silent exonic variants. Consequently, they can have a functional effect on the expression of CFTR mRNA. We studied this aspect through RT-PCR and agarose gel electrophoresis of the cDNA tracts with the variations of interest. By this approach, we were able to evaluate if their presence could in any way influence the levels of wild-type mRNA and, therefore, the functionality of the protein. The photos (**Figure 17**) show some examples of comparison between the individual variations included in the target haplotype (left for each pair) and the individual variations not included in the target haplotype (right for each pair), in homozygosis or heterozygosis (description under the photos). For each tract analyzed the length of the cDNA fragment and the haplotype sequence variations considered are indicated (description above the photos). It is evident that the individual sequence variations included in the haplotype do not produce any anomalous splicing, as also compared with the effect of individual sequence variations not included in the haplotype. Therefore, they are not able,

individually, to cause alterations in the wild type levels of CFTR mRNA.

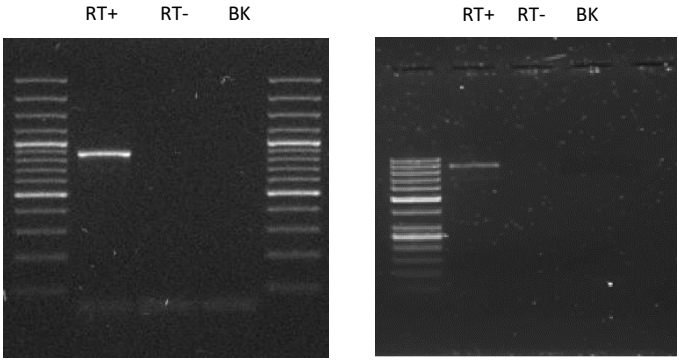
5'UTR-EX7 (1118 bp)
125 G/C; 875+40 A/G; 1001+11 C/T; GATT 6/7



125 G/G 875+40 A/A; 1001+11 T/T; GATT 6/6;

125 G/C; 875+40 A/A; 1001+11 T/T; GATT 6/7

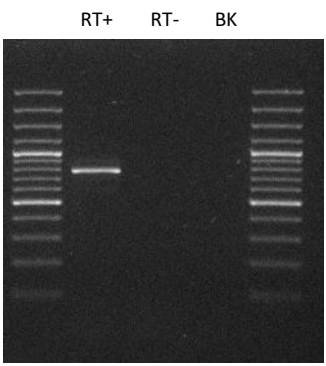
EX6a-EX10 (865 bp)
875+40 A/G; 1001+11 C/T; GATT 6/7; (TG)mTn



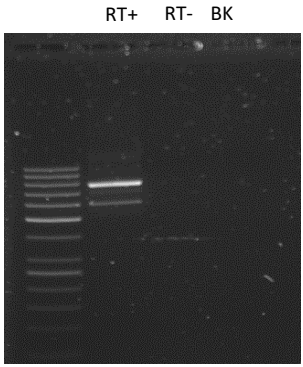
875+40 A/A; 1001+11 T/T; GATT 6/6; (TG)10T9 / (TG)10T9

875+40 A/G; 1001+11 C/C; GATT 7/7; (TG)10T7 / (TG)10T7

EX7-EX13 (780 bp)
M470V; (TG)mTn

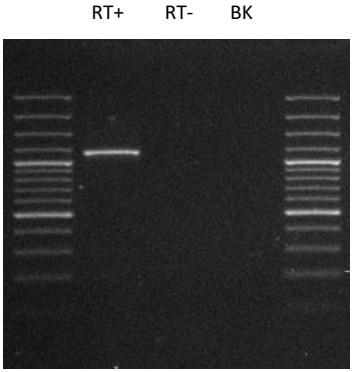


(TG)10T9 / (TG)10T9; 470 M/M

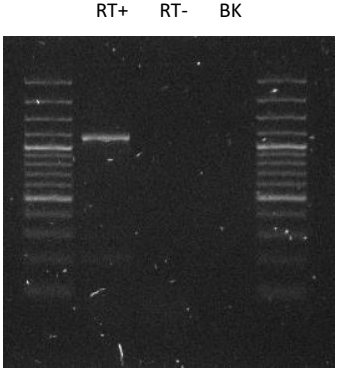


(TG)11T7 / (TG)11T7; 470 V/V

EX11-EX15 (1142 bp)
2694 T/G



2694 T/T



2694 G/G

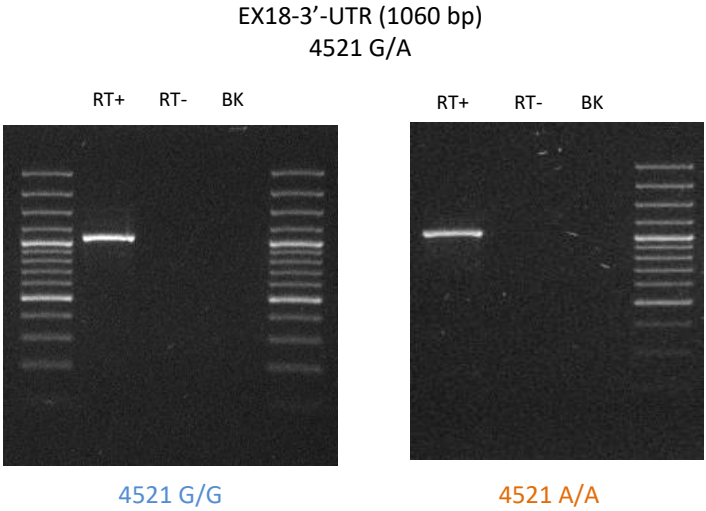


Figure 17. Study of the effect of the individual sequence variations of the target haplotype on the splicing of the corresponding mRNA tracts of the CFTR gene. On top of each pair of photos the tract considered, its specific molecular weight in bp and the variations of the haplotype that are found in the tract are indicated; on bottom of each pair of photos the variations considered are evidenced, in blue if included in the target haplotype (on left) and in orange if not (on right). Even if in some tracts more than a variation is indicated, the functional study considered them always individually.

3.2 Analysis of variant tracts (TG)mTn

Before this study, we performed a preliminary experimental functional analysis of the variants (TG)13T5, (TG)12T7,

(TG)11T5 and (TG)12T5 that, in agreement with the literature data, showed a higher percentage of abnormal splicing of exon 9. To verify if this alteration of the splicing could be partly responsible for the pathological phenotype of the populations under examination, we decided to analyze (TG)mTn variant tracts and to calculate their frequency. The graph (**Figure 18**) shows that the ISHV, CBAVD and CFTR-RD populations have a statistically significant higher cumulative frequency of these tracts than the other subjects, in particular of the (TG)12T5, which causes the highest percentage of abnormal splicing of exon 9. These patients, which are more characterized by variant tracts and sequence variations with limited functional effect rather than classical CF mutations, exhibit a CF-like clinical phenotype. In contrast, populations with a more severe CF clinical phenotype show a lower prevalence of these tracts (in agreement with a higher prevalence of severe mutations). We also noted that CBAVD patients have a higher frequency of the (TG)12T5 tract, correlated with the greater severity of their reproductive pathology, compared to ISHV patients in which the (TG)11T5 and (TG)12T7 tracts are more frequent and that show a less severe reproductive phenotype.

We decided to analyze also those (TG)mTn tracts that are not known to cause significant anomalous splicing in the CFTR gene and to calculate, also in this case, their frequency in our populations. The graph (**Figure 19**) shows that all the populations, both targets and controls, present a very similar cumulative frequency of these other tracts. Individually, the most frequent are: (TG)10T9 present in 66% of CF-PI patients; (TG)11T7 present in 59% of NORMO subjects; (TG)10T7 present in 27% of GEN subjects. In general, however, we found flat and non-statistically significant results, confirming the fact that these tracts are probably not involved in the variability of CF manifestations.

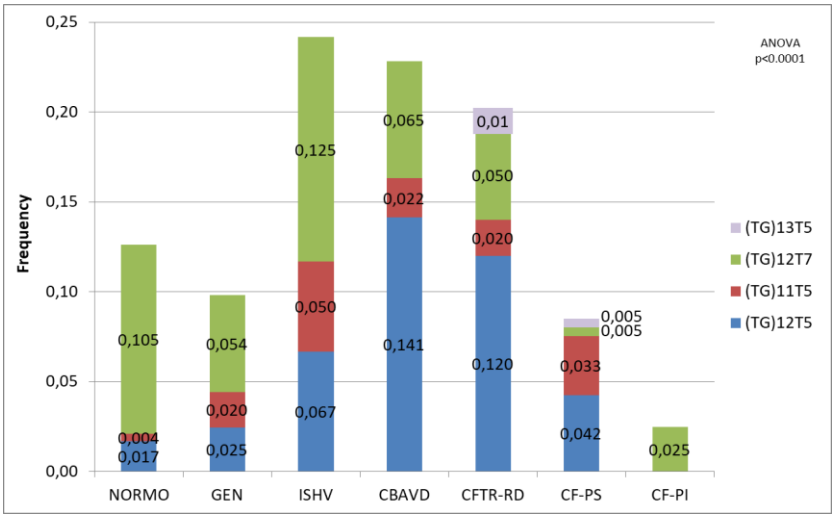


Figure 18. Frequencies of the variant tracts (TG)13T5, (TG)12T7, (TG)11T5 and (TG)12T5 in the populations under examination.

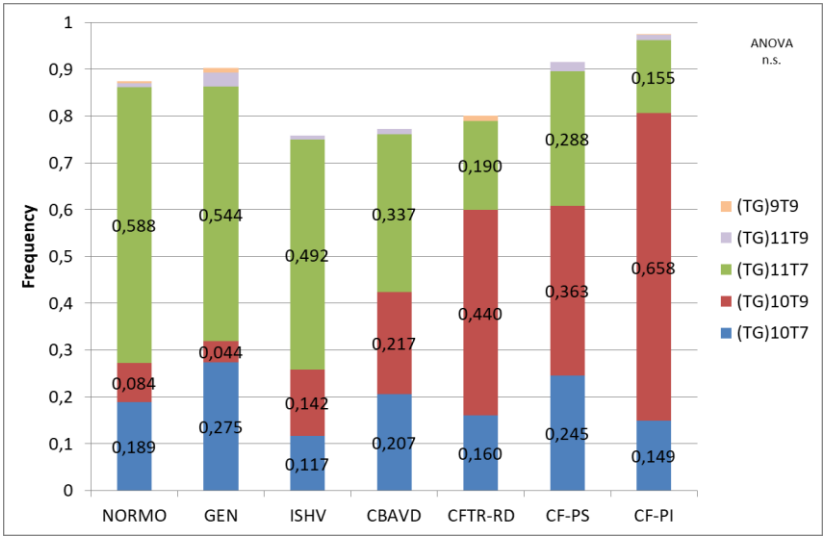


Figure 19. Frequencies of the other variant tracts (TG)mTn in the populations under examination.

At this point, we needed to characterize these variant tracts at a functional level, through RT-PCR and 1.5% agarose gel electrophoresis of the cDNA fragment in which exon 9 is present. This allowed to study how the different (TG)mTn can influence the processing of the CFTR mRNA and to assess the level of residual wild type mRNA after an anomalous splicing event of exon 9. We considered different genotypes, on which, through a specific densitometric analysis, we evaluated the percentage of wild-type mRNA (higher band that includes exon 9) and mutated mRNA (lower band that excludes the exon 9) compared to the total mRNA. For each genotype, the analysis was performed on at least 3 biological replicates and for each of them we considered at least 3 analytical replicates. In the figure below (**Figure 20**) for the genotypes (TG)12T5 / (TG)12T5 and (TG)10T9 / (TG)10T9 their different effect on the splicing of exon 9 are evidenced: (TG)12T5 / (TG)12T5 shows two bands, the higher one, wild-type and little visible, and the lower one, that resulted from a very marked exclusion of exon 9; on the contrary, (TG)10T9 / (TG)10T9 shows a single band (the wild-type one) evidencing that this tract does not cause a significant anomalous splicing of the CFTR.

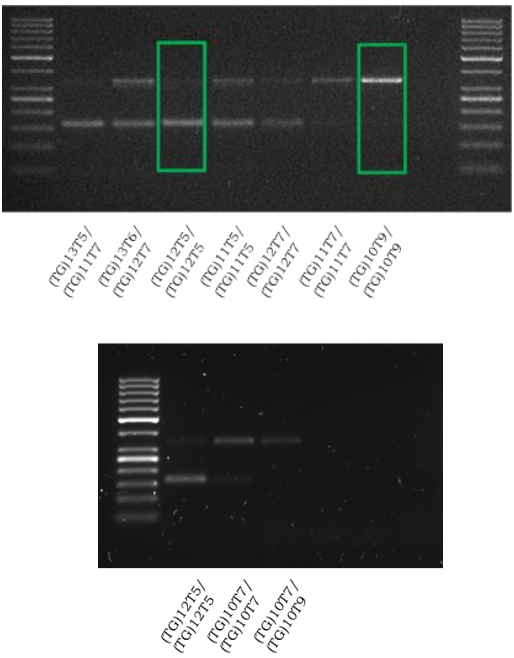


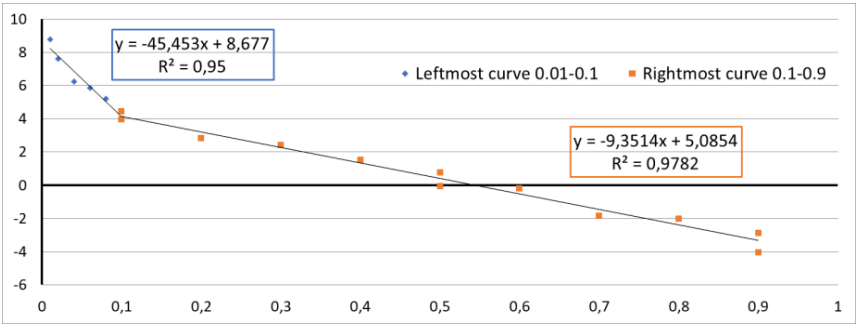
Figure 20. Examples of 1.5% agarose gel electrophoresis of RT-PCRs of different genotypes (TG)mTn of interest.

The semi-quantitative densitometric analysis revealed the average percentages of splicing, summarized in **Table 3**. The (TG)mTn genotype is shown in the first column; in the second and third columns there are the averages of the densitometric analysis of the results obtained for both the higher band (Exon 9+) and the lower band (Exon 9-). As shown, the genotypes including variant tracts (TG)11T5, (TG)13T5, (TG)12T7 and

(TG)12T5 show lower percentages of higher band (wild-type band with exon 9) and higher percentages of lower band (mutated band without exon 9); it does not happen for any of the other tracts. Even, the genotypes (TG)10T9 / (TG)10T9 and (TG)10T7 / (TG)10T9 show a percentage of abnormal splicing, respectively, equal or near to 0. This confirms and extends what is already known in literature and well correlates with our (TG)mTn frequency study. Those tracts, able to significantly reduce the amount of wild-type mRNA and consequently of functional protein, are those more frequent in the target populations. Moreover, it should be noted that the variant tract (TG)12T7, which is usually considered not to cause a high percentage of abnormal splicing, in one of our cases produces, in homozygosis, an 81.4% of mutated transcript.

The results obtained with densitometric analysis were verified through the use of specific TaqMan probes in Real-time PCR. For this quantitative study, it was initially necessary to construct the reference plasmidic curve, using mixes with known concentrations of plasmid containing exon 9 (PL CF 9+) and plasmid not containing exon 9 (PL CF 9-). The curve was based on the logarithmic relationships between plasmids: in particular we obtained two different trends, with two different formulas, and we calculate the percentage of splicing

between 1% and 10% with the leftmost curve (in blue) and the percentage of splicing between 10% and 90% with the rightmost curve (in orange) (**Figure 21**).



PL CF 9-	PL CF 9+
1%	99%
2%	98%
4%	96%
6%	94%
8%	92%
10%	90%

PL CF 9-	PL CF 9+
10%	90%
20%	80%
30%	70%
40%	60%
50%	50%
60%	40%
70%	30%
80%	20%
90%	10%

Figure 21. Plasmidic curve of the logarithmic relationships between plasmid containing exon 9 and plasmid not containing exon 9. The concentrations of each plasmid are indicated in the tables on the bottom, both for the leftmost curve (in blue) and for the rightmost curve (in orange).

The Real-time PCR results are illustrated in **Table 3**. The (TG)mTn genotype is shown in the first column; in the fourth and fifth columns there are the averages of the Real-time PCR

analysis of the results obtained for both cDNA with exon 9 (Exon 9+) and cDNA without exon 9 (Exon 9-). Also in this case, the genotypes including variant tracts (TG)11T5, (TG)13T5 and (TG)12T5 show low percentages of wild-type mRNA and high percentages of mutated mRNA. These results confirm, also by Real-time PCR, that these tracts are able to greatly influence the quantity of CFTR wild-type mRNA.

Table 3. Summary of the average percentages of Exon 9+ and Exon 9- in the functional analysis of variant tracts (TG)mTn. The (TG)mTn genotype is shown in the first column; in the second and third columns there are the averages of the densitometric analysis; in the fourth and fifth columns there are the averages of the Real-time PCR analysis. In violet the genotypes for which both densitometry and Real-time PCR analysis were performed, in pink those for which only the densitometric analysis was carried out, in light blue those for which only the Real-time PCR analysis was carried out.

(TG)mTn	DENSITOMETRIC ANALYSIS AVERAGES		REAL-TIME PCR ANALYSIS AVERAGES	
	Exon 9+	Exon 9-	Exon 9+	Exon 9-
(TG)10T9 / (TG)10T9	100,0	0,0	90,2	9,8
(TG)10T7 / (TG)11T7	87,1	12,9	87,2	12,8
(TG)10T7 / (TG)10T9	98,4	1,6	86,8	13,2
(TG)10T9 / (TG)11T7	94,9	5,1	77,0	23,0
(TG)10T7 / (TG)10T7	82,7	17,3	76,2	23,8
(TG)10T7 / (TG)12T7	87,5	12,5	73,0	27,0
(TG)11T7 / (TG)11T7	82,0	18,0	61,6	38,4
(TG)13T6 / (TG)12T7	38,7	61,3		
(TG)12T5 / (TG)12T7			36,0	64,0
(TG)11T5 / (TG)11T7			35,4	64,7
(TG)10T9 / insT	37,7	62,3	34,4	65,6
(TG)13T5 / (TG)11T7	20,1	79,9	29,8	70,2
(TG)11T5 / (TG)11T5	33,6	66,4	29,2	70,8
(TG)12T7 / (TG)12T7	18,6	81,4		
(TG)12T5 / (TG)12T5	12,7	87,3	7,2	92,8

In **Table 4** the individual percentages of Exon 9+ and Exon 9- of the subjects included in Real-time PCR functional studies were reported, for a total of 31 individuals. They were divided according to their (TG)mTn genotype, with increasing average splicing values, as in **Table 3**. In the first column the numeration is indicated; in the second the CF population to which the subject belongs; in the third the basic genotype; in the fourth the (TG)mTn genotype; in the fifth the starting material from which the RNA was extracted; in the sixth and seventh the individual percentages of Exon 9+ and Exon 9-; in the octave and in the ninth the average percentages of Exon 9+ and Exon 9- (as in **Table 3**). Some subjects are parents or partners of CF patients, on which we had already carried out other studies, and were included in this study because their RNA was available and their (TG)mTn was of interest. For three individuals we do not have Real-time PCR data and only the percentages obtained from the densitometric analysis are presented (they are indicated with a red label on the right of the table).

Table 4. List of the subjects included in Real-time PCR functional studies. In the first column there is the numeration; in the other columns it is indicated the CF population, the basic genotype, the (TG)mTn, the RNA collection, the individual and the average percentages. The values obtained only with densitometric analysis are indicated with *DENS (in red) on the right of the table.

NUMERATION	CF POPULATION	GENOTYPE	(TG)mTn	RNA COLLECTION	INDIVIDUAL PERCENTAGES		AVERAGE PERCENTAGES ± DEV. STANDARD		
					Exon 9+	Exon 9-	Exon 9+	Exon 9-	
1	CF	F508del / 1585-9412A>G	(TG)10T9 / (TG)10T9	Brushing	94.0	6.0	90.2 ± 3.6	9.8 ± 3.6	
2	CF	F508del / F508del	(TG)10T9 / (TG)10T9	CRC cells	89.9	10.1			
3	CF	[F508del;I1027T] / F508del	(TG)10T9 / (TG)10T9	CRC cells	86.8	13.2			
4	GEN	[G576A;R668C] / N	(TG)10T7 / (TG)11T7	CRC cells	87.2	12.8	87.2 ± 0.0	12.8 ± 0.0	
5	CF	W1282X / [R117L;L997F]	(TG)10T9 / (TG)10T7	Brushing	90.9	9.1	86.8 ± 9.1	13.2 ± 9.1	
6	CF	F508del / 328del538	(TG)10T9 / (TG)10T7	Brushing	91.9	8.1			
7	CF	F508del / del Ex14b-17b	(TG)10T9 / (TG)10T7	Brushing	91.2	8.8			
8	CF	F508del / UN	(TG)10T9 / (TG)10T7	Brushing	73.2	26.8			
9	CF	F508del / L558S	(TG)10T9 / (TG)11T7	CRC cells	77.0	23.0	77.0 ± 0.0	23.0 ± 0.0	
10	GEN	G576A / G576A	(TG)10T7 / (TG)10T7	Brushing	84.0	16.0	76.2 ± 11.0	23.8 ± 11.0	
11	CF FATHER	dupl Ex19 / N	(TG)10T7 / (TG)10T7	Brushing	68.4	31.6			
12	GEN	N / N	(TG)10T7 / (TG)10T7	Brushing	75.3	24.7	75.3 ± 0.0	24.7 ± 0.0	*DENS
13	CF	[G576A;R668C] / S1235R	(TG)10T7 / (TG)12T7	Brushing	73.0	27.0	73.0 ± 0.0	27.0 ± 0.0	
14	CF	R553X / del Ex2	(TG)11T7 / (TG)11T7	Brushing	58.8	41.2	61.6 ± 5.9	38.4 ± 5.9	
15	CF	UN / UN	(TG)11T7 / (TG)11T7	Brushing	70.1	29.9			
16	CF MOTHER	R553X / N	(TG)11T7 / (TG)11T7	Brushing	58.4	41.6			
17	GEN	G85E / N	(TG)11T7 / (TG)11T7	CRC cells	55.7	44.3			
18	GEN	G85E / N	(TG)11T7 / (TG)11T7	Brushing	65.2	34.8			
19	GEN	N / N	(TG)13T6 / (TG)12T7	Brushing	38.7	61.3	38.7 ± 0.0	61.3 ± 0.0	*DENS
20	CFTR-RD MOTHER	12-5 / N	(TG)12T5 / (TG)12T7	Brushing	36.0	64.0	36.0 ± 0.0	64.0 ± 0.0	
21	CFTR-RD FATHER	L1065P / N	(TG)11T5 / (TG)11T7	Brushing	35.4	64.7	35.4 ± 0.0	64.7 ± 0.0	

Table 4 follows from the previous page.

NUMERATION	CF POPULATION	GENOTYPE	(TG)mTn	RNA COLLECTION	INDIVIDUAL PERCENTAGES		AVERAGE PERCENTAGES ± DEV. STANDARD		
					Exon 9+	Exon 9-	Exon 9+	Exon 9-	
22	CF	F508del / ins317	(TG)10T9 / insT	CRC cells	34.8	65.2	34.4 ± 2.7	65.6 ± 2.7	
23	CF	F508del / ins358	(TG)10T9 / insT	CRC cells	32,0	68,0			
24	CF	F508del / ins317	(TG)10T9 / insT	Brushing	38.1	61.9			
25	CF	F508del / ins358	(TG)10T9 / insT	Brushing	32.8	67.2			
26	GEN	13-5 / N	(TG)13T5 / (TG)11T7	Brushing	33.9	66.1	29.8 ± 5.8	70.2 ± 5.8	
27	CF MOTHER	907delCins29 / 13-5	(TG)13T5 / (TG)11T7	Brushing	25.7	74.3			
28	GEN	N / N	(TG)11T5 / (TG)11T5	Brushing	29.2	70.8	29.2 ± 0.0	70.8 ± 0.0	
29	CF FATHER	N / UN	(TG)12T7 / (TG)12T7	Brushing	18.6	81.4	18.6 ± 0.0	81.4 ± 0.0	*D ENS
30	CF	[359insT;12-5] / 12-5	(TG)12T5 / (TG)12T5	Brushing	14.4	85.6	7.2 ± 10.2	92.8 ± 10.2	
31	ISHV	(TG)12T5 / (TG)12T5	(TG)12T5 / (TG)12T5	Brushing	0.0	100.0			

After evaluating the average percentages of wild-type mRNA with exon 9 and mutated mRNA without exon 9 in the genotypes (TG)mTn of interest, we compared the results obtained in the previous densitometric analysis and those obtained in Real-time PCR, as shown in **Figure 22** and in **Figure 23**. The trend of the two curves, for the (TG)mTn tracts studied, shows that there is a good correlation between the two techniques and that they can be used together to study the splicing of CFTR exon 9, although with a few exceptions.

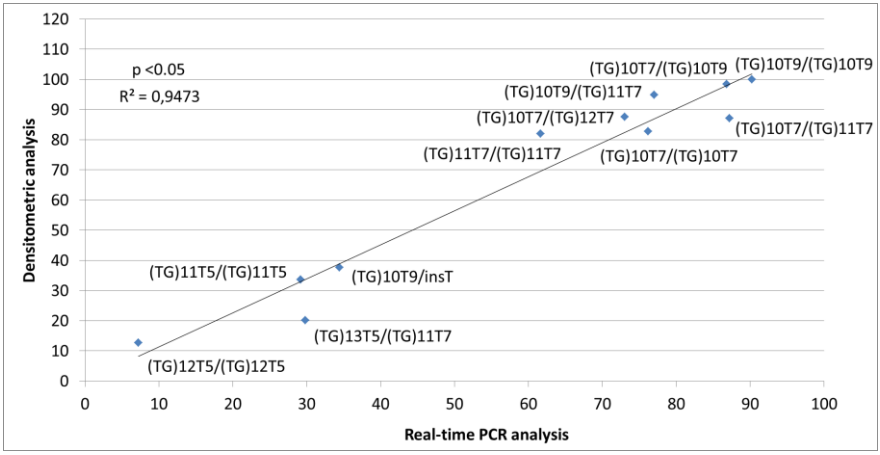


Figure 22. Comparison between densitometric analysis and Real-time PCR analysis in the study of percentage of wild-type CFTR mRNA (Exon 9+).

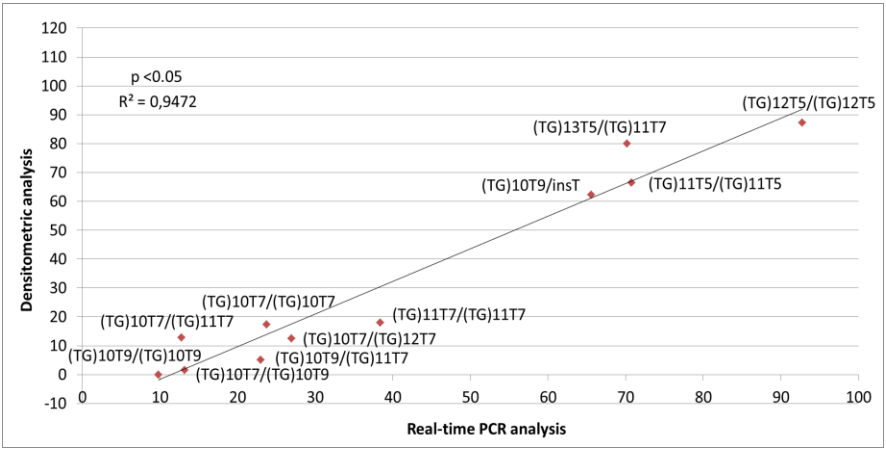


Figure 23. Comparison between densitometric analysis and Real-time PCR analysis in the study of percentage of mutated CFTR mRNA (Exon 9-).

These results allowed us to summarize the functional influence of all the (TG)mTn genotypes included in this study on CFTR splicing, estimating the final percentages of wild-type (Exon 9+) and mutated (Exon 9-) mRNA, as shown in **Figure 24** and in **Figure 25**, respectively. In general, we can assert that: genotypes with (TG)10T9, (TG)10T7, (TG)11T7 show high percentage of Exon 9+ and low percentage of Exon 9-; genotypes with (TG)11T5, (TG)12T5, (TG)13T5 show low percentage of Exon 9+ and high percentage of Exon 9-; genotypes with (TG)12T7 show a variable effect, spanning from a mild reduction of Exon 9+ to a great decrease of the percentage of Exon 9+, as for example for (TG)12T7 in homozygosis.

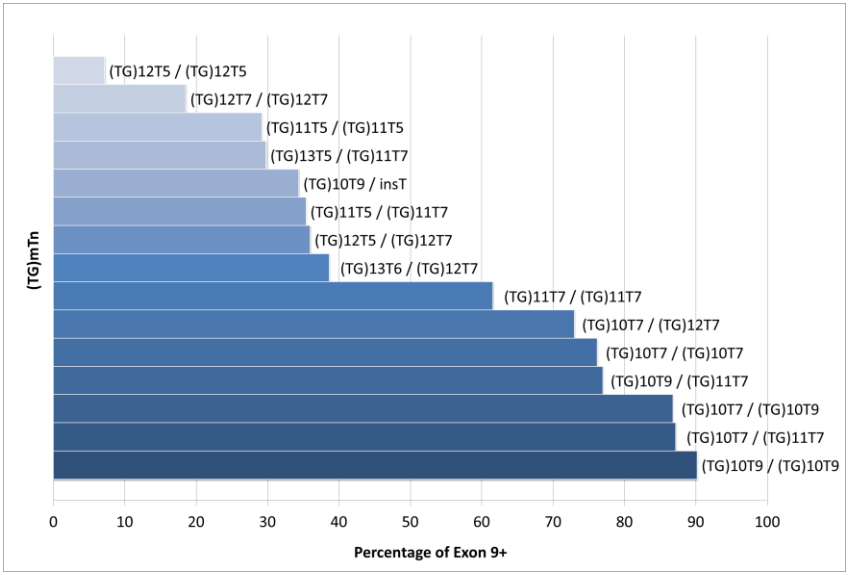


Figure 24. Percentage of wild-type mRNA (Exon 9+) in all the variant tracts (TG)mTn included in this study.

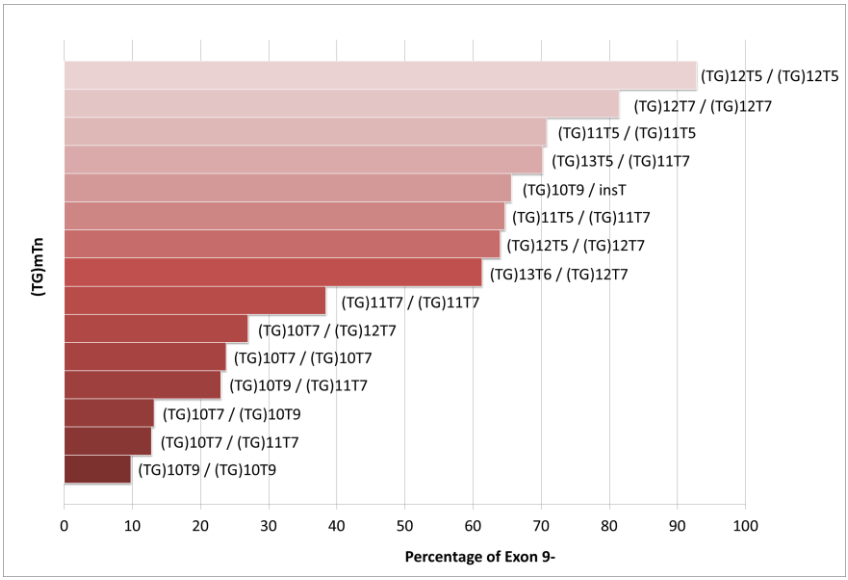


Figure 25. Percentage of mutated mRNA (Exon 9-) in all the variant tracts (TG)mTn included in this study.

After completing the analysis on (TG)mTn genotypes, we calculated the contribution of each single (TG)mTn allele on the final percentage of Exon 9+ and Exon 9-, according to our experiments. First of all, the average percentages of Exon 9+ and Exon 9- in homozygotes (TG)mTn genotypes, such as (TG)10T9 / (TG)10T9, (TG)10T7 / (TG)10T7, (TG)11T7 / (TG)11T7, (TG)11T5 / (TG)11T5 and (TG)12T5 / (TG)12T5, were evidenced. Assuming that the contribution of each single allele was 50%, the values of homozygotes were divided by two, to obtain the percentage of the contribution of (TG)10T9, (TG)10T7, (TG)11T7, (TG)11T5 and (TG)12T5 single alleles (**Table 5A**). Then the contribute of the other (TG)mTn alleles, such as (TG)12T7, insT and (TG)13T5, were derived through an indirect calculation based on the previous ones, subtracting to the total average percentages of Exon 9+ and Exon 9- those obtained from homozygotes (**Table 5B**). We decided to exclude from this analysis the (TG)12T7 / (TG)12T7, for its singular and unexpected percentages of Exon 9+ and Exon 9- (that could be evaluated only by densitometric analysis), and the (TG)13T6, because it is a rare and little known variant tract and its calculation depends only on (TG)12T7. In **Table 5C** there is a summary of the average percentages of Exon 9+ and Exon 9- for each allele. We notice that insT and (TG)13T5

produce only anomalous splicing and, therefore, not functional CFTR transcript (Exon 9-). (TG)12T7 was calculated as an average of the values obtained from (TG)10T7 / (TG)12T7 and (TG)12T5 / (TG)12T7 (having subtracted the contribute of the first allele).

Table 5A. Average percentages of Exon 9+ and Exon 9- for the alleles (TG)10T9, (TG)10T7, (TG)11T7, (TG)11T5 and (TG)12T5 (direct calculation of their contribution on CFTR functionality).

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)10T9	(TG)10T9	(TG)10T9 / (TG)10T9
Exon 9+	45,1	45,1	90,2
Exon 9-	4,9	4,9	9,8
TOTAL	50,0	50,0	100,0

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)10T7	(TG)10T7	(TG)10T7 / (TG)10T7
Exon 9+	38,1	38,1	76,2
Exon 9-	11,9	11,9	23,8
TOTAL	50,0	50,0	100,0

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)11T7	(TG)11T7	(TG)11T7 / (TG)11T7
Exon 9+	30,8	30,8	61,6
Exon 9-	19,2	19,2	38,4
TOTAL	50,0	50,0	100,0

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)11T5	(TG)11T5	(TG)11T5 / (TG)11T5
Exon 9+	14,6	14,6	29,2
Exon 9-	35,4	35,4	70,8
TOTAL	50,0	50,0	100,0

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)12T5	(TG)12T5	(TG)12T5 / (TG)12T5
Exon 9+	3,6	3,6	7,2
Exon 9-	46,4	46,4	92,8
TOTAL	50,0	50,0	100,0

Table 5B. Average percentages of Exon 9+ and Exon 9- for the alleles (TG)12T7, insT and (TG)13T5 (indirect calculation of their contribution on CFTR functionality, based on **A**).

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)10T7	(TG)12T7	(TG)10T7 / (TG)12T7
Exon 9+	38,1	34,9	73,0
Exon 9-	11,9	15,1	27,0
TOTAL	50,0	50,0	100,0

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)12T5	(TG)12T7	(TG)12T5 / (TG)12T7
Exon 9+	3,6	32,4	36,0
Exon 9-	46,4	17,6	64,0
TOTAL	50,0	50,0	100,0

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)10T9	insT	(TG)10T9 / insT
Exon 9+	45,1	-10,7	34,4
Exon 9-	4,9	60,7	65,6
TOTAL	50,0	50,0	100,0

	ALLELE 1	**ALLELE 2**	**(TG)mTn GENOTYPE**
	(TG)13T5	(TG)11T7	(TG)13T5 / (TG)11T7
Exon 9+	-1,0	30,8	29,8
Exon 9-	51,0	19,2	70,2
TOTAL	50,0	50,0	100,0

Table 5C. Summary of the average percentages of Exon 9+ and Exon 9- for each allele.

(TG)mTn ALLELE	**AVERAGE PERCENTAGES**	
	Exon 9+	Exon 9-
(TG)10T9	45,1	4,9
(TG)10T7	38,1	11,9
(TG)11T7	30,8	19,2
(TG)11T5	14,6	35,4
(TG)12T5	3,6	46,4
(TG)12T7	33,7	16,4
insT	0,0	50,0
(TG)13T5	0,0	50,0

In this work on CFTR (TG)mTn we evidenced three particular variant tracts, that were less considered or still not characterized in literature, for their complex functional activity on CFTR splicing:

i. (TG)12T7 / (TG)12T7: this (TG)mTn genotype was found once in our case series and it was studied only with densitometric analysis. Our studies underlined that this sample showed a high percentage of abnormal splicing of exon CFTR 9 (81.4%), never found in literature in patients with this (TG)mTn genotype. We couldn't examine it with Real-time PCR because the RNA analyzed was no longer available: in fact, this subject was inserted into our case series as a father of a CF patient, to study the mutations of the affected child (F508del/unknown). All the investigations were carried out following the standard protocols (using the first, the second and the third level of CFTR genetic analysis showed in **Table 1**, but also more specific and in-depth techniques), both on the father and on the son, but no other mutations were found. For these reasons we assume that the father may be a healthy carrier, with a sufficient percentage of functional protein. On the

contrary, his son, presenting a severe mutation on an allele and this particular (TG)12T7 on the other, shows the classical form of the pathology, because the reduction of functional protein is considerable. From the molecular point of view, we could speculate that this particular case may depend from the functionality of splicing machinery that, due to its peculiarity in this (TG)12T7 homozygous sample, can induce in a quote of anomalous splicing greater than usually. Obviously, additional studies are mandatory. Now we need to study this interesting sample with Real-time PCR and confirm this data, to find other (TG)12T7 / (TG)12T7 patients to study their percentage of splicing and to evaluate other causes that can contribute to this clinical picture. The (TG)12T7 was analyzed also with other variant tracts (TG)mTn in Real-time PCR: for example it shows a percentage of 27% of splicing when on the other allele there is a (TG)10T7 and a percentage of 64% of splicing when it is coupled with a (TG)12T5. These data support the hypothesis that the (TG)12T7 tract influences the splicing of CFTR exon 9 in a very variable manner.

ii. (TG)13T6 / (TG)12T7: also this (TG)mTn genotype was found just in a subject of our case series and it was studied only with densitometric analysis. It is a singular (TG)mTn because an allele presents a repetition of 13 TG followed by a repetition of 6 T: there are no case in literature in which this combination appears. Considering that the (TG)12T7 has a very variable influence on exon 9 splicing, we could not calculate the exact reduction that the single (TG)13T6 may cause on the final CFTR functional mRNA. Our purpose in the future is to study this sample also in Real-time PCR, to have a more accurate feedback regarding this genotype, and, if we have the possibility, to study other (TG)mTn genotype containing the (TG)13T6 and/or to perform *in vitro* studies by expression minigenes.

iii. (TG)10T9 / insT: two CF subjects presented this novel (TG)mTn genotype. We found that these patients had a classical (TG)10T9 on the first allele and a disrupted (TG)mTn on the second. It was not properly a variant tract (TG)mTn: in fact, we evidenced an insertion, coming from the intron 9 and containing a series of repetitions of T, introduced into the (TG)mTn (intron 8), after the duplication of a portion of DNA. In

particular a patient showed an insertion of 317 bp (ins317), while the other an insertion of 358 bp (ins358): the only difference is related to the different number of T and this allows to speculate on a molecular mechanism of divergence from a common ancestral allele. This molecular lesions appear to alter the normal splicing process of CFTR. In particular, we studied this samples both by densitometric analysis and Real-time PCR and obtained percentages of anomalous transcript over 50%. If we consider the very limited contribution of (TG)10T9 allele to exon 9 anomalous splicing, as evidenced in literature and in our experiments, we can conclude that these insertions completely suppress the physiological splicing. Consequently, it appears to cause a reduction of functional protein, if matched with a severe mutation of CFTR on the other allele, that results in a classical CF form. This allows to highlight it as a novel CF-causing mutation of CFTR.

3.3 Analysis of CFTR 3'-UTR

This part of the study focused on the analysis of CFTR 3'-UTR; in particular, we searched for the sequence variations that may influence the action of miRNAs. Initially, a bibliographic, bioinformatic and database research was carried out with the aim of highlighting the so far known miRNAs able to bind the 3'-UTR of CFTR. The results of this collection is summarized in **Table 6**. The first column shows the list of miRNAs verified with in silico analysis (prediction analysis through sites such as TargetScan, MicroRNA.org, PITA, miRDB, Microcosm, miRBase, miRanda, PicTar, EIMMo); the second and third columns show for which miRNAs the binding *in vitro* has been verified, either using a luciferase assay (qualitative binding, second column) either verifying if their presence alters in some way the levels of wild-type mRNA produced (quantitative binding, third column); the fourth column shows the only two miRNAs for which the functional effect has been currently verified also *in vivo* after nasal epithelial brushing from patients; the last column shows the miRNAs with known specific binding site on the 3'-UTR, as a result of mutagenesis assays.

Table 6. Summary of the known CFTR miRNAs. In the first column there is the list of all the miRNAs predicted with in silico analysis; in the other columns it is indicated if they were verified *in vitro*, *in vivo* and if their binding site on CFTR 3'-UTR is known.

miRNAs (In silico analysis)	Verified *in vitro*		Verified *in vivo* (after nasal epithelial brushing of patients)	Binding site known (Mutagenesis analysis)
	Qualitative binding (Luciferase assay)	Quantitative binding (Regulation of CFTR mRNA)		
mir-145	√	√		427-434
mir-377	√	√		
mir-384	√	√		1113-1125 (ARE-5698)
mir-494	√	√		1140-1146
mir-1827	√	√		
mir-939	√	√		
mir-331-3p	√	√		
mir-607	√	√		
mir-600	√	√		1498-1516 (ARE-6074)
mir-376a/b	√	√		
mir-1290	√	√		
mir-1246	√	√		
mir-101	√	√		1501-1530
mir-144	√	√		1501-1530
mir-199-3p	√			
mir-345	√			
mir-380	√			
mir-509-3p	√	√	√	1042-1066 (c.*1043A>C)
mir-645	√			
mir-223	√	√		1473-1481
mir-31				
mir-362-5p				
mir-433	√	√	√	1042-1066 (c.*1043A>C)
mir-505	√	√		
mir-943	√	√		
mir-942	√			
mir-665	√			
mir-383	√			
mir-451				
mir-150				
mir-221	√	√		
mir-126				
mir-9				
mir-93				
mir-224	√	√		
tot. 35	**tot. 28**	**tot. 20**	**tot. 2**	**tot. 9**

Once we highlighted the CFTR 3'-UTR zone potential target of known miRNAs, we planned its sequencing. To this purpose, we divided this zone into five portions (**Figure 7**). The analysis was carried out on subjects belonging to three specific categories, for a total of 108 patients, as they were the best candidates for these transcriptional regulation studies: patients presenting the same mutated genotype (two mutations found), but different phenotype (belonging to different CF populations); F508del homozygous patients, all CF-PI but with very different levels of the sweat tests; patients to whom only one of the two mutations causing the disease was found (**Table 7**). The variations found on 3'-UTR are the following:

a) *81C/T (4575+81C/T) in heterozygosis;

b) *567C/T (4575+567C/T) in heterozygosis;

c) *1251C/T (4575+1251C/T) both in heterozygosis and in homozygosis.

All the variations resulted already known: the *1251C/T has already been characterized, with a clinical significance of "likely benign" variation; the *81C/T and the *567C/T, instead, have not yet been characterized. The variations found do not fall within the miRNA binding zones shown in the last column of **Table 6**, so their meaning at the functional level is still uncertain.

Table 7. List of the patients selected for the study on CFTR 3'-UTR. In the first column the genotype considered is indicated for the three categories of patients; in the second column the number of patients with the genotype of interest is evidenced; in the third column there are the CF populations to which each patient belongs; in the last column the 3'-UTR variations found for that specific patient are indicated: in red the *81C/T (4575+81C/T) in heterozygosis, in blue the *567C/T (4575+567C/T) in heterozygosis, in black the *1251C/T (4575+1251C/T) in heterozygosis and in green the *1251T/T (4575+1251T/T) in homozygosis. For some genotypes, we had the availability of only one patient's DNA, but they were inserted the same because showed the variations of interest.

Genotype	Numeration	CF population	3'UTR variations found
G542X / (TG)12T5	1	CBAVD	*1251C/T (4575+1251C/T)
G542X / (TG)12T5	2	CFTR-RD	*567C/T (4575+567C/T)
G542X / (TG)12T5	3	FC-PS	
N1303K / (TG)12T5	1	CBAVD	
N1303K / (TG)12T5	2	CFTR-RD	
N1303K / (TG)12T5	3	CFTR-RD	
N1303K / (TG)12T5	4	CFTR-RD	
N1303K / (TG)12T5	5	FC-PS	
F508del / L997F	1	CBAVD	
F508del / L997F	2	CFTR-RD	
F508del / L997F	3	CFTR-RD	
F508del / L997F	4	FC-PS	
F508del / (TG)12T5	1	CBAVD	
F508del / (TG)12T5	2	CBAVD	
F508del / (TG)12T5	3	CBAVD	
F508del / (TG)12T5	4	CBAVD	
F508del / (TG)12T5	5	CBAVD	*1251C/T (4575+1251C/T)
F508del / (TG)12T5	6	CBAVD	
F508del / (TG)12T5	7	CFTR-RD	
F508del / (TG)12T5	8	CFTR-RD	
F508del / (TG)12T5	9	CFTR-RD	
F508del / (TG)12T5	10	FC-PS	
F508del / (TG)12T5	11	FC-PS	

Table 7 follows from the previous page.

Genotype	Numeration	CF population	3'UTR variations found
F508del / G85E	1	FC-PI	
F508del / G85E	2	FC-PI	
F508del / G85E	3	FC-PI	
F508del / G85E	4	FC-PI	
F508del / G85E	5	FC-PI	
F508del / G85E	6	FC-PI	
F508del / G85E	7	FC-PS	
F508del / G85E	8	FC-PI	
F508del / G85E	9	FC-PI	
F508del / G85E	10	FC-PI	
F508del / 2183AA>G	1	FC-PI	
F508del / 2183AA>G	2	FC-PI	
F508del / 2183AA>G	3	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2183AA>G	4	FC-PI	
G85E / G85E	1	FC-PS	
G85E / G85E	2	FC-PI	
G85E / G85E	3	FC-PI	
F508del / I1234V	1	FC-PI	
F508del / I1234V	2	FC-PS	
N1303K / G85E	1	FC-PI	
N1303K / G85E	2	FC-PS	
W1282X / 2789+5 G>A	1	FC-PI	*1251T/T (4575+1251T/T)
W1282X / Q1291R	1	FC-PI	*1251T/T (4575+1251T/T)
W1282X / Q1291R	2	FC-PS	*1251T/T (4575+1251T/T)
W1282X / Q1291R	3	FC-PS	*1251T/T (4575+1251T/T)
N1303K / 2789+5 G>A	1	FC-PS	*1251C/T (4575+1251C/T)
N1303K / 2789+5 G>A	2	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	1	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	2	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	3	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	4	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	5	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	6	FC-PI	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	7	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	8	FC-PS	*1251C/T (4575+1251C/T)
F508del / 2789+5 G>A	9	FC-PS	*1251C/T (4575+1251C/T)
1717-1 G>A / (TG)12T5	1	CFTR-RD	*1251C/T (4575+1251C/T)
1717-1 G>A / (TG)12T5	2	FC-PS	*1251C/T (4575+1251C/T)
1717-1 G>A / (TG)12T5	3	FC-PS	*1251T/T (4575+1251T/T)
1717-1 G>A / (TG)12T5	4	FC-PS	*1251C/T (4575+1251C/T)
F508del / [(TG)11T5; V562I; A1006E]	1	FC-PS	*1251C/T (4575+1251C/T)
F508del / [(TG)11T5; V562I; A1006E]	2	FC-PS	*1251T/T (4575+1251T/T)
F508del / [(TG)11T5; V562I; A1006E]	3	FC-PS	*1251C/T (4575+1251C/T)

Table 7 follows from the previous page.

Genotype	Numeration	CF population	3'UTR variations found
F508del / D1152H	1	CFTR-RD	*1251C/T (4575+1251C/T)
F508del / D1152H	2	FC-PS	*1251C/T (4575+1251C/T)
F508del / D1152H	3	FC-PS	*1251C/T (4575+1251C/T)
F508del / D1152H	4	FC-PS	*1251C/T (4575+1251C/T)
F508del / T338I	1	CFTR-RD	
F508del / T338I	2	FC-PS	
F508del / (TG)13T5	1	CFTR-RD	*81C/T (4575+81C/T)
F508del / F508del	1	FC-PI	
F508del / F508del	2	FC-PI	
F508del / F508del	3	FC-PI	
F508del / F508del	4	FC-PI	
F508del / F508del	5	FC-PI	
F508del / F508del	6	FC-PI	
F508del / F508del	7	FC-PI	
F508del / F508del	8	FC-PI	
F508del / F508del	9	FC-PI	
F508del / F508del	10	FC-PI	
F508del / F508del	11	FC-PI	
F508del / F508del	12	FC-PI	
F508del / F508del	13	FC-PI	
F508del / F508del	14	FC-PI	
F508del / F508del	15	FC-PI	
F508del / F508del	16	FC-PI	
F508del / F508del	17	FC-PI	
F508del / F508del	18	FC-PI	
F508del / F508del	19	FC-PI	
S912L / un	1	CBAVD	
[M348K;S912X] / un	2	CBAVD	*1251C/T (4575+1251C/T)
L997F / 711+5G>A (no segregation)	3	CBAVD	*1251C/T (4575+1251C/T)
G1069R / un	4	CBAVD	
(TG)12T5 / un	5	CBAVD	*1251C/T (4575+1251C/T)
[(TG)11T5;V562I;A1006E] / un	6	CBAVD	*1251T/T (4575+1251T/T)
S42F / un	7	CBAVD	*1251C/T (4575+1251C/T)
W1282X / un	8	CFTR-RD	*1251T/T (4575+1251T/T)
F508del / un	9	FC-PI	
G85E / un	10	FC-PI	
F508del / un	11	FC-PS	
F508del / un	12	FC-PS	
G85E / un	13	FC-PS	
[1249-8A>G;G576A;R668C] / un	14	FC-PS	*1251C/T (4575+1251C/T)
(TG)12T5 / un	15	FC-PS	*1251C/T (4575+1251C/T)
S549R(A>C) / un	16	FC-PS	

4. DISCUSSION

The CF phenotype is notably heterogeneous, even among subjects with identical CFTR mutations, due to a complex combination of intragenic and extragenic causes that contribute to the definition of the severity of the disease. We know that the extragenic variability is due to the existence of modifying genes that modulate the original effect of CFTR mutations, but most important may be the intragenic variability, due to the large number of mutations of the CFTR gene and to the even greater number of their possible combinations. Some examples can be the possible presence of complex alleles, that carry two or more variants in *cis* on the same allele, that constitute inheritable intragenic haplotypes; the transcriptional regulation in particular zones of the gene, like 5'-UTR and 3'-UTR; the so-called "long range interaction" of CFTR with neighboring genes.

For a better understanding of the aspects concerning the transcriptional regulation of CFTR and, consequently, the production of the messenger RNA and functional protein, is important to study the intra-CFTR variability and to compare patients with apparently identical genotypes. This can help not

only in terms of diagnosis of the disease and of definition of clinical categories of CF and related pathologies, but also to establish new approaches of screening and of personalized therapy applied on the individual patient. Therefore, there is a strong need to deepen various aspects of the pathology related to the so-called "genotype-phenotype" relationship. It results still little known and understood in CF: in fact, in many cases, there is no clear correlation between the sequence variations found in the CFTR gene and the patients' clinical manifestations. This happens in both the "classical" forms of CF and in those "non-classical", whose interpretation is more difficult.

The purpose of this study was, therefore, to give evidence to the intragenic modulators of CFTR and to investigate what may be their actual role in the context of the disease focusing the attention on three actors involved in the transcriptional regulation of the gene: intragenic haplotypes, variant tracts (TG)mTn and 3'-UTR variant. We studied them both from a structural and a functional point of view in different CF populations, to assess their contribution both quantitatively and qualitatively in defining the final clinical CF phenotype.

As already mentioned, a specific study has been carried out regarding intragenic haplotypes. The onset of a variant occurs

following a single mutational event that manifests itself with a very low frequency (in the order of 10^{-8} for each generation) compared to the number of generations that separate two individuals from their nearest common ancestor. For this reason, every new variant is initially associated with the other variants that are in the proximity of the chromosomal locus in which the mutation occurred. The specific set of variants that are on the same chromosome or in the same sub-chromosomal region is called “haplotype”. The analysis conducted, extended to CF and CF-like populations, selected two specific intragenic haplotypes. We called the first one “target”, because it was mainly represented in the target populations compared to controls: in fact, it appears mainly in ISHV, CBAVD and CFTR-RD alleles (13%, 20% and 36% respectively), while it is poorly evident in the population of normospermic subjects (6% of alleles) and in the general one (3% of alleles). This trend is even more explicit in the populations with classical CF (in 33% and 63% of CF-PS and CF-PI alleles, respectively). These evidences emphasize that the prevalence of this selected haplotype increases with the increasing severity of the clinical phenotype of the population considered. It can be seen as a “pathogenic” haplotype, for its correlation with affected populations. On the other hand, the second selected haplotype,

called “control” for its prevalence in the control populations, shows a completely opposite trend, with a prevalence in normospermic subjects (56% of alleles) and in general population (53% of alleles) compared to all the target populations. So, it can be seen as a "protective" haplotype, for its correlation with not affected populations. Searching for a functional effect of these haplotypes, we functionally analyzed the single variations included. We demonstrated that there are no differences in the splicing process of CFTR mRNA related to the individual presence of a variation which belongs or not to the target haplotype. In fact, all the experiments on CFTR mRNA evidenced that there are no anomalous forms of splicing caused by the presence of a single variation of the target haplotype. Another functional hypothesis may be that they have to be present together to influence wild-type messenger levels. Definitely, it is possible that the simultaneous presence in *cis* of the variations that compose haplotypes can influence the processing of the primary transcript of CFTR, conditioning its functionality. This aspect is now under investigation. An alternative hypothesis is that haplotypes are markers of mutated alleles. In this regard, particularly interesting is that the target haplotype is present on the majority of CF-PI alleles; in fact, since the CF-PI patients

have a high frequency of classical CFTR mutations, finding the target haplotype increases the probability that on that allele a classical mutation will be present at the same time. Consequently, haplotype analysis could be a quick approach to the detection of mutated alleles, also in broad spectrum screening programs.

The splicing process is very complex and often in many diseases, such as for example the CF, the molecular defect is due to alterations of this mechanism. The knowledge of the complexity of the splicing process revealed the existence of important elements of regulation. Variations of these elements, which are found both in coding and non-coding regions of the genes, can occur with deleterious effects on the pre-mRNA splicing. Even the so-called "silent" variations (which do not change the amino acid), as well as variations of the intronic sequences, can therefore be responsible for pathologies. In this contest, we considered very important to include in this study the analysis of the variant tract (TG)mTn, which is located on intron 8 of the CFTR gene and that regulates the splicing of exon 9. Based on literature data and results previously obtained by my research group, we decided to focus our attention on the variants (TG)13T5, (TG)12T7, (TG)11T5 and (TG)12T5. We demonstrated that they have a greater frequency in the

populations with infertility (ISHV and CBAVD) and pathologies related to CF (CFTR-RD) compared to the controls (NORMO and GEN). This is consistent with the fact that they are not populations predominantly characterized by classical CF mutations, but from variant tracts and sequence variations that result in the occurrence of mono- or oligo- symptomatic clinical phenotype. It should also be pointed out that in CBAVD, population, with a more severe clinical phenotype (from reproductive point of view), the (TG)12T5 is at higher frequency than the ISHV population that shows, from reproductive point of view, a milder phenotype. This variant tract is known in literature as being, of the three analyzed, that lowering at the greatest extent the quantity of functional mRNA, and consequently of functional CFTR protein. Instead, in the ISHV population, the other two variant tracts analyzed, which have a less severe quantitative impact, are more represented. These variant tracts are also uncommon in populations with classical CF. Other tracts, such as (TG)10T9, (TG)11T7 or (TG)10T7, that have a cumulative frequency similar in all the populations (both targets and controls), do not seem to be involved in splicing processes, as confirmed by literature and our experimental data. In fact, the results obtained on the RNA from patients confirm that only the

variant tracts (TG)13T5, (TG)12T5, (TG)12T7 and (TG)11T5 are responsible, in different percentages, of the abnormal exclusion of exon 9 from the mature transcript in a "clinically-relevant" quantity. Although the splicing observed relatively to the tracts (TG)13T5 and (TG)12T5 confirm the literature data, the results obtained for the tracts (TG)11T5 and (TG)12T7 are different from the expected ones. In fact, these tracts are usually described as responsible of low levels of aberrant splicing of exon 9. On the contrary, in our study, subjects with the (TG)11T5 / (TG)11T5 genotype show a percentage of splicing of 70% and in one (TG)12T7 / (TG)12T7 subject a splicing of the exon 9 of over 80% was evidenced, never found so high for this tract. This is surprising because until now such a high production of anomalous messenger has never been highlighted starting from these variant tracts, described so far as non-pathological. A possible explanation of the data can be based on the fact that, in some specific cases and genetic background, the variant tracts (TG)11T5 and (TG)12T7 can show an increased anomalous splicing. In addition, both variant tracts have been found in *cis*, on the same allele, with other sequence variations to constitute a complex allele. Often, also the sequence variations in *cis* can influence the maturation, levels of transcripts and amount of functional protein. It should

be considered that these variant tracts are more frequent in CF populations with mild clinical manifestations and CF-like population and that at the functional level they are able to partially influence the levels of protein that reach the apical membrane of the epithelial cells. Consequently, it seems appropriate to point out the (TG)mTn as an important modulator of the CFTR transcription and to evaluate it effectively in terms of genetic assessment.

The 3'-UTR of a gene is the zone of messenger RNA that immediately follows the translation termination codon. The 3'-UTR often contains regulatory regions that post-transcriptionally influence gene expression, both binding sites for regulatory proteins as well as microRNAs (miRNAs). By binding to specific sites within the 3'-UTR, miRNAs can decrease gene expression of various mRNAs by either inhibiting translation or directly causing degradation of the transcript. The knowledge of this particular region in CFTR is still not very detailed and, for this reason, it can be useful to study it in patients of particular clinical interest, in which the genotype-phenotype correlation is not clear. The symptomatic manifestations of a patient do not depend only from the first two sequence variations found. The 3'-UTR variants may be additional intragenic modulator important for both diagnostic

and therapeutic assessment. Our study highlighted three variations on the 3'-UTR of some subjects, which were already previously known and named *81C/T, *567C/T and *1251C/T; while the *1251C/T is already characterized and with a non-pathological functional significance, there is very little information of the other two variations in literature and they are not yet characterized at functional level. This correlates with the results we have seen in our 3'-UTR case series, that is that the variation *1251C/T is found in almost half of the subjects and is therefore very frequent, while the *81C/T and the *567C/T are present in only one patient each and therefore they are rare. Although they are not found in any of the known miRNAs binding sites, it can't be ruled out that they may be within unexplored binding sites and may interact with CFTR miRNAs that are not yet characterized. Obviously, also an alternative explanation exist for the variability possibly dependent on miRNA. It may depend on different amount of miRNAs present and able to interact with the 3'-UTR; in fact, also variations in the production machinery of miRNAs, which can cause an increased or decreased production of them, could influence the percentage of CFTR messenger present in the cell. In this case, a miRNA could act both negatively, when its presence is increased and therefore it binds more mRNA, than

positively, when it is less produced and therefore higher messenger levels can be maintained.

Our studies allow us to formulate a model that concerns the role of intragenic modifiers of CFTR, additional to that of classic mutations, in the modulation of its transcriptional activity. They can be able, in different ways, to influence the classical gene transcription scheme and to contribute to the determination of the quantity and quality of messenger RNA and, consequently, of the phenotypic characteristics of patients.

5. CONCLUSIONS

Due to its different frequency in the analyzed CF populations, the intragenic haplotype studied in this work may be considered a possible intragenic modulator, with an effect (at least in some cases) possibly additional to that of principal CFTR mutations. A mode of functional action may be the modulation of CFTR mRNA expression. Although there is no evidence of a functional role of the individual variations of the haplotype, it would be interesting to study their cumulative effect on CFTR mRNA level. Moreover, it was found that the identification of this particular haplotype marks, in 63% of cases, a mutated allele of classical severe forms of CF: so it can be seen also as a marker of these mutated alleles. The use of this haplotype can be to the basis of broad-spectrum screening programs (for example for the carrier screening), with significant cost savings and reduction of investigation times.

The analysis of the wild-type CFTR messenger allowed us to quantify the transcriptional modulation activity of the variant tracts (TG)mTn, particularly involved, in specific CF and CF-

like populations, in the defining of the functional protein amount. Our results point to the variant tract (TG)mTn as an intragenic modulator, both for the different frequency of some specific tracts in the analyzed populations and for their demonstrated functional effect. The pathological mode of action seems to be the lowering of wild-type CFTR mRNA, in some cases additional to the presence of other CFTR mutations.

Through the analysis of the 3'-UTR of CFTR gene, we improved the knowledge of poor studied CFTR portions. In some cases, we found variants within the 3'-UTR that may play a functional role. However, the 3'-UTR variants seems to be less frequent than the other intragenic variants. Therefore, 3'-UTR variations would not seem to contribute extensively to the intragenic modulation.

In CF and CF-like diseases, the intragenic variability is one of the components of the complex relationship between genotype and phenotype. The presence of intragenic modulators could have a higher prevalence than that previously known. In

particular, our study, designed on three levels of analysis, helps to shed light on the transcriptional regulation processes of CFTR. It emphasizes the importance to search for variants additional to classic mutations, at least in those cases where a discrepancy between the clinical severity of the disease and the genotype (limited to the first 2 mutations found on different alleles) is present.

The full mutational and clinical characterization of CF and CF-like patients reveals a genetic heterogeneity that implies a correlation between genotypic patterns and phenotype of macro-categories more complex than that so far known. In particular, the different prevalence of haplotypes and variant tracts in the different populations analyzed seems to contribute to the genotype-phenotype correlation.

Because of the specificity of the sequence variations studied, their functional characterization is essential, with particular regard to a possible effect on mRNA splicing. In these cases, to study the functional effect of these variations and to establish

their possible pathogenic role, the *ex vivo* approach using nasal brushing revealed to be particularly indicated.

The in-depth study on intragenic modulators of CFTR may allow to define appropriate diagnostic and therapeutic strategies, as well as to better clarify the relationship between sequence variations in DNA (genotype) and clinical manifestations (phenotype) in CF.

REFERENCES

Amaral MD. Novel personalized therapies for cystic fibrosis: treating the basic defect in all patients. J Intern Med. 2015 Feb;277(2):155-166.

Amato F, Seia M, Giordano S, Elce A, Zarrilli F, et al. (2013) Gene Mutation in MicroRNA Target Sites of CFTR Gene: A Novel Pathogenetic Mechanism in Cystic Fibrosis? PLoS ONE 8(3): e60448.

Amato F, Tomaiuolo R, Nici F, Borbone N, Elce A, Catalanotti B, D'Errico S, Morgillo CM, De Rosa G, Mayol L, Piccialli G, Oliviero G, Castaldo G. Exploitation of a Very Small Peptide Nucleic Acid as a New Inhibitor of miR-509-3p Involved in the Regulation of Cystic Fibrosis Disease-Gene Expression. Research Article, Hindawi Publishing Corporation, BioMed Research International, Volume 2014, Article ID 610718, 10 pages.

Andrieux J, Audrézet MP, Frachon I, Leroyer C, Roge C, Scotet V, Férec C. Quantification of CFTR splice variants in adults with disseminated bronchiectasis, using the TaqMan fluorogenic detection system. Clin Genet 2002: 62: 60–67.

Arduino C, Gallo M, Brusco A, Garnerone S, Piana MR, Di Maggio S, Gerbino PG, Ferrone M, Angeli A, Gaia E. Polyvariant mutant CFTR genes in patients with chronic pancreatitis. Clin Genet 1999; 56:400-404.

Auriche C, Di Domenico EG, Pierandrei S, Lucarelli M, Castellani S, Conese M, Melani R, Zegarra-Moran O, Ascenzioni F. CFTR expression and activity from the human CFTR locus in BAC vectors, with regulatory regions, isolated by a single-step procedure. Gene Ther. 2010 Nov;17(11):1341-54.

Biswas R, Kumar P, Pollard HB. Regulation of mRNA turnover in cystic fibrosis lung disease. WIREs RNA 2017, 8:e1408.

Blackledge NP, Carter EJ, Evans JR, Lawson V, Rowntree RK, Harris A. CTCF mediates insulator function at the CFTR locus. Biochem. J. (2007) 408, 267–275.

Blackledge NP, Ott CJ, Gillen AE, Harris A. An insulator element 3' to the CFTR gene binds CTCF and reveals an active chromatin hub in primary cells. 1086–1094 Nucleic Acids Research, 2009, Vol. 37, No. 4.

Bobadilla JL, Macek M Jr, Fine JP, Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations-correlation with incidence data and application to screening. Hum Mutat 2002; 19: 575-606.

Bombieri C, Claustres M, De BK et al. Recommendations for the classification of diseases as CFTR-related disorders. J Cyst Fibros 2011; 10 Suppl 2: S86-102.

Castaldo G, Tomaiuolo R. What is the role of the non-coding regions of the CFTR gene in cystic fibrosis? Expert Rev. Respir. Med. 7(4), 327–329 (2013).

Castellani C, Picci L, Tamanini A, Girardi P, Rizzotti P, Assael BM. Association between carrier screening and incidence of cystic fibrosis. JAMA 2009; 302:2573-9.

Chan HC, Ruan YC, He Q, Chen MH, Chen H, Xu Wm, Chen WY, Zhang XH, Zhou Z. The cystic fibrosis transmembrane conductance regulator in reproductive health and disease. J Physiol 2009; 587: 2187-2195.

Cheung J, Petek E, Nakabayashi K, Tsui L-C, Vincent JB, Scherer SW. Identification of the Human Cortactin-Binding Protein-2 Gene from the Autism Candidate Region at 7q31. GENOMICS Vol. 78, Numbers 1-2, November 2001.

Chillon M, Casals T, Mercier B, Bassas L, Lissens W, Silber S, Romey MC, Ruiz-Romero J, Verlingue C, Claustres M, Nunes V, Ferec C, Estivill X. Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. N Engl J Med 1995; 332:1475-1480.

Chu C-S, Trapnell BC, Curristin S, Cutting GR, Crystal RG. Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. Nature Gen 1993; 3:151-156.

Colombo C. Liver disease in cystic fibrosis. Curr Opin Pulm Med 2007;13:529-536.

Consortium for CF Genetic Analysis.

Cuppens H, Lin W, Jaspers M, Costes B, Teng H, Vankeerberghen A, Jorissen M, Droogmans G, Reynaert I,

Goossens M, Niulius B, Cassiman JJ. Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (TG)m locus explains the partial penetrance of the T5 polymorphism as a disease mutation. J Clin Invest 1998; 101:487-496.

Cutting G R. (2010). Modiefer genes in Mendelian disorders: the example of cystic fibrosis. Ann. N.Y. Acad. Sci. 1214,57-69.

Davis PB. Clinical pathophysiology and mutations of lung disease. In Cystic Fibrosis in adults 1999; Philadelphia-New Jork Lippincot-Raven Pubbl 45-67.

Dequeker E, Stuhrmann M, Morris MA et al. Best practice guidelines for molecular genetic diagnosis of cystic fibrosis and CFTR-related disorders--updated European recommendations. Eur J Hum Genet 2009; 17 (1):51-65.

Dork T, Dworniczak B, Aulehla-Scholz C, Wierczorek D, Bohm I, Mayerova A, Seydewitz HH, Nieschlag E, Meschede D, Horst J, Pander HJ, Sperling H, Ratjen F, Passarge E, Schmidtke J, Stuhrmann M. Distinct spectrum of CFTR gene mutations in congenital absence of vas deferens. Hum Genet 1997; 100:365-377.

Elia J, Delfino M, Imbrogno N, Capogreco F, Lucarelli M, Rossi T, Mazzilli F. Human semen hyperviscosity: prevalence, pathogenesis and therapeutic aspects. Asian J Androl. 2009 Sep;11(5):609-15.

Endale Ahanda ML, Bienvenu T, Sermet-Gaudelus I, Mazzolini L, Edelman A, Zoorob R, Davezac N. The hsa-miR-125a/hsa-let-7e/hsa-miR-99b cluster is potentially implicated in Cystic Fibrosis pathogenesis. Journal of Cystic Fibrosis 14 (2015) 571–579.

Estivill, Complexity in a monogenic disease. Nature genetics 1996, 12, 348-350.

Fanen P, Wohlhuter-Haddad A, Hinzpeter A. Genetics of cystic fibrosis: CFTR mutation classifications toward genotype-based CF therapies. Int J Biochem Cell Biol. 2014 Jul;52:94-102.

Farrell PM. The prevalence of cystic fibrosis in the European Union. J Cyst Fibros 2008; 7:450-3.

Ferraguti G, Pierandrei S, Maria Bruno S, Ceci F, Strom R, Lucarelli M. A template for mutational data analysis of the CFTR gene. Clin Chem Lab Med 2011; 49: 1447-1451.

Gadsby, D., Vergani, P., & Csanàdy, L. The ABC protein turned chloride channel whose falilure causes cystic fibrosis. Nature 2006, 440, 477-479.

Gambari R, Fabbri E, Borgatti M, Lampronti I, Finotti A, Brognara E, Bianchi N, Manicardi A, Marchelli R, Corradini R. Targeting microRNAs involved in human diseases: A novel approach for modification of gene expression and drug development. Biochemical Pharmacology 82 (2011) 1416–1429.

Gibson LE, Cooke RE. A test for concentration of electrolytes in sweat in cystic fibrosis of the pancreas utilizing pilocarpine by iontophoresis. Ped 1959; 23:545-549.

Gillen, A.E.; Gosalia, N.; Leir, S.H. & Harris, A. (2011). microRNA regulation of expression of the cystic fibrosis transmembrane conductance regulator gene. Biochem.J, Vol.438, No.1, pp. 25-32.

Gisler FM, von KT, Kraemer R, Schaller A, Gallati S. Identification of SNPs in the cystic fibrosis interactome influencing pulmonary progression in cystic fibrosis. Eur J Hum Genet 2013;21(4):397-403.

Goldberg, J. B., & Pier, G. B. The role of CFTR in susceptibility to Pseudomonas Aeruginosa infection in cystic fibrosis. Trends in Microbiology 2000; 8, 515-520.

Greene C and Hartl D, Developmental control of CFTR: from bioinformatics to novel therapeutic approaches. Eur Respir J 2015; 45: 18–20.

Haardt M, Benharouga M, Lechardeur D, Kartner N, Lukacs GL. C-terminal truncations destabilize the Cystic Fibrosis transmembrane conductance regulator without impairing its biogenesis. A novel class of mutation. J Biol Chem. 1999 Jul 30;274(31): 21873-7.

Hassan F, Nuovo GJ, Crawford M, Boyaka PN, Kirkby S, et al. (2012) MiR-101 and miR-144 Regulate the Expression of the CFTR Chloride Channel in the Lung. PLoS ONE 7(11): e50837.

Howard M, Frizzell RA, Bedwell DM Aminoglycoside antibiotics restore CFTR function by overcoming premature stop mutations. Nat Med 1996;4:467-469.

Jarvi K, McCallum S, Zielenski J, Durie P, Tullis E, Wilchanski M, Margolis M, Asch M, Ginzburg B, Martin S, Buckspan MB, Tsui LC. Heterogeneity of reproductive tract abnormalities in men with absence of the vas deferens: role of cystic fibrosis transmembrane conductance regulator gene mutations. Fert Steril 1998; 70:724-728.

Jean S, Pathway prediction by bioinformatic analysis of the untranslated regions of the CFTR mRNA. Genomics 94 (2009) 39–47.

Kerem E, MD, Corey M, M Sc, Kerem B, et al. The relations between genotype-phenotype in cystic fibrosis – analysis of the most common mutation (ΔF508). N Engl J Med 1990, Vol. 323, N°22, 1517-22.

Kerem, E. Atypical CF and CF related diseases. Paediatr Respir Rev. 2006; 7, 144-146.

Knowles MR, Durie PR. What is cystic fibrosis? N Engl J Med 2002: 347: 439-42.

Kredentser, J., Poktrant, C., & McCoshen, J. Intrauterine insemination for infertility due cystic fibrosis. Fertility and Sterility 1986, 45, 425-426.

Kumar P, Bhattacharyya S, Peters KW, Glover ML, Sen A, Cox RT, Kundu S, Caohuy H, Frizzell RA, Pollard HB, Biswas

R. miR-16 rescues F508del-CFTR function in native cystic fibrosis epithelial cells. Gene Therapy (2015) 22, 908–916.

Liu X, Ory V, Chapman S, Yuan H, Albanese C, Kallakury B, Timofeeva OA, Nealon C, Dakic A, Simic V, Haddad BR, Rhim JS, Dritschilo A, Riegel A, McBride A, Schlegel R (2012). ROCK inhibitor and feeder cells induce tha contidional reprogramming of epithelial cells. Am J Pathol 180: 599-607.

Lubamba B, Dhooghe B, Noel S, Leal T. Cystic fibrosis: insight into CFTR pathophysiology and pharmacotherapy, Clinical Biochemistry 45 (2012) 1132–1144.

Lucarelli M., Narzi. L., Piergentili. R., Ferraguti. G., Grandoni. F., Quattrucci. S. A 96-well formatted method for exon and exon/intron boundary full sequencing of CFTR gene. Analytical Biochemistry 2006; 353, 226-235.

Lucarelli M.; Rossi T.; Pierandrei S.; Ferraguti G.; Ciminelli B.; Modiano G.; Quattrucci S.; Mazzilli F.; Strom R. Does male fertility impairment due to idiopathic semen hyperviscosity depend on CFTR gene mutations? Journal of cystic fibrosis 2007; 10.1016/S1569-1993(07)60021-6.

Lucarelli et al. 2012 The genetics of CFTR: genotype – phenotype relationship, diagnostic challenge and therapeutic implications. In: “Cystic Fibrosis – Renewed hopes through research”, Intech open access publisher; Chap. 5, pp. 91 – 122.

Lucarelli M, Bruno SM, Pierandrei S, Ferraguti G, Stamato A, Narzi F, Amato A, Cimino G, Bertasi S, Quattrucci S, Strom R. A Genotypic-oriented View of CFTR Genetics Highlights

Specific Mutational Patterns Underlying Clinical Macro-categories of Cystic Fibrosis. Mol Med. 2015 Apr 21. doi: 10.2119/molmed.2014.00229.

Lucarelli M, Bruno S M, Pierandrei S, Ferraguti G, Testini G, Truglio G, Strom R, Quattrucci S. The impact on the genetic test of mutational patterns of CFTR gene in different clinical macro-categories of Cystic Fibrosis. J Mol Diagn. 2016 Jul;18(4):554-65.

Lucarelli M, new era of cystic fibrosis: Full mutational analysis and personalized therapy. World J Med Genet 2017 February 27; 7(1): 1-9.

Ma L, Buchold GM, Greenbaum MP, Roy A, Burns KH, Zhu H, Han DY, Harris RA, Coarfa C, Preethi H. Gunaratne PH, Yan W, Matzuk MM. (2009) GASZ Is Essential for Male Meiosis and Suppression of Retrotransposon Expression in the Male Germline. PLoS Genet 5(9): e1000635.

Margarida D. Amaral, Luka A. Clarke, Anabela S. Ramalho, Sebastian Beck, Fiona Broackes-Carter, Rebecca Rowntree, Nathalie Mouchel, Sarah H. Williams, Ann Harris, Maria Tzetis, Bernhard Steiner, Javier Sanz, Sabina Gallati, Malka Nissim-Rafinifa, Batsheva Kerem, Timothy Hefferon, Garry R. Cutting, Elisa Goina, Franco Pagani. Quantitative methods for the analysis of CFTR transcripts/splicing variants. Journal of Cystic Fibrosis 3 (2004) 17– 23.

Marx JL. The cystic fibrosis gene is found. Science 1989; 245:923-925.

Mastella G. Fibrosi cistica 1998; Edizioni CE.D.RI.M.

Masvidal l, Igreja S, Ramos MD, Alvarez A, de Gracia J, Ramalho A, Amaral MD, Larriba S, Casals T. Assessing the residual CFTR gene expression in human nasal epithelium cells bearing CFTR splicing mutations causing cystic fibrosis. European Journal of Human Genetics (2014) 22, 784–791.

McKiernan PJ, Greene CM. MicroRNA Dysregulation in Cystic Fibrosis. Review Article, Hindawi Publishing Corporation, Mediators of Inflammation, Volume 2015, Article ID 529642, 7 pages.

Megiorni F, Cialfi S, Dominici C, Quattrucci S, Pizzuti A. Synergistic Post-Transcriptional Regulation of the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) by miR-101 and miR-494 Specific Binding. (2011) PLoS ONE 6(10): e26601.

Mendeluk, g., Flecha, L. G., Castello, P., & Bregni, C. Factors Involved in the Biochemical Etiology oh Human Seminal Plasma Hyperviscosity. Journal of Andrology 2000; 262-267.

Moisan S, Berlivet S, Ka C, Le Gac G, Dostie J, Férec C. Analysis of long-range interactions in primary human cells identifies cooperative CFTR regulatory elements. 2564–2576, Nucleic Acids Research, 2016, Vol. 44, No. 6.

Morales MM, Capella MAM, Lopes AG. Structure and function of the cystic fibrosis transmembrane conductance regulator. Braz J Med Biol Res 1999; 32:1021-1028.

Niksic M, Romano M, Buratti E, Pagani F, Baralle FE. Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. Human Molecular Genetics, 1999, Vol.8, No.13, 2339-2349.

Noel S, Leal T. Emerging Roles of microRNAs in Cystic Fibrosis — From Pathogenesis to Development of New Therapies. Cystic Fibrosis in the Light of New Research, Chapter 10. Intech, 2015.

Noone PG, Knowles, MR. "CFTR-opathies": disease phenotypes associated with cystic fibrosis transmembrane regulator gene mutations. Respir Res 2001; 2:328-332.

Oglesby IK, Chotirmall SH, McElvaney NG, Greene CM. Regulation of Cystic Fibrosis Transmembrane Conductance Regulator by MicroRNA-145, -223, and -494 Is Altered in ΔF508 Cystic Fibrosis Airway Epithelium. J Immunol 2013; 190:3354-3362.

Oglesby IK, Vencken SF, Agrawal R, Gaughan K, Molloy K, Higgins G, McNally P, McElvaney NG, Mall MA, Greene CM. miR-17 overexpression in cystic fibrosis airway epithelial cells decreases interleukin-8 production. Eur Respir J 2015; 46: 1350–1360.

Oglesby IK, Agrawal R, Mall MA, McElvaney NG, Greene CM. miRNA-221 is elevated in cystic fibrosis airway epithelial cells and regulates expression of ATF6. Oglesby et al. Molecular and Cellular Pediatrics (2015) 2:1.

Ott CJ, Blackledge NP, Kerschner JL, Leir SH, Crawford GE, Cotton CU, Harris A. Intronic enhancers coordinate epithelial-specific looping of the active CFTR locus. 19934–19939, PNAS, November 24, 2009, vol. 106, no. 47.

Ott CJ, Blackledge NP, Leir SH, Harris A. Novel regulatory mechanisms for the CFTR gene. Biochem Soc Trans. 2009 August ; 37(Pt 4): 843–848.

Oueslati S, Hadj Fredj S, Belhaj R, Siala H, Bibi A, Messaoud T. Preliminary study of haplotypes linked to the rare cystic fibrosis E1104X mutation. Acta Physiol Hung. 2015 Mar;102(1):86-93.

Pagani F, Buratti E, Stuani C, Romano M, Zuccato E, Nicksic M, Giglio L, Faraguna D, Baralle FE. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a non-evolutionary conserved intronic element. J Biol Chem 2000; 275:21041-21047.

Pagani F, Stuani C, Zuccato E, Kornblihtt AR, Baralle FE. Promoter architecture modulates CFTR exon 9 skipping. J Biol Chem 2003; 278:1511-1517.

Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, Doudounakis S, Casals T, Baralle FE. New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet. 2003 May 15;12(10):1111-20.

Pier G. Role of the cystic fibrosis transmembrane conductance regulator in innate immunity to Pseudomonas aeruginosa infections; Proc. Natl. Acad. Sci.2000; 97:8822-8828.

Pignatti PF, Bombieri C, Marigo C, Benetazzo M, Luisetti M. Increased incidence of cystic fibrosis gene mutations in adults with disseminated bronchiectasis. Hum Mol Genet 1994; 4:635-640.

Ramachandran S, Karp PH, Jiang P, Ostedgaard LS, Walz AE, Fisher JT, Keshavjee S, Lennox KA, Jacobi AM, Rose SD, Behlke MA, Welsh MJ, Xing Y, McCray, Jr. PB. A microRNA network regulates expression and biosynthesis of wild-type and ΔF508 mutant cystic fibrosis transmembrane conductance regulator. 13362–13367, PNAS, August 14, 2012, vol. 109, no. 33.

Ramalho AS, Beck S, Farinha CM, Clarke LA, Heda GD, Steiner B, Sanz J, Gallati S, Amaral MD, Harris A, Tzetis M. Methods for RNA extraction, cDNA preparation and analysis of CFTR transcripts. J Cyst Fibros. 2004 Aug;3 Suppl 2:11-5.

Rendine S, Calafell F, Cappello N, Gagliardini R, Caramia G, Rigillo N, Silvetti M, Zanda M, Miano A, Battistini F, Marianelli L, Taccetti G, Diana MC, Romano L, Romano C, Giunta A et al. Genetic history of cystic fibrosis mutations in Italy. Regional distribution. Ann Hum Genet 1997; 61:411-424.

Riordan J.R. et al. Identification of the Cystic Fibrosis gene: cloning and characterization of complementary DNA. Science 1989; 245: 1066–73.

Rossi T, Grandoni F, Mazzilli F, Quattrucci S, Antonelli M, Strom R, Lucarelli M. High frequency of (TG)mTn variant tracts in the cystic fibrosis transmembrane regulator gene in men with high semen viscosity. Fert. Ster 2004; 82, 1316-1322.

Salvatore F, Scudiero O, and Castaldo G. (2002). Genotype-phenotype correlation in Cystic Fibrosis: the role of modifier genes. Am.J.Med.Genet. 111,88-95.

Salvatore D, Buzzetti R, Baldo E, Forneris MP, Lucidi V, Manunza D, Marinelli I, Messore B, Neri AS, Raia V, Furnari ML, and Mastella G, (2011). An overview of international literature from cystic fibrosis registries. Part3. Disease incidence, genotype/phenotype correlation, microbiology, pregnancy, clinical complications, lung transplantation, and miscellanea. J.Cyst.Fibros. 10,71-85.

Sato F, Tsuchiya S, Meltzer SJ, Shimizu K. MicroRNAs and epigenetics. FEBS Journal 278 (2011) 1598–1609.

Slieker MG, Sanders EA, Rijkers GT, Ruven HJ, van der Ent CK. Disease modifying genes in cystic fibrosis. J Cyst Fibros. 2005 Aug;4 Suppl 2:7-13

Sonneville F, Ruffin M, Guillot L, Rousselet N, Le Rouzic P, Corvol H, Tabary O. New Insights about miRNAs in Cystic Fibrosis. Review, The American Journal of Pathology, Vol. 185, No. 4, April 2015.

Steiner, B.; Rosendahl, J.; Witt, H.; Teich, N.; Keim, V.; Schulz, H.U.; Pfutzer, R.; Luhr, M.; Gress, T.M.; Nickel, R.; Landt, O.; Koudova, M.; Macek, M., Jr.; Farre, A.; Casals, T.; Desax, M.C.; Gallati, S.; Gomez-Lira, M.; Audrezet, M.P.; Ferec, C.; Des, G.M.; Claustres, M. & Truninger, K. (2011). Common CFTR haplotypes and susceptibility to chronic pancreatitis and congenital bilateral absence of the vas deferens. Hum.Mutat., Vol.32, No.8, pp. 912-920.

Stuhrmann M, Dork T. CFTR gene mutations and male infertility. Andrologia 2000; 32:71-83.

Suprynowicz FA, Upadhyay G, Krawczyk E, Kramer SC, Hebert JD, Liu X, Yuan H, Cheluvaraju C, Clapp PW, Boucher RC, Kamonjoh CM, Randell SH, Schlegel R (2012). Conditionally reprogrammed cells represent a stem-like state of adult epithelial cells. PNAS 109: 20035-20040.

Tazi MF, Dakhlallah DA, Caution K, Gerber MM, Chang S-W, Khalil H, Kopp BT, Ahmed AE, Krause K, Davis I, Marsh C, Lovett-Racke AE, Schlesinger LS, Cormet-Boyaka E, Amer AO. (2016) Elevated Mirc1/Mir17-92 cluster expression negatively regulates autophagy and CFTR (cystic fibrosis transmembrane conductance regulator) function in CF macrophages, Autophagy, 12:11, 2026-2037.

Teng H, Jorissen M, Van Poppel H, Legius E, Cassiman JJ, Cuppens H. Increased proportion of exon 9 alternatively spliced CFTR transcripts in vas deferens compared with nasal epithelial cells. Hum Mol Genet. 1997 Jan;6(1):85-90.

Tomaiuolo R, Fausto M, Elce A, Strina I, Ranieri A, Amato F, Castaldo G, De Placido G, Alviggi C. Enhanced frequency of CFTR gene variants in couples who are candidates for assisted reproductive technology. Clin Chem Lab Med 2011; 49: 1289-1293.

Vankeerberghen A et al. The Cystic Fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions. J Cystic Fibrosis. 2002; 1: 13–29.

Viart V, Bergougnoux A, Bonini J, Varilh J, Chiron R, Tabary O, Molinari N, Claustres M, Taulan-Cadars M. Transcription factors and miRNAs that regulate fetal to adult CFTR expression change are new targets for cystic fibrosis. Eur Respir J 2014; 45: 18–20.

Wilschanski M and Kerem E. New drugs for cystic fibrosis. Expert Opin. Investig Drugs 2011;20(9):1285-1292.

Wilschanski M. Novel therapeutic approaches for cystic fibrosis Discov Med 2013;15(81):127-33.

Yan W, Rajkovic A, Viveiros MM, Burns KH, JOHN J. Eppig JJ, Matzuk MM. Identification of Gasz, an Evolutionarily Conserved Gene Expressed Exclusively in Germ Cells and Encoding a Protein with Four Ankyrin Repeats, a Sterile-α Motif, and a Basic Leucine Zipper. Molecular Endocrinology 16(6):1168–1184.

Zarrilli F, Amato F, Morgillo CM, Pinto B, Santarpia G, Borbone N, D'Errico S, Catalanotti B, Piccialli G, Castaldo G, Oliviero G. Peptide Nucleic Acids as miRNA Target Protectors

for the Treatment of Cystic Fibrosis. Molecules 2017, 22, 1144.

Zhang L, Button B, Gabriel SE, Burkett S, Yan Y, Skiadopoulos MH, Dang YL, Vogel LN, McKay T, Mengos A, Boucher RC, Collins PL, Pickles RJ. CFTR delivery to 25% of surface epithelial cells restores normal rates of mucus transport to human cystic fibrosis airway epithelium. PLoS Biol. 2009 Jul;7(7):e1000155.

Zink D, Amaral MD, Englmann A, Lang S, Clarke LA, Rudolph C, Alt F, Luther K, Braz C, Sadoni N, Rosenecker J, Schindelhauer D. Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. The Journal of Cell Biology, Volume 166, Number 6, September 13, 2004 815–825.