

# Università degli Studi di Napoli

## “Federico II”



### Scuola di Dottorato in Genetica e Medicina Molecolare

XIX Ciclo (2003-2007)

### **Analisi di CONSERVED SEQUENCE TAGs (CSTs) del gene CFTR in soggetti affetti da Fibrosi Cistica**

**Docenti guida:**

Ch.ma Prof.ssa Lucia Sacchetti

Ch.mo Prof. Giuseppe Castaldo

**Candidato:**

Dott.ssa Ausilia Elce

**Coordinatore:**

Ch.mo Prof. Carmelo Bruno Bruni

## INDICE

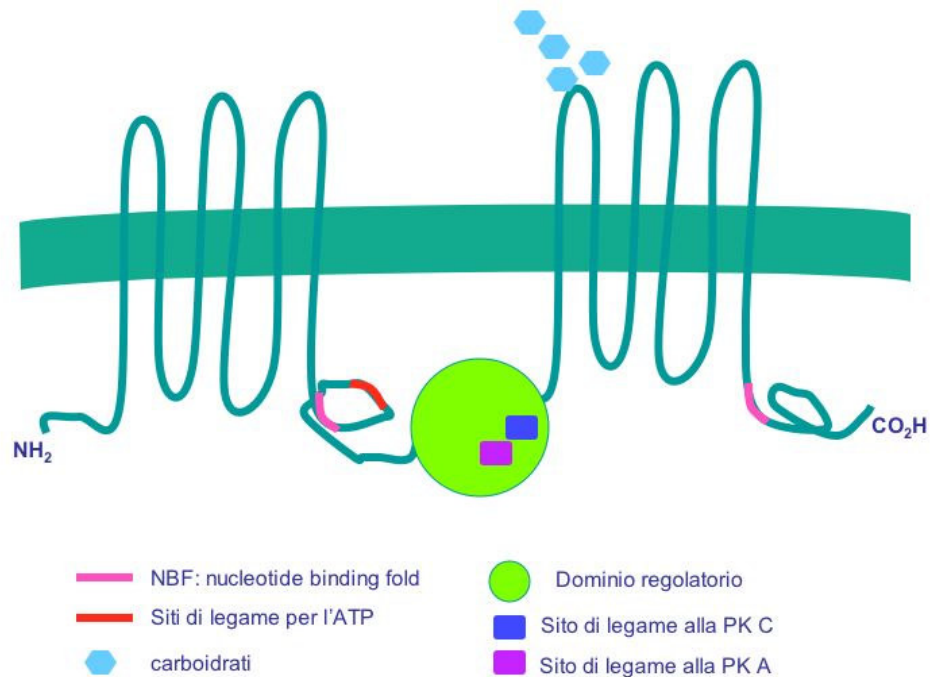
Abbreviazioni	pag. 3
1) Introduzione	pagg. 4-15
<i>La Fibrosi Cistica: il 10% degli alleli dei pazienti non presenta mutazioni nel gene-malattia</i>	pagg. 4-8
<i>Mutazioni ancora non note potrebbero alterare i meccanismi di regolazione di CFTR ed essere causali di malattia: il 5' del gene</i>	pagg. 8-14
<i>L'eterogeneità clinica della Fibrosi Cistica potrebbe dipendere da fattori genetici "modulatori del fenotipo"</i>	pagg. 14-15
2) Scopo della tesi	pag. 16-18
3) Materiali e Metodi	pagg. 19-29
<i>Pazienti</i>	pagg. 19-20
<i>Analisi molecolare del gene CFTR</i>	pagg. 20-23
<i>Analisi delle CSTs del gene CFTR mediante PCR seguita da sequenziamento capillare</i>	pagg.23-29
4) Risultati e Discussione	pagg. 30-42
<i>L'analisi delle CSTs ha portato all'identificazione di alcune varianti geniche specifiche dei pazienti CF con mutazioni non note</i>	pagg. 30-39
<i>Identificazione di tre nuovi microsatelliti nel gene CFTR</i>	pagg.39-42
5) Conclusioni	pag. 43
Bibliografia	pagg. 44-49

<b>ABBREVIAZIONI</b>	
ABC	ATP-Binding Cassette
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CSTs	Conserved Sequence Tags
PKA	Proteina Kinasi AMP ciclico dipendente
ATF/cAMP	Activating Transcription Factor/cyclic-AMP
CRE	cAMP Response Element
C/EBP	CCAAT-Enhancer Binding Protein
GRE	Glucocorticoid Response Element
DHS	DNase I Hypersensitive Site
HNF1	Hepatocyte Nuclear Factor 1
PCAF	p300/CBP-Associated Factor
CBP	CREB Binding Protein
CREB	cAMP-Response Element Binding protein
CDP/cut	CCAAT Displacement Protein/cut
AP-1	Activator Protein 1
SP-1	Specific Protein 1
EMSA	Electrophoretic Mobility Shift Assay
NF-kB	Nuclear Factor-kB
Capan1	Pancreatic adenocarcinoma 1 cells
Caco2	Colon carcinoma 2 cells
HeLa	Henrietta Lacks cells
CFPAC	Pancreatic Adenocarcinoma cells
DGGE	Denaturing Gradient Gel Electrophoresis
SSCP	Single-Strand Conformation Polymorphism
DHPLC	Denaturing High Performance Liquid Chromatography
hGCN5	Human General Control of amino acid synthesis protein 5
Calu3	Lung adenocarcinoma cells
SRF	Serum Response Factor

## INTRODUZIONE

*La Fibrosi Cistica: il 10% degli alleli dei pazienti non presenta mutazioni nel gene-malattia*

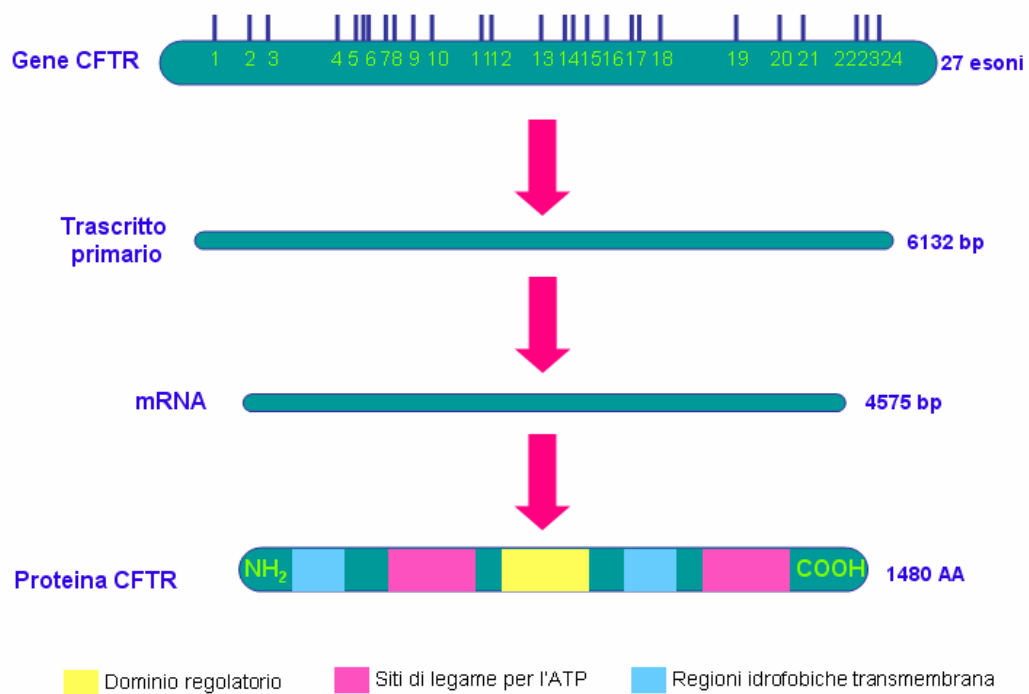
La Fibrosi Cistica (CF) è la malattia genetica ad esito letale più frequente nella popolazione caucasica: è trasmessa con meccanismo autosomico recessivo e ne è affetto un neonato ogni 2500-2700 nati vivi; la frequenza dei portatori sani è di 1:25 (1). La malattia dipende dalle alterazioni di una glicoproteina di membrana di 1480 amminoacidi, espressa e funzionante nella porzione apicale della membrana delle cellule epiteliali, denominata *Cystic Fibrosis Transmembrane conductance Regulator* (CFTR), che svolge svariate funzioni (fig. 1), di cui la principale attività è quella di fungere da canale ionico del cloro, ma nel corso degli ultimi anni è stato anche dimostrato che gioca un ruolo nell'ambito della regolazione di altri canali sodio-dipendenti (2), dell'infiammazione (3), della regolazione del pH (4) e del trasporto del glutatione (5).



**Figura 1:** rappresentazione schematica della proteina canale CFTR.

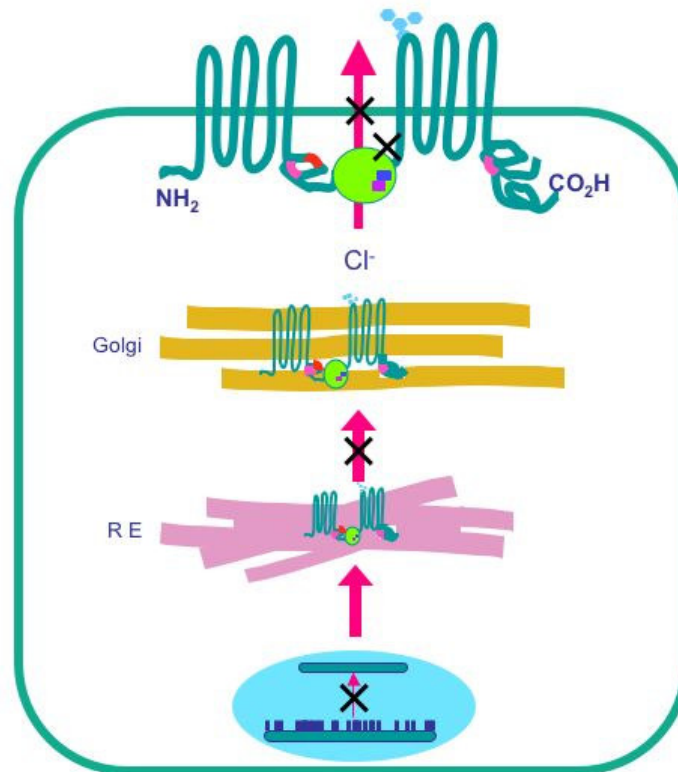
Nei soggetti in cui la proteina CFTR non presenta alterazioni, i secreti salini vengono idratati e fluidificati poiché gli ioni cloro e sodio si associano al transito di acqua. Nei soggetti in cui la proteina CFTR è alterata, le secrezioni diventano più dense, causando i sintomi sistemici della malattia (6). Le secrezioni dense ostruiscono i bronchi favorendo lo sviluppo di infezioni, i dotti pancreatici, impedendo agli enzimi digestivi di giungere all'intestino, provocando malassorbimento e deficit di accrescimento; a livello epatico rendono la bile densa, danneggiando il fegato. Anche a livello delle ghiandole sudoripare è compromesso il riassorbimento di ioni cloro e sodio. I pazienti CF eliminano una quantità eccessiva di NaCl nel sudore che può essere raccolto, dopo stimolazione delle ghiandole sudoripare con pilocarpina (test del sudore).

Il gene che codifica per la proteina CFTR, mappato nel 1989 (7) e sequenziato nel 1991 (8), è localizzato sul cromosoma 7, in posizione q31.2, ed è composto da 27 esoni (fig. 2).



**Figura 2:** dal gene alla proteina CFTR.

Ad oggi, sono state identificate oltre 1600 diverse mutazioni causative di CF: alcune sono diffuse ubiquitariamente (es. F508del, N1303K), altre sono specifiche di ristretti ambiti etnico-geografici e la maggior parte sono “private”, cioè confinate a gruppi geografici o familiari. Per esempio, la F508del (la più frequente mutazione causativa di malattia), è presente nel 90% degli alleli CF in Danimarca, ma solo nel 50% degli alleli CF in Italia (9). Pur essendo molto eterogenee, le mutazioni CF sono state raggruppate in 5 classi in base all’alterazione funzionale della proteina CFTR di cui sono responsabili (10) (fig. 3).



**Figura 3:** Le 5 classi di mutazioni del gene CFTR: la classe 1 comprende le mutazioni che inducono un difetto della produzione della proteina; la classe 2, le mutazioni che alterano il "processing" della proteina; le classi 3 e 4 comprendono mutazioni che provocano un deficit nell'attivazione del canale o un'alterazione della conduttanza; la classe 5 comprende, infine, mutazioni in siti di "splice" che riducono la quantità ma non la funzionalità della proteina CFTR.

L’analisi molecolare del gene CFTR è di fondamentale importanza per identificare i portatori sani della malattia, del tutto asintomatici, per questo motivo essa è oggetto di continui studi tesi a migliorare la sensibilità e l’accuratezza dei test genetici. Vengono sottoposti ad analisi molecolare del gene CFTR tutti quei soggetti a

maggior rischio di essere portatori (consanguinei di un affetto) e i partner di soggetti già risultati portatori sani di mutazioni del gene CFTR. Per le coppie in cui entrambi i partner risultano essere portatori sani e, quindi, presentano un alto rischio di generare un figlio affetto da Fibrosi Cistica, è possibile programmare la diagnosi prenatale, che si basa sullo studio del DNA del feto, estratto da villi coriali, in epoca precoce (dalla 9° alla 15° settimana di gravidanza), o da amniociti, in epoca più tardiva (dalla 16° alla 20° settimana di gravidanza). Sono stati messi a punto vari tipi di test di biologia molecolare per identificare le mutazioni responsabili della Fibrosi Cistica: i test più semplici (indagini di 1° livello) identificano le mutazioni più frequenti (11), mentre quelli più complessi (indagini di 2° e 3° livello) danno la possibilità di effettuare lo screening dell'intero gene o di evidenziare l'assenza di grossi tratti del gene CFTR (12). Le analisi di 1° livello sono effettuate con kit commerciali o con metodiche messe a punto in laboratorio (*home made*) e testano pannelli comprendenti le mutazioni riscontrate con maggior frequenza nell'area geografica di appartenenza del laboratorio. Con le indagini di 1° livello si raggiunge un "detection rate" (tasso di identificazione della mutazione) dell'80-85%, a seconda dell'area geografica di provenienza del paziente. Le analisi di 2° livello permettono lo scanning di tutte le regioni codificanti del gene CFTR. Tra le tecniche più utilizzate c'è il sequenziamento diretto, che consente anche lo studio delle regioni introniche più vicine agli esoni, con la possibilità, quindi, di identificare anche mutazioni implicate nell'alterazione dello splicing del trascritto. I test di 2° livello permettono di raggiungere un detection rate pari al 92% circa. Ad esempio, per determinare l'epidemiologia delle mutazioni CF nel Sud Italia, è stato condotto uno studio su una popolazione di 371 pazienti non imparentati tra loro, originari delle regioni Campania, Basilicata e Puglia. L'analisi con DGGE dell'intera regione codificante del gene CFTR ha permesso di identificare 43 mutazioni prevalenti e tre polimorfismi intragenici, con una detection rate del 91% (13). Nella nostra area geografica, circa l'8% degli alleli CF non viene caratterizzato e una situazione più o meno analoga è stata descritta nella maggior parte dei gruppi etnici studiati. Ciò costituisce un serio problema per l'identificazione di portatori sani della malattia e per la realizzazione di indagini prenatali in coppie a rischio.

Per spiegare dove possano essere localizzate le mutazioni ancora non identificate sono state formulate varie ipotesi, tra cui a) la presenza di grossi riarrangiamenti del gene CFTR (14) che, se presenti in eterozigosi, non sono

evidenziabili mediante le tecniche di scanning genomico, b) la presenza di mutazioni localizzate in regioni non ancora analizzate come gli introni o altre regioni non codificanti, le cui mutazioni potrebbero alterare siti coinvolti nella regolazione dell'espressione del gene CFTR, c) la presenza di fattori genetici che, ereditati in maniera indipendente dal gene CFTR, potrebbero modulare l'espressione fenotipica della malattia (15). Per quanto concerne il primo punto, negli ultimi anni, sono state messe a punto complesse tecniche basate sulla PCR quantitativa e sulla corsa di elettroforesi capillare, che permettono l'identificazione di grosse delezioni e/o di inserzioni; si tratta di indagini molto specialistiche ancora poco utilizzate, che consentono, in tempi relativamente rapidi, l'identificazione di mutazioni che in passato sfuggivano nel corso dell'analisi molecolare del gene (12). Per quanto riguarda i rimanenti due punti, questi costituiscono, ad oggi, i due filoni principali di ricerca per la Fibrosi Cistica, in quanto una delucidazione di tutti i meccanismi con cui diversi fattori, solo in parte noti, contribuiscono alla patogenesi della malattia, potrebbe portare al miglioramento della diagnosi molecolare, alla predizione del fenotipo clinico e alla creazione di terapie innovative per la Fibrosi Cistica. Non è escluso, infatti, che fattori genici indipendenti dal gene CFTR e fattori che si trovano all'interno del gene stesso, possano rappresentare un sistema finemente integrato, responsabile della patogenesi della malattia.

*Mutazioni ancora non note potrebbero alterare i meccanismi di regolazione di CFTR ed essere causali di malattia: il 5' del gene*

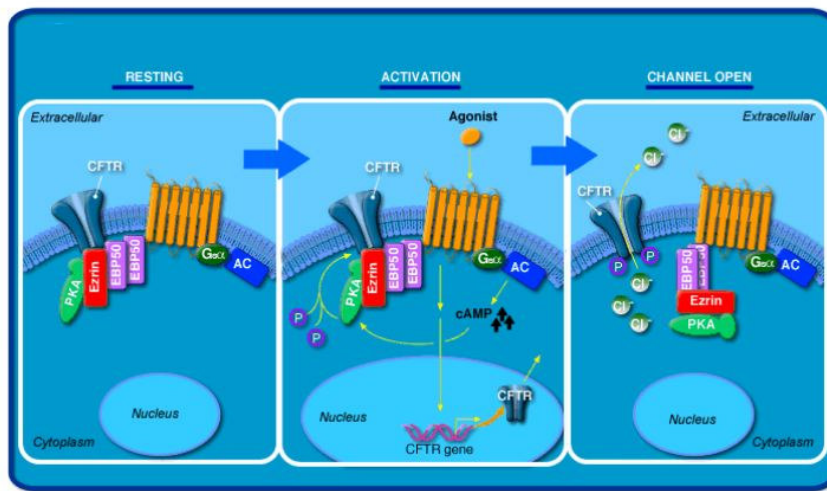
La regolazione dell'espressione del gene CFTR è stata oggetto di numerosi studi negli ultimi anni e in una parte dei pazienti CF la malattia potrebbe essere dovuta ad alterazioni dei meccanismi di regolazione dell'espressione genica.

Studi condotti a partire dal 1992 hanno rivelato come il promotore del gene non sia il principale artefice della regolazione dell'espressione genica di CFTR, avendo esso un'organizzazione che somiglia più a quella di un gene housekeeping che a quella di un gene regolato in maniera tessuto-specifica e nel corso dello sviluppo (16).

Esistono numerosi meccanismi di controllo tissutale e temporale dell'espressione genica di CFTR (17). Ad esempio, in epoca fetale è stata riscontrata una forte espressione del gene nell'epitelio dei dotti pancreatici, specialmente nei dotti



terminali, con bassi livelli riscontrati lungo i dotti intra e interlobulari (18). Alti livelli di espressione di CFTR sono stati riscontrati anche nelle cripte dell'epitelio dell'intestino fetale, con un chiaro gradiente di espressione che decresce sull'asse cripta-villo (17); bassi livelli di espressione, invece, sono stati riscontrati nell'epitelio dell'epididimo fetale. Il pattern di espressione del feto è mantenuto anche nelle prime fasi di vita postnatale, tranne che nel sistema respiratorio dove sono stati riscontrati alti livelli di espressione di CFTR che decrementano sempre più nel corso della vita (19). Il meccanismo di regolazione dell'espressione di CFTR maggiormente chiarificato è quello incentrato sul ruolo dell'AMPciclico (cAMP) e dei suoi agonisti (20). Diversi studi hanno dimostrato come il cAMP sia il principale mediatore dell'attivazione di CFTR: è capace, infatti, sia di agire a livello del promotore del gene, aumentandone la trascrizione, sia di aumentare, tramite l'attivazione della proteinkinasi AMP ciclico dipendente (PKA), l'attività di gating della proteina CFTR (fig. 4).

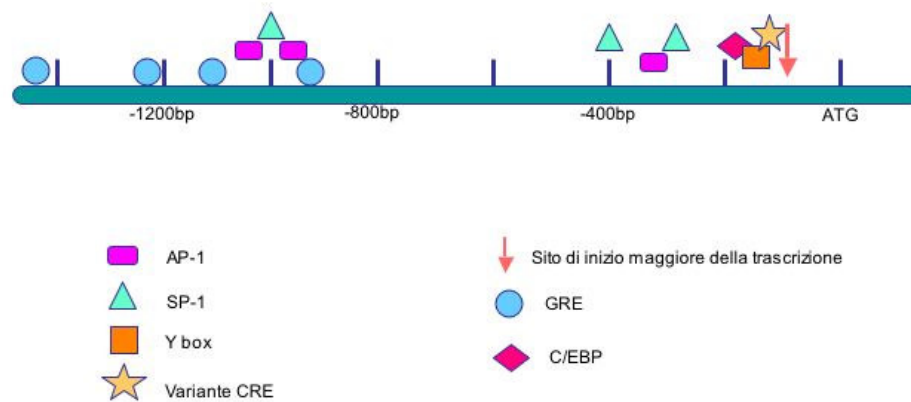


**Figura 4: il ruolo del cAMP nella regolazione dell'espressione di CFTR.**

Studi effettuati con EMSA (Electrophoretic Mobility Shift Assay) hanno dimostrato che CREB (cAMP-Response Element Binding protein) lega CRE (c-AMP Response Element) nel gene CFTR con una alta affinità in vitro. Questo elemento modula l'espressione mediata dall'AMP ciclico, quando situato a monte di un promotore omologo, suggerendo che l'espressione di CFTR è anche regolata dai livelli intracellulari di cAMP.

Mediante gli inibitori della PKA, è stato dimostrato il ruolo del cAMP nel mantenere i livelli basali di trascrizione del gene CFTR (21).

Oltre alla regolazione mediata dal cAMP, sono state identificate anche diverse proteine che agiscono il più delle volte sinergicamente per modulare l'espressione del gene. Molteplici siti di binding per fattori trascrizionali, infatti, sono stati identificati al 5', tra cui l'elemento responsivo al cAMP (CRE), un sito di legame a C/EBP e molti elementi responsivi a glucocorticoidi (fig. 5).



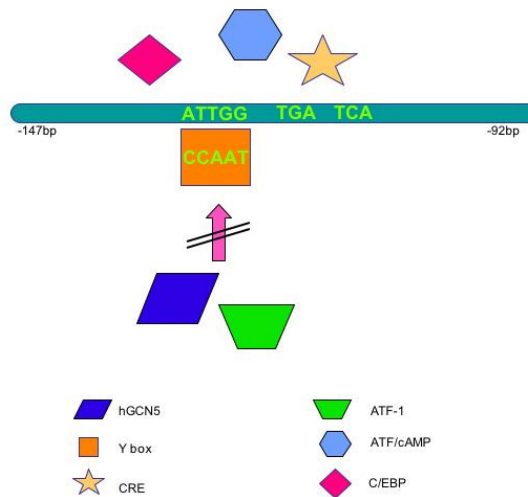
**Figura 5:** La regione del promotore di CFTR

La regione del promotore di CFTR è stata delineata mediante esperimenti in cui il promotore, progressivamente deletato, è stato messo alla guida di geni reporter: in questa maniera sono stati identificati due elementi positivi che agiscono in cis, uno dei quali corrisponde a una piccola sequenza del promotore da -226 a +98 e un elemento negativo (AP-1), considerando il +1 come il sito maggiormente utilizzato per iniziare la trascrizione. Sono state identificate, inoltre, due sequenze ricche di GC da -273 a -190 che assomigliano a siti di legame per la proteina Sp1: l'eliminazione di questi siti rende il promotore inattivo. L'elemento negativo è stato localizzato subito a monte della sequenza minima del promotore da -345 a -277, la sua inclusione nel costrutto del gene reporter elimina completamente l'espressione del gene (17).

Siti di binding per fattori trascrizionali sono stati individuati anche a distanza dalla regione del promotore e dell'esone 1. Più precisamente, in cellule Calu3 (derivanti da adenocarcinoma polmonare), è stato individuato un sito di legame per NF-κB, nella regione compresa tra il nucleotide -1103 al nucleotide -1093. Il legame

ad NF-kB sembrerebbe aumentare l'attività del promotore in questa linea cellulare. Saggi di gel shift, transfezione e supershift hanno, infatti, dimostrato che il peptide proinfiammatorio 1b attiva la proteina NF-kB, la quale è in grado di traslocare nel nucleo e legare il sito -1103 kB, collocato al 5' del gene, con un marcato aumento dei livelli di mRNA di CFTR. E' stato inoltre osservato che inibitori del pathway di NF-kB, come la pirrolidina ditiocarbonato, sono in grado di abolire l'aumento della trascrizione di CFTR dovuta all'azione di NF-kB sia in cellule Calu3 che T84 (22). Nel 1993, il gruppo di Koh (23) ha descritto come gli elementi regolatori del gene CFTR agissero in maniera cooperativa per attivare l'espressione del gene, attraverso una fitta rete di interazioni molecolari.

La regione del promotore maggiormente coinvolta in questa complessa interazione tra più fattori di regolazione dell'espressione è quella che ospita l'elemento invertito CCAAT (Y box) localizzato tra -132 e -119 (fig. 6). La sua rimozione da costrutti sotto la guida del promotore di CFTR riduce i livelli del trascritto del gene reporter (24).



**Figura 6:** *Esperimenti condotti con EMSA (Electrophoretic mobility shift assays) hanno permesso di identificare che C/EBP (CCAAT-Enhancer Binding Protein) e ATF1(Activator Transcriptional Factor 1) formano complessi con la sequenza CCAAT necessari per l'attivazione della trascrizione mediata da cAMP. La proteina CCAAT reprime la trascrizione di CFTR poiché impedisce il legame degli attivatori della trascrizione hGCN5(Human General Control of amino acid synthesis protein 5) e ATF-1. La proteina CDP/cut (CCAAT Displacement Protein/cut) recluta l'istone deacetilasi, modificando la struttura della cromatina in uno stato inattivo per la trascrizione e, inoltre, interagisce con la proteina legante CREB e con il fattore associato alla proteina legante CREB, che hanno attività istone acetiltransferasica.*

L'importanza funzionale degli elementi CRE e CCAAT è supportata dalla conservazione di una regione di 34 bp che comprende questi elementi nel gene CFTR dell'uomo, del topo, del ratto (25), della pecora (26), del maiale e della mucca (27).

L'espressione del gene CFTR sembra essere anche sotto il controllo ormonale, sia nei maschi che nelle femmine. Nei roditori l'espressione della proteina nel circolo spermatico potrebbe essere necessaria per ridurre il volume citoplasmatico degli spermatozoi e per un loro corretto sviluppo. Nelle femmine dei roditori, sono state osservate variazioni nei livelli di espressione di CFTR in relazione all'accrescimento dell'epitelio uterino (28). Inoltre, è stato osservato che gli estrogeni aumentano i livelli di CFTR nell'epitelio uterino dopo che femmine sessualmente immature e femmine mature ovariectomizzate sono state trattate con estrogeni (29). Tuttavia, non è ancora stato chiarito esattamente il pathway con cui gli ormoni sessuali influenzino i livelli di CFTR.

Un altro probabile punto di controllo per l'espressione temporale e tissutale è quello derivante dall'utilizzo di differenti siti di inizio trascrizione. Infatti, l'analisi della sequenza genomica a monte dell'esone 1 ha rivelato che non è presente l'elemento TATA box, e per questo motivo, in cellule che esprimono alti livelli del gene, possono essere utilizzati siti multipli di inizio trascrizione. In particolare, in queste linee cellulari (es., Caco2 e T84), sono state isolate molteplici varianti del trascritto, derivanti dall'utilizzo di siti alternativi di inizio trascrizione da -95 a +50 rispetto al sito maggiormente utilizzato a 72 bp a monte dell'esone 1. Complessivamente i siti minori di inizio trascrizione sono stati riscontrati in una regione di 110 bp al 5' dell'esone 1 (16). Cellule che esprimono bassi livelli di CFTR (es. HeLa), prediligono invece l'utilizzo di un sito di inizio trascrizione collocato a -32 (30). Inoltre, sono stati identificati trascritti di CFTR diversi nelle linee cellulari T84 e CFPAC che contengono a monte esoni alternativi (denominati 1a e -1a), legati direttamente all'esone 2 e che utilizzano un sito di inizio a -815 (31).

Anche in altre specie sono stati descritti trascritti con esoni alternativi al 5', che potrebbero giocare un ruolo nell'espressione tissutale e temporale di CFTR. Ad esempio, in tessuti polmonari di pecora sono stati identificati due esoni alternativi al 5', Ov1aL e Ov1aS. I livelli di espressione di Ov1aS riflettono i livelli di espressione

polmonare del gene, cioè il trascritto *Ov1aS* aumenta quando il livello globale di espressione di *CFTR* inizia a diminuire nel polmone (32, 33).

Nel topo, studi sui siti di inizio della trascrizione effettuati su RNA estratto da tessuto, hanno permesso di stabilire l'uso di un unico sito di inizio nel polmone adulto (-82), nel polmone fetale (-72), nel rene (-75) e nell'utero (-72), mentre, nell'intestino tenue sono usati indifferentemente i siti a -72 e a -75 e nel duodeno i siti a -60 e a -19 sono utilizzati con un rapporto di 4:1. Anche nel polmone umano è stata dimostrata la regolazione temporale del trascritto attraverso l'inclusione o l'esclusione degli esoni 1 e 2, con livelli che diminuiscono rapidamente tra la 14<sup>a</sup> e la 26<sup>a</sup> settimana di gravidanza, ma i meccanismi con cui questi esoni alternativi potrebbero influenzare l'espressione di *CFTR* sono ancora oggetto di studio (33). In conclusione, la variazione nel sito di inizio della trascrizione potrebbe determinare un livello di regolazione tessuto-specifica dell'espressione di *CFTR*, mentre l'uso di esoni alternativi al 5' potrebbe determinare la regolazione di *CFTR* nel polmone con un meccanismo ancora non completamente definito.

Infine, studi con linee cellulari che esprimono bassi livelli di *CFTR*, hanno rivelato la presenza di una isola CpG che si estende da -400 a +300 nell'introne 1. Questa regione è altamente metilata in cellule non epiteliali non esprimenti *CFTR*, suggerendo che il controllo dell'espressione tessuto specifica del gene sia appannaggio di molteplici eventi tra cui anche fenomeni epigenetici su DNA e cromatina (23).

Da quanto descritto finora, si può stabilire che risultano ancora poco chiari i meccanismi con cui *CFTR* viene espresso e che la maggior parte degli elementi che conferiscono una regolazione temporale e tessuto-specifica all'espressione risiedono solo in parte nella regione del promotore. Per questo motivo sono ancora in corso numerosi studi mirati ad identificare le regioni di DNA potenzialmente coinvolte nell'espressione del gene e i meccanismi con cui avviene tale espressione.

Nel 1995 l'analisi di una regione di 400 kb di DNA genomico del gene *CFTR* umano ha portato alla caratterizzazione di molteplici siti ipersensibili alla DNasi I (DHS) (25). Rappresentati da sequenze dinucleotidiche intersperse nel genoma, i DHS si trovano generalmente in regioni che vengono rapidamente degradate nel saggio di digestione con l'enzima DNasi I. Poiché la degradazione è associata all'assenza di legame ai nucleosomi, che li proteggerebbero dal taglio, è ipotizzabile che questi siti

siano potenzialmente coinvolti nella regolazione dell'espressione genica. Queste osservazioni (che saranno maggiormente dettagliate nella discussione) hanno spostato l'interesse degli studi sui meccanismi di regolazione dell'espressione genica dalla regione del promotore alle regioni introniche.

*L'eterogeneità clinica della Fibrosi Cistica potrebbe dipendere da fattori genetici "modulatori" del fenotipo*

A causa del differente impatto sulla produzione della proteina, le numerose mutazioni causative fino ad ora riscontrate nel gene CFTR sono state oggetto di studio, da parte di molti gruppi, al fine di generare una possibile correlazione tra il difetto molecolare del gene CFTR e l'eterogeneità clinica della malattia.

E' noto, infatti, che la Fibrosi Cistica sia una malattia cronica, evolutiva, con prognosi severa, che si manifesta, però, in maniera variabile da paziente a paziente, con un decorso clinico non sempre prevedibile sulla base degli esami clinici e molecolari. Vi sono, infatti, casi di pazienti che vanno incontro a morte nei primi giorni di vita per occlusione intestinale da ileo da meconio (presente nel 10% dei casi di Fibrosi Cistica), in altri invece la malattia ha una progressione più lenta, tale da consentire il trattamento dei sintomi clinici attraverso protocolli terapeutici sempre più innovativi. A partire dagli anni '50 l'aspettativa di vita dei pazienti CF è andata costantemente aumentando fino a raggiungere, attualmente, un'età media di sopravvivenza di circa 30 anni, che continua ad aumentare di anno in anno. Alcuni pazienti con Fibrosi Cistica e un'espressione clinica meno severa, grazie alle terapie sviluppate negli ultimi anni, hanno già superato i 50, i 60 ed anche i 70 anni. Tuttavia, l'estrema variabilità fenotipica della malattia costituisce ancora oggi un serio problema ai fini della comprensione dei meccanismi patogenetici e della generazione di terapie sempre più mirate alla risoluzione dei sintomi clinici dei pazienti CF. L'insufficienza pancreatica, presente nell'85-90% dei pazienti CF, sembra essere chiaramente associata all'alterazione molecolare del gene CFTR (34), mentre risultano ancora non chiari tutti i meccanismi patogenetici responsabili delle altre manifestazioni cliniche della malattia.

L'insufficienza respiratoria (che rappresenta il 95% delle cause di mortalità per i pazienti CF) ha un'insorgenza variabile, prevalentemente nei primi mesi o anni di vita, con una severità molto eterogenea e non riconducibile al tipo di difetto

molecolare del gene CFTR. Uno studio condotto dal nostro gruppo su coppie di fratelli affetti da CF, con genotipo identico e differente espressione polmonare, i quali venivano sottoposti agli stessi protocolli terapeutici e che seguivano lo stesso stile di vita, ha rivelato come risultati estremamente difficoltosi ricondurre la severità della manifestazione polmonare ad una specifica classe di mutazioni del gene CFTR (15).

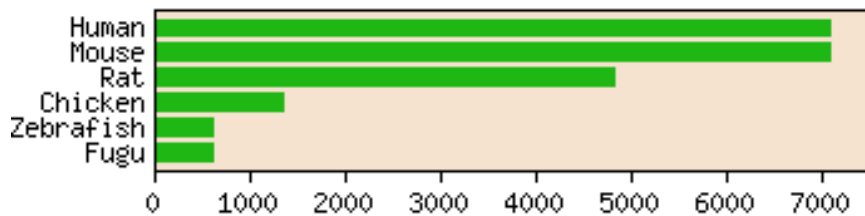
Anche per il coinvolgimento del fegato nella Fibrosi Cistica non esiste una chiara correlazione tra manifestazione clinica e difetto molecolare del gene CFTR. La compromissione epatica, la cui prevalenza è compresa tra il 4% e il 30% nei pazienti CF, ha una patogenesi non chiara. Il quadro clinico epatico, infatti, può spaziare da una lieve epatopatia cronica a forme più gravi di cirrosi giovanili che spesso richiedono il trapianto di fegato in età pediatrica. Anche per l'espressione epatica sono stati condotti studi su coppie di fratelli CF con lo stesso genotipo e fenotipo epatico discordante (15) che hanno rivelato l'assenza di una chiara correlazione di questa manifestazione clinica con uno specifico difetto molecolare del gene CFTR.

Alla diagnosi della malattia, che è basata su criteri clinici e sul test del sudore, non è possibile, dunque, associare una prognosi specifica. La mancanza di correlazione tra genotipo e fenotipo clinico della malattia costituisce ancora un punto di dibattito e di sviluppo per nuovi studi da parte dei gruppi impegnati nella ricerca della Fibrosi Cistica.

## SCOPO DELLA TESI

Dai dati presenti in letteratura è emerso che molti meccanismi di controllo dell'espressione genica di CFTR coinvolgono le regioni introniche del gene, così come il 5' e il 3'.

Recenti dati pubblicati in letteratura hanno dimostrato come, effettuando un'analisi comparativa dei genomi di varie specie animali (Homo Sapiens, Topo, Ratto, Zebrafish, Fugu), si possa rilevare la presenza di sequenze strettamente conservate nel corso dell'evoluzione e disperse in tutto il genoma (35) (Fig. 7).



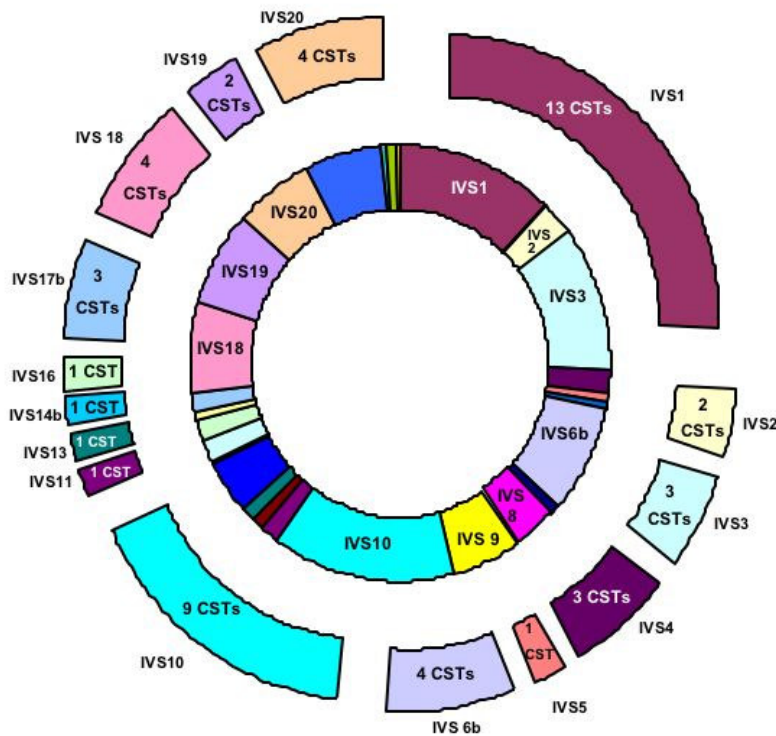
*Figura 7: CST in Homo Sapiens, Topo, Ratto, Pollo, Zebrafish, Fugu*

Nel 2005 il gruppo di ricerca in bionformatica del CEINGE (36) ha effettuato un'analisi di comparazione di sequenze tra il genoma umano e quello murino, utilizzando il tool bioinformatico BLASTZ. Sono state così identificate numerose regioni genomiche conservate denominate Conserved Sequence Tags (CSTs), la cui lunghezza supera i 100 bp e la cui omologia tra i due genomi è circa 70% (36). Le CSTs identificate con questo metodo sono state raggruppate in un database di sequenze che contiene soprattutto informazioni riguardanti i principali geni responsabili di malattia nell'uomo, con l'obiettivo di ricavare informazioni utili alla comprensione dei meccanismi molecolari ancora sconosciuti alla base di queste patologie. Con la realizzazione di questo database è emerso che il 40,5% di tali sequenze conservate sembrerebbe ricadere nelle regioni intrageniche, il 31,8% nelle regioni esoniche e il 27,7% nelle regioni introniche.

Sorprendentemente, quindi, esiste un gran numero di elementi conservati tra i diversi genomi, la cui funzionalità risulta essere ancora sconosciuta. Questi elementi, non essendo regioni codificanti, potrebbero fungere da elementi di controllo e di regolazione dell'espressione genica, potrebbero essere elementi strutturali



cromosomali o dar vita a RNA non codificante. Mutazioni in queste regioni potrebbero essere responsabili di malattie genetiche. Nel gene CFTR è stata riscontrata la presenza di numerose CSTs sia negli esoni sia negli introni. Le regioni che ospitano le CSTs sono lunghe mediamente 500 bp e consistono di una regione centrale con un'omologia tra uomo e topo uguale o superiore al 70% e le regioni circostanti con un'omologia compresa tra il 50 e il 70%. Le CSTs non sono distribuite in tutti gli introni del gene, ma principalmente nell'introne 1 (13 CSTs) e nell'introne 10 (9 CSTs). Le rimanenti si trovano invece nell'introne 2 (2 CSTs), nell'introne 3 (3 CSTs), nell'introne 4 (3 CSTs), nell'introne 5 (1 CST), nell'introne 6b (4 CSTs), nell'introne 11 (1 CST), nell'introne 13 (1 CST), nell'introne 14b (1 CST), nell'introne 16 (1 CST), nell'introne 17b (3 CSTs), nell'introne 18 (4 CSTs), nell'introne 19 (2 CSTs) e infine nell'introne 20 (4 CSTs) (fig. 8).



**Figura 8:** localizzazione delle 52 CSTs all'interno del gene CFTR.

Da quanto descritto nell'introduzione, la genetica della Fibrosi Cistica pone oggi due principali problemi: 1) anche dopo aver effettuato l'analisi molecolare dell'intera regione codificante del gene CFTR, il 10% circa degli alleli CF resta ancora non caratterizzato; 2) l'espressione clinica della malattia, anche a livello dei principali organi coinvolti, può dipendere da fattori genetici non ancora identificati.

Questo lavoro di tesi è incentrato sull'analisi delle CSTs introniche identificate nel gene CFTR per stabilire la presenza di possibili varianti alleliche in soggetti clinicamente affetti da CF, ma con genotipo non ancora definito dopo scanning genomico delle regioni codificanti. L'obiettivo è quello di ricavare informazioni utili alla comprensione dei meccanismi molecolari implicati nella regolazione dell'espressione del gene CFTR, non ancora ben definiti. Le ipotesi formulate sono che le regioni introniche altamente conservate nell'evoluzione, potrebbero contenere siti di legame per elementi di controllo e di regolazione dell'espressione genica, oppure essere elementi strutturali cromosomali o rappresentare RNA non codificante. Mutazioni in queste regioni, alterando tali meccanismi, potrebbero rivelarsi causative di malattia, o comunque modulare l'espressione clinica.

Le informazioni ottenute dall'analisi del DNA di soggetti affetti da CF, una volta analizzate e confrontate con quelle relative ad una popolazione di controllo potrebbero contribuire in maniera essenziale alla diagnostica molecolare, dando la possibilità da un lato di aumentare il "detection rate" e dall'altro di effettuare valutazioni prognostiche sempre più accurate.

Per l'analisi delle regioni introniche del gene CFTR è stata utilizzata la seguente strategia:

- Selezione delle CSTs del gene CFTR mediante tool bioinformatico.
- Selezione della popolazione da arruolare nello studio
- Disegno dei primer
- Amplificazione del gene CFTR mediante Polymerase Chain Reaction
- Analisi diretta della sequenza nucleotidica con il metodo di Sanger
- Analisi dei prodotti amplificati mediante elettroforesi capillare

## MATERIALI E METODI

### *Pazienti*

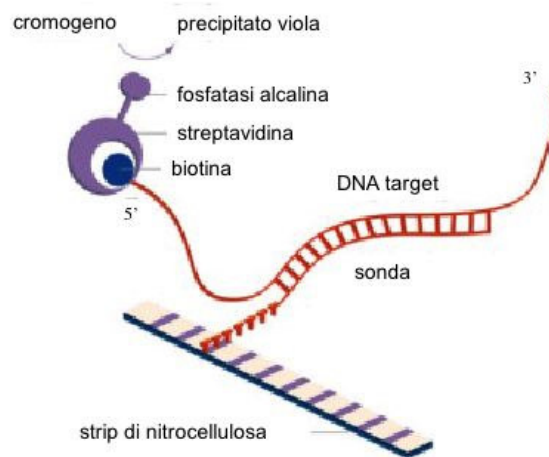
Lo studio è stato effettuato su 90 campioni di DNA conservati presso la genoteca “Fibrosi Cistica” del CEINGE di Napoli. Tutti i campioni di DNA utilizzati appartengono a pazienti che hanno dato il consenso informato all’utilizzo del proprio campione di sangue periferico per fini di ricerca. Su questi soggetti è stata ricercata la presenza di nuove varianti geniche introniche potenzialmente coinvolte nella patogenesi della malattia.

I pazienti affetti da CF, inclusi nello studio, provengono dal Sud-Italia e, in particolare, dalle regioni della Campania, della Puglia, della Basilicata e del Molise. Questi hanno ricevuto la diagnosi della malattia presso i rispettivi centri regionali deputati allo screening e al trattamento della Fibrosi Cistica, tra cui alcune strutture di diagnosi e cura dell’Azienda Ospedaliera Universitaria “Federico II” di Napoli, con le quali il nostro gruppo di ricerca collabora da diversi anni. La diagnosi della malattia è stata effettuata valutando parametri clinici (presenza di insufficienza pancreatica, valutazione della funzionalità polmonare in relazione all’età del paziente, ritardo nella crescita, valutazione della funzionalità epatica) e di laboratorio. Il test di laboratorio utilizzato per la diagnosi della malattia è il test del sudore, che consente di effettuare la diagnosi di Fibrosi Cistica attraverso la misurazione del Cloro e del Sodio presenti nel sudore, sempre elevati nei pazienti con Fibrosi Cistica. L'esame viene eseguito con la metodica di Gibson e Cooke, la quale è attualmente quella più accreditata. L'esame dura circa 40 minuti, è indolore, non richiede alcuna preparazione e può essere effettuato a partire dal 40° giorno di vita. Il test consiste nella stimolazione della sudorazione sulla parte volare dell'avambraccio tramite iontoforesi pilocarpinica della durata di 6 minuti. Successivamente il sudore prodotto viene raccolto con carta assorbente per 30 minuti. Al termine di tale periodo, il campione di sudore prelevato viene opportunamente preparato ed analizzato. Si considerano "patologiche" le concentrazioni maggiori di 60 mEq/l (milliequivalenti per litro) e "normali" le concentrazioni inferiori a 40 mEq/l. Tuttavia negli adulti si tende a considerare la soglia "patologica" sui 70 mEq/l, mentre nei neonati la soglia di "normalità" viene spostata in basso a 30 mEq/l e quella di "patologia" a 50 mEq/l. I valori compresi tra

40 e 60 (negli adulti tra 40 e 70 e nei neonati tra 30 e 50) sono considerati "borderline", cioè di incerto valore per la diagnosi.

#### *Analisi molecolare del gene CFTR*

Oltre alla diagnosi clinica della malattia, i pazienti inclusi in questo studio hanno anche ricevuto la diagnosi molecolare, che è stata effettuata inizialmente mediante la metodica del Reverse Dot Blot, che viene eseguita presso il nostro laboratorio mediante l'utilizzo di 3 kit commerciali INNO-LIPA (INNO-LiPA *CFTR* 19 e INNO-LiPA *CFTR* 17+Tn e INNO-LiPA *CFTR* Italian Regional), prodotti dalla Innogenetics (Innogenetics, Netherlands). Questa metodica si basa sull'amplificazione contemporanea di più tratti del gene *CFTR*, mediante multiplex PCR, e sulla successiva ibridazione del prodotto con sonde molecolari depositate su strisce di nitrocellulosa. Per ogni mutazione da evidenziare sono costruite due sonde, una relativa alla sequenza normale e l'altra alla sequenza mutata. Una reazione colorimetrica (biotina/streptavidina/fosfatasi alcalina) permette la visualizzazione di un pattern di bande diverso a seconda che il soggetto in esame sia omozigote normale, omozigote mutato, eterozigote o eterozigote composito (Fig. 9).



**Figura 9:** *schema della metodica Reverse Dot-Blot.*

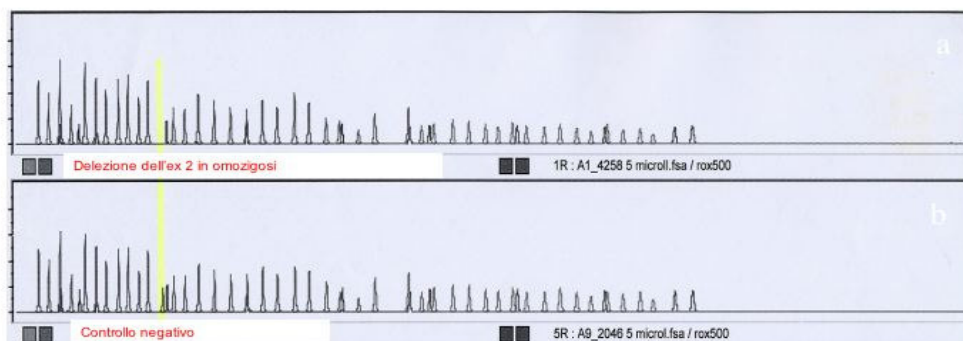
Mediante l'analisi diretta di 50 mutazioni causative, con la metodica del Reverse Dot Blot, il nostro laboratorio di diagnostica molecolare è in grado di individuare l'85%

circa degli alleli CF più frequenti nella nostra area geografica (Campania, Puglia, Basilicata e Molise).

I pazienti risultati negativi al Reverse Dot Blot e con diagnosi clinica confermata per la Fibrosi Cistica sono stati sottoposti a scanning dell'intera regione codificante, mediante amplificazione e sequenziamento diretto dei 27 esoni del gene CFTR. Nel caso in cui anche questa metodica non rivelasse la presenza di mutazioni è stata eseguita la ricerca di macrodelezioni del gene CFTR, mediante la metodica della MLPA. Il DNA di ogni paziente è stato analizzato mediante un kit commerciale basato su PCR quantitativa seguita da elettroforesi capillare (MLPA SALSA kit, MRC-Holland, Netherlands) (37). La tecnica MLPA (multiplex ligation-dependent probe assay) permette di analizzare i 27 esoni del gene CFTR mediante l'amplificazione sequenza specifica e simultanea effettuata con un set di 44 probe sintetici, essa può essere così schematizzata:

- denaturazione del DNA: 98°C per 5'
- ibridazione con i SALSA-probe: i probe ibridizzano alla sequenza da analizzare mediante un'incubazione di 16 ore a 60°C.
- reazione di ligasi: la ligasi termostabile (Ligasi 65) si attiva a 54°C per 15' e si inattiva una volta raggiunti i 98°C.
- reazione di amplificazione:
- analisi mediante elettroforesi capillare: con l'utilizzo di Rox 500 come standard interno.

Con questo metodo, il numero di copie della sequenza target è confrontato con quello di un campione di controllo; la presenza di una macrodelezione comporta una riduzione del 35-50% dell'area del picco relativa al prodotto di amplificazione della sequenza target. Inoltre i dati ottenuti sono confrontati con quelli di un campione eterozigote per una macrodelezione del gene CFTR, di un campione wild-type ed, infine, di un controllo negativo. Nella figura 10 è riportato un esempio della macrodelezione dell'esone 2 in omozigosi in confronto con l'elettroferogramma di un campione wildtype.



**Figura 10:** elettroferogramma dei 27 esoni del gene *CFTR*, ottenuto con la metodica dell'MLPA

I risultati ottenuti mediante MLPA sono stati anche confermati tramite l'utilizzo di una metodica *home made*, basata sull'amplificazione, mediante multiplex PCR e successiva corsa elettroforetica, dei breakpoint delle macrodelezioni individuate con la MLPA.

Tutte le metodiche fino ad ora descritte consentono di raggiungere un "detection rate" pari al 92%. Tuttavia, in 37 pazienti, originari del Sud-Italia e non imparentati tra loro, che hanno ricevuto una diagnosi clinica e di laboratorio positiva per Fibrosi Cistica non è stato possibile identificare una o entrambe le mutazioni causative di malattia, per un totale di 43 alleli con mutazioni non caratterizzate. Tali pazienti sono risultati essere un interessante gruppo di studio per l'analisi molecolare delle 52 CSTs introniche identificate nel gene *CFTR*.

Come popolazione di controllo sono stati selezionati 4 gruppi di soggetti.

In particolare, i primi due gruppi sono costituiti da:

- a) 10 pazienti affetti da CF, non imparentati tra loro, con genotipo F508del/F508del e fenotipo severo. Questi soggetti presentavano grave insufficienza pancreatica e un'espressione epatica severa della malattia, caratterizzata da cirrosi multilobulare e ipertensione portale (con varici esofagee o splenomegalia) tali da richiedere il trapianto di fegato, tutte le altre cause (genetiche, metaboliche e virali) di malattia epatica sono state escluse;
- b) 10 pazienti affetti da CF, non imparentati tra loro, con genotipo F508del/F508del e fenotipo lieve. Anche questo gruppo di soggetti presentava

insufficienza pancreatica, ma non era presente alcuna manifestazione epatica della malattia.

Questi due gruppi, pur avendo lo stesso genotipo, presentano una diversa manifestazione clinica della malattia e, quindi, potrebbero essere portatori di ulteriori varianti geniche introniche correlabili allo sviluppo di uno specifico fenotipo.

Il terzo gruppo è costituito da 13 pazienti affetti da CF con genotipo caratterizzato dalla presenza di mutazioni note di varia tipologia, tra cui macrodelezioni e fenotipo grave, la selezione di questi pazienti è stata effettuata allo scopo di valutare l'eventuale associazione delle varianti introniche ad alleli con comuni mutazioni causative di malattia.

Nell'ultimo gruppo sono stati inseriti 20 soggetti, fratelli o sorelle di pazienti affetti con genotipo noto, che dopo l'analisi molecolare sono risultati non portatori di mutazioni del gene CFTR.

*Analisi delle CSTs del gene CFTR mediante PCR seguita da sequenziamento capillare*

Per l'amplificazione delle CSTs del gene CFTR sono state disegnate 52 diverse coppie di primer, con la stessa temperatura di annealing (54°C). Nella tabella 1 sono riportate le sequenze degli oligonucleotidi utilizzati e la lunghezza in base pair (bp) relativa al prodotto amplificato.

	CST ID*	Sequenze dei primer	Amplificato
IVS1	594018	F:GCCTAATGTGATGAGCCACA R:AGGTGGAGTAACAAAAGGTGGA	479bp
	594020-594022	F:GGGGGAAGGAATTTAGCTTTC R:GCAAACCAGCCTGATAGCAT	484bp
	594024	F:TGATCTGTCAAAGTGGGTTTTT R:CAAGGTGCATGTACTGTGATGA	577bp
	594026	F:GGATCCCTTCAGTTTGAGAGTC R:TTCAGGGCATGAGAGAAGAGA	534bp
	594028	F:GACTTCTACACCAGTCCCTCAA R:GGAAAACCTGGAAAACCGTGA	531bp
	594030	F:GTGTTATGGGAGGGAGTGGA R:AAAGCCGTAAAGCCACTCTG	532bp
	594032	F:TTTCTTGATCCTACATGGTTGTG R:AGCCTTCCAAAAAGAAGCA	620bp
	594034	F:GCTCTGAAAAGGCTGAGAGTGA R:GGGCTGTCAGGCTTGTTTAG	519bp
	594036	F:CCATGCCCAAGTGATGGTAAT R:TGGCTCAAGGTTTACCAAGG	490bp
	594038	F:CCCCAAAGAGAAATGTCACC R:TTGTGCAACTGTCAAAACGTC	499bp
	594040	F:CCAAGGAGAGGTGGAACCTG	485bp

		R:GCATCCTCAGAGTGCCTTTC	
	594042	F:TCATTGTTGGGCAGCATAAG	550bp
	594044	R:AAAGGAAAATGTGCTCCCACT	
		F:TTGCCACCTGTATGTTTCGAC	503bp
		R:GGCCCTCATATCCAATCTCA	
IVS2	594048	F:GCCAGGCTTGTCTTCAAATC	570bp
		R:TGCCACTGATTTCTGTTAATGA	
	594050	F:CCTAGTTCTTCCTTGTGGTTCA	400bp
		R:TCAGGTGGCTTTTTCTGTGA	
IVS3	594054	F:GGATGAATCCAACCTCAAACA	509bp
		R:TGGGGAACATACAAGGCACT	
	594056	F:TGCCATCATGAAGGAGCTAA	499bp
		R:CAACAGCCAGTTTCTCTTCCA	
	594058	F:TCAGGACGCTTCTTGTGTCT	490bp
		R:TCCTCAACCGAAGAGGACAT	
IVS4	594062-594064	F:GCTGTTTTGATGGACTTATTTGG	634bp
		R:TTGAAGCAGTGTTACACCA	
	594066	F:GCTTTGGTTTGCTCCCTAAA	503bp
		R:TCTGGAAGGATTTTCTTGAGGA	
	594068	F:GCCTTGAAGAAGGGATTGTG	536bp
		R:CATTGATATTGTGCGGGAGTT	
IVS5	594072	F:AACTGGAAAAGGCGGAGTTTT	520bp
		R:GCATAAAGCAAGAACCACAGC	
IVS6b	594078	F:AACTCGGGAATGCATAGGG	494bp
		R:ACCTGCAGCTCAGTTTCCAT	
	594080	F:TCAGCACATGGGAATTTCA	493bp
		R:AAGAATGACCCAGCCATGAG	
	594082	F:GTTCTTGGGGGAATTGTTCA	507bp
		R:CTTGCACAGCCTTTTGACC	
	594084	F:GGAGCACTGAAGCCTCATAT	499bp
		R:TTTGTCTTGTGGTTGGTTG	
IVS10	594094	F:GGGCAAACCGTGGATTAGAT	401bp
		R:CTTCCGCAACTTTTCCACTC	
	594096	F:CGATACCAAATGCAAACAT	547bp
		R:TTAATGGGCACAGGGATACG	
	594098	F:TCCTGCACTTGACAATACCG	513bp
		R:CCCTCCAAGCTTTTTTCATA	
	594100	F:AACTCTGCTTGGCAGTTTCA	380bp
		R:CTTGGCTACTGAATGCAAATGT	
	594102	F:CATTTGCATTTCAGTAGCCAAGA	729bp
		R:TGACTGCAAACCTCTGGGTGA	
	594104	F:TTTCATTTGAGCCCTGGAAC	499bp
		R:TCAAGAGGGTCGAAACATCG	
	594106-594108	F:TGTCCATTTTCTTGTGCAT	570bp
		R:GGAGTGCATAACAACATATCCAA	
	594110	F:TGGTTTTGTTGGTCCAGTTG	549bp
		R:ATCCTCCACTGCCATTTCAA	
	594112	F:ACTCCAAGGGAACCTCGAAGG	489bp
		R:TGAAGCTGCTGAATCTGGAA	
IVS11	594116	F:CCATCATAGACCCTGCCACT	459bp
		R:TCAGCTTTTGGAAAGCAACA	
IVS13	594200	F:AAGGAGGCAGTCTGTCTGA	514bp
		R:CCACAGAGGGTTCAAGATGG	
IVS14b	594128	F:TGTTTCATTTCTCCCAAGCA	461bp
		R:GAGCATGCACCCTTAACCTC	
IVS16	594134	F:TGGAGGAGGTGAAAGTGACC	501bp
		R:CATGAATTCAACGCATGGAT	
IVS17b	594140	F:GGAAAATGAAACATCCGCATT	506bp
		R:ACCTTCTCGGTGTTCCATGA	
	594142	F:CTACCAGGGCTATGCCACAT	492bp
		R:CGGCTGAGCATCTGATTTG	
	594144	F:CCCTGCTTTTGGTCTCTTTG	500bp
		R:GCTGGAGTTTACAGCCCACT	
IVS18	594148	F:GTTGGGAAACAGCAGGAGAA	523bp
		R:CCAGGAAAGGCTACTTGTGC	



	594150	F:TCTCCAAGTGAAGCAGCAGA R:TGACTCCTCCCCTCAGATTC	418bp
	594152	F:TTGTGCAGTGCTGCTCATAG R:GGCATTTCACCTTCTGTGT	532bp
	594202	F:GTGTCCTGTGGGCCATAGTT R:AATATGCTGGTGCCCATTTTC	487bp
IVS19	594156	F:TTGCGGGAAGTACTCTAGCC R:CTTGAATATGCCCAGTGCAG	487bp
	594158	F:TGCTTCTGGCTTGAGCCTAT R:TTCCCTGATCCAGTTCTTCC	398bp
IVS20	594162	F:AGCCTTGGGCTTTTGTTTT R:TCCATTAGCACAAAGGGAGAA	504bp
	594164-594166	F:GGACCAAATCTGGACAGCA R:TGGGACCACAAGACTTAGGTT	554bp
	594168	F:ATGGGATTTGGAGCCTTGTT R:AACCACCACAGCACAGCATA	524bp
	594170	F:CTTGGGTGACCTCCAAAGAC R:TAATTGCTTGGGAAATGGTG	472bp

**Tabella 1:** Ad ogni CST è stato attribuito un unico codice numerico identificativo ricavato dal database DG-CST (<http://dgcst.ceinge.unina.it/>)

Tutte le reazioni di amplificazione sono state eseguite su Thermal Cyclers 2700 e 9700 della Applied Biosystems seguendo il protocollo di seguito riportato:

Reagenti	Volume (µl)
H <sub>2</sub> O	24,7
Hot Master Mix 2,5X	20
Oligo F (8µM)	1
Oligo R (8µM)	1

Il volume finale di reazione è pari a 50 µl (3.3 µl di DNA + 46.7 µl della miscela di reazione). L'Hot Master Mix (HMM, Eppendorf) contiene tutti i reagenti necessari alla reazione di amplificazione, quali MgCl<sub>2</sub>, dNTPs, tampone di reazione ed enzima Taq polimerasi termostabile. Inoltre, è presente un inibitore complessato all'enzima DNA polimerasi, che lo inattiva a temperature inferiori a 40°C, impedendone il legame al DNA stampo ed aumentandone, in tal modo, la specificità. A temperature maggiori di 60°C, l'inibitore è spiazzato dal legame del DNA all'enzima Taq polimerasi. Nel tampone di reazione è contenuto un ulteriore reattivo, un agente chelante gli ioni Mg<sup>2+</sup>, che ne regola la concentrazione in maniera ottimale. L'amplificazione di ciascuna CST prevede l'allestimento di un controllo negativo,

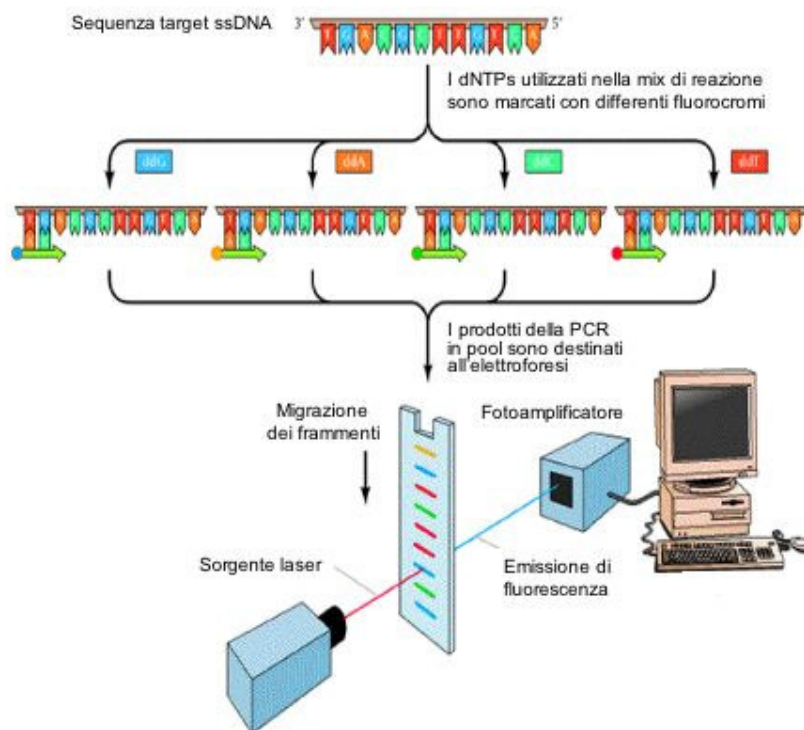
costituito dalla presenza di tutti i reagenti escluso il DNA; tale controllo è necessario per valutare eventuali contaminazioni avvenute durante l'allestimento della reazione.

Il programma di amplificazione è costituito dalle seguenti fasi:

94°C x 90''	
94°C x 30''	
54°C x 30''	35 cicli
65°C x 60''	
65°C x 5'	

Per il controllo della reazione di amplificazione è stata eseguita un'elettroforesi su gel d'agarosio al 2% in TBE 1X, utilizzando 4 µl della miscela di reazione.

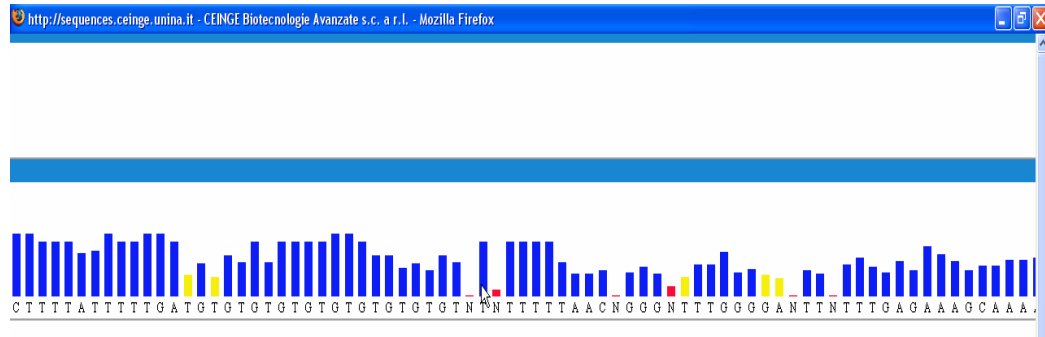
Successivamente, i prodotti di amplificazione sono stati preparati per essere processati presso il servizio di sequenziamento presente nel nostro istituto. Il sequenziamento diretto, che viene eseguito presso il CEINGE di Napoli, è stato standardizzato mediante una procedura automatizzata basata sul metodo a terminazione di catena ideato da Sanger nel '77 (fig. 11) (38).



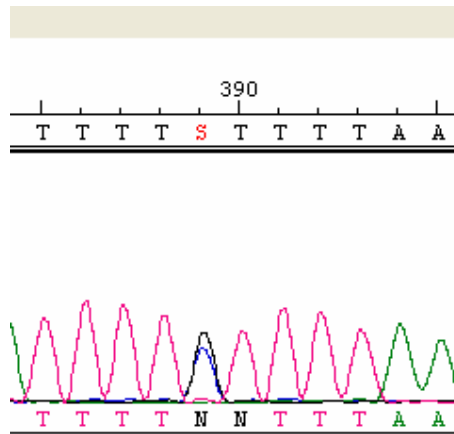
**Figura 11:** *schema della metodica del sequenziamento capillare (metodo di Sanger modificato).*

La sequenza di una molecola di DNA a singolo filamento viene determinata tramite la sintesi enzimatica di catene polinucleotidiche che terminano ciascuna con una diversa base azotata modificata (ddNTP), in seguito all'incorporazione e alla coniugazione ad un fluorocromo. I frammenti ottenuti sono analizzati tramite elettroforesi capillare e rilevati mediante un segnale di fluorescenza caratteristico della specifica base azotata incorporata. Per ciascun amplificato è stata determinata la sequenza su entrambi i filamenti di DNA. Dopo opportuni controlli di qualità operati dal servizio di sequenziamento (fig. 12), la sequenza può essere visualizzata sia sottoforma di elettroferogramma (fig. 13), in cui la rappresentazione grafica della composizione in basi del campione di DNA in esame appare in una successione di picchi colorati,

ciascuno dei quali è rappresentativo di una base azotata, sia in formato FASTA (fig. 14 ), cioè in una successione di lettere che corrispondono alla sequenza di basi azotate del campione, che può essere facilmente confrontata ed immessa nei database di sequenze disponibili on line.



*Figura 12: ogni sequenza generata dal servizio di sequenziamento viene sottoposta a controlli di qualità che confermino l'attendibilità dell'attribuzione del segnale luminoso del fluorocromo alla base azotata corrispondente.*



*Figura 13: esempio di un elettroferogramma relativo alla sequenza di un campione risultato eterozigote per una sostituzione nucleotidica di una G (picco nero) in una C (picco blu).*

CEINGE Biotecnologie Avanzate s.c. a r.l. - Mozilla Firefox

File Modifica Visualizza Vai Segnalibri Strumenti ?

http://sequences.ceinge.unina.it/viewseq.php?varcode=8101.01&varrequestid=8101&selectedproject=

**REQUEST**

<b>Request number:</b> 8101	<b>Accept date:</b> 2007-05-04	<b>Finish date:</b> 2007-05-07	<b>Sequences:</b> 30
<b>Institution:</b> dbbm	<b>User:</b> vanacore@dbbm	<b>Email:</b>	<b>Telephone:</b> 860
<b>Notes:</b>			

**SEQUENCE**

<b>Name</b> 6406_9	<b>Code</b> 8101.01	<b>Template</b> PCR
<b>Vector</b>	<b>Primer</b> CF9For	<b>Status</b> done
<b>Do purification</b> yes	<b>Length</b> 364	

```

LOCUS       Sample: 6406_9
DEFINITION Code: 8101.01
BASE COUNT   74 a    39 c    62 g    119 t    70 others
ORIGIN
1  NNNNNNNNNN NNNNNNNNNN NNNNANATCT GACAAACTCA TCTTTTATTT TTGATGTGTG
61  TGTGTGTGTG TGTGTNTNTT TTTAACNGGG NTTTGGGGAN TINTTTGAGA AAGCAAAAACA
121 AAACAATAAC AATANAAAAA NTTCTANNGG NGANGACNNC CTNTTCTTCA GTANTTTCTC
181 ANTTCTTGGT ACTCCTGTCC TGAAAAGATAT TANTTTCAAG ATANAAAAGAN GACANNTGTT
241 GGNNGNTGCT GGATCCANTG GANCNGGCAN GGTANTTNTT TTGTCCTTCN CTATTAANAA
301 CTTANTTTGG TGCCNTGTC TTTTTTTTTT TCTANTTTGT AGNGNNGGAN GGTNTTTTTNN
361 GANN
//

```

[Back to sequences list](#) [Back to control panel](#)

**Figura 14:** esempio di una sequenza in formato FASTA, ottenuta dal servizio di sequenziamento del CEINGE di Napoli. La sequenza ottenuta in tale formato può essere immessa direttamente dal sito del servizio di sequenziamento nei database on line ed essere confrontata con le sequenze di riferimento.

## RISULTATI E DISCUSSIONE

*L'analisi delle CSTs ha portato all'identificazione di alcune varianti geniche specifiche dei pazienti CF con mutazioni non note*

L'analisi delle 52 CSTs introniche del gene CFTR è stata condotta su un totale di 90 soggetti ed ha comportato la realizzazione di circa 9.360 sequenze. Tutti i dati derivanti dall'analisi di sequenza sono stati immessi in un database, creato ad hoc dal gruppo di ricerca in bioinformatica del CEINGE, attraverso l'utilizzo del linguaggio PostgreSQL (fig. 15).

The screenshot shows the CEINGE database interface. At the top, there is a navigation bar with 'HOME LOCAL WEBMAIL BIOINFORMATICS' and a user profile 'User: elce'. The main header features the CEINGE logo and the text 'BIOTECNOLOGIE AVANZATE' with the tagline 'listening to biomolecules to silence disease'. Below this, there are language options for 'English' and 'Italiano'. The interface is divided into a left sidebar with navigation menus for 'Sequences', 'Subjects', 'Regions', 'Mutations', 'Detected muts', and 'Analysed genes'. The main content area is titled 'Subjects' and displays a list of patient data. Below this, there is a section for 'Detected mutations' with a table listing specific variations.

ID	Variation	Sequence	Diagnostic	This study	Locus	Origin	Controversial	
7	F508del		<input checked="" type="checkbox"/>	<input type="checkbox"/>	Homozygous		<input type="checkbox"/>	View
8	Unknown		<input type="checkbox"/>	<input type="checkbox"/>	Heterozygous		<input type="checkbox"/>	View

**Figura 15:** una delle schermate del database in cui sono stati immessi i dati relativi all'analisi di sequenza delle CSTs (<http://web.ceinge.unina.it/ceinge/cftr/>).

In primo luogo, nel database sono stati importati i dati relativi all'analisi molecolare del gene CFTR e ai parametri clinici dei soggetti inclusi nello studio. Di volta in volta, sono state archiviate le sequenze sia sotto forma di elettroferogramma che in formato FASTA e i dati derivati dall'analisi delle sequenze sono stati elaborati mediante un programma specifico che attribuisce alla variante il corretto nome, in accordo alla nomenclatura standard (39). Una volta identificate, le varianti sono state confrontate con una lista di polimorfismi già noti in letteratura ([www.ensembl.org](http://www.ensembl.org)) e quindi classificate come nuove o già conosciute (fig. 16).

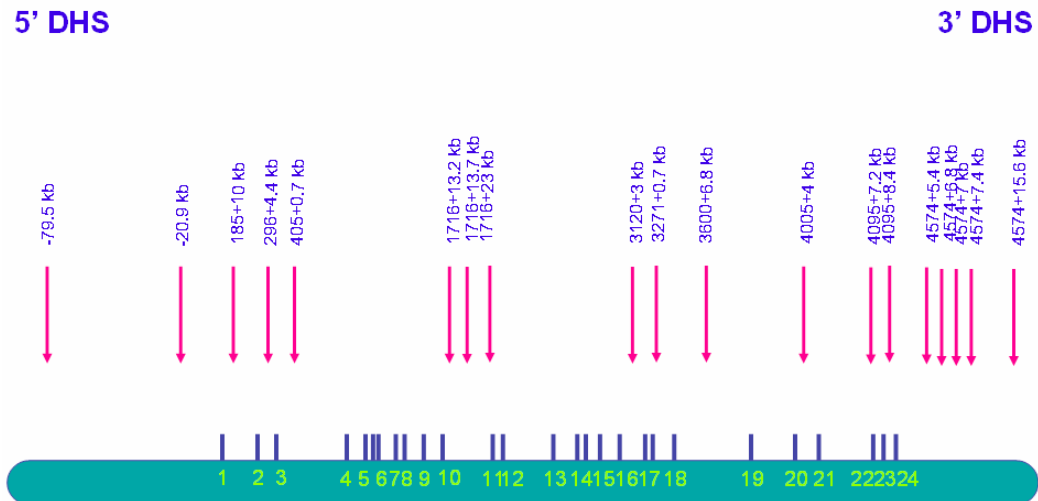


**Figura 16:** esempio dell'identificazione d una variante genica in una CST con l'utilizzo di un programma in linguaggio Java, creato dal gruppo di bioinformatica del CEINGE di Napoli. Il programma compara la sequenza immessa in formato FASTA con quella di riferimento, individuando le differenze ed annotandole nel database.

Dall'analisi delle 9.360 sequenze è scaturito che parte delle CSTs (57,6%) risulta essere conservata nei vari gruppi di soggetti analizzati, e parte (42,3%) presenta un certo grado di variabilità, in alcuni casi tale da poter essere sfruttata in ambito diagnostico. In particolare, il 57,6% delle regioni CSTs analizzate non ha presentato varianti geniche in nessun gruppo di soggetti in esame. Questo dato, una volta rafforzato dall'ampliamento della casistica e rapportato al grado di conservazione inter-specie delle regioni introniche, potrebbe essere di notevole interesse nel costituire un ulteriore tassello per la comprensione del ruolo degli introni nell'ambito della funzionalità del gene. E' stato ipotizzato, infatti, che l'alto grado di conservazione delle CSTs potrebbe derivare dal fatto che queste regioni possano essere essenziali per il legame con fattori trascrizionali, o essere implicate nei meccanismi di splicing, per cui la mancanza o l'alterazione di queste regioni potrebbe rivelarsi incompatibile con la vita.

Molteplici dati derivanti dalla letteratura confermano il ruolo delle regioni introniche nella funzionalità di svariati geni. Soprattutto dallo studio dei siti ipersensibili al taglio della DNAsi I (DHS), è emerso che gli introni sono coinvolti nei meccanismi di controllo dell'espressione genica. Molti geni presentano siti DHS sia in regioni codificanti che non codificanti. Alcuni DHS sono costitutivi, come quelli nel

gruppo dell' $\alpha$ -globina, cioè presenti in tutte le cellule, mentre altri hanno una specificità tessuto-temporale. Per quanto concerne il gene CFTR, ad oggi, sono stati identificati dodici DHS all'interno delle regioni introniche del gene CFTR, due al 5' e cinque al 3' (fig. 17).



**Figura 17:** DHS identificate sul gene CFTR.

I siti intronici sono localizzati negli introni 1, 2, 3, 10, 16, 17a, 18, 20, 21 e sono stati denominati in base alla loro distanza in kb dall'ultima base del precedente esone (numero di accesso alla CFTR cDNA GenBank M28668) seguendo le linee guida del Consorzio di Analisi Genetica CF (<http://www.genet.sickkids.on.ca>).

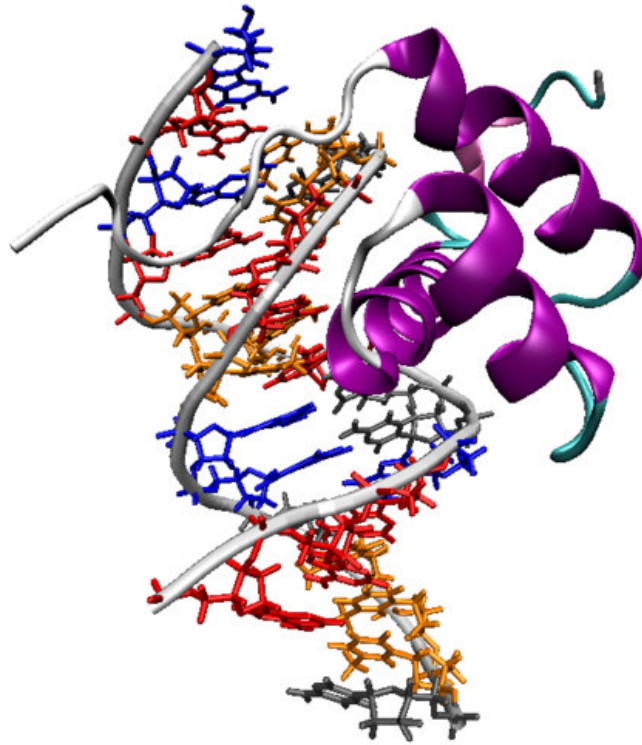
Numerosi studi hanno confermato che alcuni siti DHS del gene CFTR siano presenti in maniera selettiva soltanto in alcuni tipi di linee cellulari, suggerendo una certa specificità del sito in relazione al tessuto. Ad esempio, è stato osservato che i DHS negli introni 2, 16, 17a e 18 sono presenti esclusivamente in cellule derivanti dai dotti pancreatici e nelle linee cellulari intestinali (40). E' stata testata la capacità dei DHS di influenzare l'attività del promotore di CFTR (41), ed è risultato che solo i DHS degli introni 1, 20 e 21 ne incrementano l'attività. Nello stesso studio, inoltre, è stato osservato che i DHS negli introni 10 e 18 sono responsivi ad attivatori dell'espressione di CFTR in cellule Caco2 e Capan1 da adenocarcinoma pancreatico (es., forskolin, un induttore di cAMP intracellulare, e mitomycin C) (41).



L'analisi genomica comparativa del gene CFTR umano, di pecora, di mucca e maiale ha permesso di mappare più dettagliatamente i DHS negli introni 16 e 20 del gene umano (42). L'analisi comparativa di specie omologhe ha anche rivelato che i DHS negli introni 2, 3, 10, 17a, 18 e 21, insieme ai DHS nella regione del gene al 3', mostrano un alto grado di conservazione tra le specie (42). Nella regione al 3' del gene CFTR sono state mappate cinque DHS che mostrano un certo grado di tessuto-specificità solo nelle linee cellulari epiteliali (43).

Dati di notevole interesse sono scaturiti soprattutto dallo studio del DHS dell'introne 1, per il quale è stata osservata la capacità di agire sinergicamente al promotore, aumentando l'espressione del gene reporter di oltre il 60% rispetto al costrutto privo di questo sito, esclusivamente in linee cellulari intestinali (44).

Tramite DNaseI footprinting, usando estratti nucleari Caco2, che esprimono alti livelli di CFTR, sono stati analizzati fattori trascrizionali che potenzialmente potrebbero legarsi alle regioni che contengono DHS negli introni 10, 17a e 20 (45). La sequenza è stata analizzata con il database Transfac 4.0 e sono stati individuati sette siti leganti il fattore nucleare epatocitico 1 (HNF1) (fig. 18): tre nell'introne 10, due nell'introne 17a, e due nell'introne 20. Sebbene originariamente sia stato descritto come un fattore epatico-specifico, HNF1 è espresso anche nel pancreas, nel rene, nello stomaco e nell'intestino tenue; inoltre regola alcuni geni intestinali e pancreatico-specifici come il gene  $\alpha_1$ -antitripsina e il gene dell'insulina attraverso l'interazione con la regione del promotore o tramite i siti intronici (46,47).



**Figura 18:** HNF1 è in grado di interagire con il DNA tramite il dominio elica-giro-elica, inducendo un ripiegamento della regione a cui è legato e contemporaneamente reclutando sia l'apparato trascrizionale che le proteine responsabili della acetilazione degli istoni, e quindi dello svolgimento della cromatina.

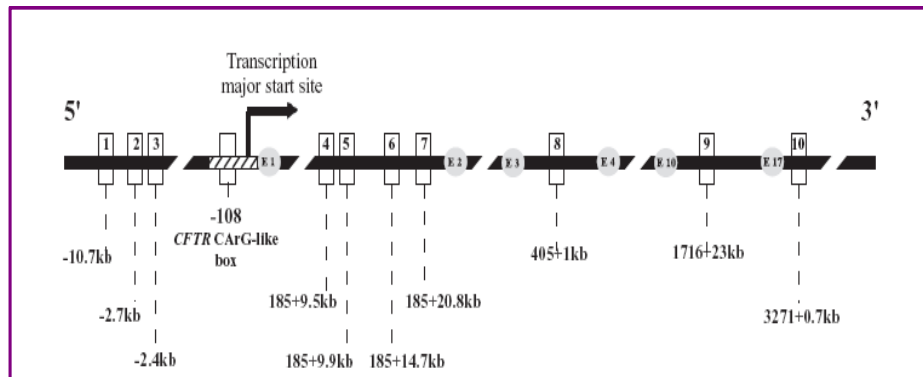
Dunque, è possibile che HNF1, legandosi ad alcuni siti intronici, possa modulare l'espressione di CFTR. Nelle linee cellulari Caco2, l'mRNA di CFTR è espresso a bassi livelli prima che le cellule arrivino a confluenza, con la differenziazione spontanea post-confluenza si ha un parallelo aumento dei livelli di espressione di CFTR e di HNF1 (48). Inoltre, l'analisi dei livelli di espressione di CFTR in topi knockout per HNF1 ha mostrato una significativa diminuzione ( $P=0.037$ ) dei livelli di CFTR nell'intestino tenue se paragonata a quella riscontrata nei topi wildtype. Interessante è il fatto che il ceppo di topi negativo a HNF1, oltre a mostrare un nanismo di tipo Laron e un diabete mellito non insulino-dipendente, presenta sterilità per la presenza di tratti genitali maschili rudimentali. La diminuzione dell'espressione di CFTR in questi topi potrebbe essere ricondotta all'infertilità maschile causata dall'assenza o parziale presenza dei vasi deferenti (49).

Dei sette siti HNF1 identificati nel gene CFTR, solo quello nell'introne 10 è omologo nel topo e nell'uomo. Un possibile meccanismo attraverso il quale HNF1 regola la

trascrizione di CFTR potrebbe essere dovuto al particolare ripiegamento del DNA indotto da HNF1, grazie al quale esso è capace, dai siti intronici a cui è legato, di interagire con il promotore: è stata già dimostrata l'interazione di HNF1 con proteine che hanno effetti sull'espressione di CFTR, come la proteina legante CREB (CBP/p300), che quando è legata al dominio N-terminale di HNF1 aumenta rapidamente legando un altro coattivatore, il fattore associato CBP (P/CAF), all'estremità C-terminale. La capacità nel reclutare l'istone acetiltransferasi ai geni target, in aggiunta al rimodellamento della cromatina, attraverso la demetilazione del promotore, potrebbe attivare la trascrizione di CFTR tramite modificazione nucleosomale e reclutamento dei meccanismi generali di trascrizione (50, 51). Mentre HNF1 $\alpha$  è coinvolto nella regolazione di CFTR nell'intestino tenue e nell'epitelio dei dotti pancreatici, il preciso meccanismo che porta ad una regolazione temporale e tessuto-specifica di CFTR sembra coinvolgere diverse proteine agenti *in trans* che si legano ad elementi multipli regolatori situati all'interno del gene.

Nel 2005, ulteriori studi sui siti DHS di CFTR hanno portato all'identificazione di un fattore trascrizionale che, come HNF1, è in grado di modulare l'espressione di CFTR tramite il legame a regioni introniche del gene. Il fattore trascrizionale, denominato Serum Response Factor (SRF), è capace di interagire con una regione del promotore di CFTR, definita Serum Response Element (SRE) (52). Questa regione, chiamata anche CArG box, è costituita dalla sequenza nucleotidica CC(A/T)<sub>6</sub>GG e è stata identificata per la prima volta a -108 dall'inizio del sito di inizio trascrizione. In questa regione era stata precedentemente identificata dal gruppo di Romey un sito di binding per il fattore trascrizionale Ying Yang 1 (YY1), per il quale è stato dimostrato il ruolo di repressore trascrizionale per l'espressione di CFTR (53). SRF sembrerebbe, invece, fungere da attivatore trascrizionale di CFTR, poiché è in grado di competere con YY1, sia *in vitro* che *in vivo*, per il legame al promotore del gene. Il fattore SRF, che appartiene ad una famiglia di proteine con motivo CArG-like strettamente conservato tra le specie, è un importante attivatore trascrizionale per numerosi geni tessuto-specifici, inoltre, sembra svolgere un ruolo cruciale nell'ambito dello sviluppo delle vie aeree di *Drosophyla*. Per questo motivo è stato ipotizzato che esso possa avere un ruolo analogo nello sviluppo del sistema polmonare umano, determinando l'attivazione trascrizionale del gene CFTR nel corso dello sviluppo embrionale (52). Studi computazionali hanno rivelato come SRF abbia siti di binding non solo nel

promotore di CFTR, ma anche negli introni (fig. 19), quindi è ipotizzabile che, come HNF1, induca un ripiegamento nel DNA necessario per l'avvicinamento al promotore e il reclutamento di fattori trascrizionali e proteine implicate nel rimodellamento della cromatina in corrispondenza del promotore stesso.



**Figura 19:** SRF ha molteplici siti di binding al 5', negli introni e al 3' di CFTR

Non è stato completamente chiarito il meccanismo con cui SRF agisca attivando la trascrizione di CFTR, per cui ulteriori studi potrebbero da un lato definire in maniera più esaustiva i meccanismi di regolazione del gene CFTR che coinvolgono gli introni e dall'altro individuare target più specifici per il trattamento di questa patologia, dato che anche un lieve aumento dell'espressione funzionale di CFTR è in grado di migliorare il fenotipo della malattia, soprattutto in quei tessuti, come il polmone, che necessitano di una attività residua minima di CFTR per una corretta funzionalità.

Basandoci sulle informazioni ottenute dalla letteratura, che confermano il coinvolgimento degli introni nella regolazione della funzionalità del gene CFTR, intendiamo proseguire lo studio analizzando quelle regioni introniche conservate che non presentano varianti geniche. Dopo aver compiuto un'accurata analisi statistica dei risultati a nostra disposizione sarebbe interessante approfondire le basi molecolari che da un lato hanno portato alla conservazione di alcune CSTs e dall'altro alla loro variabilità. A tale scopo potremmo avvalerci dell'analisi computazionale, al fine di effettuare una comparazione tra le sequenze da noi ottenute e le sequenze dei principali motivi regolatori attualmente conosciuti e pubblicati in letteratura. Questo tipo di analisi potrebbe rivelare la presenza di meccanismi coinvolti nell'espressione genica o nella organizzazione della cromatina ad oggi sconosciuti nel gene responsabile della Fibrosi Cistica.

Dal nostro studio sulle CSTs introniche del gene CFTR è anche emerso che 42,3% delle regioni analizzate presenta un certo grado di variabilità: in totale sono state annotate 43 variazioni di sequenza di cui 17 non note e 26 note in letteratura (<http://www.genet.sickkids.on.ca/cftr/>) (tab. 2). In particolare, si tratta di 39 sostituzioni nucleotidiche, un'inserzione, una STR e 2 polyA. Le varianti identificate ricadono sia nella porzione delle CSTs con un'omologia del 70% tra uomo e topo, sia nelle regioni circostanti. La distribuzione di tali varianti nei soggetti studiati è abbastanza complessa: alcune sono state riscontrate con la stessa frequenza in tutti i gruppi testati, dimostrando quindi un'assenza di correlazione tra la presenza della variante e la presenza della malattia; 11 varianti sono state, invece, riscontrate solo in pazienti affetti da CF con alleli caratterizzati da assenza di mutazioni nelle regioni codificanti. Per queste 11 varianti è stato effettuato uno studio aggiuntivo, ampliando la casistica di partenza con una popolazione di 90 soggetti sani originari del Sud-Italia (180 alleli), scelti in maniera casuale nella popolazione generale. Questo ulteriore studio ha permesso di restringere l'interesse a sole 2 varianti alleliche la cui prevalenza è stata osservata esclusivamente nella popolazione dei 37 soggetti affetti con genotipo non definito. Le 2 varianti non sono state descritte da altri gruppi di studio e ricadono nella regione di alta omologia tra uomo e topo, per cui sarebbe auspicabile mettere a punto un protocollo sperimentale che consenta di determinare la funzione di queste 2 regioni del gene, analizzando, ad esempio, l'interazione di eventuali interattori molecolari con la sequenza wild type e con la sequenza mutata.

CST	Variante	Note
594026	185+2253T/G	Non descritta
594032	185+8741 A/C 185+9203 T/C	
594034	185+9290 A/G	All'interno della CST
594038	186-9673 G/A 186-9574 G/A 186-9531 T/A	All'interno della CST, non descritta All'interno della CST, non descritta
594040	186-9282 G/C 186-9095 C/T	All'interno della CST
594054	405+296 A/T 405+565 T/C	All'interno della CST, non descritta
594056	405+3213 C/T 405+3479 G/A	Non descritta
594058	406-7999 A/G 406-7920 G/C 406-7906 T/A 406-7868 T/A	All'interno della CST All'interno della CST, non descritta All'interno della CST Non descritta
594062-64	621+1053 G/A 621+1089 G/A 621+1101 T/A	Non descritta Non descritta
594066	622-1274 C/T 622-1100 T/C 622-1216 T/C	Non descritta
594072	711+385 T/A	
594078	1001+732insT 1001+623 A/G	All'interno della CST, non descritta
594082	1001+1481 A/G	Non descritta
594104	1716+7233 T/G 1717-13704 T/C	All'interno della CST
594112	1717-13668 T/C 1717-13667 G/A 1717-13420 G/C	All'interno della CST, non descritta
594140	3499+467 C/T	Non descritta
594144	3500-140 A/C 3600+1795 C/T	Non descritta
594148	3600+1945 G/A 3600+1991 T/C	All'interno della CST, non descritta
594150	3600+5179 C/T	All'interno della CST
594152	3601-65 C/A	All'interno della CST
594202	3601-5913 G/A	All'interno della CST, non descritta

**Tabella 2:** le 41 variazioni di sequenza identificate nelle CSTs sono state confrontate con la lista di polimorfismi noti in letteratura, 17 di esse risultano non descritte da nessun gruppo, 15 ricadono nella zona di alta omologia tra uomo e topo (>70%), le rimanenti ricadono nelle regioni circostanti (omologia uomo-topo compresa tra il 50% e il 70%).

Le restanti varianti sembrano viaggiare in associazione a specifiche mutazioni (non senso, missenso e delezioni) presenti nel gruppo di pazienti affetti da CF, con genotipo noto. La segregazione notata per alcune varianti con le principali mutazioni causative di CF, verificata attraverso ulteriori studi computazionali, potrebbe contribuire alla generazione di un pannello di marker genetici ereditati in diverse

combinazioni (aplotipi) insieme alla mutazione causativa di malattia; in maniera analoga potremmo anche costituire un aplotipo per quei pazienti i cui alleli risultano ancora non caratterizzati. In questi casi lo studio dell'aplotipo potrebbe, infatti, sostituire la ricerca diretta di alterazioni a carico del gene CFTR, e quindi rendere possibile la realizzazione di diagnosi di portatori e diagnosi prenatali complicate dalla mancata identificazione della mutazione causativa di malattia.

#### *Identificazione di tre nuovi microsatelliti del gene CFTR*

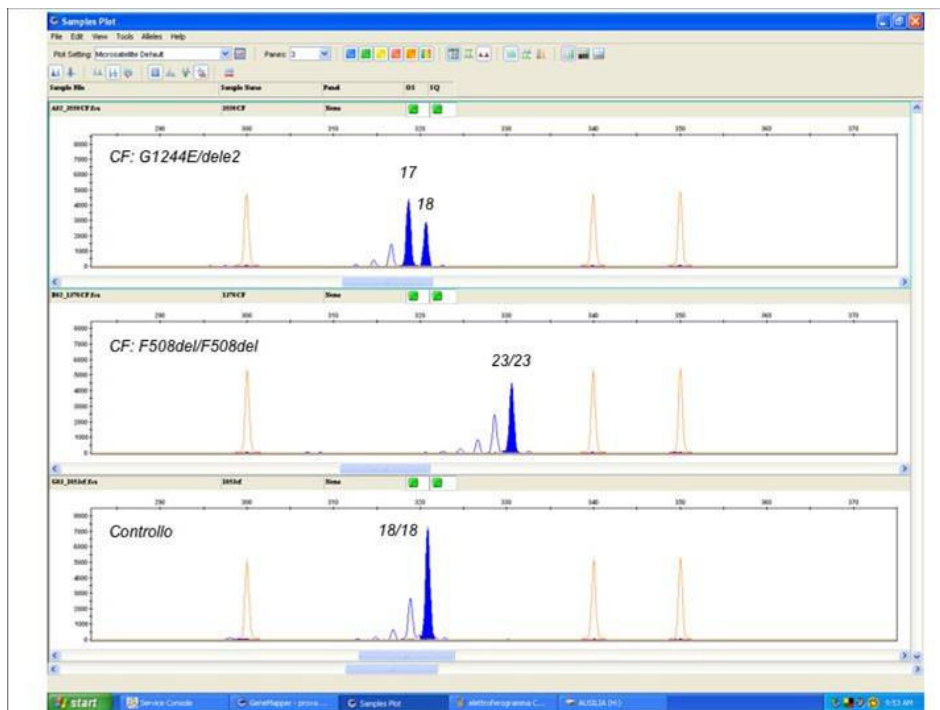
Per le 3 regioni altamente polimorfiche identificate con lo studio delle CSTs (1 STR e 2 polyA), invece, è stato messo a punto un protocollo sperimentale che prevede l'utilizzo dell'elettroforesi capillare, poichè questa tecnica è più sensibile delle convenzionali tecniche di separazione elettroforetiche e consente di separare frammenti di DNA che differiscono anche di un singolo nucleotide. La STR della CST 594102, una ripetizione di CA nell'introne 10 non è stata precedentemente descritta in letteratura ed è caratterizzata da un'estrema variabilità di lunghezza (fig. 20). Finora, per questo microsatellite, sono stati analizzati 118 soggetti (236 alleli), suddivisi in:

- 20 pazienti affetti da CF con genotipo F508del/F508del e fenotipo grave o lieve della malattia. In questi pazienti la STR sembra avere una lunghezza ricorrente di circa 23 ripetizioni di CA, tale lunghezza sembra essere associata all'allele F508del;
- 32 soggetti sani facenti parte della popolazione di controllo, nei quali è stata riscontrata una notevole variabilità allelica rappresentata da espansioni o delezioni del dinucleotide CA, per un totale di 8 varianti alleliche differenti finora identificate;
- 66 pazienti affetti da CF con mutazioni di vario tipo su entrambi gli alleli, nei quali è stata notata una stretta associazione tra il tipo di mutazione e la specifica lunghezza della STR.

Anche per i due polyA dell'introne 3 (CST 594056) e dell'introne 4 (CST 594062-64) sono stati osservati alleli differenti in lunghezza, caratterizzati da inserzioni o delezioni del nucleotide A. La strategia di analisi è stata la stessa impiegata per la STR: anche in questo caso, infatti, sono stati analizzati 118 soggetti (236 alleli) suddivisi in:

- 20 pazienti affetti da CF con genotipo F508del/F508del e fenotipo grave o lieve della malattia. In questi pazienti i polyA sembrano avere una lunghezza ricorrente di circa 18 A per il polyA dell'introne 3 e 19 A per il polyA dell'introne 4, questi alleli sembrano essere associati all'allele F508del;
- 32 soggetti sani facenti parte della popolazione di controllo, nei quali è stata riscontrata una certa variabilità allelica rappresentata da inserzioni o delezioni del nucleotide A, per un totale di 3 varianti alleliche differenti finora identificate;
- 66 pazienti affetti da CF con mutazioni di vario tipo su entrambi gli alleli, nei quali è stata riscontrata una certa variabilità allelica associabile al tipo di mutazione portata dai vari soggetti su entrambi gli alleli.

Una volta estesa la casistica ad una popolazione di controllo di numerosità elevata e calcolata l'eterozigosità dei vari alleli identificati tramite questo studio, la STR dell'introne 10 e i 2 polyA degli introni 3 e 4 potrebbero essere impiegati come marcatori nell'analisi di linkage per CF, dato l'esiguo numero di marcatori polimorfici intragenici conosciuti.



**Figura 20:** Il risultato della corsa di elettroforesi capillare consente di determinare il preciso numero di ripetizioni di CA per entrambi gli alleli del soggetto analizzato.



Per quanto riguarda le CSTs che presentano varianti geniche, siamo principalmente interessati a capire se mutazioni che ricadono in queste CSTs compromettano lo splicing del trascritto oppure la regolazione della trascrizione mediante attivazione o disattivazione di regioni regolatorie positive o negative.

Sono state individuate alcune possibili metodiche che potrebbero essere messe a punto per proseguire lo studio, tra cui, l'allestimento di studi in vivo per analizzare le variazioni di sequenza delle CSTs di nostro interesse, utilizzando ad esempio topi mutagenizzati per le mutazioni evidenziate dall'analisi di sequenza; oppure, mediante il campionamento di cellule dell'epitelio nasale dei soggetti con mutazioni nelle CSTs, con un comune tampone nasale, potrebbero essere allestite colture cellulari su cui, successivamente, si potrebbe valutare la funzionalità della proteina CFTR mediante saggi di elettrofisiologia, allo scopo di osservare la presenza di eventuali variazioni del flusso ionico della proteina a seconda che il soggetto in esame sia portatore in omozigosi o in eterozigoti di una mutazione intronica.

Un altro approccio potrebbe essere rappresentato dalla messa a punto di metodiche che consentono di analizzare direttamente le interazioni DNA-Proteina. A tal fine potremmo ricorrere all'utilizzo dell'Electrophoretic Mobility Gel Shift Assay (EMSA) oppure all'innovativa metodica del Fishing for Partner. Quest'ultima è una tecnica di recente adozione presso il nostro istituto e consiste in una cromatografia su colonna e una successiva analisi di spettrometria di massa. Il DNA di interesse viene amplificato e legato chimicamente ad una molecola detta "flag", che ha il ruolo di immobilizzare il DNA sulla colonna cromatografica. Successivamente all'interno della colonna viene fatto passare un estratto cellulare (citoplasmatico o nucleare), derivante da linee cellulari che esprimono il gene di interesse. Il passaggio di questo estratto in colonna determina il legame di possibili interattori al DNA. In seguito il complesso viene eluito dalla colonna e le molecole che hanno legato il DNA vengono analizzate tramite spettrometria di massa. Questa tecnica potrebbe rivelare come mutazioni delle CSTs possano favorire o meno il legame di specifici interattori molecolari (proteine, RNA, etc.).

Parallelamente allo sviluppo delle tecniche descritte, saremmo anche interessati ad avviare uno studio per valutare lo stato di metilazione del promotore del gene CFTR, poiché è noto che parte del meccanismo con cui il gene è regolato dipende

anche da fenomeni epigenetici. In letteratura ci sono pochi dati in merito all'argomento, ma dal lavoro del gruppo di Koh è stato osservato che il silenziamento del gene, mediante la metilazione dell'isola CpG al 5' del gene in cellule non esprimenti CFTR, potrebbe rappresentare un importante punto di regolazione per l'espressione tessuto specifica (23). Dunque, una disregolazione dei meccanismi epigenetici, anche dettata da un'alterazione dell'interazione tra DNA e fattori trascrizionali, potrebbe influire sull'espressione del gene in differenti tessuti ed essere responsabile, quindi, della compromissione di specifici organi nei pazienti CF.

## CONCLUSIONI

Da questo lavoro di tesi, il primo studio incentrato sull'analisi di regioni introniche conservate del gene CFTR, sono emersi tre principali punti di riflessione. Il primo punto riguarda l'identificazione di varianti geniche nelle CSTs, potenzialmente coinvolte nella patogenesi della malattia, poiché presenti esclusivamente nel gruppo di soggetti affetti da Fibrosi Cistica e con genotipo non ancora caratterizzato. La conferma che queste varianti siano effettivamente responsabili dello sviluppo della malattia potrebbe portare ad un miglioramento della diagnosi molecolare della malattia, con un aumento del "detection rate" e la possibilità di effettuare un maggior numero di diagnosi di portatori e prenatali, complicate dalla mancata identificazione del difetto molecolare del gene CFTR. Il secondo punto riguarda, invece, l'identificazione di regioni introniche conservate del gene CFTR, in cui non è stata osservata la presenza di alcuna variante in tutti i gruppi di soggetti inclusi nello studio. La forte conservazione di queste regioni, probabilmente dovuta ad una alta pressione selettiva nel corso dell'evoluzione, suggerisce che esse ospitino elementi regolatori o strutturali non ancora caratterizzati, il cui studio potrebbe portare ad una delucidazione dei meccanismi regolatori e della funzionalità del gene CFTR. Il terzo punto, infine, riguarda l'identificazione di tre nuovi microsatelliti altamente polimorfici all'interno del gene CFTR. Dati preliminari hanno portato all'osservazione di un'elevata eterozigosità dei tre microsatelliti, per cui è auspicabile che essi possano essere in breve tempo utilizzati come marcatori nelle analisi di linkage per la diagnosi molecolare della Fibrosi Cistica.

In conclusione, tutti i risultati ottenuti grazie a questo studio potrebbero essere utili per chiarire i meccanismi ancora sconosciuti che portano alla produzione della proteina CFTR e al suo corretto funzionamento e, allo stesso tempo, potrebbero fornire informazioni utili per il miglioramento delle tecniche di diagnostica molecolare e della prognosi e terapia della Fibrosi Cistica.

## BIBLIOGRAFIA

- 1) McIntosh I, Cutting GR. Cystic Fibrosis transmembrane conductance regulator and etiology and pathogenesis of cystic fibrosis. *FASEB J.* 1992; 6: 2775-2782.
- 2) Huang P, Gilmore E, Kultgen P, Barnes P, Milgram S, Stutts MJ. Local regulation of cystic fibrosis transmembrane regulator and epithelial sodium channel in airway epithelium. *Proc Am Thorac Soc.* 2004;1(1):33-7.
- 3) Pilewski JM, Frizzell RA. Role of CFTR in airway disease. *Physiol Rev.* 1999;79:S215-55.
- 4) Morales MM, Capella MA, Lopes AG. Structure and function of the cystic fibrosis transmembrane conductance regulator. *Braz J Med Biol Res.* 1999 Aug;32(8):1021-8.
- 5) Hudson VM. New insights into the pathogenesis of cystic fibrosis: pivotal role of glutathione system dysfunction and implications for therapy. *Treat Respir Med.* 2004;3(6):353-63.
- 6) Friedman KJ, Silverman LM. Cystic fibrosis syndrome: a new paradigm for inherited disorders and implications for molecular diagnostics. *Clin Chem* 1999; 45: 929-931.
- 7) Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 1989; 245: 1066-73.
- 8) Gregory RJ, Cheng SH, Rich DP, Marshall J, Paul S, Hehir K, Ostedgaard L, Klinger KW, Welsh MJ, Smith AE. Expression and characterization of the cystic fibrosis transmembrane conductance regulator. *Nature* 1990; 347: 382-386.
- 9) 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, Padoan R, Pianaroli A, Raia V, De Ritis G, Battistini A, Grzincich G, Japichino L, Pardo F, Antonelli M, Quattrucci S, Lucidi V, Castro M, Santini B, Castello M, Guanti G, Leoni GB, Cao A, Toffoli C, Lucci E, Vullo C, Torricelli F, Sbernini F, Romeo G, Ronchetto P, Seia M, Rossi A, Ferrari M, Cremonesi L, Salvatore F, Castaldo G, D'Alcamo E, Maggio A, Sangiuolo F, Dallapiccola B, Maceratesi P, Bisceglia L, Gasparini P, Carbonara A, Bonizzato A, Cabrini G, Bombieri C, Pignatti PF, Borgo G, Castellani I C, Villani A, Arduino

- C, Salvatore D, Mastella G, Piazza A. Genetic history of cystic fibrosis mutations in Italy. I. Regional distribution. *Ann Hum Genet* 1997; 61: 411-424.
- 10) Bienvenue T. Les bases moléculaires de l'hétérogénéité phénotypique dans la mucoviscidose. *Am Biol Clin* 1997; 55: 113-21.
- 11) Tomaiuolo R, Spina M, Castaldo G. Molecular diagnosis of cystic fibrosis: comparison of four analytical procedures. *Clin Chem Lab Med* 41 (1):26-32
- 12) Hantash FM, Redman JB, Starn K, Anderson B, Buller A, McGinniss MJ, Quan F, Peng M, Sun W, Strom CM. Novel and recurrent rearrangements in the CFTR gene: clinical and laboratory implications for cystic fibrosis screening. *Hum Genet*. 2006 Mar;119(1-2):126-36. Epub 2005 Dec 17. Erratum in: *Hum Genet*. 2006 Apr;119(3):352.
- 13) Castaldo G, Polizzi A, Tomaiuolo R, Cazeneuve C, Girodon E, Santostasi T, Salvatore D, Raia V, Rigillo N, Goossens M, Salvatore F. Comprehensive cystic fibrosis mutation epidemiology and haplotype characterization in a southern Italian population. *Ann Hum Genet* 2005; 69: 15-24.
- 14) Audrézet MP, Chen JM, Raguénès O, Chuzhanova N, Giteau K, Le Maréchal C, Quéré I, Cooper DN, Férec C. Genomic rearrangements in the CFTR gene: extensive allelic heterogeneity and diverse mutational mechanisms. *Hum Mutat* 2004; 23: 343-57.
- 15) Salvatore F, Scudiero O, Castaldo G. Genotype-Phenotype Correlation in Cystic Fibrosis: the role of modifier genes. *Am J Med Genet* 2002, 111:88-95
- 16) Yoshimura K, Nakamura H, Trapnell BC, Dalemans W, Pavirani A, Lecocq JP, Crystal RG. The cystic fibrosis gene has a "housekeeping"-type promoter and is expressed at low levels in cells of epithelial origins. *J Biol Chem* 1991; 266: 9140-9144.
- 17) McCarthy VA, Harris A. The CFTR gene and regulation of its expression. *Pediatr Pulmonol* 2005; 40: 1-8.
- 18) Trezise AE, Chambers JA, Wardle CJ, Gould S, Harris A. Expression of the cystic fibrosis gene in human foetal tissues. *Hum Mol Genet* 1993; 2: 213-218.
- 19) Engelhardt JF, Zepeda M, Cohn JA, Yankaskas JR, Wilson JM. Expression of the cystic fibrosis gene in adult human lung. *J Clin Invest* 1994; 93: 737-749.

- 20) McDonald RA, Matthews RP, Idzerda RL, McKnight GS. Basal expression of the cystic fibrosis transmembrane conductance regulator gene is dependent on protein kinase A activity. *Proc Natl Acad Sci USA* 1995; 92: 7560-7564.
- 21) Matthews RP, McKnight GS. Characterization of the cAMP response element of the cystic fibrosis transmembrane conductance regulator gene promoter. *J Biol Chem* 1996 271:31869-31877.
- 22) Cafferata EG, Guerrico AM, Pivetta OH, Santa-Coloma TA, Brouillard F, Bouthier M, Leclerc T, Clement A, Baudouin-Legros M, Edelman A. NF-kappaB activation is involved in regulation of cystic fibrosis transmembrane conductance regulator (CFTR) by interleukin-1beta. *J Biol Chem* 2001; 276: 15441-15444.
- 23) Koh J, Sferra TJ, Collins FS. Characterization of the cystic fibrosis transmembrane conductance regulator promoter region. Chromatin context and tissue-specificity. *J Biol Chem* 1993; 268: 15912-15921.
- 24) Pittman N, Shue G, Leleiko NS, Walsh MJ. Transcription of cystic fibrosis transmembrane conductance regulator requires a CCAAT-like element for both basal and cAMP-mediated regulation. *J Biol Chem* 1995; 270: 28848-28857.
- 25) Vuillaumier S, Dixmeras I, Messai H, Lapoumeroulie C, Lallemand D, Gekas J, Chehab FF, Perret C, Elion J, Denamur E. Cross-species characterization of the promoter region of the cystic fibrosis transmembrane conductance regulator gene reveals multiple levels of regulation. *Biochem J* 1997; 327: 651-662.
- 26) Mouchel N, Tebbutt SJ, Broackes-Carter FC, Sahota V, Summerfield T, Gregory DJ, Harris A. The sheep genome contributes to localization of control elements in a human gene with complex regulatory mechanisms. *Genomics* 2001; 76: 9-13.
- 27) Ellsworth RE, Jamison DC, Touchman JW, Chissoe SL, Braden Maduro VV, Bouffard GG, Dietrich NL, Beckstrom-Sternberg SM, Iyer LM, Weintraub LA, Cotton M, Courtney L, Edwards J, Maupin R, Ozersky P, Rohlfsing T, Wohldmann P, Miner T, Kemp K, Kramer J, Korf I, Pepin K, Antonacci-Fulton L, Fulton RS, Minx P, Hillier LW, Wilson RK, Waterston RH, Miller W, Green ED. Comparative genomic sequence analysis of the human and mouse cystic fibrosis transmembrane conductance regulator genes. *Proc Natl Acad Sci U S A*. 2000 Feb 1;97(3):1172-7

- 28) Trezise AE, Linder CC, Grieger D, Thompson EW, Meunier H, Griswold MD, Buchwald M. CFTR expression is regulated during both the cycle of the seminiferous epithelium and the oestrous cycle of rodents. *Nat Genet* 1993; 3: 157-164.
- 29) Rochwerger L, Buchwald M, Dho S, Parker L, Foskett JK, Ellsworth RE, Jamison DC, Touchman JW, Chisoe SL, Braden Maduro VV. Stimulation of the cystic fibrosis transmembrane regulator expression by estrogen in vivo. *Endocrinology* 1993; 133: 921-930.
- 30) Zhang L, Wang D, Fischer H, Fan PD, Widdicombe JH, Kan YW, Dong JY. Efficient expression of CFTR function with adeno-associated virus vectors that carry shortened CFTR genes. *Proc Natl Acad Sci* .1998; 18;95(17):10158-63.
- 31) Mouchel N, Broackes-Carter F, Harris A. Alternative 5' exons of the CFTR gene show developmental regulation. *Hum Mol Genet* 2003 12:759-769.
- 32) Broackes-Carter FC, Williams SH, Wong PL, Mouchel N, Harris A. Alternative splicing of the ovine CFTR gene. *Mamm Genome* 2003 14:778-787.
- 33) White NL, Higgins CF, Trezise AE. Tissue-specific in vivo transcription start sites of the human and murine cystic fibrosis genes. *Hum Mol Genet* 1998; 7: 363-369.
- 34) Borgo G, Mastella G, Gasparini P, Zorzanello A, Doro R, Pignatti PF. Pancreatic function and gene deletion F508del in cystic fibrosis. *J. Med Genet* 1990; 27:665-669.
- 35) Williams SH, Mouchel N, Harris A. A comparative genomic analysis of the cow, pig, and human CFTR genes identifies potential intronic regulatory elements. *Genomics* 2003; 81: 628-639.
- 36) Boccia A, Petrillo M, Di Bernardo D, Guffanti A, Mignone F, Confalonieri S, Luzi L, Pesole G, Palella G, Ballabio A, Banfi S. DG-CST (Disease Gene Conserved Sequence Tag), a database of human-mouse conserved elements associated to disease genes. *Nucleic Acids Research* 2005; 33: 505-510.
- 37) Alibakhshi R, Kianishirazi R, Cassiman JJ, Zamani M, Cuppens H. Analysis of the CFTR gene in Iranian cystic fibrosis patients: Identification of eight novel mutations. *J Cyst Fibros.* 2007; 25.
- 38) Wilson K, Walker J. *Metodologia biochinica. Le bioscienze e le biotecnologie in laboratorio.* 2005. Edizioni Cortina.

- 39) Dequeker E, Cuppens H, Dodge J, Estivill X, Goossens M, Pignatti PF, Scheffer H, Schwartz M, Tummeler B, Cassiman JJ. Recommendations for quality improvement in genetic testing for cystic fibrosis. European Concerted Action on Cystic Fibrosis. *Eur J Hum Genet.* 2000; 8: 2-24.
- 40) Smith DJ, Nuthall HN, Majetti ME, Harris A. Multiple potential intragenic regulatory elements in the CFTR gene. *Genomics* 2000; 64: 90-96.
- 41) Phylactides M, Rowntree R, Nuthall H, Ussery D, Wheeler A, Harris A. Evaluation of potential regulatory elements identified as DNase I hypersensitive sites in the CFTR gene. *Eur J Biochem* 2002; 269: 553-559.
- 42) Williams SH, Mouchel N, Harris A. A comparative genomic analysis of the cow, pig, and human CFTR genes identifies potential intronic regulatory elements. *Genomics* 2003; 81: 628-639.
- 43) Nuthall HN, Moulin DS, Huxley C, Harris A. Analysis of DNase-I-hypersensitive sites at the 3' end of the cystic fibrosis transmembrane conductance regulator gene (CFTR). *Biochem J* 1999; 341: 601-611.
- 44) Mogayzel PJ Jr, Ashlock MA. CFTR intron 1 increases luciferase expression driven by CFTR 5-flanking DNA in a yeast artificial chromosome. *Genomics* 2000; 64: 211-215.
- 45) Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, Schacherer F. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* 2000; 28: 316-319.
- 46) Ryffel GU. Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families: functional and pathological consequences. *J Mol Endocrinol* 2001; 27: 11-29.
- 47) Hu C, Perlmutter DH. Regulation of alpha1-antitrypsin gene expression in human intestinal epithelial cell line caco-2 by HNF-1alpha and HNF-4. *Am J Physiol* 1999; 276: 1181-1194.
- 48) Mouchel N, Henstra SA, McCarthy VA, Williams SH, Phylactides M, Harris A. HNF1alpha is involved in tissue-specific regulation of CFTR gene expression. *Biochem J* 2004; 378: 909-918.
- 49) Lee YH, Sauer B, Gonzalez FJ. Laron dwarfism and non-insulin-dependent diabetes mellitus in the Hnf-1alpha knockout mouse. *Mol Cell Biol* 1998; 18: 3059-3068.



- 50) Parrizas M, Maestro MA, Boj SF, Paniagua A, Casamitjana R, Gomis R, Rivera F, Ferrer J. Hepatic nuclear factor 1-alpha directs nucleosomal hyperacetylation to its tissue-specific transcriptional targets. *Mol Cell Biol* 2001; 21: 3234-3243.
- 51) Boj SF, Parrizas M, Maestro MA, Ferrer J. A transcription factor regulatory circuit in differentiated pancreatic cells. *Proc Natl Acad Sci USA* 2001; 98: 14481-14486.
- 52) Renè C, Magali T, Iral F, Doudament J, L'Honorè A, Gerbon C, Demaille J, Claustres M, Romey MC. Binding of serum response factor to cystic fibrosis transmembrane conductance regulator CArG-like elements, as a new potential CFTR transcriptional regulation pathway. *Nucleic Acid Res*, 2005; 33: 5271-5290.
- 53) Romey MC, Pallares-Ruiz N, Mange A, Mettling C, Peytavi R, Demaille J, Claustres M. A naturally occurring sequence variation that creates a YY1 element is associated with increased cystic fibrosis transmembrane conductance regulator gene expression. *J Biol Chem*. 2000 Feb 4;275(5):3561-7.