UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

# Characterization of mRNA Dysfunctional Mechanisms Associated with the Genetic Disease Cystic Fibrosis

**Doutoramento em Bioquímica**

Especialidade de Genética Molecular

Verónica Manuela Rôxo Felício

Tese orientada por:

Professora Doutora Margarida Amaral

Documento especialmente elaborado para a obtenção do grau de doutor

2018

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

# Characterization of mRNA Dysfunctional Mechanisms Associated with the Genetic Disease Cystic Fibrosis

**Doutoramento em Bioquímica**

Especialidade de Genética Molecular

Verónica Manuela Rôxo Felício

Tese orientada por:

Professora Doutora Margarida Amaral

Júri:

Presidente:

- Doutora Amélia Pilar Grases dos Santos Silva Rauter, Professora Catedrática e Presidente do Departamento de Química e Bioquímica da Faculdade de Ciências da Universidade de Lisboa (FCUL)

Vogais:

- Doutora Luísa Maria Ferreira Romão Loison, Investigadora Principal do Instituto Nacional de Saúde Doutor Ricardo Jorge;
- Doutora Luísa Maria Quental Mota Vieira, Investigadora Principal da Unidade de Genética e Patologia Molecular do Hospital do Divino Espírito Santo de Ponta Delgada
- Doutra Margarida Sofia Pereira Duarte Amaral, Professora Catedrática da Faculdade de Ciências da Universidade de Lisboa;
- Doutora Rita Maria Pulido Garcia Zilhão Aranha Moreira, Professora Auxiliar da Faculdade de Ciências da Universidade de Lisboa;
- Doutora Margarida Henriques da Gama Carvalho, Professora Auxiliar da Faculdade de Ciências da Universidade de Lisboa;

Documento especialmente elaborado para a obtenção do grau de doutor

Fundação para a Ciência e Tecnologia do Ministério da Ciência, Tecnologia e Ensino Superior, SFRH/BD/87478/2012

2018

*Aos meus filhos*

*Francisco e Xavier*

*E ao meu marido*

*Gonçalo*

*OS três pilares da minha vida*

# *Acknowledgments/Agradecimentos*

Agradeço a todos os que contribuíram, de forma direta ou indireta, para a realização deste trabalho. Salientando o meu agradecimento:

Em especial à minha orientadora Professora Doutora Margarida Amaral, por todo o apoio e suporte, pela transmissão de conhecimentos, por me ter proporcionado a excelente oportunidade de trabalhar no seu laboratório e por toda a confiança em mim depositada ao longo destes anos. Agradeço toda a paciência, motivação, desafios propostos e a oportunidade de trabalhar com alguém com tanta dedicação, entusiasmo, determinação e rigor.

À Doutora Anabela Ramalho, uma grande impulsionadora do meu doutoramento e que me deu todas as bases para o mesmo. Os seus ensinamentos de trabalho de bancada foram uma mais valia na realização do trabalho laboratorial, tudo o que aprendi sobre genotipagem, PCR, mini-genes, mutações é a si que devo. Obrigada pela sua disponibilidade para as minhas dúvidas e pela sua ajuda na resolução de problemas.

Ao ministério da Ciência e Ensino Superior e à Fundação para a Ciência e a Tecnologia, por terem possibilitado a realização deste trabalho, através da concessão da bolsa de doutoramento de que fui recipiente. Ao Departamento de Química e Bioquímica da Faculdade de Ciências da Universidade de Lisboa e ao BioISI - Biosystems & Integrative Sciences Institute por me acolherem durante o meu trabalho de doutoramento.

Quero agradecer à Dra. Celeste Barreto, à Dra. Luísa Pereira, ao Dr. Carlos Lopes, e ao Dr. José Cavaco pela colaboração preciosa com pacientes de FQ.

Comecei este doutoramento rodeada de pessoas maravilhosas que foram seguindo as suas vidas ao longo do tempo, mas a quem devo muito. Marta Palma, Marisa Sousa, Inna Uliyakina e Simão Luz não estiveram presentes fisicamente até ao fim, mas mesmo longe ajudaram. Muito obrigada pelo constante apoio, amizade e disponibilidade para ajudar em qualquer questão. Os vossos conselhos e sugestões foram absolutamente fulcrais para a minha evolução enquanto pessoa e cientista.

Quero agradecer todos os meus colegas de laboratório: À Susana Igreja pela ajuda e apoio nas experiências, projetos, relatórios e na minha tese de doutoramento; ao Hugo Botelho muito obrigada por todos os ensinamentos de microscopia, análise de resultados, pelas folhas de excel super cromas para os ensaios de efluxo e pela maravilhosa capacidade que tens para tornares tudo mais fácil de perceber. Mas principalmente, muito obrigada pelas pausas,

pelos momentos de desanuviar a cabeça e organizar ideias; ao Luka Clarke pela ajuda, partilha de conhecimentos e companhia no trabalho com pacientes. ao Professor Carlos Farinha pelos ensinamentos; ao Luís Marques pela paciência no microscópio e pelas pausas para café; à Sofia Correia, à Madalena, à Margarida, à Sofia Ramalho, à Iris Silva, à Ines Pankonien, ao Luís Sousa obrigada pelas conversas de laboratório que ajudaram na resolução de alguns problemas e também proporcionaram um ambiente divertido.

Em último, mas muito importante, pessoas que considero mais que colegas de laboratório, pessoas que considero amigas. Sara Canato, obrigada por tudo, obrigada pelo apoio, pela ajuda, pela cumplicidade, foste mais que uma colega de trabalho. Nikhil Awatade, thank you for the teachings, for the opportunity to work closer to the patients, for the conversations, for the long hours of work always with good disposition. Remain the memories of days and nights making lungs, that October will never be forgotten. João Santos, Joana Lérias, Arsénia Massinga, Iris Silva e José Múrias vocês foram suportes importantes no laboratório, proporcionaram momentos maravilhosos no trabalho, mas principalmente estiveram presentes na minha vida. Todos vocês deixaram marca em mim como cientista pelas conversas e sugestões, mas marcaram-me muito mais como pessoa, obrigada por isso.

Não posso deixar de agradecer às três da vida airada, Ana Luísa e Sara Fernandes, pelas conversas, pelas fugidas ao café, pelo apoio, carinho e amizade que sempre transmitiram. Pelas loucuras e pelos momentos sérios, um muito obrigado, foram um grande pilar nesta fase da minha vida, em muitas alturas pensei que não aguentava, mas vocês mostraram que tudo se faz e que em grupo é muito mais fácil.

Á minha família devo muito mais que um simples obrigado. Mãe e pai vocês sempre foram a minha inspiração, a minha força ajudaram-me a remar em direção aos meus objetivos. A vocês devo ter chegado ao fim desta etapa. Ao meu irmão agradeço as vezes que me deu na cabeça e me deu força para chegar ao fim. Às minhas avós, aos meus tios, aos meus padrinhos, aos meus priminhos e afilhada, à minha cunhada e sobrinha agradeço os momentos de distração e alegrias que me ajudaram a superar momentos mais difíceis deste doutoramento. Aos meus sogros agradeço a motivação e a insistência e ajuda para terminar esta fase da minha vida. Foram uma grande ajuda neste processo final, sem vocês não teria conseguido.

Aos meus filhos, Francisco e Xavier e ao meu marido Gonçalo, os pilares da minha vida, que me fazem levantar todos os dias com um sorriso e me mostram que eu sou capaz, aos meus 3 homens a quem eu devo tudo…OBRIGADA por compreenderem as minhas ausências, a minha falta de tempo e por vezes de paciência. Gonçalo sem o teu encorajamento e compreensão, teria sido impossível para mim terminar o meu doutoramento.

Verónica Manuela Rôxo Felício foi bolseira de doutoramento da Fundação para a Ciência e Tecnologia do Ministério da Ciência, Tecnologia e Ensino Superior.

SFRH/BD/87478/2012

# *SUMMARY*

Cystic Fibrosis (CF) is the most common autosomal disease in Caucasians, with an estimated incidence of 1:6000 births in Portugal. The most relevant clinical aspect of its classic manifestation is chronic lung disease, which is the main cause of morbidity and mortality. Other symptoms include, pancreatic dysfunction, male infertility and high concentrations of chloride ($Cl^-$) in sweat. However, even though the classic form of the disease is well defined, its pathophysiology is not completely understood with this pleiotropic disease having highly variable manifestations of clinical phenotypes. Novel therapies aim to correct the basic defect, specifically focusing on the rescue of Cystic fibrosis transmembrane conductance regulator (CFTR) function in CF airways. Most of these CFTR modulator strategies target the F508del, the most common mutation.

Nevertheless, widespread evidence has demonstrated that a significant number of CF-causing mutations affect splicing efficiency and the stability of mRNA molecules. Here, we propose to elucidate the regulatory mechanisms underlying these CF-associated mutations. To this end, our aims are: 1) to identify *CFTR* gene mutations in individuals with non-CF chronic lung diseases, namely chronic obstructive pulmonary disease (COPD), asthma and disseminated bronchiectasis (DB); 2) to identify CFTR gene mutations in individuals with a suspicion of CF, followed by the analysis of CFTR expression in their native tissues to characterize the impact of CFTR splicing or premature termination codon (PTC) mutations in the structure and levels of mRNA; 3) to identify key factors in the nonsense/mediated decay (NMD) pathway by automated microscopy screens using a cell model expressing a novel CFTR NMD-PTC/read-through mini-gene reporter; and 4) to screen for novel compounds suppressing PTC mutations by automated microscopy screens using the previous cell-based model, as potential corrective therapies for CF. The expected results will provide knowledge on RNA-processing dysfunction and on the efficacy of novel RNA modulator compounds towards a "personalised medicine" approach.

Regarding the **first objective**, our data show that 7 (out of 136) patients with non-CF respiratory diseases presented CFTR mutations in one allele, in contrast with the control group, in which no mutations were detected. We analysed the association of CFTR gene mutations with each of the three respiratory diseases considered. For asthma our data did not show an increase in mutation frequency when compared to the control group. For DB, we found an increase in the frequency of *CFTR* gene mutations, albeit with no statistical significance, which is in agreement with previous reports. For COPD however, we found a statistically significant increase in *CFTR* gene mutation incidence, relative to the control group.

Our data reinforce the importance of characterizing *CFTR* gene mutations on non-CF respiratory diseases in Portuguese patients, to gain a better understanding of the epidemiology and etiology of these diseases. The results also lead to the identification of groups of patients who may benefit from the new therapeutic compounds currently under development to correct the basic CFTR protein defect in CF.

Concerning the **second objective**, we have developed a novel RNA-based approach to detect unknown CFTR mutations [Felício *et al* (2017) *Clin Genet* **91**: 476-481]. We are currently using this protocol for patients with a suspicion of CF and none or just one CFTR mutation identified. With this method we identified one mutation (711+3A>T) which had been previously reported but had not been characterized. We can conclude that this is a rapid, robust and inexpensive method to detect rare mutations, and therefore a method that can be easily used after a first screen.

Regarding the **third objective**, we used this CFTR-NMD reporter to identify novel NMD factors by screening a previously validated shRNA library – a subset of The RNAi Consortium (TRC) – which is enriched in shRNAs targeting genes with a known or predicted involvement in transcript processing (425 genes), using HT microscopy. We selected the 24 top hits for the confirmation: 11 genes with NMD score ≥ 2 and more than 2 shRNAs with the same phenotype; 2 genes from the screen that showed read-through activity; and 11 other genes resulting from a high-throughput screen (HTS) aimed at the identification of CFTR splicing regulators (unpublished data). We chose these genes related to splicing because this is a process known to be required for NMD to occur and thus the knock-down of such genes can lead to NMD inhibition. The confirmation screen was performed using a library of siRNAs targeting the previously selected genes, however the results were inconclusive due to the low NMD score obtained. We have identified 4 genes with higher values but NMD score ≤ 1, three (eIF4A3, SREK1 and RPS19) are related to splicing, elF4A3 is directly and RPS19 is indirectly related to NMD, SREK1 is related only to splicing, with the fourth gene, ADIRF, having unknown functional properties. Some of the hits identified within this screen may be potential drug targets by their effects in inhibiting NMD, when knocked-down. With the results obtained in the confirmation screen we decided to follow the 24 genes identified in the primary screen for validation studies using two different techniques: WB and qRT-PCR (study in progress).

Lastly, the **fourth objective** was to restore functional protein production to PTC mutations using read-through compounds, with the ultimate aim of CF patient treatment. The CFTR-NMD construct used has the G542X nonsense mutation, mCherry at the N terminus and eGFP fused at the CFTR C-terminus. Through the screen of the library, we identified new small-

molecule compounds that induced PTC read-through. To confirm the read-through efficiency future experiments are needed using additional techniques, such as, WB and transcript analysis by semi-quantitative PCR and qRT-PCR. Finally, to validate the top hit compounds, it is necessary to test them in patient's materials, including nasal primary cells for functional activity and intestinal organoids to determine a dose response and to test these compounds in combination with potentiators and correctors.

The studies presented in this dissertation had the overall aims of advancing the current knowledge on RNA-processing dysfunction and of identifying novel RNA modulator compounds towards a "personalised medicine" approach. The results obtained have indeed provided new insights into: 1) the relationship of other respiratory diseases with the presence of CFTR mutations; 2) new approaches to detect CFTR mutations in DNA and RNA; and 3) our understanding of key factors in NMD and read-through activity in relation to CFTR nonsense mutations.

Keywords: CF diagnosis; genotyping; intronic mutations; rare mutations; splicing; nonsense mutations; premature termination codons (PTC); Nonsense mediated decay (NMD): Read-through activity

# *RESUMO*

A Fibrose Quística (FQ) é a doença autossómica recessiva letal mais comum na população Caucasiana, afetando cerca de 1 em 2500-6000 nados vivos, dependendo da região geográfica do globo, e com uma frequência de portadores de 1 em cada 25-40 indivíduos. Esta doença é causada por mutações no gene CFTR (do inglês Cystic Fibrosis Transmembrane Conductance Regulator) localizado no cromossoma 7. A proteína CFTR é expressa na membrana apical das células epiteliais onde funciona como um canal de cloreto ($Cl^-$), regulando o transporte iónico e de água. Clinicamente, a FQ é caracterizada por múltiplas manifestações em diferentes órgãos, sendo, contudo, a doença a nível pulmonar a principal causa de morbilidade e mortalidade. Outros sintomas clássicos da FQ incluem insuficiência pancreática exócrina, obstrução intestinal, infertilidade masculina e concentração elevada de eletrólitos no suor (característica utilizada no principal teste de diagnóstico).

Foram já descritas mais de 2000 mutações no gene *CFTR*, supondo-se ser grande parte delas causadoras de FQ. Assume-se que cerca de 13% das variantes não sejam patogénicas, mas para a grande maioria (~85%) ainda não estão definidas as respetivas consequências funcionais e clínicas. Existem estudos que sugerem estarem as mutações no gene *CFTR* também associadas a outras doenças respiratórias, como Bronquiectasias difusas (BD), Doença Pulmonar Obstrutiva Crónica (DPOC) ou Asma Brônquica. A elucidação dos efeitos moleculares e funcionais das mutações é uma importante fonte de informação para prognóstico clínico, mas também sobre a estrutura e função da proteína *CFTR*, nomeadamente para desenvolver compostos terapêuticos para correção do defeito específico de cada mutação. Assim, a classificação de mutações CFTR de acordo com seu defeito funcional permite conceber uma estratégia terapêutica comum para mutações na mesma classe, tendo sido propostas sete classes de mutações, que mais recentemente evoluíram para "teratipos".

Porém, muitos desafios permanecem na conceção de terapias para pacientes com FQ que apresentam mutações raras e que não podem ser tratadas com as terapêuticas aprovadas. Uma fração significativa (~ 24%) gera um codão de terminação prematura ou PTC (do inglês, "*Premature Termination Codon*"), incluindo: 8,4% nonsense e 15,7% mutações *frameshift*. As mutações PTC levam em geral a uma redução significativa, ou ausência total, de expressão de mRNA e proteína normais, e estão normalmente associadas a um fenótipo clínico grave. A redução nos níveis de transcritos ocorre através do mecanismo de degradação do mRNA denominado NMD (do inglês "*Nonsense-Mediated Decay*") que degrada seletivamente os transcritos portadores de PTCs. Até ao momento, não há terapias para tratar especificamente

pacientes portadores de mutações PTC. O NMD é um mecanismo essencial de regulação genética conservado em eucariotas, desempenhando um papel importante na modulação dos fenótipos clínicos, bem como na resposta a terapias. Espera-se que os pacientes que expressam níveis mais altos de mRNA contendo PTC devido a NMD menos eficiente, respondam de forma mais robusta às terapias de supressão / *read-through*.

Assim, o NMD representa assim um potencial alvo terapêutico para doenças causadas por PTCs, sendo de grande relevância identificar drogas alvo que inibam/ reduzam o NMD, levando ao amento dos níveis de transcritos contendo PTC. Esta redução do NMD acoplada à administração de agentes de *read-through* podem levar à produção de proteína CFTR completa ("*full-length*") funcional.

O objetivo geral deste trabalho centrou-se na compreensão do efeito das mutações CFTR a nível fenotípico e posteriormente na caracterização do efeito das mutações que afetam os transcritos quer ao nível *splicing* quer da estabilidade (NMD). Assim, os objetivos específicos do presente trabalho eram: 1) identificar mutações no gene *CFTR* em indivíduos com doenças respiratórias não-FQ; 2) identificar mutações *CFTR* em indivíduos com suspeita de FQ (e apenas com uma ou nenhuma mutação identificada) e analisar tecidos nativos de alguns desses pacientes para caracterizar o impacto de mutações que afetam a estrutura/níveis de mRNA; 3) identificar novos fatores de NMD usando um inovador modelo celular expressando um triplo repórter de CFTR com uma mutação PTC baseado na fluorescência e caracterizar os respetivos mecanismos de NMD; e por fim, 4) testar novos compostos quanto à sua atividade de *read-through* (supressão de mutações de PTC) naquele modelo celular.

O **primeiro objetivo** do presente trabalho pretendia determinar se a presença de mutações CFTR está associada a um risco elevado de asma, bronquiectasias disseminadas (BD) ou doença pulmonar obstrutiva crónica (DPOC). Foram analisados, para as mutações mais comuns em Portugal, 187 indivíduos portugueses, divididos em dois grupos distintos: o primeiro incluiu indivíduos com doenças respiratórias não-FQ (n=136): asma (n = 54), DPOC (n = 51) ou BD (n = 31); e o segundo incluiu indivíduos controlo saudáveis (n = 51). Os nossos dados mostraram que 7/136 pacientes com doenças respiratórias não-FQ apresentaram mutações *CFTR* em um alelo vs nenhum no grupo controlo, sendo que não identificámos diferenças significativas na frequência de mutações vs o grupo controlo para a asma, nem para DB (embora aqui tenha sido encontrada uma tendência para um aumento na frequência de mutações). Para DPOC, no entanto, encontramos um aumento estatisticamente significativo na frequência de mutações do gene *CFTR* vs o grupo controlo. Confirmámos assim a importância das mutações *CFTR* em DPOC (e uma tendência em BD) em pacientes portugueses, levando não só a uma melhor perceção da epidemiologia e etiologia destas

doenças, mas também à identificação de grupos de pacientes que podem vir a beneficiar das novas terapias moduladoras da proteína CFTR para a FQ.

No estudo acima referido, identificámos numa paciente com DPOC (fumadora) com 67 anos uma mutação FQ grave (G542X) e mais duas variantes no outro alelo (G576A-R668C), uma delas associada a fenótipo FQ variável por *splicing* alternativo (por *skipping* do exão 13). Assim, analisamos esta paciente para averiguar os níveis de CFTR funcional em biópsia retal por camara de Ussing e confirmar/excluir um possível diagnóstico de FQ. Estes testes demonstraram níveis de função normais, revelando assim que o alelo complexo de prognóstico desconhecido, não era neste caso causador de FQ, em concordância com o fenótipo clínico, levando à exclusão do diagnóstico de FQ. Mostrou-se assim que em casos selecionados, o estudo funcional da proteína CFTR pode ser uma ajuda preciosa para confirmar / excluir um diagnóstico de FQ.

Quanto ao **segundo objetivo**, desenvolvemos uma nova abordagem baseada em RNA para deteção de mutações desconhecidas no gene *CFTR* [Felício et al (2017) *Clin Genet* **91**: 476-481]. Atualmente este protocolo é usado para pacientes com suspeita de FQ e nenhuma ou apenas uma mutação CFTR identificada. Com este método, identificámos uma mutação (711 + 3A> T) já relatada, mas não caracterizada anteriormente. Podemos concluir que este é um método rápido, robusto e barato para detetar mutações raras que podem ser facilmente utilizadas após uma primeira deteção dum painel de mutações mais comuns.

O **terceiro objetivo** do presente trabalho, consistiu na identificação de novos fatores que modulam o NMD de transcritos CFTR com mutações PTC. Para tal, construímos um novo modelo celular expressando um mini-gene CFTR (isto é, o cDNA CFTR completo contendo ainda os intrões IVS14, 15 e 16) e com a mutação G542X (exão 12), que promove NMD. Dado conter mCherry em N-terminal, permite detetar uma redução no NMD bem como atividade de *read-through* através do eGFP em C-terminal. Este modelo foi então usado para rastrear uma biblioteca de 425 genes, envolvidos no processamento de transcritos. Foram selecionados 13 *top hits* para confirmação: 11 genes com de NMD scrore ≥ 2 e mais de 2 shRNA com o mesmo fenótipo. Para além destes 11 genes foram identificados 2 genes que mostraram atividade *read-through.* Aos 13 genes identificados no *screen* de NMD foram adicionados 11 genes que demonstram ter impacto no processo de *splicing* (S Igreja, MD Amaral, resultados não publicados) pelo que considerámos serem assim relevantes para o estudo de fatores que inibem NMD. Em resumo, o estudo de confirmação utilizando uma biblioteca de siRNA foi inconclusivo, visto que os NMD scores foram muito baixos, mas podemos mesmo assim identificar 4 genes com maior relevância nos resultados. Assim, tendo em conta os resultados do *screen* de confirmação, seguimos para um *screen* de validação

com os mesmos 24 genes selecionados, utilizando duas técnicas distintas: análise dos transcritos por qRT-PCR e da proteína resultante por *Western blotting*.

O **quarto e último objetivo** do presente trabalho, focou-se na identificação de compostos com atividade de *read-through* para o tratamento final de pacientes com FQ. Para este fim utilizámos o modelo em cima descrito, utilizando a fluorescência de eGFP em C-terminal para quantificar a proteína completa (*full-length*) produzida. Assim, por HTS analisámos 24 compostos análogos ou derivados do Ataluren/ PTC124, bem como outros não análogos, que foram sintetizados quimicamente por um grupo com quem colaboramos (Prof. Cristina Moiteiro, FCUL, Lisboa). Identificámos nesta biblioteca alguns compostos que levaram ao aparecimento de fluorescência verde, no entanto para serem considerados verdadeiros compostos de *read-through*, mais estudos têm de ser realizados para confirmar o efeito dos mesmos ao nível do RNA e da proteína.

Os resultados obtidos nesta dissertação contribuíram para uma maior compreensão da relação entre doenças respiratórias não-FQ e a presença de mutações CFTR, bem como para o desenvolvimento duma nova metodologia para detetar mutações a nível do mRNA. Contribuíram ainda para o avanço do conhecimento dos mecanismos de processamento de mRNA com mutações PTC, ao identificar novos fatores moduladores de NMD e compostos potencialmente moduladores deste processo para abordagens terapêuticas inovadoras de "medicina personalizada".

Palavras-chave: Diagnóstico FQ; genotipagem; mutações intrónicas; mutações raras; mutações nonsense; codão stop prematuro (PTC); Nonsense mediated decay (NMD); atividade "*read-through*".

De acordo com o disposto no artigo 31º do Regulamento de Estudos Pós-Graduados da Universidade de Lisboa, Despacho nº 2950/2025, publicado no Diário da República - 2ª Série - nº 57 – 23 de Março de 2015, foram utilizados nesta dissertação resultados incluídos nos seguintes artigos:

1. Felício V, Ramalho AS, Clarke LA, Awatade NT, Lopes C, Barreto C, Amaral MD. Incidence of mutations in the CFTR gene in Portuguese individuals with respiratory diseases. [Manuscript in preparation].

2. Gonska T, Avolio J, Ip W, Griffin K, Dupuis A, Felício V, Souza CG, Dalge S, Ribeiro MA, Ribeiro AF, Amaral MD, Ribeiro JD, Tullis E, Sousa M, Servidoni M. Comparing β–adrenergic sweat secretion to rectal biopsy and nasal potential difference measurements as adjunctive diagnostic test for Cystic fibrosis. Manuscript submitted.

3. Clarke LA, Awatade NT, Felício V, Calucho M, Pereira L, Azevedo P, Cavaco J, Barreto C, Bertuzzo C, Gartner S, Beekman J, Amaral MD. The effect of PTC mutations on CFTR mRNA abundance: A basis for read-through therapies in Cystic Fibrosis. Manuscript submitted.

4. Felício V, Ramalho AS, Igreja S, Amaral MD (2017) mRNA-based Detection of Rare CFTR Mutations Improves Genetic Diagnosis of Cystic Fibrosis in Populations with High Genetic Heterogeneity. *Clin Genet* **91**: 476-481. [PMID: 27174726]

No cumprimento do disposto da referida deliberação, a autora esclarece serem da sua responsabilidade, exceto quando referido em contrário, a execução das experiências que permitiram a elaboração dos resultados apresentados, assim como a interpretação e discussão dos mesmos. Os resultados obtidos por outros autores foram incluídos com autorização dos mesmos para facilitar a compreensão dos trabalhos e estão assinalados nas respetivas figuras e metodologias. Durante o presente doutoramento, foram ainda produzidos resultados incluídos em outros artigos publicados/ submetidos em revistas internacionais, nomeadamente:

1. Ramalho AS, Clarke LA, Sousa M, Felício V, Barreto C, Lopes C, Amaral MD (2016) Comparative *ex vivo*, *in vitro* and *in silico* analyses of a CFTR splicing mutation: Importance of functional studies to establish disease liability of mutations *J Cyst Fibros* **15**: 21-33.

2. Awatade NT, Ramalho S, Silva IAL, Felício V, Botelho HM, De Poel E, Vonk A, Beekman JM, Farinha CM, Amaral MD (2018) R560S: a class II CFTR mutation that is not rescued by current modulators. J Cyst Fibros. Epub 18 July. https://doi.org/10.1016/j.jcf.2018.07.001

3. Marcão A, Barreto C, Pereira L, Vaz L, Cavaco J, Casimiro A, Felix M, Silva T, Barbosa T, Freitas AC, Nunes S, Felício V, Lopes L, Amaral M, Vilarinho L (2018) Cystic Fibrosis newborn screening: PAP value in populations with stringent rules for genetic studies. Intl J Neonatal Screen. Issue No: Vol. 4, No. 3 (2018) DOI: 10.3390/ijns4030022

4. Uliyakina I, Da Paula AC, Afonso S, Lobo MJ, Felício V, Botelho HM, Farinha CM, Amaral MD (2018) Full Rescue of F508del-CFTR Processing and Function by CFTR Modulators can be Achieved by Removal of Two Unique Regulatory Regions. BioRxiv. Preprint 10.1101/320630 [www.biorxiv.org/content/early/2018/05/11/320630]

# *Contents*

# *ABBREVIATIONS*

| | |
|---|---|
| aa | Amino acid Ab Antibody |
| ABC | ATP-binding cassette |
| ASL | Airway surface liquid |
| ATP | Adenosine 5’-triphosphate |
| Band B | core-glycosylated or immature form of CFTR |
| Band C | fully-glycosylated or mature form of CFTR |
| CaCC | Calcium-activated chloride channel |
| cAMP | cyclic adenosine 5’-monophosphate |
| CBAVD | Congenital bilateral absence of vas deferens |
| CBC | CAP-binding complex |
| CBP | CAP-binding protein |
| CCH | Carbachol |
| cDNA | complementary DNA |
| CF | Cystic fibrosis |
| CFTR | Cystic fibrosis transmembrane conductance regulator |
| CFTR | Gene encoding CFTR protein |
| *CFTR-RD* | CFTR-Related disorders |
| CFSPID | CF screen positive with an inconclusive diagnosis |
| C-terminus | Carboxy terminus |
| CRMS | CFTR-related metabolic syndrome |
| DEAD | Asp-Glu-Ala-Asp motif |
| DECID | decay-inducing complex |
| DHX34 | DEAH box polypeptide 34 |
| DNA | Deoxyribonucleic Acid |
| DMSO | Dimethyl sulfoxide |
| EDTA | Ethylenediamine tetra acetic acid |
| EJC | exon junction complex |
| eIF | eukaryotic initiation factor |
| eIF4A3 | eukaryotic translation initiation factor 4A3 |
| EMEM | Eagle’s minimum essential medium |
| ENaC | Epithelium sodium channel |
| ER | Endoplasmic reticulum |
| eRF | eukaryotic release factors |
| ERQC | ER quality control |
| F508del | Deletion of phenylalanine residue at position 508 |
| FBS | Foetal bovine serum |
| FDA | Food and Drug Administration |
| FEE | Fecal Elastase E1 |
| FEV1 | Forced expiratory volume in first second |
| Flag tag | sequence motif DYKDDDDK |
| FLP | flippase |
| FRT | flippase recombination target |
| Fsk | Forskolin |
| GFP | green fluorescent protein |
| HBE | Human Bronchial Epithelial cells |

| | |
|---|---|
| HTS | High throughput screen |
| ICM | intestinal current measurement |
| IRT | Immunoreactive trypsinogen |
| Isc | Short-circuit current |
| IVS | intervening sequence, intron |
| MAGOH | mago-nashi homolog |
| MCC | mucociliary clearance |
| mRNA | messenger ribonucleic acid |
| mRNP | messenger ribonucleoprotein particle |
| MSD | Membrane-spanning domain |
| MVCC | mutation of varying clinical consequence |
| N-terminal | amino-terminus |
| NBD | Nucleotide binding domain |
| NBS | newborn screening |
| NGS | Next-Generation Sequencing |
| NMD | Nonsense mediated decay |
| nt | nucleotides |
| NPD | Nasal Potential Difference |
| ORF | open reading frame |
| PABP | poly(A)-binding protein |
| PABPC1 | poly-A binding protein complex 1 |
| PAP | Pancreatitis Associate Protein |
| PBS | Phosphate buffered saline |
| PCR | polymerase chain reaction |
| PD | potential difference |
| PFA | Paraformaldehyde |
| PI | pancreatic insufficiency |
| PKA | cAMP-dependent protein kinase |
| PKC | Protein kinase C |
| PM | Plasma membrane |
| PTC | Premature termination codon |
| PTC124 | (3-[5-(2-fluorophenyl)-[1,2,4]oxadiazol-3-yl]-benzoic acid |
| QC | Quality control |
| RD | Regulatory domain |
| RFLP | Restriction Fragment Length Polymorphisms |
| rpm | rotations per minute |
| RT | Room temperature |
| RT-PCR | Reverse transcriptase polymerase chain reaction |
| SD | standard deviation |
| SEM | Standard error of the mean |
| siRNA | small interfering RNA |
| SMG | suppressor of morphological defects on genitalia |
| SMG-1 | Serine/threonine-protein kinase 1 |
| sq-RT-PCR | semi quantitative reverse transcriptase polymerase chain reaction |
| SURF | SMG1-UPF1-eRFs complex |
| TM | Transmembrane segments |

| | |
|---|---|
| uORF | upstream open reading frame |
| UPF | up-frameshift (proteins) |
| VX-770 | N-(2,4-Di-tert-butyl-5-hydroxyphenyl)-4-oxo-1,4-dihydroquinoline-3-carboxamide |
| VX-809 | 3-[6-[[1-(2,2-difluoro-1,3-benzodioxol-5-yl) cyclopropanecarbonyl]amino]-3-methylpyridin-2-yl] benzoic acid |
| % v/v | Percentage expressed in volume per volume |
| WB | Western blot |
| wt-CFTR | Wild-type CFTR |
| % w/v | Percentage expressed in weight per volume |

# *I GENERAL INTRODUCTION*

# *1 Cystic Fibrosis- Clinical Phenotype, Diagnosis & Prognosis*

## *1.1 Clinical Phenotype*

Cystic Fibrosis (CF, MIM#2109700) is the most common lethal autosomal recessive disorder in the Caucasian population, affecting 1 in 2,500-6,000 newborns in Europe with currently ~34,000 patients listed in the European Registry. The disease frequency is variable among ethnic groups, being highest in North-eastern European populations and quite rare among oriental populations. In Portugal it is estimated that 1 in 6000 newborns are affected (De Boeck and Amaral, 2016; Santos et al., 2017).

The disease was first described in 1938, based on autopsy studies of malnourished infants, who showed a recessive autosomal pattern of inheritance (Andersen, 1938). Although the biochemical basis of CF was not identified for 50 years, the disease was known to be associated with abnormalities of chloride ($Cl^-$) and sodium ($Na^+$) transport in several epithelia (Quinton, 1983). In 1989, using a positional cloning strategy, it was determined that a single gene was responsible for the disease onset (Kerem et al., 1989; Riordan et al., 1989; Rommens et al., 1989). The discovery of the CF gene led to the demonstration that the impaired chloride transport is due to failure of a cAMP regulated $Cl^-$ channel. This gene was shown to be expressed at the apical cell membrane in a large number of epithelial tissues, and was cautiously named Cystic Fibrosis Transmembrane conductance Regulator (CFTR; MIM*602421) (Kerem et al., 1989; Riordan et al., 1989; Rommens et al., 1989).

The pathophysiology of CF can be described by the so-called "CF pathogenesis cascade" which starts with the presence of two defective *CFTR* genes associated with the CF phenotype (Figure 1.1). At the cellular level, this is characterised by the lack of, or reduced, cAMP-mediated $Cl^-/HCO3^-$ secretion. In other words, the cells from CF patients do not have the ability to transport cAMP-mediated $Cl^-/HCO3^-$ to the same extent as normal cells (Amaral and Kunzelmann, 2007; Amaral, 2014).

By acting as a negative regulator of the epithelial $Na^+$ channel (ENaC), CFTR is also a general regulator of ion and water transport in epithelia (Rich et al., 1990). Due to the fact that water follows the direction of $Na^+$ transport across the epithelia, the CF defect is also associated with enhanced water reabsorption. This event, according to some authors leads to increased salt concentration in the airway surface liquid (ASL) (Puchelle et al., 2002).

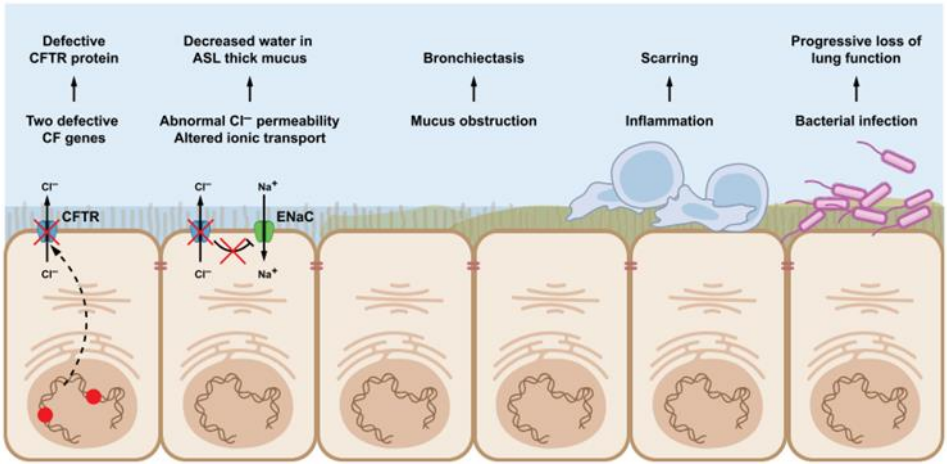



**Figure 1.1 - CF Pathogenesis Cascade: From primary gene defect to lung disease.** The mechanism of CF dysfunction starts with the primary CFTR gene defect and ultimately leading to severe lung deficiency. CFTR, cystic fibrosis transmembrane conductance regulator; ASL, airway surface liquid; ENaC, epithelial $Na^+$ channel. Adapted from (Amaral, 2014)

CF is a multi-system disease but the main cause of morbidity and mortality is chronic lung disease (Collins, 1992). In the respiratory tract, CF is manifested by the obstruction of the airways by thick, dehydrated mucus that prevents proper mucociliary clearance (MCC). This leads to recurring bacterial infections, especially by *Pseudomonas aeruginosa* and *Staphylococcus aureus*, that generate a hyper inflammation environment in the lungs of patients from a very early age, which leads to exacerbated tissue remodelling and fibrosis (Amaral and Kunzelmann, 2007).

Other CF symptoms include exocrine pancreatic insufficiency (PI, 85%), intestinal obstruction, elevated sweat electrolytes and male infertility (95%) (Collins, 1992; Jarvi et al., 1995; Welsh and Smith, 1995; Zielenski and Tsui, 1995). Additional (rarer) clinical manifestations include hepatobiliary disease (7%) and diabetes mellitus (Welsh and Smith, 1995).

Even though CF is a monogenic disease, the genotype-phenotype correlation is not straightforward. Nevertheless, homozygosity for the most common mutation F508del is usually associated with the 'classical' CF phenotype: progressive obstructive lung disease, PI, male infertility and very high concentrations of $Cl^-$ in sweat (Boucher, 2004). Each organ affected in CF requires a different level of CFTR function. Decreasing levels of CFTR function are associated with progressive involvement of more organs or systems and with a more severe phenotype (Rowe et al., 2005). The primary organ systems involved are the airways, in which mucus plugging and inflammation lead to progressive small airways obstruction and bronchiectasis; the gastrointestinal tract, liver, and exocrine pancreas, in which inspissation of viscid secretions leads to intestinal obstruction, cholestasis, and malabsorption of fat and protein; and the male reproductive tract, in which obstructive azoospermia leads to infertility.

However, some patients can have only one (or some) of these clinical phenotypes and still have no CF, being diagnosed with a CFTR-related disorder (CFTR-RD) or "CFTR-opathy" (Amaral, 2014).

A CFTR-RD is thus defined as: "*a clinical entity associated with CFTR dysfunction that does not fulfil the diagnostic criteria for CF*" (Bombieri et al., 2011). Three main clinical entities illustrate these phenotypes when associated with CFTR dysfunction: congenital bilateral absence of the *vas deferens* (CBAVD), acute recurrent or chronic pancreatitis and disseminated bronchiectasis. Careful attention should be paid to exclude other known aetiologies, to the degree of screening for CFTR mutations and to the evaluation of CFTR function in these patients (Paranjape and Zeitlin, 2008; Michl et al., 2016).

## *1.2 Diagnosis*

Initially a diagnosis of CF relied on the clinical phenotype, with recognition of the characteristic signs and symptoms of "classical" CF. There has been a surprising degree of difficulty encountered worldwide in establishing the diagnosis in a minority of cases and consequently, healthcare providers continue to face uncertain cases and challenging diagnostic dilemmas (Farrell et al., 2017c). However, the increased use of prenatal screening for maternal CF carrier status, prenatal ultrasound screening (that might reveal meconium ileus, meconium peritonitis, bowel obstruction, or echogenic bowel), and newborn screening (NBS) has resulted in the routine diagnosis of asymptomatic or minimally symptomatic infants and provided an opportunity to foster their normal growth and development (Farrell et al., 2017b). Nevertheless, the extensive NBS programmes implemented have identified increasing numbers of asymptomatic CF patients merely by identifying elevated serum concentrations of immunoreactive trypsinogen (IRT). This, poses new challenges to the CF diagnosis paradigm, especially when associated with borderline sweat $Cl^-$ (see below) and/or inconclusive CFTR genotypes (Farrell et al., 2017c). These inconclusive CF diagnosis through NBS, are designated as "*CF screen positive with an inconclusive diagnosis*" (CFSPID), also called "*CFTR-related metabolic syndrome*" (CRMS) (Farrell et al., 2017a). Diagnosis of CF in non-screened populations can be challenging because the age of onset and severity of symptoms can differ greatly as a result of highly variable degrees of CFTR dysfunction.

Figure 1.2 provides a simplified algorithm for the application of these consensus criteria to individuals with a suspicion of CF due to a positive NBS result, the appearance of signs or symptoms, or recognition of an immediate family member with history of CF (most often a sibling but may also include a parent or child). It should be noted that a positive NBS result does not mean the infant has CF; the probability of a CF diagnosis following a positive result varies greatly depending on the NBS method used (Farrell et al., 2017c).

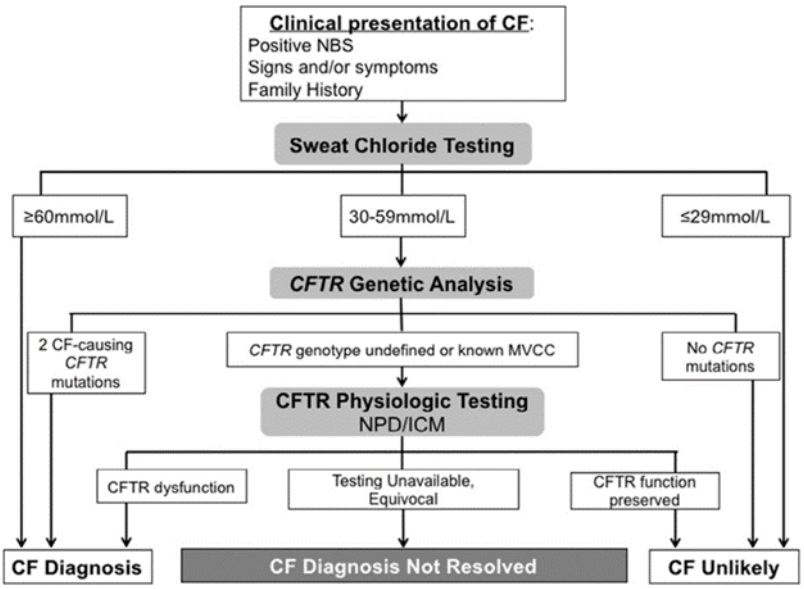



**Figure 1.2- Schematic representation of CF Diagnosis.** Established guidelines to diagnose CF according to clinical presentation and other types of evidence. Adapted from (Farrell et al., 2017c)

A confirmed CF diagnosis can only be established when an individual has both a clinical presentation of the disease and evidence of CFTR dysfunction. The tests conducted to establish CFTR dysfunction are done hierarchically, with sweat chloride being usually the first, followed by CFTR genetic analysis, and then CFTR functional tests like altered transepithelial nasal potential difference (NPD) or intestinal current of voltage measurements (ICM/IVM) (Sousa et al., 2012). For a positive diagnosis of CF two positive sweat tests are required and/or two CF-causing mutations identified. Only very rarely can an individual with a negative sweat chloride <30 mmol/L (see below) be considered to have CF and this diagnosis is only considered if all possible alternative diagnoses are excluded and if additional tests (genetic, functional tests) demonstrate dysfunctional CFTR. If only 1 CFTR gene variant is identified, further ("extended") *CFTR* gene analysis should be performed. CF is confirmed if both alleles possess CF-causing mutations. If two (or one) of these mutations is undefined, or if a mutation of varying clinical consequence (MVCC) is present CF is possible but not confirmed; CF is unlikely if the only mutations found have been described as non-CF-causing. If a CF diagnosis is not resolved, it can be considered as CRMS or CFSPID, if following NBS, or CFTR-RD (Farrell et al., 2017c).

However, depending on the ethnic background of the populations tested (Bobadilla et al., 2002), there are cases positively diagnosed with CF, from infancy to adulthood, that present "non-classical", or milder forms of the disease, representing 2% to 10% of all CF diagnoses. Most non-classical CF patients meet the same diagnostic criteria as classic CF patients, however, they may have a later presentation or have involvement of only one or two organ systems and usually with milder symptoms (Paranjape and Zeitlin, 2008). The majority of

patients with non-classical CF carry one severe and one mild or two mild CFTR mutations. A mild mutation is categorized as a gene alteration that still allows the resulting CFTR protein to partially function to a variable degree (Groman et al., 2005; Sousa et al., 2012).

Certain individuals show conflicting results from the diagnostic tools available, such as inconclusive genetic testing, borderline sweat tests with both non-CF and CF features. For such individuals with clinical phenotypes not fully meeting the CF diagnostic criteria it is also difficult to exclude CF and as such additional functional tests (NPD or ICM/IVM) should be carried out.

**Sweat test**

Although NBS is now widely implemented, to confirm a diagnosis of CF two positive sweat tests are required, with values ≥ 60 mmol/L indicating a diagnosis of CF and < 30 mmol/L indicating CF as very unlikely. For individuals with a clinical phenotype or positive NBS and sweat chloride levels 30-59 mmol/L further testing for CFTR mutations and function (such as NPD and ICM/IVM are indicated.

**Genetic tests for CFTR**

The discovery of the CF gene in 1989 (Riordan et al., 1989) provided a valuable tool for the genetic diagnosis of the disease. However, there are a large number of identified variants of the *CFTR* gene (currently more than 2000), most of which of unknown disease prognosis (Cystic Fibrosis Mutation Database, 2017). A more recent CFTR mutation database (CFTR2: https://cftr2.org) reports just the mutations of known disease prognosis.

All patients with clinical suspicion of CF and positive or borderline sweat test should be evaluated through genotype analysis. Traditionally, identification of CFTR mutations was performed by testing a panel of a given number of mutations (usually 5-150 mutations), usually customized for the most common mutations present in the local populations. In Europe, it has become clear which mutations should be analysed in an initial screening, since 90% of CF related mutations can be detected using specific methods for only 30 mutations (De Boeck, et al., 2011). However, there is a wide geographic variability and in the South of Europe the spectrum of mutations is much wider than in the North. So, other strategies can be used to identify mutations, with the direct DNA sequencing of the complete gene (or at least of all 27 exons and the respective flanking intronic regions) being the ultimate gold standard.

**Nasal Potential Difference (NPD)**

Active transport of ions across the nasal epithelium results in a potential difference (PD), which can be measured by a voltmeter between two electrodes: 1) an exploring electrode placed on the surface of the nasal epithelium and 2) a reference electrode (generally subcutaneous), which represents the internal surface (Keenan et al., 2015). NPD measurements should be performed in both nostrils and reported as an averaged result. This approach reduces variability between both nostrils and repeated measurements. This diagnosis method has become the standard reporting method in CF diagnostic and is widely used in research applications. Basal values for NPD are usually interpreted as normal (< -12 mV), intermediate (-12 to -7.7 mV) or abnormal (> -7.7 mV) (Ooi et al., 2012). CF patients usually present an increased magnitude of baseline PD (reflecting $Na^+$ hyperabsorption), a dramatic response to amiloride, a transient or blunted response to low $Cl^-$ solution and isoproterenol (β-adrenergic agonist) and an exaggerated response to ATP (reflecting $Cl^-$ secretion through Calcium-activated chloride channels (CaCC)) (Taylor et al., 2009).

**Intestinal Current/Voltage Measurements (ICM/IVM)**

Bioelectrical measurements on rectal epithelia can be applied as a functional diagnostic tool to aid in establishing a diagnosis of CF, if sweat test results are equivocal and/or if CFTR-disease causing mutations are not readily identified by genetic testing. For Intestinal Current/Voltage Measurement (ICM/IVM), an intestinal biopsy (obtained by suction or forceps) and a special micro-Ussing chamber are needed. Through this method the *ex vivo* transepithelial short-circuit current ($I_{sc}$) or transepithelial voltage ($V_{te}$) is measured and reflects the net ion fluxes across the tissue, which is calculated as equivalent current ($I_{eq\text{-}sc}$) by Ohm's law in the case of voltage measurements. In CF, the intestinal CFTR-mediated $Cl^-$ secretion is impaired, while absorptive processes remain unchanged or may be enhanced. The values of $I_{sc}$ (or $I_{eq\text{-}sc}$) in response to forskolin (Fsk), an activator of CFTR, can lead to normal, absent or reduced currents. Under normal conditions, the $I_{sc}$ responds to Fsk alone and to Fsk with carbachol (CCH), which enhance the driving force for $Cl^-$ to exit the cell by stimulating basal $K^+$ channels. This response consists of two components: a first lumen-negative current that is caused by the apical chloride efflux, followed by a larger lumen-negative current, caused by apical chloride secretion under the higher driving force (Sousa et al., 2012). In ICM/IVM of CF individuals, the apical $Cl^-$ efflux in response to Fsk and CCH does not occur, instead only a lumen-positive peak in response to CCH is observed corresponding to $K^+$ efflux. A biphasic response may also be observed under CCH+Fsk (first lumen-positive corresponding to $K^+$ channels efflux and then lumen-negative corresponding to $Cl^-$ efflux) due to residual CFTR-mediated $Cl^-$ efflux in milder forms of CF (Sousa et al., 2012).

**Neonatal Screening**

Crossley and colleagues (Crossley et al., 1979) from Auckland, New Zealand, established the foundation for CF NBS in the 1970s by measuring immunoreactive trypsinogen (IRT) on a dried blood spot specimen, using a radioimmunoassay. In CF, the trypsinogen release into the circulation appears to be enhanced by abnormal pancreatic duct secretions. Thus, IRT levels are found to be elevated in CF and determination of IRT levels in newborns constitutes the assay used in most CF neonatal screening programmes (Rosenfeld et al., 2016). IRT values tend to remain raised for several months in newborns with CF, whereas in false positives they usually return to normal within the first weeks of life. Screening for CF can have different algorithms and use an increased IRT level from the dried blood spot specimen as the first stage. Most frequently, the second stage incorporates molecular testing for CFTR mutations on the first specimen (IRT/DNA). In some programmes, IRT concentrations are analysed on a second specimen collected between 1 and 2 weeks of life (IRT/IRT), whereas other programmes combine the two strategies, following 2 increased IRT results with CFTR mutation analysis performed on the second specimen (IRT/IRT/DNA) (Farrell et al., 2017b). An important outcome of recent studies is the discovery that Pancreatitis Associate Protein (PAP) may provide advantages for programmes that routinely collect samples after 36h of age and can be incorporated into these programmes in different ways (Farrell et al., 2017c). The careful selection of the mutations on the CFTR screening panel is critical in order to ensure adequate coverage for the racial and ethnic composition of the population being screened. The most important is that each programme adopts an algorithm that is adapted to the characteristics of the population being tested and also matches the health system conditions of each country. This approach has demonstrated long-term benefits such as early nutritional treatment, reduced hospitalizations (due to pulmonary exacerbations) and improved survival.

In the 1950s, median life expectancy for patients with CF was a few months; the main causes of death were meconium ileus and malnutrition subsequent to pancreatic malabsorption. During the past six decades, median age of survival has increased progressively, and is now more than 40 years in developed countries (Farrell et al., 2017b). Replacement of pancreatic enzymes and intensive therapy guided by multidisciplinary teams have brought a major improvement in the treatment of CF and survival. Besides, the new guidelines for diagnosis of CF are very important to identify the patients as CF and begin health care at the earliest possible stage (Table 1.1).

**Table 1.1** – **Guidelines for CF diagnosis.** Summary from the 2015 CF Foundation guidelines for diagnosis of CF. Adapted from (Farrell et al., 2017c)

| **Guidelines for screened populations** |
|---|
| • Sweat testing: should be done in everyone |
| • Sweat $Cl^-$: <30 mmol/L is normal threshold for all ages (exceptions occur) |
| • NPD/ICM: useful; testing should be conducted in a validated lab |
| • CFTR mutations: use CFTR2 mutation list, with guidelines given for mutations not included in CFTR2 |
| • Presumptive diagnosis of CF: can be made ($NBS^+$ and 2 CF mutations or signs and symptoms of CF; or meconium ileus) and treatment started; diagnosis must be confirmed with a sweat test |
| • Genetic analysis: recommended in addition to that done during NBS |
| **Revisions to guidelines for CRMS/CFSPID** |
| • CRMS = CFSPID: now harmonized definition |
| • Repeat sweat testing recommended; NPD/ICM testing may be considered |
| • Clinical assessment: by age 2 months; duration and frequency of follow-up remains to be determined |
| **Revisions to guidelines for nonscreened population with inconclusive sweat chloride values** |
| • Sweat $Cl^-$: <30 mmol/L is normal threshold for all ages (exceptions occur) |
| • Ancillary testing: NPD/ICM |
| **Revisions to general definitions** |
| • CFTR-related disorder: a symptomatic entity that does not meet diagnostic criteria for CF |
| • Avoid terms like "atypical" or "nonclassic" CF because there is no consensus definition of these terms |

## 2. CFTR: from Gene to Protein and Function: CFTR Mutations & Functional Classes

### 2.1 – CFTR: From Gene to Protein and Function

The *CFTR* gene is located on the long (q) arm of chromosome 7, bands 31-32 (7q31-7q32), and is one of the largest human genes, spanning ~190 kb. After transcription and splicing, the mRNA comprising 6.5 Kb is translated into the CFTR protein (or ABCC7) with 1480 amino acid (aa) residues (Riordan et al., 1989). The gene consists of 27 exons and 26 introns (Figure.1.3). The sizes of both exons and introns vary greatly, with the exons ranging from 38 bp in exon 16 to 724 bp in exon 14. Regarding intron length, the smallest intron (intron 25) comprises 600 bp, and the largest (intron 11) 28,085 bp (Ellsworth et al., 2000).

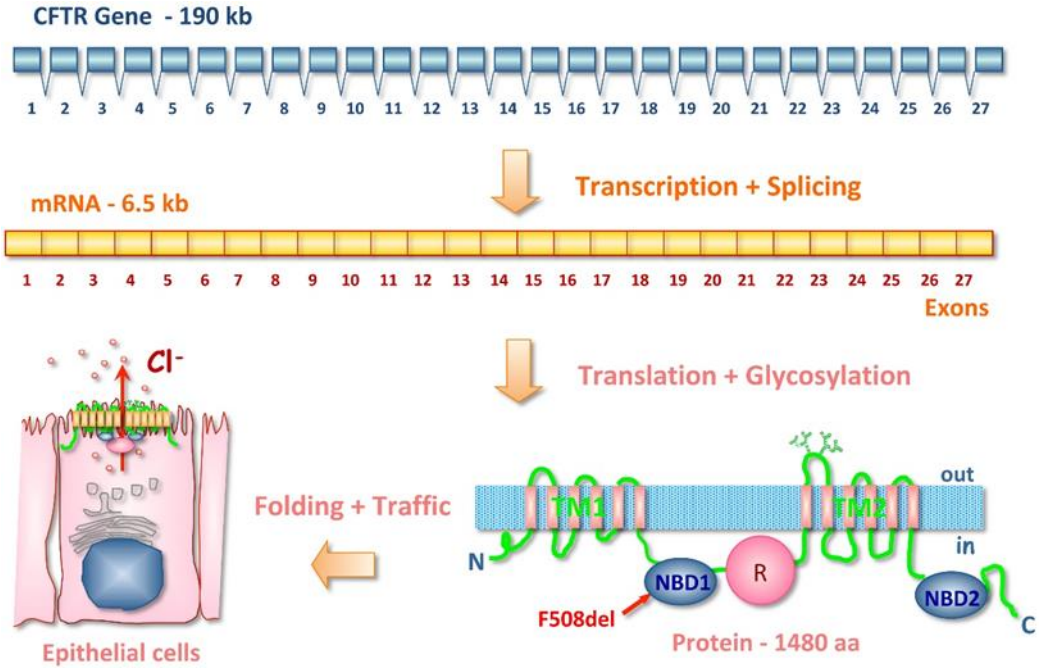



**Figure.1.3** - **Schematic representation of the CFTR gene, mRNA and protein**. N– N-terminus; TM– transmembrane domain; NBD – nucleotide-binding domain; R – regulatory domain; C – C terminus CFTR gene consist of 27 exons and introns. After transcription and mRNA splicing, a 1480 amino acid sequence is translated. Post-transcriptional modification of the protein glycosylation, occurs before translation to the cell membrane. The CFTR protein act at the apical membrane of epithelial cells (more details in section 2.2.1). [Image from MD Amaral, with permission].

As mentioned above, CFTR is a multidomain protein containing 1480 aa residues, that functions as a cAMP-activated and phosphorylation-regulated $Cl^-$ and bicarbonate channel at the apical membrane of epithelial cells. CFTR is a member of the ATP-binding cassette (ABC) transporter superfamily, also known as ABCC7 (Gadsby et al., 2006). Like other ABC transporters, it has two membrane spanning domains (MSDs) with six transmembrane segments each, portions of which form the pore through which anions pass. Additionally, it has two nucleotide binding domains (NBD1 and NBD2) and a regulatory domain (RD). Both NBDs bind and hydrolyse ATP, which drives channel opening and closure, respectively. The

RD, which is absent in all other ABC transporters, contains multiple consensus sites for phosphorylation by various kinases. Phosphorylation of the RD by protein kinase A (PKA) in response to cyclic AMP is regarded as the major determinant for opening of the channel (Riordan et al., 1989; Riordan, 2005).

Proteins belonging to the ABC transporter family are responsible for active transport of substrates across cell membranes, where ATP hydrolysis serves as the source of energy to drive the transport. However, CFTR is functionally distinct from other ABC transporters as it enables bidirectional permeation of anions, rather than vectorial transport of solutes (Riordan, 2005). Several hypotheses were raised regarding the function of CFTR protein. The model currently accepted for the mechanism of channel function and regulation proposes that (Figure 1.4) the first step is the phosphorylation of the RD by cAMP-dependent PKA (and by PKC) then ATP binds to the NBDs, which subsequently dimerize opening the channel (Sheppard and Welsh, 1999; Hwang and Sheppard, 2009) with the consequent alteration in the MSDs conformation, leading to the opening of the channel pore. It has also been shown that intramolecular interaction between the RD and both the N-terminus and NBD1 are required to regulate CFTR function (Baker et al., 2007). Even though the mechanism of CFTR channel gating is not fully understood, opening and closing of this $Cl^-$ channel is tightly regulated by the cellular balance of kinase and phosphatase activity and by ATP levels. Furthermore, the open probability (Po) of the channel is controlled by the extent of RD phosphorylation at its multiple sites (Hwang and Sheppard, 2009).

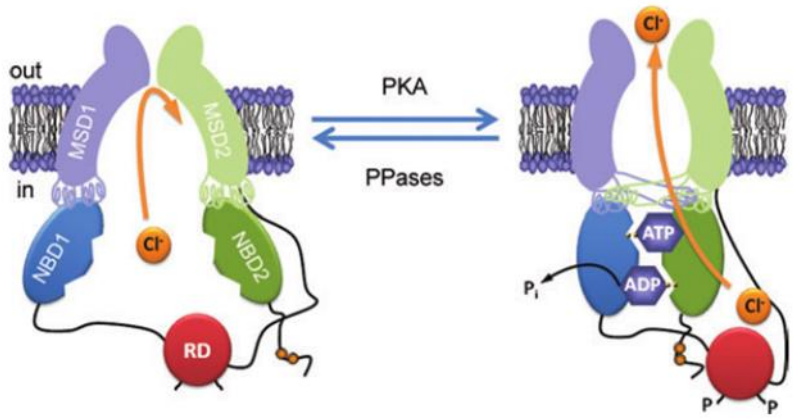
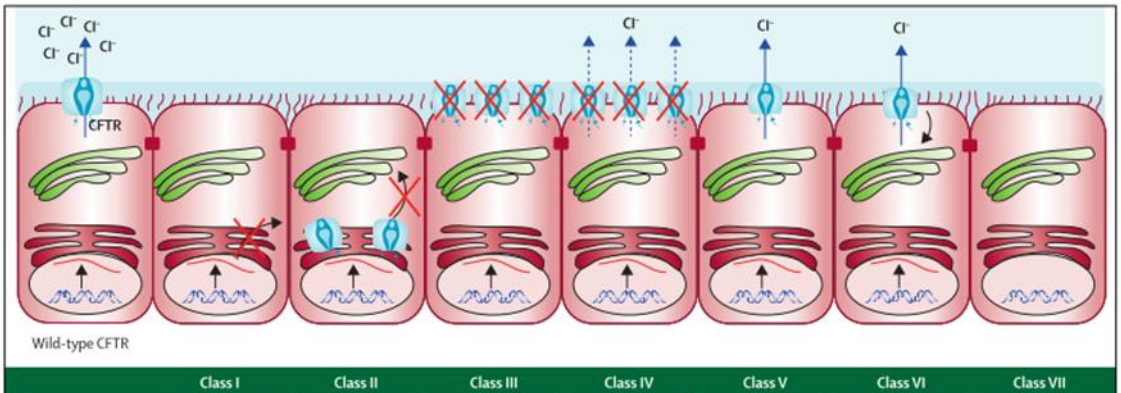



**Figure 1.4 - Conformational changes of the CFTR $Cl^-$ channel during channel gating.** The simplified model shows a CFTR channel under quiescent (left) and activated (right) conditions. Communication between the NBDs and MSDs via the intracellular loops is either orthogonal (e.g. NBD1–MSD2) according to the most recent CFTR structural models based on Sav168834. P (phosphorylation of the RD); Pi (inorganic phosphate); PKA (cAMP-dependent protein kinase); PPase (protein phosphatase). In and Out denote the intra- and extracellular sides of the membrane, respectively. [Image reproduced from Hwang and Sheppard, 2009].

Besides its function as $Cl^-$ channel, CFTR has also been shown to regulate several other channels and transporters, thus being a general regulator of ion transport in epithelia. CFTR was found to be expressed in luminal membranes of both secretory and absorptive epithelia, playing a predominant role in both cAMP- and, as more recently shown (Faria et al., 2009; Lérias et al., 2018) in $Ca^{2+}$-activated $Cl^-$ secretion.

## *2.2 – CFTR mutations & Functional Classification*

As to date, 2026 mutations have been reported in the CFTR gene (Cystic Fibrosis Mutation Database, 2017), a number that comprises both disease causing mutations and polymorphisms that do not affect the CF phenotype. From these mutations, 39% are missense mutations, 16% frameshift mutations, 11% splicing mutations, 8% nonsense mutations, 2% In frame in/dels, 3% large in/dels, 0.8% mutations in the promoter, 13% sequence variation and 7% are of unknown consequences (Cystic Fibrosis Mutation Database, 2017). Currently these mutations are grouped into 7 functional classes defined according to the molecular changes caused by the different CFTR variants, with these evolving into theratypes, i.e., groups of mutations which can be corrected by the same therapeutic strategy (Figure 1.5).




| | Class I | Class II | Class III | Class IV | Class V | Class VI | Class VII |
|---|---|---|---|---|---|---|---|
| CFTR defect | No protein | No traffic | Impaired gating | Decreased conductance | Less protein | Less stable | No mRNA |
| Mutation examples | GLy542X, Trp1282X | Phe508del, Asn1303Lys, Ala561Glu | Gly551Asp, Ser549Arg, Gly1349Asp | Arg117His, Arg334Trp, Ala455Glu | Ala455Glu, 3272-26A→G, 3849+10 kg C→T | c. 120del23, rPhe508del | dele2,3(21 kb), 1717-1G→A |
| Corrective therapy | Rescue synthesis | Rescue traffic | Restore channel activity | Restore channel activity | Correct splicing | Promote stability | Unrescuable |
| Drug (approved) | Read-through compounds (no) | Correctors (yes) | Potentiators (yes) | Potentiators (no) | Antisense oligonucleotides, correctors, potentiators? (no) | Stabilisers (no) | Bypass therapies (no) |

**Figure 1.5** – **Functional classes of CFTR mutations/ theratypes**. Wild-type CFTR protein at the plasma membrane of epithelial cells properly functioning as a $Cl^-$ channel; (I) class I mutations: prevent translation; (II) class II: defective processing; (III) class III: defective regulation; (IV) class IV: defective conductance; (V) class V: reduced synthesis; (VI) class VI: decreased stability; (VII) class VII: no mRNA. Adapted from (De Boeck and Amaral, 2016)

**Class I – Mutations that lead to no protein production**

Class I mutations are those that can give origin to premature terminations codons (PTCs), such as nonsense mutations or some frameshift mutations, which are thus associated with more severe CF phenotypes (Elborn, 2016). Examples of class I mutations include modification of the codons that codify for the glycine in the position 542, the arginine in position 553, or the glutamine in position 637, into a stop codon (G542x, R553X and Q637X, respectively). G542X and R553X are two of the most common mutations in Europe following the F508del mutation (WHO, 2004). These mutations lead to significant degradation of the respective transcripts by nonsense mediated decay (NMD) so as to prevent production of truncated proteins that might be translated from these transcripts (De Boeck and Amaral, 2016). These mutations will be discussed in greater detail in section 3.

**Class II – Mutants that affect intracellular traffic**

Class II mutations are those that lead to incorrect folding and processing and thus prevent the protein from reaching the apical membrane of epithelial cells. F508del, the most common CF causing mutation, is a member of this class. By not being correctly folded, F508del CFTR protein is retained in the endoplasmic reticulum (ER) and targeted for degradation, thus not being fully-glycosylated in the Golgi nor transported to the apical membrane.

**Class III - Mutations affecting channel gating**

These mutations cause a defective response of CFTR to the phosphorylation of the RD by PKA after activation by cAMP. They are often located in the NBDs, affecting their interaction with ATP, and thus interfering with the correct gating of the channel. An example of a class III mutation is G551D (glycine to aspartate) which is quite frequent in Ireland (14%) and is thus also called the "Celtic mutation".

**Class IV - Mutations that lead to defective chloride transport**

Class IV mutations, such as R334W (mutation of an arginine to a histidine), lead to the synthesis of proteins which can be transported to the membrane and respond to stimulus, but have a reduced ion transport function due to defective conductance. Patients with these mutations have a milder CF phenotype.

**Class V – Mutations that lead to reduced levels of protein**

This class includes mutations that lead to reduced levels of protein, such as those affecting splicing of pre-mRNA such as 3272-26A>G or the TG12T5 sequence in intron 8 or missense mutations such as A455E (alanine to glutamate), which lead to reduced levels of transcripts

and thus low levels of functional protein. These mutations are also associated with milder CF phenotypes.

**Class VI – Reduced stability or altered regulation of separate ion channels**

In this group are included membrane-rescued mutations such as F508del, 120del23 (the deletion of the first codon), N287Y (asparagine to tyrosine). This class also includes variants with nonsense or frameshift mutations that originate PTCs in the last exon of the CFTR gene and lead to the production of carboxyl terminus (C-terminus)-truncated proteins. These proteins while being correctly processed, transported to the membrane and presenting normal function, are unstable and have an increased turnover at the cell surface. This is another class of mutations that leads to a milder CF phenotype.

**Class VII – No mRNA, unrescuable mutations**

Class VII mutations are the so-called "unrescuable mutations", because CFTR cannot be pharmacologically rescued *per se*, and include for example large deletions such as the del2,3(21kb) mutation.

# *3 Nonsense-mediated mRNA decay and CFTR nonsense mutations*

The expression of protein-coding genes in eukaryotes involves the orchestration of transcriptional and posttranscriptional processes. To ensure the fidelity of these processes, the eukaryotic cell has evolved several quality control mechanisms. The best characterized mRNA quality control system in eukaryotes is the nonsense mediated decay (NMD) of mRNA. This surveillance mechanism detects and ensures the rapid degradation of faulty mRNAs harbouring premature translation termination codons (PTCs) that would otherwise have resulted in the synthesis of C-terminally truncated proteins. Hence, the quality control function of NMD relies on protecting the cell from the deleterious dominant-negative or gain-of-function effects of these truncated proteins (Kurosaki and Maquat, 2016; Nickless et al., 2017).

**Nonsense-mediated mRNA decay**

First discovered in yeast and then extensively studied in *Caenorhabditis elegans*, *Drosophila*, mouse, human cells, and other model systems, NMD is a RNA surveillance pathway that acts at the interface between transcription and translation (Nickless et al., 2017). Three key steps constitute the NMD pathway: 1) the identification of the PTC; 2) the assembly of the surveillance complex on the mRNA and 3) the degradation of the targeted RNA. The mechanism by which PTCs are recognized and discriminated from natural stop codons is translation-dependent and leads to the recruitment of NMD *trans*-acting factors, called upframeshift (UPF) proteins, such as UPF1, UPF2 and UPF3 (Lykke-Andersen and Heick Jensen, 2015).

During translation, stop codons are recognized by eukaryotic release factors (eRFs) eRF1 and eRF3, which are present in the translation apparatus, and in turn recruit UPF1, followed by the recruitment of the serine/threonine protein kinase SMG1. These proteins, together with the eRFs form the surveillance complex (SURF), and eventually lead to translation termination involving ribosomal dissociation (Figure 1.6). However, recognition of PTCs leads to the interaction of the SURF complex with UPF2 and 3 present at a downstream exon junction complex (EJC) leading to UPF1 phosphorylation that triggers transcript degradation (Hug et al., 2015; Nickless et al., 2017).

Within the nucleus, pre-mRNAs are synthesized and immediately undergo *cis* modifications binding various proteins during pre-mRNA processing, prior to mRNA formation. These include capping at the 5′ end, i.e., addition of a 7-methylguanosine residue (Cap) to protect the transcript from 5′-to-3′ exoribonucleases and provide a binding platform for the Cap-binding protein (CBP) complex (CBC), composed of CBP80–CBP20 (Da Costa et al., 2017).

In mammalian cells, and according to the classical EJC-dependent model, NMD depends on the interaction of the translation termination complex with a dynamic multiprotein assembly, the so-called EJC (Lykke-Andersen and Heick Jensen, 2015). NMD is tightly intertwined with mRNA biogenesis even at its earliest stages, since it is promoted by both the CBC and EJC. The latter consists of four core proteins- eukaryotic translation initiation factor 4A3 (eIF4A3), cancer susceptibility candidate 3 (CASC3), RNA-binding motif protein 8A (RBM8A or Y14), and mago-nashi homolog (MAGOH)- which are deposited ~20-24 nucleotides (nt) upstream of ~80% of all exon-exon junctions (Hir et al., 2016; Kurosaki and Maquat, 2016). This splicing-dependent "mark" serves to guide the NMD machinery once the mRNA is exported to the cytoplasm for translation-dependent inspection by NMD factors.

EJCs upstream of and within mRNA coding regions are removed by ribosomes during translation in the cytoplasm, however, because PTCs shorten the length of the coding region, any downstream EJCs that normally reside within the coding region would fail to be removed from what becomes the 3′UTR thus triggering NMD (Kurosaki and Maquat, 2016).

Initially, UPF1 associates with SMG1 and acts as a "clamp", interacting directly with eRF1 and eRF3 to form the SURF complex in the vicinity of the PTC (Hug et al., 2015; Da Costa et al., 2017) (Figure 1.6). UPF1 is the central NMD factor, it is an ATP-dependent RNA helicase containing an abundance of serine and threonine residues located within its N- and C- termini that function as phosphate-acceptor sites (Popp and Maquat, 2017). SMG1 is a phosphatidylinositol 3-kinase-related protein kinase that phosphorylates many of these residues, and this activity is initially held in check by the SMG8-SMG9 complex (Yamashita, 2013). Two subunits of the SMG1c complex, the SMG8 and SMG9 associate tightly with SMG1 and regulate its activity through the induction of conformational changes, with SMG8 binding to the preformed SMG9-SMG1 complex and maintaining the kinase in its inactive state (Yamashita et al., 2009; Arias-Palomo et al., 2011). Subsequently, the SURF complex interacts with UPF2, UPF3b and an EJC downstream of the PTC to form the decay-inducing complex (DECID), that triggers UPF1 phosphorylation and dissociation of eRF1 and eRF3 (Figure 1.6) (Da Costa et al., 2017). UPF1 phosphorylation has two consequences: first, phosphorylated UPF1 represses further translation initiation through an interaction with eIF3 that is critical for mRNA decay; second, phosphorylated UPF1 residues form platforms exposing new regions of UPF1 onto which RNA-degradative enzymes are either directly or indirectly recruited, resulting in targeted, rather than indiscriminate, mRNA destruction (Lykke-Andersen and Heick Jensen, 2015). As a consequence of the remodelling of NMD complexes, UPF1 adopts its active helicase conformation, due to the reorganization of its inhibitory domains through association with UPF2. The activated NMD complex (UPF1, UPF2 and UPF3b) is translocated from its position upstream of the EJC towards the 3’ of the EJC (Hug

et al., 2015). Subsequently, phosphorylated UPF1 associates with the phospho-binding proteins SMG5, SMG6 and SMG7, and general mRNA degradation factors. Further rearrangements of this complex lead to mRNA degradation. SMG5, SMG6 and SMG7 share a common 14-3-3-like domain, which is usually associated with protein-protein interactions. This domain in SMG7 is thought to promote the dephosphorylation of UPF1 by binding phospho-UPF1 (Fukuhara et al., 2005). SMG6 itself is an endonuclease, which can form both phospho-dependent and phospho-independent interactions with UPF1. SMG6 cleaves NMD targets in the vicinity of the PTC, leading to the initiation of NMD-mediated RNA degradation. This canonical mammalian NMD pathway is not universal, since alternative NMD branches that are independent of UPF2, UPF3b or the EJC have been described (Lykke-Andersen and Heick Jensen, 2015; Raxwal and Riha, 2016; Popp and Maquat, 2017).

The EJC is involved in translation, mRNA transport, turnover, and the identification of UPF2 and UPF3 as EJC components suggests an additional role of this complex in NMD. The heterodimer Y14/MAGOH stabilizes the high affinity of eIF4A3 for RNA by preventing the ATP hydrolysis of eIF4A3 and further primes it to bind to CASC3 (Barentsz homologue). This arrangement allows eIF4A3 to clamp several proteins onto RNA in a stable and sequence independent manner, which in turn is used by the NMD machinery to recognize an aberrant PTC (Hug et al., 2015). Thereby, a short motif of the core NMD factor UPF3b binds to the EJC core factors and serves as a platform to assemble an active NMD complex. Two other DEAD box proteins, DDX5 and DDX17, have also recently been shown to interact with UPF3b and this interaction appears to be crucial for the degradation of a limited subset of NMD substrates. It was recently demonstrated that the RNA helicase DHX34, a member of the DExH/D box family of proteins, associates with the SURF complex and promotes the transition to the DECID complex (Hug et al., 2015; Kurosaki and Maquat, 2016; Raxwal and Riha, 2016). Thereby, DHX34 probably affects the remodelling of the SURF complex by promoting the dissociation of the eukaryotic release factor eRF1 and eRF3 in an ATP dependent manner from the SURF complex (López-Perrote et al., 2015).

**Pioneer round of translation**

The 5’end of mRNAs acquires a modified guanine residue during transcription that binds to the nuclear cap-binding protein (CBP) complex CBP80-CBP20 (CBC), which supports the pioneer round of translation (Ishigaki et al., 2001). The Cap is then replaced by eIF4E that directs subsequent rounds of translation. Spliced CBC-bound mRNAs, unlike eIF4E bound mRNAs, are typically associated with one or more EJCs. This feature suggests a role for CBC in NMD. Indeed, it has been proposed that NMD occurs exclusively during the pioneer round of translation and that transcripts bound to eIF4E are NMD-insensitive (Ishigaki et al., 2001;

Trcek et al., 2013). Additionally, CBC-bound complexes contain the nuclear poly(A) binding protein 1 (PABPN1) (Chiu et al., 2004; Kashima et al., 2006). The distinguishing features between the pioneer round of translation complexes and the steady-state translation initiation complexes contribute to the specialized functions of each complex. Although translation contributes to remodelling of the pioneer translation complexes by augmenting the displacement of EJC components, the replacement of CBC by eIF4E is promoted by the binding of the nuclear transport receptor importin β (Kurosaki and Maquat, 2016). In support of a role for CBC in NMD, a recently study has shown that interaction of CBP80 with UPF1 promotes NMD in two sequential steps (Hwang et al, 2011). First CBP80 chaperones SMG1-UPF1 interacts with eRF1-eRF3 at the PTC, promoting SURF complex assembly; and subsequently CBP80 physically joins the distal-EJC while still chaperoning SMG1-UPF1, which results in SMG1-mediated UPF1 phosphorylation (Hwang et al., 2011). Phosphorylated-UPF1 binds eIF3 of the 43S preinitiation complex positioned at the translation initiation codon and inhibits 60S subunit joining, thereby eliciting translation repression (Isken et al., 2008).

The decision of whether an mRNA will be targeted or not for degradation by the NMD pathway is made when the ribosome is poised at the termination codon. During normal translation termination, when the Stop codon enters the A-site, it is recognized by the termination factor eRF1, which forms a complex with eRF3 that involves the C-terminal domains of both proteins. In turn, eRF3, through its N-terminal domain, also interacts with the C-terminal domain of PABPC1 (Peixeiro et al., 2012) and this interaction results in stimulation of efficient translation termination (Brogna and Wen, 2009).

**CFTR nonsense mutations**

PTCs, account for ~8% of all CFTR variations (Cystic Fibrosis Mutation Database, 2017). These are classified as class I mutations and include both nonsense and frameshift, or even splicing mutations that result in a frameshift of the CFTR open reading frame (ORF) and ultimately introduce a premature stop codon. Thus, leading to total, or almost total, degradation of the respective transcripts by NMD, and therefore preventing any production of full length CFTR protein. These mutations are associated with "classical" severe CF phenotypes.

CFTR transcripts bearing mutations such as G542X and W1282X are prone to degradation by the NMD pathway. However, transcripts bearing the nonsense mutation Q39X, do not undergo NMD. Previous *in vitro* studies have shown that β-globin nonsense mutations in close proximity to the initiation codon also protect messengers from undergoing NMD (Inácio et al., 2004). It has been proposed that this might have two causes. Firstly, the interaction of the PABPC1 with the translation termination complex at the nonsense mutation site thus inhibiting

the activation of NMD pathway intermediaries and their interaction with the EJC complex present downstream of the PTC (Nagy and Maquat, 1998). Secondly, alternatively mRNA protection from NMD could result from re-initiation of translation at a Kozak consensus downstream of the mutation which would stabilize these PTC-containing transcripts (Neu-Yilik et al., 2011). The presence of alternative initiation codons downstream of this mutation was indeed shown for CFTR (Ramalho et al., 2009) with this being the most plausible reason why E39X mRNA is not an NMD substrate. Inhibiting NMD should increase the abundance of PTC-containing mRNAs, which could then potentially be targeted by read-through drugs to upregulate the production of full-length proteins during translation.

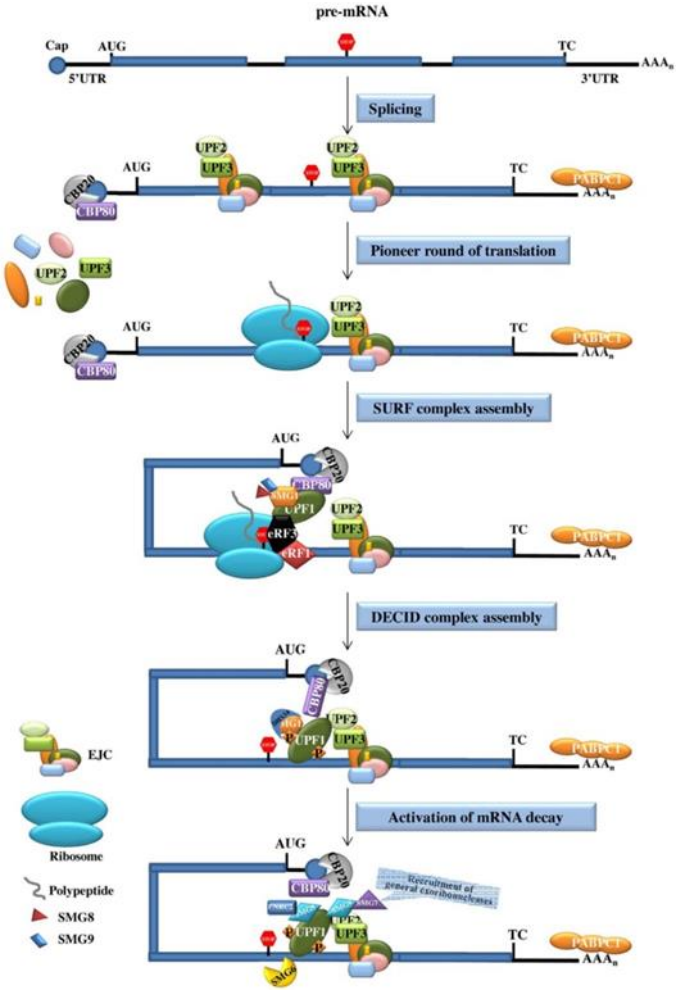



**Figure 1.6 - The nonsense-mediated mRNA decay pathway.** During splicing, exon junction complexes (EJCs) assemble 20-24 nucleotides (nt) upstream of exon-exon junctions. The translating ribosome displaces the EJCs during the pioneer round of translation. If a PTC is located more than 50-54 nt upstream from the last EJC, the termination complex is able to interact with EJC-associated NMD factors. The CBP80 protein at the mRNA 5'end can bridge UPF1-SMG-1 to eRF1-eRF3 at the PTC, leading to the formation of the SURF complex. The interaction of UPF1 with EJC-bound UPF2 induces the SMG1-mediated phosphorylation of UPF1 and dissociation of eRF1-eRF3. Phosphorylated UPF1 recruits SMG-7 and SMG-5 proteins and EJC recruits SMG-6, triggering the decay of the PTC containing transcript [Adapted from da Costa et al., 2017].

## *3.1 – PTC-definition and read-through compounds*

The mechanism of discrimination between a premature and a natural termination codon has been one of the central questions regarding the NMD mechanism. As mentioned in the previous sections, not all PTC-containing transcripts undergo NMD. To trigger NMD, at least one EJC must be present downstream of the PTC, in fact mRNAs from intronless genes or with a PTC in the last exon or with a PTC that spans two exons are immune to NMD (Lykke-Andersen and Heick Jensen, 2015). The involvement of splicing and the EJC in NMD led to a formulation of a model for PTC definition where the EJCs would function as a "mark" to discriminate PTCs from natural termination codons. It is assumed that during the first round of translation the elongating ribosomes have the ability to displace the EJC that were assembled during splicing. The PTC must also be located >50-55 nts upstream of the 3'-most EJC in order to reduce the mRNA abundance (Nagy and Maquat, 1998). The presence of an EJC downstream of the terminating ribosome would allow the interplay between UPF2 and/or UPF3b at the EJC with the SURF complex at the ribosome, culminating in NMD triggering.

In most organisms, three codons do not code for amino acids but are signals to indicate the end of an amino acid chain. These nonsense codons are UAG (amber), UAA (ochre) and UGA (opal). Under normal conditions, translation termination is a very efficient process with an estimated error rate of approximately 0.1% (Keeling et al., 2014a). However, eRF1 and near-cognate aminoacyl-tRNAs (aminoacyl-tRNAs with an anticodon complementary to two of the three nucleotides of the stop codon) normally compete for the ribosome A site binding. Under certain conditions, the rate at which near-cognate aminoacyl-tRNAs successfully compete with eRF1 at a stop codon can be increased, resulting in incorporation of an amino acid carried by a near-cognate aminoacyl-tRNA into the nascent polypeptide (Roy et al., 2015). This process is termed "termination suppression" or "read-through". In the case of a premature stop mutation, read-through normally results in the continued elongation of the polypeptide chain in the correct reading frame and the production of "full-length" protein (Kulyté et al., 2005; Keeling et al., 2014b).

Therapeutic approaches aimed at promoting translational read-through of the PTCs, and thus enabling the synthesis and expression of "full-length" functional proteins, have been developed with relatively positive results. Several pharmacologic approaches to induce read-through have been discovered, yet none has yielded an optimal combination of efficacy and safety (Mutyam et al., 2016). For instance, *in vitro* work has demonstrated that certain aminoglycosides (such as G418, gentamicin or tobramycin) can promote read-through (Keeling et al., 2014b; Solomon et al., 2016) but clinical trials studies have shown contradictory results (Clancy et al., 2001, 2007). In addition, these drugs have been shown to be unsuitable for long-term use. Synthetic aminoglycoside derivatives optimized for translation suppression

of the eukaryotic ribosome have exhibited improved read-through and reduced toxicity when compared *in vitro* (Mutyam et al., 2016).

A high-throughput screening (HTS) aimed at the identification of non-aminoglycosides capable of promoting read-through, revealed a small molecule, Ataluren (also known as PTC124), capable of inducing PTC read-through without severe side effects (Welch et al., 2007). Unfortunately, clinical trials for this molecule in CF patients, did not shown the expected effects and the trial was terminated.

## *4 CFTR Modulator Treatments*

CF is a very severe disease with a current median age of life expectancy in the fourth decade of life. The mainstay of treatment in CF is symptomatic and focuses on compensating for exocrine pancreatic insufficiency with pancreatic enzymes, fat- soluble vitamins and high caloric intake; and slowing lung disease progression with inhaled and physical therapies that improve airway clearance, and antibiotic therapy (Amaral, 2014; Michl et al., 2016).

An enormous research effort and progress has been made in understanding the consequences of these mutations on the CFTR protein structure and function. Several studies, have led to the approval of two new drug therapies, ivacaftor (vx770) and Orkambi (ivacaftor/lumacaftor; vx770/vx809), that are able to bind to defective CFTR proteins and partially restore their function. At present time these drugs are mutation-specific therapies. They are the first personalized medicine for CF with a possible disease-modifying effect. Other strategic approaches include the PTC read-through drugs, antisense oligonucleotides that correct the basic defect at the mRNA level or gene editing to restore the defective gene as well as gene therapy. All these strategies are needed to develop disease-modifying therapies for all patients with CF (Harman et al., 2018).

The goals of innovative approaches to therapy include the development of drugs that correct the basic defect in CF, that in turn can delay the progression, or even prevent respiratory disease, if administrated in the onset of the patient's life (Figure 1.7).

**Mutation-Specific Approaches**

The advent of CFTR modulator therapy in the last decade has revolutionized the approach to CF treatment by targeting the underlying CFTR defect. The fast progression from preclinical trials to clinical use of these novel agents, and the expanding research effort in identifying more potent drugs targeting a broader range of mutation classes brings, at the same time, excitement and challenge to field.

**Correction of class I mutations**

High-throughput drug screens have been carried out to identify compounds that inhibit NMD (Durand et al., 2007; Dang et al., 2009; Iwatani-Yoshihara et al., 2017). Aminoglycoside antibiotics have some ability to suppress PTCs by promoting the ribosomal read-through activity and restoring "full-length" CFTR production (Wilschanski et al., 2003; Altamura et al., 2013). They interact with the ribosome and instead of the PTC being read, they facilitate the random addition of an amino acid to the polypeptide chain allowing the translation to continue on to the normal Stop codon. However, aminoglycosides have severe side effects, such as nephrotoxicity and ototoxicity, when used at high concentrations and/or when used long-term (Durante-Mangoni et al., 2009; Fajac and Wainwright, 2017). Ataluren was also developed as a read-through agent by PTC Therapeutics, with no structural similarities to aminoglycosides. Studies with oral administration of Ataluren in patients with CF, bearing at least one CFTR nonsense mutation, have shown improved nasal epithelial $Cl^-$ transport. However, the results from a Phase III clinical trial for CF showed no significant benefit and thus the drug was discontinued (PTC Therapeutics, 2017). Other strategies have been attempted, such as suppressor tRNAs i.e., modified tRNAs whose anticodons were altered to recognize a stop codon, thus allowing the incorporation of an amino acid at the stop codon and a bypass of translation termination (Keeling et al., 2014b).

**Correction of class II mutations**

Mutations in this class have a folding defect that results in the rapid targeting of the resulting protein for degradation in the proteasome by the ER quality control (ERQC), thus preventing it from reaching the plasma membrane (PM). The most common CFTR mutation, F508del, belongs to this class. Pharmacological rescue of F508del-CFTR requires two types of compounds: correctors that promote folding and restore mislocalized protein to the PM, and also potentiators to increase CFTR $Cl^-$ channel function as this mutant also displays a gating defect similarly to those in class III (see below). Combination therapy with lumacaftor (corrector) and ivacaftor (potentiator) was assessed in phase 3 trials involving patients homozygous for F508del aged 12 years and older. Significant, although modest, improvement in lung function and nutrition, and reduction in the rate of pulmonary exacerbations were observed (Hubert et al., 2017). The combined therapy using lumacaftor and ivacaftor (Orkambi) was approved by the Food and Drug Administration (FDA) in 2015 for patients aged 12 years and older who are homozygous for Phe508del-CFTR and for patients with 6 years in 2016 (Fajac and Wainwright, 2017). Recently, the combination of another corrector, (VX-661, tezacaftor) and potentiator ivacaftor (Symdeko), has been also approved for F508del-

homozygous patients. There are also several ‘next generation’ molecules which are in Phase III trials as a triple combination.

**Correction of class III mutations**

CFTR mutants in this class have reduced response to channel agonists. The most common mutation in this class is G551D. Also included in this group is the F508del mutation (see above). CFTR potentiators such as genistein and other flavonoids can activate $Cl^-$ conductance and overcome mutations in this class. The potentiator VX-770 has been shown to increase CFTR activity in CF patients carrying the G551D mutation. This was the first CFTR-modulator drug approved that corrected the basic defect of a CFTR-mutant (Amaral, 2014; Michl et al., 2016). Moreover, VX-770 was also shown to correct other gating defect CFTR mutants located at the cell surface and consequently it was approved for clinical use in CF patients carrying those mutations.

**Correction of class IV mutations**

Class IV mutants are associated with reduced $Cl^-$ conductance. Stimulation of CFTR conductance can be achieved either by increasing the overall cell surface density of these mutants, thus compensating for the reduced conductance, and by augmenting the existing activity of channels with potentiators, namely the CFTR-potentiator VX-770 (lumafactor) which was approved for the treatment of some class IV mutants (Amaral, 2014).

**Correction of class V mutations**

This class includes mutations that reduce the amount of functional CFTR, such as alternative splicing mutations that generate both aberrant and correctly spliced mRNAs. The major therapeutic goal in these cases is to increase the levels of correctly spliced transcripts. However, modulation of the splicing process is a very complex approach, and most likely would have to be tissue specific. Nevertheless, combined therapy using both CFTR correctors (e.g., VX-809), and potentiators (e.g., VX770) can also bring clinical benefit by increasing both trafficking of CFTR to the PM and channel activity. Ivacaftor has been approved for five splicing mutations, for people age 2 or older with one of these splicing mutations. Another strategy would be to use the RNA-based antisense oligonucleotide (AON), which was shown to be promising *in vitro* in correcting alternative splicing (Igreja et al., 2016).

**Correction of class VI mutations**

Class VI mutants are less stable being associated with fast internalization and degradation of the protein. Therapies aimed at stabilizing the CFTR protein at the PM would be successful in correcting these mutants. These could include agents such as hepatocyte growth factor

(HGF) previously shown to increase PM stability of F508del-CFTR on its own or rescued by VX809 (Moniz et al., 2013).

**Correction of class VII mutations**

Patients with so-called unrescuable mutations are those in which there is no RNA production, the most straightforward approach would be to target alternative non-CFTR anion channels (Figure 1.7). Besides CFTR, other $Cl^-$ channels exist in epithelia, namely in the airways, being thus possible to envisage the development of strategies to modulate these other channels so as to compensate for the absence of functional CFTR. Since CFTR is a negative regulator of the epithelial $Na^+$ channel (ENaC), its absence from the apical membrane leads to enhanced ENaC activity in the surface airway of epithelial cells and thus excessive $Na^+$ absorption. Therefore, one pursued strategy for CF pharmacotherapy targets ENaC, specifically ENaC blocking by specific inhibitors such as amiloride, benzamil or phenamil and probably by PKC activation (Amaral and Kunzelmann, 2007). Stimulation of alternative $Cl^-$ channels, such as the calcium ($Ca^{2+}$)-activated $Cl^-$ channels (CaCCs), could be an alternative approach. Stimulation of the CF airway epithelial cells by luminal P2Y2 purinergic receptors with ATP or UTP was shown to restore $Cl^-$ secretion (Knowles et al., 1991). Anoctamin 1 (ANO1; also known as TMEM16A) is a CaCC present in many epithelial cells, namely in the airways and ANO6 (also known as TMEM16F) is an essential component of the outwardly rectifying $Cl^-$ channel. Both constitute good alternative channel candidates to compensate for the absence of functional CFTR.

The $Cl^-/HCO_3^-$ transporter SLC26A9 is associated with susceptibility to meconium ileus and diabetes (De Boeck and Amaral, 2016) and has been shown to function as an anion conductance channel itself in the apical membranes of HBE cells, as it contributes to transepithelial $Cl^-$ currents under basal and cAMP/protein kinase A–stimulated conditions, although its activity in HBE cells is stimulated by functional CFTR (Bertrand et al., 2009). SLC26A9-mediated $Cl^-$ secretion is essential for preventing airway obstruction in allergic airway disease (Anagnostopoulou et al., 2012).

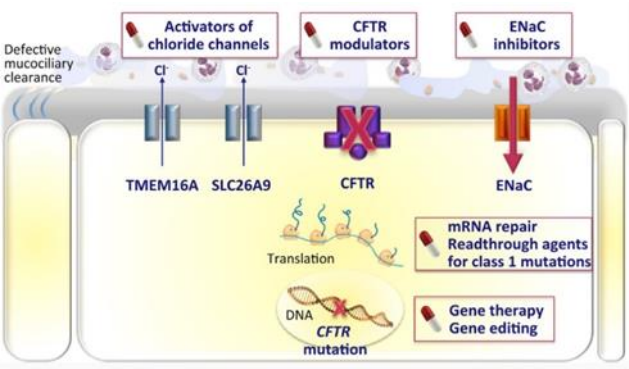



**Figure 1.7 - Gene therapy, gene editing and mRNA repair are strategies using oligonucleotides to target the DNA or the RNA in order to restore CFTR function**. Read-through agents for class 1 mutations and CFTR modulators are small pharmacological molecules aiming at restoring CFTR function. ENaC inhibitors and activators of non-CFTR chloride channels could bypass the dysfunctional CFTR and restore normal sodium absorption and chloride secretion. Several of these approaches could be combined [Adapted from Fajac and Wainwright, 2017].

# *5 Objectives of the present work*

The current work was aimed at understanding the mechanisms underlying CF-causing mutations that affect mRNA structure and/or stability and at testing novel therapeutic strategies aimed at correcting such basic defects. It comprises the following specific objectives:

1) To identify CFTR gene mutations in individuals with non-CF lung diseases, namely chronic obstructive pulmonary disease (COPD), asthma and disseminated bronchiectasis (DB);

2) To identify CFTR gene mutations in individuals with a suspicion of CF and to analyse CFTR expression directly in native tissues (respiratory cells and colon) from some of these patients, in order to identify alterations in the structure and levels of mRNA;

3) To identify key factors in the NMD pathway by automated microscopy screens using a cell model expressing a novel CFTR NMD-PTC/ read-through mini-gene reporter;

4) To screen for novel compounds suppressing PTC mutations by automated microscopy screens using the above cell-based model, as potential corrective therapies for CF.

# *II MATERIALS & METHODS*

# *1 Characterization of Subjects in Study*

## *1.1 Ethical*

These studies were part of a research project that was approved by the ethical committee of Santa Maria Hospital, Lisbon, and included the establishment a DNA/RNA bank. This bank, as well as the anonymised DNA bank, followed the international ethical guidelines, which included informed consent, confidentiality, anonymity of personal data, and an abandonment option in case of expressed will.

## *1.2 Subjects*

Data presented in Chapter 1 (section III. Results & Discussion), refers to the analysis of 187 blood samples from Portuguese individuals divided in two distinct groups. The first group comprehends samples from individuals with different respiratory diseases and distributes as follow: Asthma (n=54), COPD (n=51) and Bronchiectasis (n=31). The second group comprehends samples from healthy individuals, without any symptom of respiratory disease, used as control (n=51).

As for the data included in Chapter 2 (section III. Results & Discussion) we selected eight individuals with suspicion of CF (followed up at the Lisboa University CF Centre) that presented one or no CFTR mutation determined by DNA analysis. We also analysed non-CF subjects and p.F508del-homozygous patients, used as controls. Following informed consent, RNA samples were collected from CFTR-expressing tissues, including: *(i)* nasal cells obtained by nasal brushing; or *(ii)* colon epithelium obtained from rectal biopsies.

The mutation nomenclature used in this dissertation is in accordance with the HGVS guidelines using as reference the CFTR sequence Genbank NM_000492 (legacy name is often also given for better recognition of the mutation) and the CFTR exons are numbered from 1–27.

## *1.3 Study Design*

In Chapter 1 (section III. Results & Discussion) we intend to carry out a screening of 12 more frequent CFTR mutations in Portuguese patients with CF in individuals with chronic lung diseases, such as Diffuse Bronchiectasis, Bronchial Asthma and COPD. Detection of CF mutations will be performed by the ARMS, Tetra-ARMS and RFLP technique in our laboratory

to allow rapid detection in DNA or RNA samples or by sequencing. For the cases in which CFTR mutations are detected, the possible analysis of RNA samples and functional analysis of ionic transport mediated by CFTR in rectal biopsies are also foreseen.

This experimental work will be performed by Verónica Felício, in the laboratory of the Functional Genomics and Proteostase Unit, coordinated by Prof. Margarida Amaral and integrated in BioISI (BioSystems and Integrative Sciences Institute) located at the Faculty of Sciences of the University of Lisboa.

The laboratory studies will be carried out in collaboration with the Clinical University of Pneumology of the Faculty of Medicine of Lisboa (Dr. Carlos Lopes and Dr. Pilar Azevedo), where patients will be recruited (51 patients with smoking COPD, 54 patients with non-smoking bronchial asthma , 31 patients with diffuse bronchiectasis, and 51 healthy individuals). After approval of the study by the ethics committee of the Santa Maria Hospital, informed consent will be obtained for participation in the study and collection of the necessary materials. Patients will be clinically analyzed according to a specific protocol, which will include review of clinical records, clinical, analytical, imaging and respiratory function evaluation, in order to obtain a precise clinical characterization of the different pathologies studied here. At any time patients can leave the study without any consequence for the clinical follow-up of their disease. The rectal biopsies will be performed by Dr Ana Isabel Lopes (HSM).

This project will provide important data on the influence of CFTR mutations on respiratory diseases in Portuguese patients leading to a better perception of the epidemiology of these diseases. The results may also lead to the identification of patient groups that may benefit from the new therapeutic compounds currently under development to correct the basic CFTR protein defect in CF.

# *2 Analysis of Patients' Materials*

## *2.1 Rectal biopsies and Nasal epithelial cells*

Nasal epithelial cells were collected by gently brushing the inferior turbinate of both nostrils using a 2.5 or 3 mm interdental brush (Harris et al., 2004). Colon tissue was obtained by small superficial rectal forceps biopsies described in Sousa et al.,2012.

### *2.1.1 DNA and RNA Analysis*

#### *2.1.1.1 DNA Extraction*

Blood samples were collected in EDTA tubes and anonymously coded, before being stored at 4ºC. The rectal biopsies collected were immediately frozen in liquid nitrogen, and then transferred to –80◦C until DNA extraction.

For DNA extraction the Wizard® Genomic DNA Purification Kit Technical (Promega, USA) was used as follow:

Before DNA extraction the tubes containing the blood, samples were inverted 2-3 times. Cell Lysis Solution was added to the Blood samples in a ratio of 2:1, Cell Lysis Solution: Sample, in a sterile 15 ml centrifuge tube. Next, the two solutions were mixed by inverting the tube 5–6 times, followed by 10 min incubation at room temperature. After lysis of the red blood cells, samples were centrifuged at 2,000 × g for 10 min at room temperature. The supernatants were discarded without disturbing the visible white pellet. The white blood cells were lysed after incubation with Nuclei Lysis Solution. Protein Precipitation Solution was added to the nuclear lysate and samples were centrifuged at 2,000 × g for 10 min at room temperature. The supernatants were transfered to a 15 ml centrifuge tube containing 3 ml of isopropanol at room temperature. Afterwards, the solutions were mixed by inversion, until the white thread-like strands of DNA formed a visible mass. The DNA was pelleted by centrifugation at 2,000 × g for 1 min at room temperature. After discarding the supernatant, the DNA pellets were washed with 70% ethanol. After centrifugation at 2,000 × g for 1 min at room temperature, the supernatant was discarded, and the DNA pellets were air dried, 10 to 15 min. DNA Rehydration Solution was added according to the pellet size and incubated for 1 hour at 65°C.

For the rectal biopsies, 600µl of Nuclei Lysis Solution was added to 10–20 mg of fresh or thawed tissue in a 1.5 ml tube, followed by 30 min incubation at 65°C. Before proceeding with the DNA extraction, 3µl of RNase Solution was added to the nuclear lysate, and samples were mixed by inverting the tube 2–5 times, followed by a 30 min incubation at 37°C. DNA extraction was performed as above, starting at the Protein Precipitation Solution step.

#### *2.1.1.2 RNA Extraction*

RNA from Rectal Biopsies and Nasal cells was extracted using the Nucleo Spin RNAII Kit (Macherey-Nagel, Duren, Germany).

Before collecting the Nasal Epithelial Cells, a 1.5 ml microcentrifuge tube was prepared per sample containing 350 µl extraction Buffer RA1 supplemented with 1% (v/v) β-mercaptoethanol. After collecting the nasal cells, the nasal brush was immersed in the supplemented RA1 buffer, with the brush being passed through a cut p200 tip.

The frozen rectal biopsies samples were immersed in Buffer RA1 supplemented with 1% (v/v) β-mercaptoethanol and 30 mg of tissue were disrupted using a p1000 tip.

The cell lysates were cleared after filtering the samples in a NucleoSpin® Filter, and centrifuged for 1 min at 11,000 x g. To homogenize the flow-through 350 µl of 70% ethanol were added. The homogenized lysate was loaded in a NucleoSpin® RNA Column, followed by sample centrifugation at 11,000 x g for 30 sec. Subsequently, 350 µl Membrane Desalting Buffer (MDB) were added and samples were centrifuged at 11,000 x g for 1 min, to allow the membrane to dry. A DNase reaction mix was added directly to the membrane and incubated at room temperature for 15 min. Three washes were performed, one with Buffer RAW2 and the other two with Buffer RA3, all of them followed by centrifugation at 11,000 x g for 30 sec. The RNA samples were eluted in 60 µl RNase-free $H_2O$, and centrifuged at 11,000 x g for 1 min.

Every step was executed in a RNAse free environment to avoid RNA degradation. RNA extracted samples were quantified using the Nanodrop 200 spectrophotometer (Thermo Fisher Scientific, USA), and their purity was assessed by the A260/A280 ratio. Only the RNA samples with a A260/A280 ratio above 2.0 were used for following analysis.

#### *2.1.1.3 cDNA synthesis*

cDNA was generated from 1 µg of RNA sample. On ice, in a sterile, nuclease-free microcentrifuge tube, a reaction mixture was prepared containing 1 µg of RNA, 300 ng of random hexamer primers, 0.5 mM of dNTP mix (10 mM each) and Nuclease-free water up to 16 µl. Denaturation of secondary structures of the RNA molecule was performed by incubating the reaction at 65°C for 5 min, followed by a brief incubation of the reaction on ice. To each RNA sample mix 4 µL of a Reverse transcription mix was added, containing 1x reaction Buffer (10x – 2 µL), 40 Units of Ribonuclease Inhibitor (1 µL), and 200 Units of M-MuLV Reverse Transcriptase (1 µL). The reactions were briefly centrifuged and incubated at 25°C for 10 min, followed by 50 min incubation at 50°C. Heat inactivation of the reverse transcriptase was performed by incubating the samples at 85°C for 5 min. The cDNA products were stored at -20°C. All the reagents used for cDNA synthesis were from NZYtech (Lisbon, Portugal).

### *2.1.2 PCR reactions to identify CFTR mutations*

To detect CFTR mutations both the DNA and cDNA samples were analysed as follows: a pre-reaction mix was prepared containing 10 pmol/µL of forward and reverse primer mix, 1 Unit of Taq Polymerase (NZYtech), 1x of Polymerase Buffer solution (10x, NZYtech), 3 µL MgCl2 (50 mM), 25 µM dNTPs mix, and water up to a final volume of 49 or 45 µL, for DNA and cDNA reactions, respectively.

All these steps were performed in a DNA/RNA free environment to avoid contamination of the reagents, and samples and reaction mix were kept on ice to avoid unspecific amplification. The PCR reactions were incubated in a T-Professional thermo-cycler (Biometra), using the heating-cooling cycles described in Table 2.2. After the thermo-cycling, each reaction was analysed by an agarose gel electrophoresis. Samples were separated in a 2% (m/v) agarose (Lonza) gel in TAE buffer with 1 Unit of RedSafe (Chembio), for 40 min at 120 Volts. After electrophoresis the PCR products were visualized by exposing the gel to UV light in a dark container coupled to a Kodak photographic machine (Kodak EDAS 290 Electrophoresis Documentation System). Image analysis was performed using the Kodak EDAS 290 Electrophoresis Documentation System software.

**Table 2.2 -** PCR program used for the detection of CFTR and β-actin transcripts by RT-PCR

| Cycle Step | Temperature (ºC) | Time (min) | Number of cycles |
|---|---|---|---|
| **Pré-denaturation** | 94 | 2 | 1 |
| **Denaturation** | 94 | 1 | |
| **Annealing** | (according to the primers) | 1 | 35 |
| **Extension** | 72 | 2 | |
| **Final extension** | 72 | 5 | 1 |

PCR analysis of the cDNA samples was performed using nine different primer combinations, in order to cover the complete CFTR coding region (Table 2.4). For each PCR reaction 5µl of cDNA sample were used.

Different mutations were identified through three different techniques: Amplification Refractory Mutation System (ARMS), Tetra-ARMS and Restriction Fragment Length Polymorphisms (RFLP). For each PCR reaction 50 ng of DNA sample were used.

### *2.1.2.1 ARMS Technique*

Identification of F508del, G542X and N1303K mutations was carried out using the ARMS PCR technique. Different primer combinations were used, as shown in Table 2.3, with different PCR fragment sizes expected. For each mutation two separate PCR reactions were done with primers set to detect the wildtype or mutant allele.

**Table 2.3 -** PCR primers and condition for ARMS-PCR

| Mutation | Sequence 5' to 3' | Primer | Annealing (°C) | Size (bp) |
|---|---|---|---|---|
| **F508del** | F: GGC ACC ATT AAA GAA AAT ATC ATC TT | CF10N | | |
| | F: GGC ACC ATT AAA GAA AAT ATC ATT GG | CF10M | 58 | 222 |
| | R: CAT TCA CAG TAG CTT ACC CA | CF10ZL | | |
| **G542X** | F: CAA CTG TGG TTA AAG CAA TAG TGT | CF11ZR | | |
| | R: GTG TGA TTC CAC CTT CTC C | G542X_N | 56 | 174 |
| | R: GTG TGA TTC CAC CTT CTC A | G542X_M | | |
| **N1303K** | F: GAT CAC TCC ACT GTT CAT AGG GAT CCA AG | N1303K_N | | |
| | F: GAT CAC TCC ACT GTT CAT AGG GAT CCA AC | N1303K_M | 50 | 328 |
| | R: CTC AAT TTC TTT ATT CTA AAG ACA TTG G | N1303K_C | | |

**Table 2.4 -** Primer sets used for the nine PCR reactions and F508del-ARMS, and respective sequence, position in CFTR mRNA and expected PCR product size

| PCR reaction | cDNA Region | Primer name | Sequence | Temperature Annealing | cDNA Position | Fragment Length (bp) |
|---|---|---|---|---|---|---|
| 1 | 5'NC- | 5'NC | 5'- GCATTAGGAGCTTGAGCCCA - 3' | 62ºC | 77-96 | 743 |
| | Ex6a | AC1L | 5' – GAAACCAAGTCCACAGAAGGC – 3' | 64ºC | 819-799 | |
| 2 | Ex5 | CF Ex5.F | 5' – CTCCTTTCCAACAACCTGAAC – 3 | 62ºC | 679-699 | 751 |
| | Ex9 | CF9AS | 5' – GTACTGAAGAAGAGGCTGTC – 3' | 60ºC | 1429-1408 | |
| 3 | Ex8 | B3R | 5' – AATGTAACAGCCTTCTGGGAG – 3' | 52ºC | 1318-1338 | 431 |
| | Ex11 | Ex11L | 5' – TATATTGTCTTTCTCTGCAAACTTGG – 3' | 70ºC | 1724-1749 | |
| 4 (see also below) | EX10 | CF10N | 5' – GGCACCATTAAAGAAAATATCATCTT – 3' | 58ºC | 1630-1655 | 207 |
| | EX10 | CF10M | 5' – GGCACCATTAAAGAAAATATCATTGG – 3' | 58ºC | 1630-1655 | |
| 5 | Ex11 | C1R | 5' – GTGGAGGTCAACGAGCAAGA – 3' | 62ºC | 1778-1797 | 904 |
| | Ex14a | CF14a.R | 5' – AAGGAATGTGGTCCATGTAG – 3' | 58ºC | 2663-2682 | |
| 6 | Ex13 | D2R | 5' – GCAAACTTGACTGAACTGGA – 3' | 58ºC | 2530-2549 | 730 |
| | Ex17a | D2L | 5' – GTTGTTTGAGTTGCTGTGAG – 3' | 58ºC | 3240-3259 | |
| 7 | Ex16 | E2R | 5' – GATGACCTTCTGCCTCTTAC – 3' | 54ºC | 3082-3101 | 700 |
| | Ex19 | E2L | 5' – CTGTGAGATCTTTGACAGTC – 3' | 58ºC | 3762-3781 | |
| 8 | Ex18 | Ex18.F | 5' - AACTCCAGCATAGATGTGG – 3' | 56ºC | 3574-3592 | 829 |
| | Ex22 | Ex22.R | 5' – GCCAAGCACATCAACTGCA – 3' | 58ºC | 4184-4202 | |
| 9 | Ex21 | Ex21.F | 5' – CTATGAACAGTGGAGTGATC – 3' | 58ºC | 4050-4069 | 549 |
| | Ex24 | F2L | 5' – CATGTCAACATTTATGCTGC – 3' | 56ºC | 4579-4598 | |
| **Specific primers for detection of the F508del mutation using the ARMS technique** | | | | | | |
| 4. ARMS for normal sequence at F508 | EX10 | CF10N | 5' – GGCACCATTAAAGAAAATATCATCTT – 3' | 58ºC | 1630-1655 | 207 |
| 4. ARMS for mutant F508del sequence | EX10 | CF10M | 5' – GGCACCATTAAAGAAAATATCATTGG – 3' | 58ºC | 1630-1655 | |
| 4. Common ARMS reverse primer | Ex12 | 561AL | 5' – CAAATCAGCATCTTTGTATACTG – 3' | 62ºC | 1814-1836 | |

### *2.1.2.2 Tetra – ARMS Technique*

The DNA samples were further analysed for an additional 7 CFTR mutations, as show in Table 2.5. For each sample, PCR reactions were set, comprising the wildtype allele, mutant allele and control primers.

**Table 2.5 -** PCR and condition for tetra-primer AMRS-PCR.

| Mutation | Primer Sequence 5' to 3' | $T_m$ | Annealing (°C) | Amplicon size |
|---|---|---|---|---|
| **A561E C>A** | **Forward inner primer (A allele):** AAT TTC CAT TTT CTT TTT AGG GA | 51°C | 54 | 327 bp (A allele) |
| | **Reverse inner primer (C allele):** AAT ATA CAA AGA TGC TGA TTT GTA TTG | 51°C | | 148 bp (C allele) |
| | **Forward outer primer (5' - 3'):** GTG AAT CGA TGT GGT GAC CA | 60°C | | 426 bp (from two outer primers) |
| | **Reverse outer primer (5' - 3'):** CTG GTT TAG CAT GAG GGG GT | 62°C | | |
| **R334W C>T** | **Forward inner primer (T allele):** CAC TAA TCA AAG GAA TCA TCC GCT | 56°C | 52 | 137 bp (T allele) |
| | **Reverse inner primer (C allele):** GAA TGA GAT GGT GGT GAA TAT TTT ACG | 56°C | | 215 bp (C allele) |
| | **Forward outer primer (5' - 3'):** GGA AGG CAG CCT ATG TGA GA | 60°C | | 410 bp (from two outer primers) |
| | **Reverse outer primer (5' - 3'):** GCA AAG TTC ATT AGA ACT GAT C | 62°C | | |
| **R1066C C>T** | **Forward inner primer (C allele):** CAA GCT TAA AAG GAC TAT GGA CAC GTC | 58°C | 50 | 278 bp (C allele) |
| | **Reverse inner primer (T allele):** CTG CCG TCC GAA GGC CCA | 62°C | | 228 bp (T allele) |
| | **Forward outer primer (5' - 3'):** TTC AAA GAA TGG CAC CAG TGT | 60°C | | 462 bp (from two outer primers) |
| | **Reverse outer primer (5' - 3'):** ATA ACC TAT AGA ATG CAG CA | 54°C | | |
| **G576A G>C** | **Forward inner primer (G allele):** GAT TTG TAT TTA TTA GAC TCT CCT TTA GG | 52°C | 52 | 288 bp (G allele) |
| | **Reverse inner primer (C allele):** TTT TTC TGT TAA AAC ATC TAG GTT TG | 51°C | | 192 bp (C allele) |
| | **Forward outer primer (5' - 3'):** GTG AAT CGA TGT GGT GAC CA | 60°C | | 426 bp (from two outer primers) |
| | **Reverse outer primer (5' - 3'):** CTG GTT TAG CAT GAG GGG GT | 62°C | | |
| **P205S C>T** | **Forward inner primer (C allele):** GCA CAT TTC GTG TGG ATC GAT C | 57°C | 52 | 96 bp (C allele) |
| | **Reverse inner primer (T allele):** AGG AGT GCC ACT TGC AAC GA | 56°C | | 185 bp (T allele) |
| | **Forward outer primer (5' - 3'):** TTA GTG TGC TCA GAA CCA CG | 60°C | | 240 bp (from two outer primers) |
| | **Reverse outer primer (5' - 3'):** GAA ACC AAG TCC ACA GAA GGC | 64°C | | |
| **I148N T>A** | **Forward inner primer (T allele):** CCA TTT TTG GCC TTC ATC ATA T | 53°C | 52 | 156 bp (T allele) |
| | **Reverse inner primer (A allele):** AAA CAT AGC TAT TCT CAT CTG CAT TCT AT | 54°C | | 270 bp (A allele) |
| | **Forward outer primer (5' - 3'):** TGT GTT GAA ATT CTC AGG GT | 56°C | | 376 bp (from two outer primers) |
| | **Reverse outer primer (5' - 3'):** TTG TAC CAG CTC ACT ACC T | 56°C | | |
| **711+1 G>T** | **Forward inner primer (G allele):** AAC CTG AAC AAA TTT GAT GGA G | 58ºC | 54 | 89 bp (G allele) |
| | **Reverse inner primer (T allele):** CCT AAA AGA TTA AAT CAA TAG GTA CAC AA | 58ºC | | 356 bp (T allele) |
| | **Forward outer primer (5' - 3'):** TTG TAC CAG CTC ACT ACC T | 56ºC | | 395 bp (from two outer primers) |
| | **Reverse outer primer (5' - 3'):** AAC TCC GCC TTT CCA GTT GT | 60ºC | | |

### *2.1.2.3 RFLP Technique*

The G85E and 3272-26 A>G mutations were analysed using the Restriction fragment length polymorphisms (RFLP) PCR method. The PCR reactions were performed as described above (Materials and Methods, section 2.1.2). For analysis 10 µl of the amplified product were used, after verification of expected PCR product in an agarose gel: 318 bp for G85E and 267 bp for 3272-26 A>G. Each PCR product was incubated with 1 U of HinfI and BtsI restriction enzymes (New England BioLabs, UK) for 1h at 37°C and 55°C, respectively. Digested products were analysed on a 2% agarose gel.

For the G85E Wildtype allele three fragments are expected with 55, 91, and 172 bp, while only two fragments are expected for the mutant allele, a 91 and a 227 bp fragments. For the 3272-26 A>G Wildtype allele, only one fragment 318 bp is expected, and two fragments of 134 and 137 bp are expected in a mutant allele, as shown in Table 2.6.

**Table 2.6 -** PCR primers and condition for RFLP technique

| Mutation | Sequence 5' to 3' | Annealing (°C) | Enzyme |
|---|---|---|---|
| **G85E** | F: AGA TAT CTG GCT GAG TGT TT<br>R: TGT GAT ACA TAA TGA ATG TAC | 56 | **HinfI** |
| **3272-26 A>G** | F: TTC AAA GAA TGG CAC CAG TGT<br>R: TAA ATT CAG AGC TTT GTG GAA C | 56 | **BtsI** |

## *2.1.3 Quantitative Analysis of CFTR Transcripts*

The analysis of the transcripts is important to understand mutation effect on the RNA. One technique used to measure transcripts levels is quantitative real-time PCR (qRT-PCR) which allows us to monitor the progress of a PCR reaction as it occurs in real time, and this combination of amplification and detection not only increases the accuracy of PCR product quantification but also eliminates the need for post-PCR manipulations. Two main experimental procedures can be followed, based on the use of either fluorescently labelled gene-specific probes (e.g., Taqman) or dyes that intercalate double-stranded DNA during product extension (e.g., SYBR Green). We used the SYBR Green method for relative quantification of CFTR transcripts from a group of experimental vs. control samples, using a housekeeping gene as an internal control. This technique can be used to measure changes in CFTR mRNA abundance following experimental manipulations in cell cultures or to confirm altered expression of the CFTR gene in native tissues from patients vs. controls.

qRT-PCR was performed in a 96-well plate, with experimental and control samples. Each cDNA sample was analysed in triplicate and reactions were prepared as previously described (Ramalho et al., 2011). For G542X/G576A-R668C patient analysis, primers were designed using the Primer3 software (http://frodo.wi.mit.edu/). The primers were designed to identify the

full-length exon 13 (5′- TTAGCAAGTATACAAAGATGCTGATTT-3′), and the exclusion of exon 13 promoted by the G576A mutation (5′- AATTTCTTTAGCAAGCTGTGTCTG-3′) products of the G576A-R668C allele (common reverse primer: 5′-GCTGCTACCTTCATGCAAAA-3′).

A set of G542X specific primers (G542X reverse primer: 5′-GTGTGATTCCACCTTCTCC-3′; wt reverse primer: 5′- GTGTGATTCCACCTTCTCA-3′; common Forward primer: 5′-CAACTGTGGTTAAAGCAATAGTGT-3′) was also used to compare abundance of transcripts from the two separate alleles. Quantification of each cDNA was performed using the Evagreen system (Biorad) in a BioRad Cx96. After normalization to β-actin the samples were analysed according to the ddCT method, and fold change between full length and alternatively spliced products was calculated using the following formula: $FC = 2^{(-ddCT)}$.

### *2.1.4 Statistic Analysis*

Statistical analyses were performed with SPSS software (v. 19; SPSS Inc, Chicago, IL, USA) and a p value $< 0.05$ was accepted to indicate a statistically significant difference.

Data presented in Chapter 1 (section III. Results & Discussion) included 187 Portuguese participants, 136 with respiratory disease and 51 healthy controls. Results are presented stratified by individuals, gender, age and as totals.

Associations between different respiratory diseases and a control group were assessed using Fisher's exact test, that can assess for independence between two variables when the comparing groups are independent and not correlated. The Fisher's exact test runs an exact procedure especially for small-sized samples.

### *2.1.5 Ussing Chambers*

To assess the intestinal CFTR-mediated $Cl^-$ secretion in the patient, rectal biopsy specimens were mounted and analysed in micro-Ussing chambers perfused in Ringer solution under open-circuit conditions, as previously described (Sousa et al., 2012). Briefly, Amiloride (Amil) was added to block electrogenic sodium ($Na^+$) absorption through ENaC and indomethacin was applied for 40-60 min to inhibit endogenous cAMP formation through prostaglandins (Sousa et al., 2012). As reported, cAMP-dependent and cholinergic $Cl^-$ secretion in human rectal tissues relies on functional CFTR. Thus, we used 3-isobutyl-1-methylxantine (IBMX basolateral) and forskolin (Fsk basolateral) to activate cAMP-dependent $Cl^-$ secretion and carbachol (CCH, basolateral) for cholinergic co-activation (Mall et al., 2000). The percentage of CFTR function was calculated for maximal CFTR activation and normalized to the corresponding mean value for the reference non-CF control group.

Values were obtained from quantification of the equivalent short-circuit (Ieq-sc-CCH/IBMX/Fsk) from the original tracings.

## *3 CFTR NMD-PTC mini-gene model*

A mCherry-wt-intron14/15/16-CFTR-eGFP (hereinafter referred to as mCherry-wt-eGFP) mini-gene and a mCherry-G542X-intron14/15/16-CFTR-eGFP (hereinafter referred to as mCherry-GX-eGFP) mini-gene model were constructed. This mini-gene model allowed us to identify factors involved in the NMD pathway and to detect read-trough compounds. The fused mCherry at the N-terminus of the mini-gene allowed us to detect NMD inhibition by detection of red fluorescence using fluorescence microscopy and the eGFP (enhanced green fluorescent protein) fused at the C-terminus allowed us to detect read-trough compounds by detection of green fluorescence as illustrated in Figure 2.8.

The CFTR mini-genes were previously produced using a "sticky feet PCR" strategy to introduce the CFTR introns - IVS14, IVS15, and IVS16 - consecutively into the pCDNA5/FTR/CFTR mammalian expression vector carrying the complete wt-CFTR (Masvidal et al., 2014). Thus, the "triple mini-gene" consists of a full 4.5 kb CFTR cDNA plus full length of the intron 14 (2274 bp), inserted between cDNA exons 14/15; a short length version of the intron 15 618 bp– IVS15art (containing the first 5' 259 bp and the last 3' 350 bp of the normal IVS15) inserted between exons 15/16; and the full length intron 16 (670 bp) inserted between exons 16/17. Herein, cloning of the mCherry coding sequence and eGFP coding sequence, upstream and downstream of the CFTR mini-gene, respectively, was done using the same cloning method as for the IVS's.

To induce NMD the nonsense mutation, in this study the G542X mutation, must be recognized as a PTC and this only occurs if an intron is downstream of the nonsense mutation (Da Costa et al., 2017).

NMD and PTC read-through studies described in the third and fourth chapters were performed using this triple-colour CFTR mini-gene model.

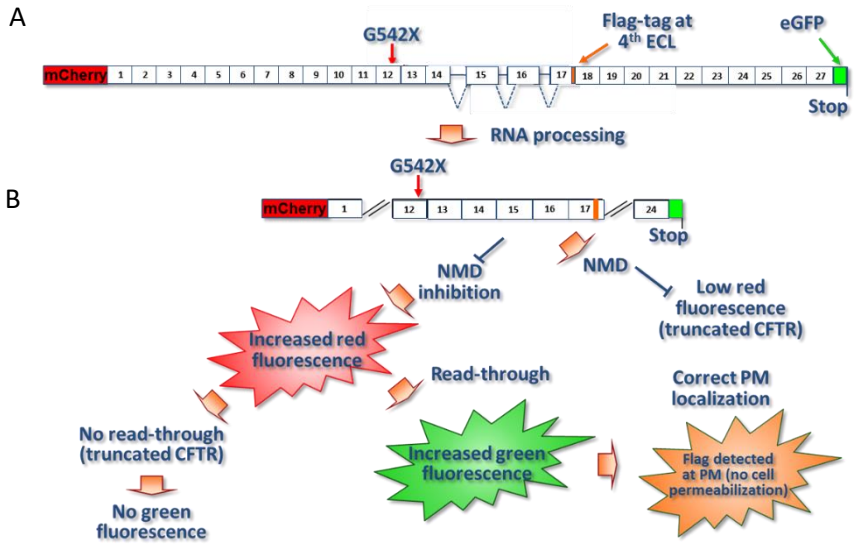



***Figure 2.8- Schematic representation of the three-colour CFTR NMD reporter. (A)*** *The CFTR mini-gene construct used here (inserted into the pCDNA5-CFTR vector) with the G542X mutation (from herein called mCherry-wt-eGFP) contains IVS14, 15 and 16, mCherry (red) at the N-terminus, and eGFP (green) at the C-terminus as well as the Flag tag at exon 17 (for plasma membrane localization).* ***(B)*** *Scheme of NMD fluorescence assay: when NMD takes place, no green fluorescence (right) occurs and only basal red fluorescence is detected and if the transcript escapes NMD, red fluorescence will be increased (middle), whereas if PTC read-through also takes place, green fluorescence will be detected, along with an increase in red fluorescence (left).*

## *3.1 Production of the CFTR G542X mutation by site-directed mutagenesis*

The nonsense mutation G542X in exon 12 (the most common nonsense mutation in CF patients) was introduced in the mCherry-wt-eGFP using the KOD Hot Start Kit (Invitrogen, USA), with a complementary pair of mutagenic primers described in Table 2.7.

After site-directed mutagenesis plasmids were isolated (using the kit NZYMiniprep from NZYtech (Lisbon, Portugal)) and sequenced by automatic DNA sequencing to confirm the introduction of the mutation.

**Table 2.7 -** Primers and PCR conditions

| Primer | Direction | Sequence (5'→ 3') | Temperature (ºC) | Sequence Position |
|---|---|---|---|---|
| G542X.F | Forward | GAC AAT ATA GTT CTT TGA GAA GGT GGA ATC | 48 | Exon 11 |
| G542X.R | Reverse | GAT TCC ACC TTC TCA AAG AAC TAT ATT GTC | 48 | Exon 11 |

## *3.2 Production of stable cell lines and cell culture maintenance*

Stable cells lines carrying the wt and the G542X mini-genes were generated using the Human embryonic kidney 293 Flp-In (HEK293 Flp-In) cells (Invitrogen, Carlsbad, CA), hereinafter referred to as HEK Flp-In, a cell line derived from HEK293 cells and containing both Flp-In and T-REx systems.

The creation of each of the isogenic stable cells is made possible by the Flp-In system (Life Technologies). In this study the cells used were the HEK Flp-In that already contain the flippase recombination target (FRT) site linked to zeocin resistance gene in their genome. By co-transfecting these cells with both the flippase (Flp) recombinase plasmid, pOG44, and an expression plasmid containing the gene of interest flanked by FRT site in both end, and the hygromycin resistance gene. The presence of the Flp recombinase induces the integration of the fragment in the plasmid flanked by the FRT site in recombinase target site present in the cells genome leading to the stable expression of the gene of interest, loss of the zeocin resistance and acquisition of resistance to Hygromycin B. Using this system only a single copy of the fragment of the plasmid of interest is integrated per cell, originating a cell population with isogenic expression of the gene of interest, after selection with Hygromycin B.

Parental HEK Flp-In cells were cultured in Eagle Minimal Essential Medium supplemented with L-glutamine (EMEM, Lonza) supplemented with 10% (v/v) foetal bovine serum (FBS, Gibco) and supplemented with 200 µg of zeocin (Invitrogen).

HEK Flp-In cells stable cell lines were generated by transfection with the pOG44 plasmid and each mini-gene, simultaneously, in a proportion of 3:1 for a total 2 µg of DNA, using the Lipofectamine 2000 Transfection Reagent (Invitrogen # 12566014). Stable cell lines were selected 48h after transfection with 200 µg/ml of Hygromycin B (Life Technologies). supplemented medium. Cultures were maintained at 37ºC in a humidified atmosphere of 5% (v/v) $CO_2$.

# 4 HTS – NMD and PTC read-through screens

## 4.1 Screen Assay for NMD and PTC read-through

HEK Flp-In mCherry-wt-eGFP and mCherry-GX-eGFP cells were grown to confluence and split 24 h before seeding. For the shRNA screen $1x10^3$ cells/well were seeded in 384 well plates. For the PTC read-through compound assay and for siRNA confirmation, $1x10^3$ cells/well were seeded in 96 well plates. Cells were seeded using a MulitdropTM Combi peristaltic dispenser (Thermo Scientific #5840300).

### 4.1.1 NMD Screen of the RNA-related shRNA library

A subset of the shRNA library from the RNAi Consortium (TRC), Broad Institute (kindly provided by Dr L. Moita) Oberdoerffer et al., 2009; Mishra et al., 2010) was used to generate loss-of-function cellular phenotypes for a collection of genes with a known or predicted involvement in transcript processing. In this screen 425 different genes were evaluated, with an average of 5 shRNAs per gene. shRNA information is described in Table S 4.3 and the target sequences are provided in The RNAi Consortium webpage (The RNAi Consortium, 2017).

The cell lines stably expressing the wt and mutant mini-genes were transduced using 3 µl of non-concentrated shRNA lentiviral supernatant, and 8 mg ml $^{-1}$ polybrene was added per well. The plates were spun at 2200 rpm for 90 min at 37°C. After spin, cell media was replaced with fresh media. Forty-eight hours post infection, staining was performed (refer to section 4.2).

#### *4.1.1.1 Confirmation of selected hits from the shRNA screen by additional siRNAs*

Candidate genes related to NMD were defined as a ‘hit’ if two or more shRNAs increased mCherry expression by standard deviation from the mean of the controls (cells transduced with lentivirus encoding a scrambled-shRNA). Hit candidate genes were further validated in a confirmation screen, using a siRNA library targeting the candidates.

96-well plates were coated with customized siRNAs (Dharmacon™ siRNA Libraries) for solid-phase reverse transfection. The protocol used is adapted from a previously reported (Erfle et al., 2007). An aqueous 0.2% (w/v) gelatine solution was prepared and filtered with

0.45µM pore size filter, and a 0.4M glucose solution was prepared in Opti-MEM (Gibco #51985). The transfection mix was prepared by mixing 277 µL of the sucrose/Opti-MEM solution, 323 µL of Lipofectamine® 2000 (Invitrogen # 12566014) and 161.5 µL doubly distilled water. This transfection mix was distributed into a 96- conic well plate (6 µL/well, "Plate A"). In parallel, 10 µL of 0.2% gelatine solution was distributed into another 96-conic well plate (30 µL/well, "Plate B"). In a new 96-well plate, 3.5 µL of the transfection mix ("Plate A") was mixed with 2.5 µL of a 20 µM siRNA solution per well ("Plate C"). After 20 min incubation, 3.5 µL of the 0.2% gelatine solution (from "Plate B") was added. From each well of the "Plate C" 8.5 µL of the transfection + siRNA + gelatine was diluted 50-fold with doubly distilled water. Finally, 50 µL of each well were transferred to a 96-well imaging plate, lyophilized and stored in an anhydrous atmosphere before cell seeding. HEK Flp-In cells were seeded into the 96-well imaging plate at a cell density of 2 x $10^4$ cells/well. After 48 hours of incubation, staining was performed (refer to section 4.2).

### *4.1.2 Screen of compounds for read-through of the G542X mutation*

As described above, for the read-through compounds screen 96 well plates were coated with poly-L-lysine solution and 1x$10^3$ cells/well were seeded on top. A library of 24 chemically synthesized compounds were tested, synthesised and kindly provided by C. Moiteiro lab (FCUL, Lisboa, Portugal). After 24h, the compounds were added to the cells at three different concentrations: 5 µM, 1 µM, and 0.5 µM. The cells are incubated for 48 h. Each assay plate included several controls: the vehicle (DMSO, 0.5%), three positive controls: 1 µM SMG1 (NMD Inhibitor; kindly provided by Martin Mense), 600 µM geneticin (G418; Sigma-Aldrich, #A-1720), and 10 µM PTC124. After 48h of incubation with the compounds, the cells were stained (refer to section 4.2).

## *4.2 Staining*

Cell culture medium was removed, and cells were washed with ice cold PBS+/+ (PBS supplemented with 0.7mM $CaCl_2$ and 1.1mM $MgCl_2$) using a manual 96 channel pipette liquidator (LiquidatorTM 96, Mettler Toledo #17010335). The cells were then incubated with 3% (w/v) paraformaldehyde for 20 min at 4ºC and transferred to room temperature for the remaining staining procedure. Cells were then washed three times with PBS and incubated 1 h with a Hoechst 33342 solution (200 ng/mL, Sigma #B2261). The cells were washed 3 times with PBS. Before imaging the cells were kept overnight with PBS at 4ºC.

## *4.3 Image Acquisition*

Cell imaging was performed with an automated Leica DMI 6000B widefield epifluorescence microscope equipped with a metal halide light source (EL6000), a 16 bit 2048 x 2048 pixel resolution CMOS camera (Hamamatsu Orca-Flash4.0) and a 10x HC PL APO objective with a numerical aperture of 0.4. Three fluorescence channels were acquired: Hoechst (cell nuclei), mCherry (N-terminal CFTR) and GFP (C-terminal CFTR). The Hoechst channel was used for contrast-based autofocus. Each well was imaged at 5 non-overlapping fields, to obtain a representative coverage of the cell population.

## *4.4 Image Analysis*

Automated high content image analysis was performed with open source software tools: image quantification was performed using CellProfiler [PMID: 21349861] and statistical analysis was performed using shinyHTM (https://github.com/hmbotelho/shinyHTM), an application based in R (Team, 2017). Using CellProfiler, each cell was segmented by thresholding on the nuclear staining and allowing for a 1 pixel dilation (estimating the cytoplasm location). Cells touching the image border, containing too large or too small nuclei, having elongated nuclei (eccentricity > 0.8) or having more than ~1% of their pixels saturated were not considered for the analysis. The morphological filters aim at disregarding damaged, dead or mitotic cells, which do not represent a physiological epithelium. After background correction using the dark frame / flat field method, mCherry and GFP fluorescence intensities were measured in every cell. In this system, the integrated mCherry fluorescence signal in each cell is proportional to the extent of NMD inhibition, which is the amount of CFTR transcripts which escape NMD and produce a truncated protein:

$$NMD\ Inhibition \propto \text{mCherry Integrated Fluorescence}$$

The efficiency of PTC read-through in each cell is proportional to the eGFP fluorescence and represents the full-length CFTR:

$$PTC\ Readthrough\ Efficiency \propto eGFP\ Integrated\ Fluorescence$$

The integrated fluorescence is represented according with the above formula:

$$A = \sum_{i=1}^{n} I_{pixel}$$

$$B = Mediana\ (A), image\ X$$

$$Integrated\ fluorescence = B$$

Each image was characterized by the median value of “NMD inhibition” and “PTC read-through efficiency”. After imaging at least 4 image fields in triplicate, the shinyHTM software was used to perform quality control (QC) and statistical analysis. The QC aimed at excluding out of focus images and images with high background fluorescence. After averaging the NMD Inhibition and PTC Read-through efficiency for all images relating to the same shRNA/siRNA/compound treatment passing QC, the effect of each shRNA/siRNA and compound towards NMD Inhibition/read-through efficiency was expressed as a deviation score versus the appropriate negative control:

$$NMD\ Score\ = \frac{\text{NMD Inhibition}_i - \text{NMD Inhibition}_{NegCtrl}}{SEM_{\text{NMD Inhibition},Neg\ Ctrl}}$$

$$PTC\ Readthrough\ Score\ = \frac{PTC\ Readthrough_i - PTC\ Readthrough_{NegCtrl}}{SEM_{PTC\ Readthrough,NegCtrl}}$$

The negative control (NegCtrl) for compound-based assays is the vehicle (DMSO) and for the shRNA or siRNA assays is a non-targeting “Scrambled” version of the nucleic acid. In the formula, *NMD Inhibition* and *PTC read-through* are the median NMD inhibition or PTC read-through values obtained in individual images; $SEM_{XXX,Neg\_control}$ is the standard error of the mean for median ‘NMD Inhibition‘ or ‘PTC read-through‘ recorded upon ‘Scrambled’ treatment. We consider the effects those whose magnitude is larger than twice the negative control SEM to be significant (|Deviation Score| > 2).

# *III RESULTS & DISCUSSION*

# *Chapter 1 - Incidence of mutations in the CFTR gene in Portuguese individuals with respiratory diseases.*

## 1. Abstract

**Background:** It is estimated that 4% of the Caucasian population are carriers of *CFTR* gene mutations. However, there are studies suggesting that there is a higher incidence of *CFTR* mutations in individuals with several other respiratory diseases such as Disseminated Bronchiectasis (BD), Bronchopulmonary Aspergillosis, Sinusitis, Chronic Obstructive Pulmonary Disease (COPD) or Bronchial Asthma (BA). While in BD, several studies have shown an increased incidence of *CFTR* gene mutations in comparison to control groups, in the case of BA, previous studies have yielded controversial results.

**Objective:** Our goal is to screen individuals with chronic non-CF lung diseases, namely DB, BA and COPD for the presence of the 12 most frequent CFTR mutations found in Portuguese patients with CF.

**Methods**: A total of 136 individuals with non-CF respiratory diseases (asthma, n=54; COPD, n=51) and DB, n=31) referred by Hospital Lisboa-Norte and 51 controls were assessed for the presence of CFTR mutations, using three different techniques: 1) ARMS (amplification refractory mutation system) technique; 2) Tetra- ARMS; 3) and Restriction Fragment Length Polymorphisms (RFLP), optimized in our laboratory to allow rapid detection in DNA or RNA samples followed by sequencing. In cases where CFTR mutations were detected, we further analyzed RNA samples and performed CFTR-mediated ion transport measurements in rectal biopsies.

**Results and Discussion**: Our data show that 7 (out of 136) patients with non-CF respiratory diseases displayed CFTR mutations in one allele vs none in the control group. We analysed the correlation of CFTR gene mutations with each of the three respiratory diseases investigated. Regarding BA our data did not show an increase in mutation frequency. For DB, we found a non-statistically significant increase in the frequency of *CFTR* gene mutations, in agreement with previously published independent reports. Regarding COPD however, we found a statistically significant increase in *CFTR* gene mutations.

**Conclusion:** This work confirmed the importance of *CFTR* gene mutations on non-CF respiratory diseases in Portuguese patients leading to a better perception of the epidemiology and etiology of these diseases. These results also lead to the identification of groups of patients who may benefit from the new therapeutic compounds currently under development to correct the basic CFTR protein defect in CF.

## 2. Introduction

Cystic Fibrosis (CF) (OMIM #219700) is the most common autosomal recessive disease in the Caucasian population, with an estimated incidence of 1: 3,500-4,000 births in Portugal (WHO, 2004). The most relevant clinical aspects of CF classical manifestation are: progressive lung disease brought about by dehydration of the epithelial airway surface liquid (ASL) and a failure in MCC, leading to recurrent bacterial infections and a chronic inflammatory response; pancreatic dysfunction; male infertility and elevated concentrations of chloride ($Cl^-$) in sweat. However, although the classic CF manifestation is well defined, its pathophysiology is not yet fully understood. This disease has pleotropic manifestations presenting a great variability of clinical phenotypes (Bergougnoux et al., 2015; Clarke et al., 2015; De Boeck and Amaral, 2016).

CF is caused by mutations in the Cystic Fibrosis Transmembrane Conductance Regulator gene (*CFTR*, MIM*602421). Currently, more than 2,000 variants have been identified for this gene (Cystic Fibrosis Mutation Database, 2017), of which the p.Phe508del mutation is the most frequent, with a 70% worldwide incidence of CF chromosomes (in ~ 90% of CF patients). About 1 in 30 Caucasians, 1 in 65 Africans and 1 in 90 Asians carry a mutation in the *CFTR* gene (Nielsen et al., 2016). Due to the high number of sequence variants found in the *CFTR* gene, their consequences in terms of disease manifestation are not yet fully understood. Moreover, there are studies that suggest the association of CFTR mutations with several other respiratory diseases (Bombieri et al., 1998) such as diffuse Bronchiectasis (BD) (Pignatti et al., 1995), bronchopulmonary aspergillosis (Miller et al., 1996), Sinusitis (Wang et al., 2000), Chronic Obstructive Pulmonary Disease (COPD) (Pignatti et al., 1996; Dahl and Nordestgaard, 2009) and Bronchial Asthma (Dahl et al., 1998a).

A detailed screening of the *CFTR* gene revealed the presence of CFTR mutations, or rare variants, in patients with DB and in patients with COPD (Ngiam et al., 2006; Bienvenu et al., 2010). However, studies that assess the association between asthma and the defective *CFTR* gene have produced conflicting results (Ngiam et al., 2006; Douros et al., 2008). Indeed, the main challenge of large genetic analyses is the interpretation of unknown CFTR sequence variations, this problem is heightened when rare or unique variants are identified and when the molecular consequences are very mild or unknown. Carriers of CFTR mutations have been identified in a higher number among patients with DB and bronchial hypersecretion compared to the general population (Bombieri et al., 1998). Increased frequency of the p.Phe508del allele has been reported in French patients with bronchial hypersecretion, in German patients with DB, and in French patients with DB and elevated sweat chloride concentration (Gervais et al., 1993).

The high prevalence of asthma in patients with CF has been known for a long time. However, it is still unclear whether carriers of CFTR mutations are susceptible to asthma. Some studies have reported a positive association between the presence of CFTR mutations and asthma (Dahl et al., 1998b; Tzetis et al., 2001; Ngiam et al., 2006), while other studies have shown no association (de Cid et al., 2001; Douros et al., 2008; Dixit et al., 2013).

Due to the presence of some similarities in the clinical manifestations of COPD and CF, investigators have sought to determine whether there is a link between *CFTR* gene mutations and the risk of COPD. Much attention has been given to the most common CF-causing mutation, p.Phe508del and several studies have failed to demonstrate that carriers are at increased risk for COPD (Dahl and Nordestgaard, 2009; Raju et al., 2014). However, a meta-analysis of 13 microarray data sets focused on CF-related differential gene expression has confirmed striking similarities in gene expression profiles between CF and COPD (Clarke et al., 2015). These studies strongly suggest that loss of CFTR function in CF and COPD may define common pathogenic pathways for both diseases. Individuals with COPD not only present a decrease in respiratory epithelial cell CFTR function but also display significant correlations between the clinical manifestations of COPD and the level of CFTR suppression in their lower respiratory tract (Cantin, 2016).

These chronic respiratory diseases are responsible for increasing morbidity and mortality in developed countries, with excessive costs in terms of health care and reduced labour productivity. Thus, it is important to identify possible therapeutic targets and to better characterize the pathophysiology associated with each of them.

Therefore, it is of interest to determine if the presence of CFTR mutations is associated with an elevated risk of asthma, DB and COPD. For that purpose, we carried this study by screening the most frequent CFTR mutations in Portuguese CF-patients in a cohort of Portuguese individuals with chronic lung diseases. Subsequently, we wanted to determine whether the CFTR variations found in these subjects have an influence on their lung disease and possibly identify potential genotype-phenotype correlations.

The sample population analysed consisted of 187 Portuguese individuals, divided in two distinct groups. The first group included subjects with one of the following respiratory disease: asthma (n=54), COPD (n=51) and bronchiectasis (n=31) and the second group included healthy control individuals (n=51). We have also used a group of 27 CF patients, as a control for the methodologies used to detect the CFTR mutations. The most common CFTR mutations in Portugal chosen to be analysed in this study and the different techniques used to detect these mutations are shown in Table 3.8. For detailed information on the techniques and protocols used in this study see chapter II (Materials and Methods).

**Table 3.8 –** The 12 most common CFTR mutations in Portuguese CF patients (WHO, 2004) and techniques used for their detection

| Mutation | Detection method |
|---|---|
| **F508del** | ARMS-PCR |
| **G542X** | ARMS-PCR |
| **N1303K** | ARMS-PCR |
| **A561E** | Tetra-primer ARMS-PCR |
| **R334W** | Tetra-primer ARMS-PCR |
| **R1066C** | Tetra-primer ARMS-PCR |
| **G576A** | Tetra-primer ARMS-PCR |
| **P205S** | Tetra-primer ARMS-PCR |
| **I148N** | Tetra-primer ARMS-PCR |
| **711+1 G>T*** | Tetra-primer ARMS-PCR |
| **3272-26 A>G** | RFLP |
| **G85E** | RFLP |

*mutation analysed jointly with Arsénia J. Massinga.

## 3. Results

The clinical phenotypes of the group of patients with respiratory diseases and control subjects analysed in this study are shown in Table 3.9. A total of 136 individuals with non-CF respiratory diseases such as asthma, DB and COPD were analysed. The control group (n=51) consisted of subjects presenting no chronic respiratory symptoms or diseases, named here healthy control.

**Table 3.9** – The different groups of individuals analysed in this study.

| Group/Disease | Individuals (N) | Age (Average ± SD) | Gender (F/M) |
|---|---|---|---|
| **Asthma** | 54 | 44.72 ± 13.92 | 39/15 |
| **DB** | 31 | 46.00 ± 13.91 | 17/14 |
| **COPD** | 51 | 67.95 ± 11.47 | 13/39 |
| **Healthy control** | 51 | 34.02 ± 9.76 | 41/10 |

A total of 187 individuals (patients and controls) were screened for the 12 most common *CFTR* mutations previously described to occur in Portuguese CF-patients, using three different techniques, ARMS, Tetra-ARMS and RFLP (Table 3.8). We used three different techniques, because for two mutations (F508del and G542X) we already had the optimized protocol implemented and for the others we chose the most appropriate technique to identify the DNA alteration. For all the patients under study blood samples were taken, DNA was extracted and diluted to 50 ng/µl for use in the different PCR techniques (see Materials and Methods). The results shown in Figure 3.9, were obtained using the ARMS methodology for the detection of the p.Phe508del mutation. In this technique two reactions were performed: *i)* reaction I which is specific for the wt sequence in the region of the p.Phe508del mutation, where PCR amplification is only expected if the sequences in the CFTR alleles are normal;

and *ii)* reaction II which is specific for the recognition of the p.Phe508del sequence, for which PCR product amplification is only detected if this mutation is present. From the analysis of these PCR products, it is possible to observe that individuals *a*, *b* and *d* do not carry the p.Phe508del mutation since there is only amplification for reaction I. Individuals *c* and *e* are carriers of the p.Phe508del mutation, since there was product amplification for both reactions. Finally, individual *f* is homozygous for the p.Phe508del mutation since amplification is only visible in reaction II, and no amplification was observed for reaction I.

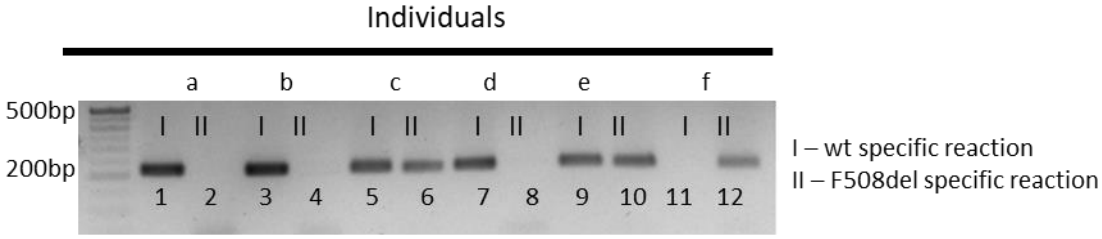



**Figure 3.9. Detection of the p.Phe508del mutation using the ARMS technique.** Representative results obtained for 5 patients from the current study and internal control do the technique (CF patient homozygous for p.Phe508del). I corresponds to the PCR reaction with primers specific for the normal sequence; II corresponds to the PCR reaction with primers specific for the allele with the mutation F508del; Lanes 1, 2, 3, 4, 7 and 8 from individuals a, b and d do not have the F508del mutation; lanes 5, 6, 9 and 10 from individuals c and e are heterozygous for the F508del mutation; lane 11 and 12, individual F is homozygous for F508del. M – marker 100bp. Electrophoresis on 1% agarose gel.

Figure 3.10A shows an example for the detection of the p.G85E mutation by Restriction Fragment Length Polymorphisms (RFLP) analysis (see details in the Material and Methods) in which differences in the length of DNA fragments obtained following restriction digestion indicate the presence/absence of a mutation. After PCR amplification of a segment of genomic DNA, a restriction enzyme (HinfI - New England BioLabs, UK) was used to cut this fragment and the resulting DNA fragments were separated according to their size by agarose gel electrophoresis. The technique relies on differences in the normal and mutant DNA sequences at particular restriction sites, which lead to DNA fragments of different lengths when the DNA is cut by this restriction enzyme. In Figure 3.10, individuals a and c do not carry the G85E mutation, in contrast with individual b, who is a carrier of this mutation and individual e who has this mutation in both CFTR alleles. For the p.G576A mutation we use the tetra-ARMS technique. In contrast to conventional ARMS PCR that amplifies the two alleles in two different PCR reactions, tetra-ARMS PCR amplifies both wt and mutant alleles together with a control fragment, in a single PCR reaction tube. The method employs four primers to amplify a non-allele specific larger fragment containing the mutation site and allele specific amplicons representing each of the two allelic forms. Figure 3.10B shows an example of the results obtained for the analysis of DNA from 5 individuals with a suspicion of CF. Only individual d carries this mutation, while the other four individuals do not.

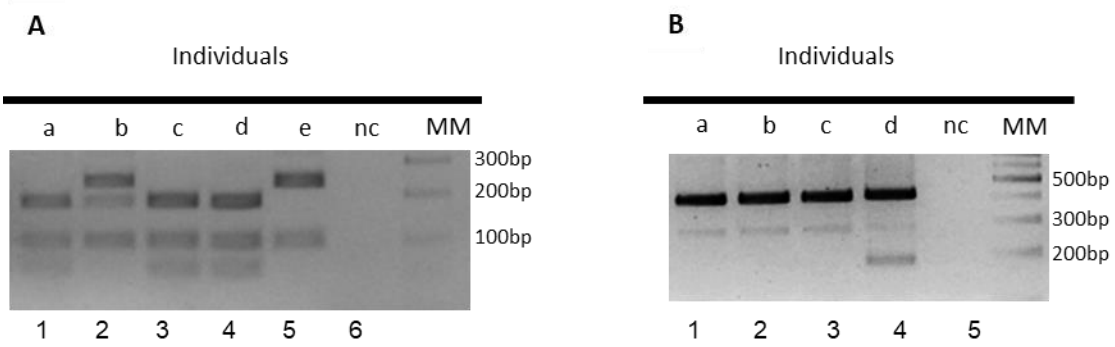



**Figure 3.10. Detection of the p.G85E and p.G576A mutations using the RFLP and Tetra-ARMS techniques, respectively. (A)** Representative results obtained for detection of G85E mutation in 3 individuals from the current study and 2 internal controls (CF patients heterozygous (b) and homozygous (e) for p.G85E). Product amplification of the wt allele shows the appearance of three fragments of 55, 91 and 172 bp respectively, while the mutant alleles shows two fragments of 91 and 227 bp. Individuals *a*, *c* and *d* (lanes 1, 3 and 4, respectively) do not have the p.G85E mutation. Individual *b* (lane 2) is heterozygous for G85E mutation, as four different fragments of 55, 91,172, 227 bp are observed. This indicates that one allele has wt sequence and the other one has the mutated sequence. Individual *e* (lane 5) is homozygous for the p.G85E mutation, as the two expected fragments for this mutation are observed. PCR negative control (nc, lane 6-) using water. (**B).** Results obtained for detection of p.G576A mutation in 3 patients from the current study and 1 internal control (CF patient heterozygous (d) for p.G576A). With this technique three different fragments with different sizes are expected, one from the outers primers (PCR control), another from the wt sequence and a third one from the mutated sequence. In this case, PCR control – 426bp; wt – 288bp; mut – 192bp, respectively. Individuals *a*, *b* and *c* (lanes 1, 2 and 3, respectively) do not have the p.G576A mutation, as two fragments corresponding to wt and PR control were observed. Individual *d* (lane 4), is heterozygous for the p.G576A mutation, with detection of three different fragments. PCR negative control using water (nc, lane 5). Both electrophoresis was run on 2% agarose gels and MM – marker 100bp.

Analysis of the 12 CFTR mutations in all the 187 individuals, led to the identification of CFTR mutations in 7 subjects (6 with 1 mutation and 1 with 2 mutations) with non-CF respiratory diseases and none in the healthy control group. Specifically, CFTR mutations were found in 1 out of 54 patients with asthma ($p>0.05$), in 2 out of 31 patients with DB ($p>0.05$) and in 4 out of 51 patients with COPD ($p\leq0.05$) one of which was a compound heterozygote (see Table 3.10).

**Table 3.10-** Global results for the CFTR mutation identification on the 136 patients and 51 controls

| | Asthma | Bronchiectasis | COPD | Healthy control |
|---|---|---|---|---|
| Individuals | 54 | 31 | 51 | 51 |
| Individuals with mutation | 1 | 2 | 4 | 0 |
| Number mutated alleles | 1 | 2 | 5 | 0 |
| Individuals with 2 CFTR mutations | 0 | 0 | 1 | 0 |

Four CFTR variants were found, F508del, G542X, G576A and R668C (see Table 3.11).

**Table 3.11-** CFTR variants identified in the subjects analysed in this study.

| | Asthma | Bronchiectasis | COPD | Healthy control | Total |
|---|---|---|---|---|---|
| F508del | 1 | 2 | 2 | - | 5 |
| G542X | - | - | 1 | - | 1 |
| G576A-R668C | - | - | 2 | - | 2 |
| **Total** | 1 | 2 | 5 | 0 | 8 |

Based on the CFTR2 database (https://www.cftr2.org), these CFTR variants are in two categories: severe mutations (CF) and unclassified variants (UV). The two CF mutations were: p.Phe508del which was found in five patients, and p.Gly542X which was detected in only one patient. The UVs were: p.Gly576Ala and p.Arg668Cys, both found in two patients with COPD. These have been reported to segregate as a complex allele p.[Gly576Ala; p.Arg668Cys] (Casals et al., 2004; El-Seedy et al., 2012). In one case, through the analysis of the CFTR gene in other family members, it was possible to confirm that these two variants are indeed present in the same allele. To address the effect of these two UVs, we conducted a separate analysis of the CFTR RNA levels and protein function of this specific COPD patient (see case report below).

A statistical analysis using a Chi-square test was performed to assess whether there was a correlation between CFTR mutations and the respiratory diseases analysed in this study. The incidence of CFTR mutations in these patients showed no significant correlation between CFTR mutations and asthma or bronchiectasis (p=0.33 and 0.07, respectively). However, we found a positive association between CFTR mutations and COPD (p=0.03) – see Table 3.12.

**Table 3.12-** Statistical analysis for the results obtained in this study.

| | | Individuals | | | | Alleles | | |
|---|---|---|---|---|---|---|---|---|
| | | Total | | Mutated | | Total | Mutated | Fisher's |
| **Disease/group** | | (n) | (%) | (n) | (%) | (n) | (n) | p-value |
| | Asthma | 54 | 29% | 1 | 1.9% | 108 | 1 | 0.33 |
| | Bronchiectasis | 31 | 17% | 2 | 6.5% | 62 | 2 | 0.07 |
| | COPD | 51 | 27% | 4 | 7.8% | 102 | 5 | **0.03** |
| **Healthy controls** | | 51 | 27% | 0 | 0.0% | 102 | 0 | |

## 4. Discussion

This study was conducted in 187 individuals, namely 136 subjects with chronic non-CF respiratory diseases and 51 healthy control individuals, to determine if there is a possible association between CFTR mutations and asthma, disseminated bronchiectasis (DB) and COPD in Portuguese patients. Genotyping was performed for 12 of the most common CFTR mutations in Portugal, of which the p.F508del, the p.G542X mutations and the p.G576A-R668C complex allele were found in heterozygosity in 7 of these patients (5%).

We investigated the association of CFTR gene mutations in heterozygosity with asthma and we have not found an association in contrast with previous studies (Dahl et al., 1998a, 2005; Lowenfels et al., 1998). Dahl and colleagues (Dahl et al., 2005), studying a Danish population, reported that there is a positive association between F508del mutation carriers and asthma. Lowenfels, et al. (Lowenfels et al., 1998) conducted a multinational survey in 1113 obligate CFTR gene mutation heterozygotes and 688 controls, where they found that Incidence of asthma in CF heterozygotes was 9.6%, similar to was reported by Dahl, et al. (Dahl et al., 1998a). Contradictory results were obtained by Schroeder et al (Schroeder et al., 1995), in the United States, since these authors claim that carriers of p.F508del are somewhat protected against asthma and in this study they found that the odds ratio (OR) for asthma in carriers of F508del is 0.31 (95% CI: 0.107-0.909). Finally, the Epidemiology on the Genetics and Environment of Asthma (EGEA) study in France showed that 3.2% of asthma cases and 2.9% of controls were F508del heterozygotes. This difference was not statistically significant (de Cid et al., 2001). Our results showed that the group of asthma patients did not have a significant difference in CFTR mutations compared to the control group.

Similarly, we did not find an association of CFTR gene mutations in DB patients vs the control group, despite the high percentage of carriers (6.5%) that we found in this group. Indeed, the incidence of the p.F508del mutation was found to be 2-fold increased in our group of DB patients when compared to the control group, however this difference was not statistically significant. Previous reports from other groups did not encounter a statistically significant higher incidence of CFTR mutations in DB patients either (Bombieri et al., 1998; Ngiam et al., 2006). According to published data, heterozygosity for p.F508del appears to predispose for DB and asthma, but its involvement in other pulmonary diseases remains unproven. It is thus possible that a larger cohort could have confirmed this observation.

The most significant proportion of CFTR mutations carriers that we identified was in the COPD group (6.8%) and this observation was found to be statistically significant ($p \leq 0.03$) namely for the p.F508del mutation (3.9%). Thus, our study showed that CFTR mutations represent a statistically significant risk factor for development of COPD disease. This is not

surprising as one major cause of COPD is cigarette smoke (CS) and several studies have demonstrated that it induces CFTR internalization and dysfunction (Clunes et al., 2012; Hassan et al., 2014). It is thus likely that having only one functional copy of CFTR increases the susceptibility to not just CS but also to other respiratory pollutants.

In our study we focused only on the analysis of common CFTR gene mutations/variants in Portugal and their association with COPD, asthma and DB. We have shown an association between CFTR mutations and COPD. Future studies should be performed using a larger cohort so as to detect minor to modest genotype disease association. Additionally, other CFTR mutations should be included in the screening panel.

## *Chapter 1.1 – Case Report*

Case Report Analysis – Identification of the influence of the unknown variant complex allele p.[Gly576Ala; p.Arg668Cys] in the COPD phenotype.

### 1. Abstract

**Background:** Cystic fibrosis (CF) is a genetic disease caused by mutations in the CFTR gene, whose classic phenotype is the presence of bronchiectasis, chronic sinusitis and pancreatic insufficiency. Diagnostic criteria are based on evidence of CFTR dysfunction (e.g., sweat $Cl^-$ test) and/or identification of 2 CF-causing mutations in the *CFTR* gene together with a consistent clinical phenotype. In cases of difficult diagnosis, other functional studies of CFTR may be used.

**Aim:** The main objective of this study was to correlate clinical phenotype with levels of functional CFTR in patient-derived tissue (rectal biopsy) of a 67-year old COPD (smoker) female patient carrying two CFTR mutations: a nonsense mutation (G542X) that leads to transcript degradation by NMD and a complex allele (G576A-R668C), with G576A being responsible for alternative splicing (exon 13 skipping) and leading to the production of normal and aberrant transcripts to variable extents.

**Methods:** Various parameters (FEV1, sweat chloride) were determined for this patient. For the functional characterization we used freshly isolated rectal biopsies to assess: measurements of CFTR-mediated chloride secretion in micro-Ussing chamber, Forskolin-Induced Swelling (FIS) assay in intestinal organoids and levels of the G542X/G576A-R668C-CFTR transcripts by quantitative RT-PCR (qRT-PCR).

**Results and Discussion:** This patient was followed up in Pulmonology consultation after hospitalization at the age of 56 due to an exacerbation of chronic bronchitis. The patient was medicated with ipratropium bromide (normalizing symptoms) and antibiotic therapy. Her respiratory function was stable (FEV1 / FVC-79%, FEV1prev.-104%), and her diffusing capacity for carbon monoxide (DLCO) was decreased but normalized when corrected for alveolar volume. The patient's CT scans showed no changes and microbiological isolates were persistently negative. A sweat test was also negative and faecal elastase assay excluded pancreatic insufficiency. Rectal biopsies were performed to assess CFTR-mediated $Cl^-$ secretion by Ussing chamber and to generate intestinal organoids. As both excluded major CFTR dysfunction, a diagnosis of CF was rejected.

**Conclusion**: Two CFTR mutations were identified in this COPD patient and the complex-allele is of unknown prognosis, the clinical phenotype was not suggestive of CF, which was corroborated by the functional tests of the CFTR protein leading to the exclusion of CF.

Analysis of the CFTR gene may lead to the identification of mutations of uncertain pathological significance. In selected cases the functional study of the CFTR protein may be a valuable aid to confirm/ exclude a diagnosis of CF.

## 2. Introduction

Cystic fibrosis (CF) and chronic obstructive pulmonary disease (COPD) are chronic airway diseases with different causes that share some features in pathologic changes and clinical presentation. Both these disorders are characterized by chronic mucosal and airway inflammation, each with distinct pathophysiologic features, but a common increase in the infiltration of neutrophils triggered by a variety of inflammatory mediators (Sehnert et al., 2002; Chilvers, 2010). The pathologic processes in these diseases seem to involve progressive inflammatory responses with elements of tissue remodelling, airway obstruction and reduction in expiratory flow rates. COPD is a rising cause of morbidity and mortality worldwide, and Portugal is no exception. In 1990, COPD was the 6$^{th}$ most common cause of death worldwide and was ranked 12$^{th}$ for DALYs (disability- adjusted life years) (Cardoso et al., 2013). It was projected that by 2020 it will have risen to the 5$^{th}$ rank, and be the third most common cause of death (Cardoso et al., 2013). The prevalence of COPD rises steeply to above 10% amongst people who are aged over 40 (Bárbara et al., 2013).

The national prevalence for COPD was estimated to be 9.0% and 5.3% in two previous studies conducted in selected age groups (≥40 years old in one study and between 35 and 69 years old in the other); there was an additional study carried out in the Lisbon area that showed a prevalence of 14.2% (in (≥40 years old patients ) (Bárbara et al., 2013; Cardoso et al., 2013; Araújo et al., 2017).

Several groups have reported that smokers with COPD have decreased CFTR function in both the upper and lower airways, suggesting that dysfunctional CFTR may also play a role in the pathogenesis of COPD ((Ghosh et al., 2015)). In addition, CFTR dysfunction was found to be associated with chronic bronchitis symptoms and dyspnea, indicating that abnormal CFTR may be particularly important towards causing mucus retention. These findings are supported by *in vitro* studies showing reduced airway surface liquid (ASL) height and impaired MCC caused by cigarette smoke exposure (Eickmeier et al., 2010; Dransfield et al., 2013; Muthuswamy et al., 2014).

Here, we present a clinical case of a 67-year-old woman with COPD who is a smoker. She was followed-up by the Pulmonology consultation since the age of 56, after hospitalization due to an exacerbation of chronic bronchitis. During this time, the patient was medicated with ipratropium bromide, (normalizing symptoms) and antibiotic therapy.

The main objective of this work was to correlate clinical phenotype with levels of functional CFTR in patient-derived tissue (rectal biopsy) in this 67-year COPD (smoker) female patient who carries two CFTR mutations, namely: a nonsense mutation (G542X), that leads to degradation by NMD and one complex allele (G576A-R668C), with G576A leading to

alternative splicing (exon 13 skipping) and the production of normal and aberrant transcripts to variable extents. To this end, various parameters (FEV1, sweat chloride) were determined for this patient. For the functional characterization we used freshly isolated rectal biopsies to assess: CFTR-mediated $Cl^-$ secretion in micro-Ussing chamber, Forskolin-Induced Swelling (FIS) assay in intestinal organoids; and assessed the levels of G542X/G576A-R668C-CFTR transcripts by quantitative RT-PCR (qRT-PCR).

Altogether, results from this study may help to provide a better understanding of the impact of the G576A-R668C variation on clinical phenotype and to establish/ exclude a CF diagnosis for this patient, with key clinical implications for personalized therapeutic intervention.

## 3. Results

### *Patient clinical presentation*

This patient was followed up in Pulmonology consultation since the age of 56, after hospitalization due to an exacerbation of chronic bronchitis, being medicated with ipratropium bromide (normalizing symptoms) and antibiotic therapy. Her respiratory function was stable (FEV1 / FVC-79%, FEV1prev.-104%), and her diffusing capacity for carbon monoxide (DLCO) was decreased but normalized when corrected for alveolar volume, the microbiological isolates were persistently negative and CT scans showed no changes. A sweat test was negative and faecal elastase assay excluded pancreatic insufficiency. (Table 3.13).

**Table 3.13 - Clinical Phenotype of 67-yr old female smoker COPD patient**

| | |
|---|---|
| Respiratory Function | FEV1 / FVC-79%, |
| | FEV1prev.-104% |
| Fecal elastase test | Normal |
| Sweat test | Negative |
| DLCO* | SB – 65% |
| | VA – 76% |

*Diffusing capacity for carbon monoxide

### *Assessment of CFTR-mediated $Cl^-$ secretion in rectal biopsies by Ussing chamber*

Rectal biopsies were performed to assess CFTR-mediated Cl- secretion by Ussing chamber since these measurements were previously shown to constitute a valuable biomarker for CF diagnosis and prognosis (Mall et al., 2000; Sousa et al., 2012). To assess the intestinal CFTR-mediated $Cl^-$ secretion in the patient, rectal biopsy specimens were mounted and analysed in micro-Ussing chambers perfused in Ringer solution under open-circuit conditions, as previously described (Sousa et al., 2012) (see also the Materials & Methods section).

Briefly, Amiloride (Amil) was added to block electrogenic sodium ($Na^+$) absorption through ENaC and indomethacin was applied for 40-60 min to inhibit endogenous cAMP formation through prostaglandins (Sousa et al., 2012). As reported, cAMP-dependent and cholinergic $Cl^-$ secretion in human rectal tissues relies on functional CFTR. Thus, we used 3-isobutyl-1-methylxantine (IBMX basolateral) and forskolin (Fsk basolateral) to activate cAMP-dependent $Cl^-$ secretion and carbachol (CCH, basolateral) for cholinergic co-activation (Mall et al., 2000).

Values obtained from quantification of the equivalent short-circuit (Ieq-sc-CCH/IBMX/Fsk) from these original tracings (Figure 3.11, Table 3.14) were in agreement with those reported previously for non-classical CF (Sousa M et al, 2012). These results indicate that the G576A-R668C variant can be considered as a non-classical CF mutation.

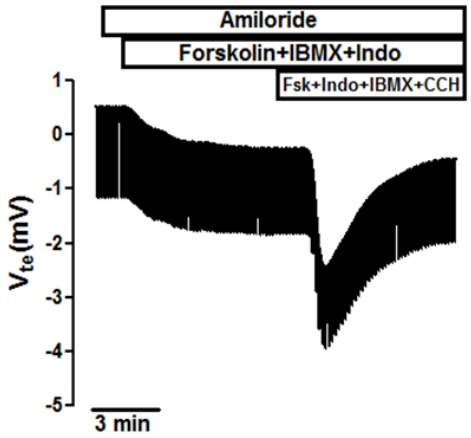



**Figure 3.11 - Original Ussing chamber tracings representing CFTR-mediated $Cl^-$ secretion in rectal biopsies from a patient with the G542X/G576A-R668C genotype**. Representative tracing of the subject with the G542X/G576A-R668C genotype showing the effect on the transepithelial voltage of cholinergic (CCH, 100µM), and cAMP-dependent (IBMX, 100µM + Fsk, 2µM, basolateral) stimulation. [Data obtained in collaboration with Nikhil T Awatade, included here with permission]

**Table 3.14 - Summary of CFTR-mediated Cl- secretion in rectal biopsies with the G542X/G576A-r668c genotype.**

| Patient | Basal Rte($\Omega.cm^2$) | $I_{eq\text{-}sc\text{-}CCH/IBMX/Fsk}$ ($\mu A/cm^2$) |
|---|---|---|
| G542X/G576A-R668C (67 years old) | 16.18±1.77 | 187.1±1.75 |

*(Rte - Transepithelial resistance). (n=2). Data ± SEM.*

### *FIS assay in rectal organoids with the G542X/G576A-R668C genotype*

In order to further assess CFTR function as well as the efficacy of CFTR modulators *ex vivo*, i.e., to assess whether this combination of CFTR mutations will respond to existing and novel CFTR modulators (in case of a possible diagnosis of CF), the technique of intestinal

organoids, as described by the Beekman group (Dekkers et al., 2013), namely the forskolin swelling (FIS) assay was used.

Firstly, we carried out the FIS assay in organoids from the COPD patient without any treatment (Figure 3.12) to determine the levels of CFTR channel activity. To this end, we stimulated organoids with increasing doses (0.05 to 3 µM) of the CFTR channel agonist Fsk (Figure 3.12A). These FIS assay data were then used to calculate the absolute area under the curve (AUC), assuming the baseline=100% for 60min (Figure 3.12B).

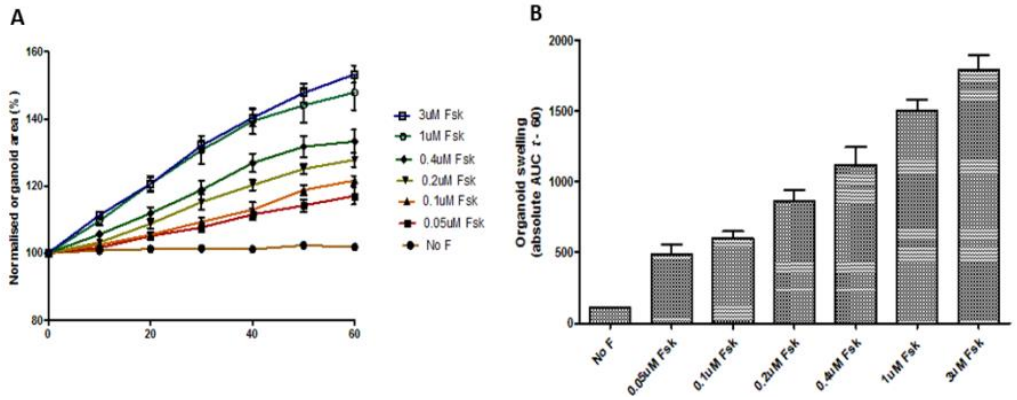



**Figure 3.12– FIS assay in G542X/G576A-R668C rectal organoids. (A)** Time-course tracings of the Fsk- induced surface area increase relative to t=0 (normalized area) of G542X/G576A-R668C organoids at different Fsk concentrations (0.05 to 3 µM) averaged from three independent experiments. **(B)** Organoids swelling results from the FIS assay at different Fsk concentrations expressed as the area under the curve (AUC) (baseline=100%, t=60min). Data are mean±SD (n=3). [Data obtained in collaboration with Nikhil T Awatade, included here with permission].

These data indicate that these organoids exhibit normal CFTR activity, as shown by enhanced swelling during increasing Fsk concentration, consistently to the data obtained for the fresh biopsies in the Ussing chamber. These data area also consistent with what has been previously described for patients with one CFTR causing mutation and no mutations in the other allele (Dekkers et al., 2013).

### *Relative quantitative transcript analysis*

To determine the effect of the G542X mutation and the G576A-R668C polymorphism at the CFTR transcript level, RNA was extracted from both colonic tissue and organoids from the COPD patient and analysed by quantitative Real Time PCR (qRT-PCR). This analysis allows us to assess both the degradation of transcripts resulting from the G542X mutation through nonsense-mediated decay (NMD) and the skipping of exon 13 (Δ13) resulting from the G576A-R668C complex allele, as previously described for G576A (Pagani et al., 2003). In this analysis we used two control samples: RNA from a patient with the F508del/G542X genotype and RNA from a non-CF individual, to validate the set of primers for each mutation/polymorphism.

Relative qRT-PCR analysis using primer pairs specific for the G542X mutation (Figure 3.13) shows that transcript levels from the G542X and the non-G542X alleles in patient-

derived materials are present in a ratio of 19:81 in biopsies and of 18:82 in organoids. Relative qRT-PCR analysis using primer pairs specific for the exon 13 skipping (occurring for the transcript from the G576A-R668C complex allele) shows that Δ13 and full-length transcripts (including those from G542X, since skipping of this exon occurs normally for all individuals) are present in a ratio of 30:70 (in the intestinal biopsy) and of 51:49 (in organoids) (Figure 3.14, Table 3.15).

These data suggest that two abnormal CFTR transcripts are present: the G542X transcript and the alternatively spliced transcript derived from the G576A-R668C allele. With the G542X specific primer  relative transcript levels are expected to be very low, because this is a nonsense mutation. This mutation leads to significant degradation of the respective transcripts by nonsense-mediated mRNA decay (NMD) which acts as a surveillance mechanism to prevent production of truncated (non-functional) proteins. From these data we estimate a degradation of around 80% for the G542X transcripts. Using primers specific for the exon 13 skipping we obtained more full-length transcripts derived from the control samples (HC – healthy control and F508del/G542X) than from the index patient which indicates that the G576A/G668C is implicated in the exclusion of the exon 13 from the CFTR transcripts.

**Table 3.15 - Relative abundance of G542X/G576A-R668C CFTR transcripts as a % of total CFTR mRNA.**

| Individual | Percentage of CFTR transcripts from G542X transcripts | | Percentage of CFTR transcripts from ex13 skipping | |
|---|---|---|---|---|
| | Biopsy | Organoids | Biopsy | Organoids |
| COPD Patient | 19% | 18% | 30% | 51% |
| F508del/G542X | -- | 20% | | 7% |
| wt/wt | -- | 0.2% | | 4% |

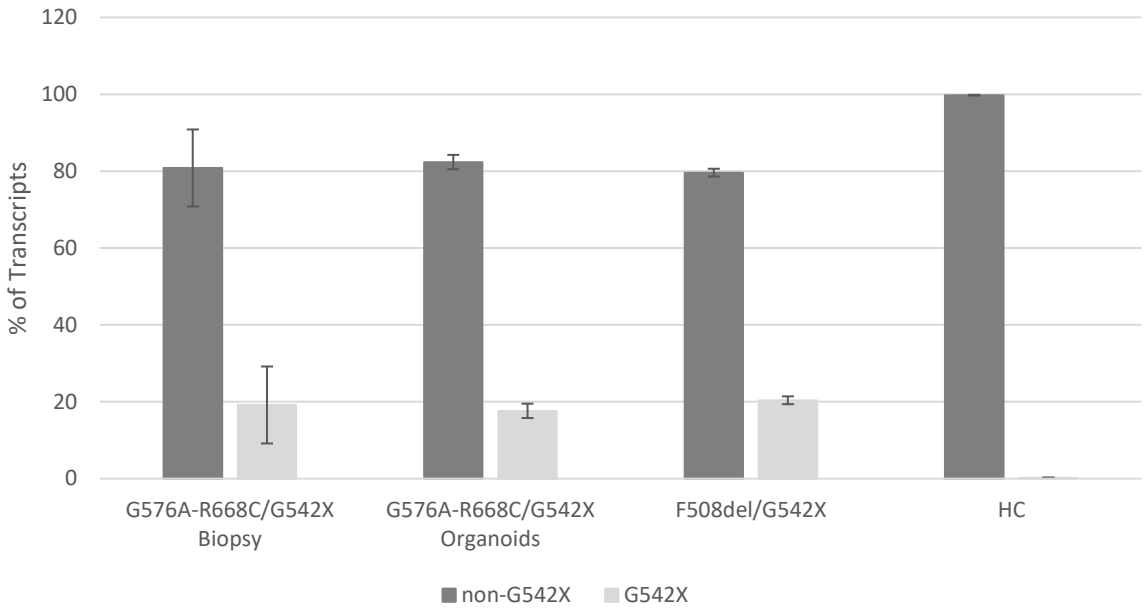



**Figure 3.13 - Relative abundance of CFTR transcripts resulting from the G542X mutation quantified by qRT-PCR in in rectal biopsies and organoids from the COPD patient with the G542X/G576A-R668C genotype**. To assess the abundance of the transcripts we used primers specific for the G542X mutation. RNA from another patient with the F508del/G542X genotype and a non-CF individual were used as controls. Data shown here are mean values (see Methods) using RNA extracted from rectal biopsies (n=3) or intestinal organoids (n=3).

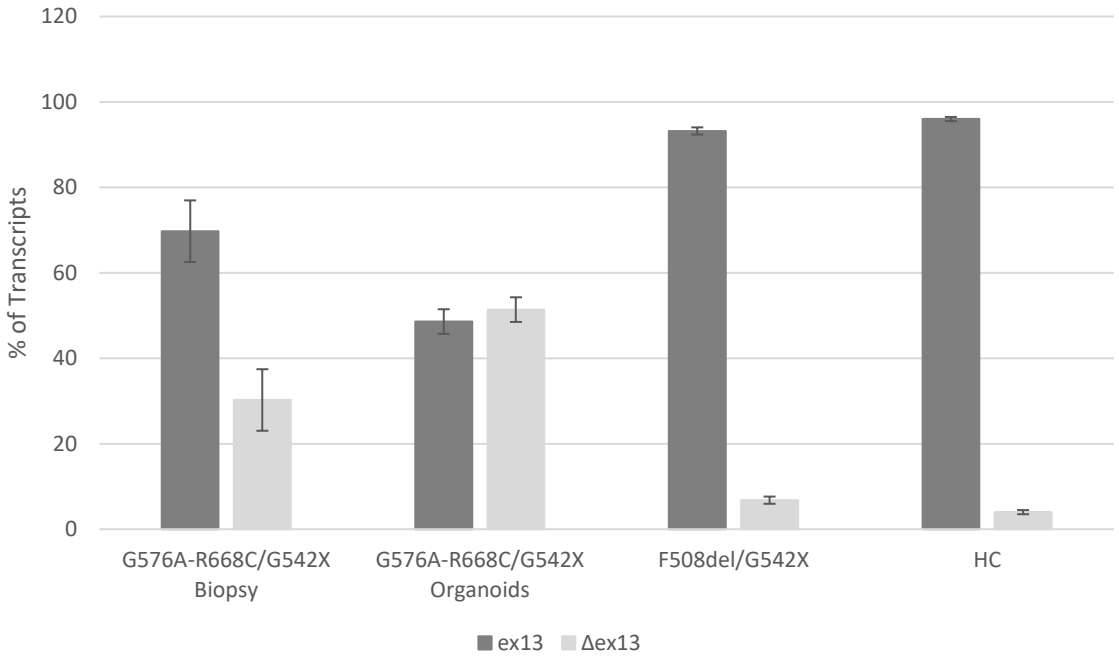



**Figure 3.14 - Relative abundance of CFTR transcripts resulting from exon 13 skipping quantified by qRT-PCR in rectal biopsies and organoids from the COPD patient with the G542X/G576A-R668C genotype**. To assess the abundance of the transcripts we used primers specific for exon 13 skipping. RNA from another patient with the F508del/G542X genotype and a non-CF individual were used as controls. Data shown here are mean values from qRT-PCR experiments (see Methods) using RNA extracted from rectal biopsies (n=3) or intestinal organoids (n=3).

## 4.Discussion

The G576A variant is currently described as a non-CF causing mutation in the CFTR2 database (www.cftr2.org/2017) (Figure 3.14 **- Relative abundance of CFTR transcripts resulting from exon 13 skipping quantified by qRT-PCR in rectal biopsies and organoids from the COPD patient with the G542X/G576A-R668C genotype**. To assess the abundance of the transcripts we used primers specific for exon 13 skipping. RNA from another patient with the F508del/G542X genotype and a non-CF individual were used as controls. Data shown here are mean values from qRT-PCR experiments (see Methods) using RNA extracted from rectal biopsies (n=3) or intestinal organoids (n=3).

, Table 3.16). However, some studies have detected this mutation in individuals with both classical CF symptoms and in patients who have evidence of a clinical disease only in a subgroup of the organ systems qualifying to have a CFTR-related disorder (CF-RD) or "CFTR-opathies". These forms, including late-onset pulmonary disease, congenital bilateral absence of vas deferens (CBAVD), or idiopathic pancreatitis frequently show a genetic diagnostic challenge due to the unclear genotype–phenotype correlation (Pignatti et al., 1995). Pagani and colleagues (Pagani et al., 2003) have showed that the G576A mutation leads to exon 13 skipping and an increase of aberrant transcripts when the R668C change in exon 14 is also present. The authors describe exonic regulatory sequences involved in the splicing process and highlighted the importance of analysing missense changes, even if they have been found in normal chromosomes, because they can affect the levels of CFTR expression.

**Table 3.16 – Number of alleles and Classification of studied mutations in the CFTR2 database.**

| Variant legacy name | # alleles in CFTR2 | % pancreatic insufficient | Variant final determination |
|---|---|---|---|
| R668C | 68 | 44% | Non CF-causing |
| G542X | 3595 | 98% | CF-causing |
| F508del | 98735 | 98% | CF-causing |
| G576A | 74 | 33% | Non CF-causing |

In summary, in this study we have analysed a 67-year COPD (smoker) female patient with the G542/G576A-R668C genotype for CFTR function and CFTR mRNA expression in patient-derived materials, so as to correlate clinical phenotype with functional CFTR levels. Despite the 2 CFTR mutations identified, the expressed phenotype is not suggestive of CF, which was confirmed by the functional analysis of the CFTR protein: normal activity results for CFTR-mediated $Cl^-$ secretion assessed in micro-Ussing chamber and by FIS assay in intestinal organoids. Although about 2,000 CFTR mutations have been reported in the *CFTR* gene, most do not have an established/ excluded pathogenic role. While the Mediterranean mutation G542X is described as having a deleterious role, the G576A mutation was initially considered neutral and was not associated with classical CF. However, this variant induces changes in

the splicing process (exon 13 skipping), leading to the production of normal and aberrant transcripts. Thus, the variable levels of resulting abnormal transcripts can be related to different phenotypes and CFTR-related diseases. It is estimated that only about 10% of the ionic transport preserved is necessary to ensure that there are no manifestations of the CF disease.

Analysis of the CFTR gene may lead to the identification of mutations of uncertain pathological significance and their study can help lessen the therapeutic and psychological burden. The phenotypic heterogeneity of CF, disease-modifying genes and environmental factors give rise to different expressions that render the interpretation of the genotype challenging. In select cases the functional study of the CFTR protein may be a valuable aid.

In conclusion, despite the observation that two CFTR mutations were identified in this COPD patient and that the complex-allele is of unknown prognosis, analysis of the CFTR function was very relevant, since we could observe that the levels of CFTR activity were within the normal range leading to the exclusion of a diagnosis of CF. Nevertheless, it is very likely that the patient's CFTR genotype (that greatly reduces the expression of normal CFTR transcripts) potentiated the pathological effects of tobacco, and can be related to the development of more acute respiratory symptoms than those that would be observed in an individual with normal copies of the CFTR gene.

# *Chapter 2 - mRNA-based Detection of Rare CFTR Mutations Improves Genetic Diagnosis of Cystic Fibrosis in Populations with High Genetic Heterogeneity*

Data included in this chapter are included in (with minor alterations)

Felício V, Ramalho AS, Igreja S, Amaral MD (2017) mRNA-based Detection of Rare CFTR Mutations Improves Genetic Diagnosis of Cystic Fibrosis in Populations with High Genetic Heterogeneity. *Clin Genet* **91**: 476-481. [PMID: 27174726]

## 1. Abstract

Even with the advent of next generation sequencing complete sequencing of large disease-associated genes and intronic regions is economically not feasible. This is the case of cystic fibrosis transmembrane conductance regulator (*CFTR*), the gene responsible for cystic fibrosis (CF). Yet, to confirm a CF diagnosis proof of CFTR dysfunction needs to be obtained, namely by the identification of two disease-causing mutations. Moreover, with the advent of mutation-based therapies, genotyping is an essential tool for CF disease management. There is, however, still an unmet need to genotype CF patients by fast, comprehensive and cost-effective approaches, especially in populations with high genetic heterogeneity (and low p.F508del incidence), where CF is now emerging with new diagnosis dilemmas (Brazil, Asia, etc). Herein, we report an innovative mRNA-based approach to identify CFTR mutations in the complete coding and intronic regions. We applied this protocol to genotype individuals with a suspicion of CF and only one or no CFTR mutations identified by routine methods. It successfully detected multiple intronic mutations unlikely to be detected by CFTR exon sequencing. We conclude that this is a rapid, robust and inexpensive method to detect any CFTR coding/intronic mutation (including rare ones) that can be easily used either as a primary approach or after a routine DNA analysis.

**Keywords:** CF diagnosis; genotyping; intronic mutations; rare mutations; splicing.

## 2. Introduction

Cystic fibrosis (CF; MIM# 219700) the most frequent, life-threatening monogenic disorder in Caucasians, is dominated by the respiratory disease, the main cause of morbidity and mortality. Other CF symptoms include pancreatic insufficiency (PI), gastrointestinal, elevated chloride ($Cl^-$) in sweat (the basis of the most common diagnostic test) and male infertility, with wide clinical variability in organ involvement (reviewed in (Rowe et al., 2005)).

CF is caused by mutations in a single gene the cystic fibrosis transmembrane conductance regulator gene (CFTR; MIM #602421; GenBank NM 000492.3), one of the largest human genes with 27 exons spanning ~190 kb and encoding a Cl− channel. Although one single mutation – p.F508del – accounts for ~70% of CF chromosomes, ~2000 *CFTR* gene variants were reported (Cystic Fibrosis Mutation Database, 2017), posing considerable challenges to the establishment of a genetic CF diagnosis, particularly in populations of low p.F508del incidence.

Classical CF diagnosis is suggested by characteristic clinical features, a CF sibling or, more recently, by a positive (Newborn screening) NBS result (Levy et al., 2016). Nevertheless, to confirm a CF diagnosis, evidence of CFTR protein dysfunction is required (De Boeck et al., 2006; Farrell et al., 2008; Amaral, 2014), namely by identification of two CFTR mutations of established disease liability (Sosnay et al., 2013; CFTR2, 2018), two abnormal sweat-$Cl^-$ tests (≥ 60 mEq/l), distinctive transepithelial nasal potential difference measurements or absence/reduced CFTR-mediated $Cl^-$ secretion in rectal biopsies (Ratjen and Döring, 2003).

However, non-classical CF, with borderline sweat-Cl− levels (30–60 mEq/l) symptoms suggestive of CF and/or a familial CF history, often lead to no (or just one) CFTR mutation identified by routine panels (Feldmann et al., 2003). This is particularly striking in populations with high genetic heterogeneity (and low p.F508del incidence), where CF is now emerging with new diagnosis dilemmas (Brazil, Asia, etc), due to these regional differences. Also, extended NBS programmes which identify a significant proportion of asymptomatic NBS-positive individuals (Farrell et al., 2008; Levy et al., 2016), require fast and cost-effective genetic analyses to identify the CF-causative mutations, along with clinical follow up. Moreover, new mutation-specific therapies using CFTR modulators, are now in the clinic (De Boeck and Amaral, 2016), with knowledge of the genotype being thus critical for adequate patient management.

Herein, we report a fast and inexpensive mRNA-based protocol to identify CFTR mutations in the complete coding and intronic regions using just nine polymerase chain reactions (RT-PCR), and one assay for p.F508del. Robustness is shown by its successful application to

genotype eight individuals with a suspicion of CF but with only one or no CFTR mutation identified by routine methods.

## 3. Results

### *CFTR mutation analysis using nine RT-PCR reactions and p.F508del by amplification-refractory mutation system (ARMS)*

We established an mRNA-based protocol to detected CFTR mutations which covers the complete coding region using nine RT-PCR reactions (Figure 3.15A, Table 2.4). In case no previous panel of mutations had been analysed before, reaction no. 4 should be performed in duplicate with two different forward primers (4N, 4M), both using the same reverse primers (see end of Table 2.4) so as to identify/exclude the most prevalent mutation. This p.F508del-ARMS specifically amplifies either the normal or transcripts containing the p.F508del sequence (Figure S 3.1). The fragments resulting from the nine RT-PCR reactions (Figure 3.15B) cover the complete coding analysis of the CFTR gene and through their DNA sequencing any mutation lying on this region can be detected. Also, comparing the size of the RT-PCR products obtained with the expected ones, identifies aberrant transcripts resulting from splicing mutations and/or deletions and insertions, as shown here in transcripts from four CF individuals (Figure 3.16 and Figure 3.17).

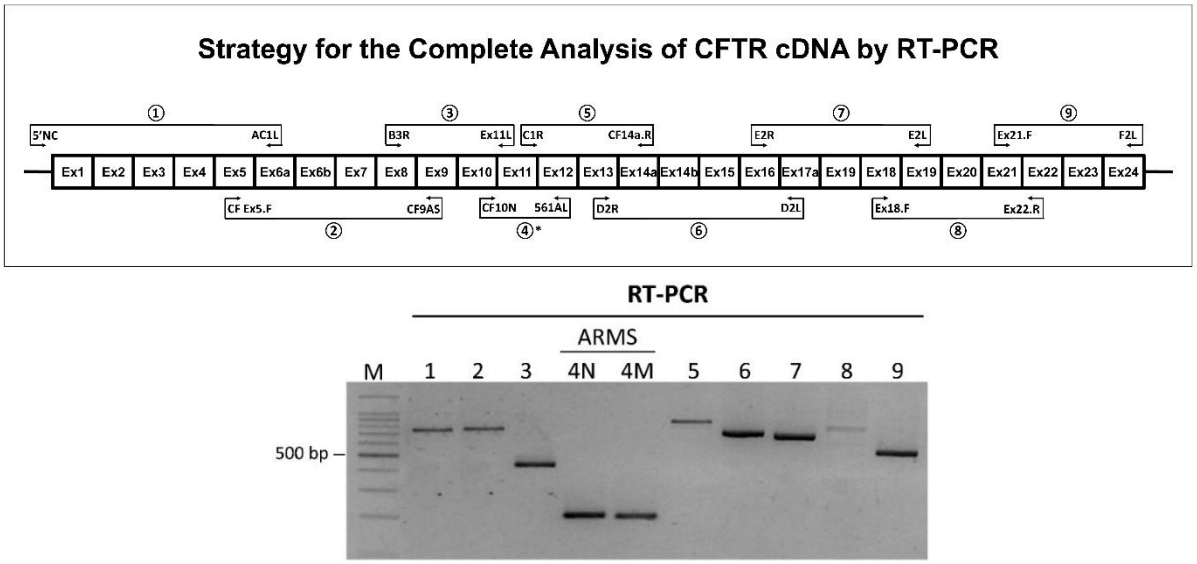



**Figure 3.15** - **mRNA-based cystic fibrosis transmembrane conductance regulator (CFTR) mutation analysis. (A)** Diagram showing the distribution of nine reactions used to analyse the complete CFTR coding region and also to detect consequences of splicing mutations that lead to alternative transcripts. *Reaction 4 can be performed in duplicate with two different forward primers so as to detect the p.F508del mutation by the amplification-refractory mutation system (ARMS) technique (see text and Figure 3.16 for details). **(B)** Representative result of the nine polymerase chain reaction (RT-PCR) products plus the two p.F508del-ARMS reactions (see also Table 2.4). Products of the nine RT-PCR reactions were analysed by 2% (w/v) agarose gel electrophoresis (lanes 1–9) covering the entire coding region of CFTR RNA. Results from the two ARMS reactions are also shown: one specific for the normal sequence at the F508 site and another specific for p.F508del (lanes 4N and 4M, respectively). M – 100bp molecular weight markers.

### *Detection of a CFTR mutation deep in intron 10*

Applying the exon 10-12 PCR strategy (Figure 3.16A) to individual P8 (Figure 3.16B, lane 1) led to observation of an upper band (311 bp) in addition to the expected one (207 bp) revealing the presence of a cryptic exon (104 bp) resulting from the inclusion of a region within IVS10.

Sequencing of genomic DNA (gDNA) from individual P8 in the IVS10 region confirmed the presence of 1584+18672A>G, a previously reported mutation (Costantino et al., 2010) which lies deep in IVS10 and thus would escape identification by extended DNA analysis of exons and flanking sequences.

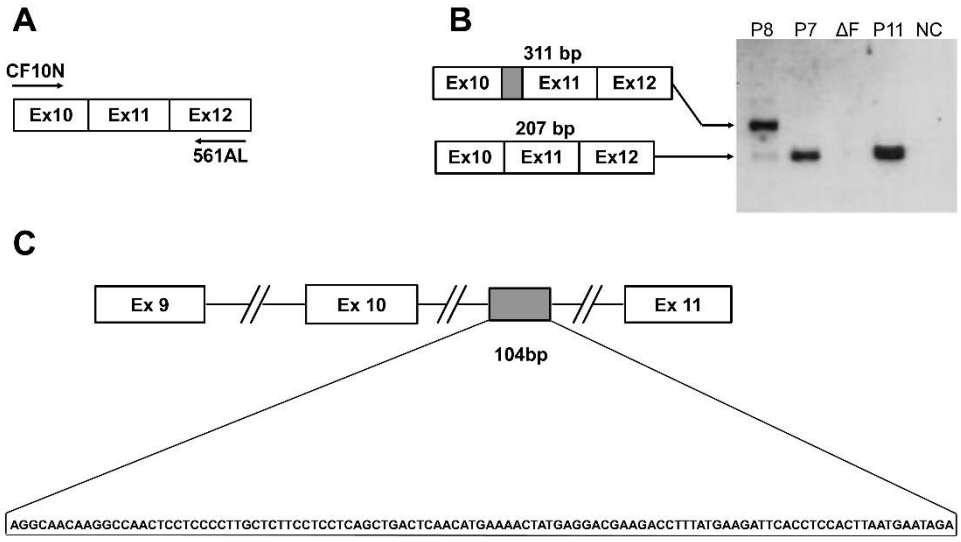



**Figure 3.16** - **Detection of a mutation leading to the generation of a cryptic exon from IVS10.** **(A)** Polymerase chain reaction (PCR) strategy for the cystic fibrosis transmembrane conductance regulator (CFTR) transcripts in the exon 10-12 region using the non-p.F508del specific forward primer (see Figure 3.15A). **(B)** Results of 3% (w/v) agarose gel electrophoresis of PCR products in this region from three individuals with a suspicion of CF or no CFTR mutations identified (P7, P8, P11) and one p.F508del-homozygous patient (ΔF) as a control. For the RT-PCR products of individual P8 (lane1), an upper band (of 311bp) is observed in addition to the faint band of the expected size (207bp). As expected this RT-PCR resulted in no amplification products for the p.F508del-homozygous individual (lane 3) because only the primer specific for the normal sequence was used (see Figure 3.15A). NC corresponds to the negative control (H2O) of the PCR. **(C)** Sequencing of the upper band in lane 1, revealed the presence of a cryptic exon (of 104bp) resulting from inclusion of a region within IVS10.

### *Detection of various splicing mutations*

Figure 3.17 provides examples of mutations successfully identified in the regions of: (a) exons 1-6a; (b) exons 11-14a; (c) exons 13-17a; and (d) exons 16-19. P1 mRNA had an additional lower band in the exon 1-6a region (Figure 3.17A). cDNA sequencing confirmed that this results from exon 5 skipping (not shown), and is thus probably caused by an IVS5 mutation. Sequencing of P1 gDNA IVS5 identified the 711+3A>T (HGVS: c.579+3A>T)

mutation.

Transcripts from all individuals showed a lower band in the exon 11-14a region (Figure 3.17B), confirming that low levels of exon 12 skipping are a common occurrence even in non-CF individuals due to the presence of neutral single-nucleotide polymorphisms (Steiner et al., 2004). However, because in P4 this band was of high intensity, we performed IVS12 gDNA sequencing which identified the 1812-1G>A (HGVS: c.1680-1G>A) mutation.

For P1 and P4, an additional lower band in the exon 13-17a region was detected (Figure 3.17C) and its sequencing confirmed that it corresponds to exon 16 skipping (not shown), which is suggestive of an intronic mutation. This was confirmed by IVS16 gDNA sequencing, which identified 3120+1G>A (HGVS: c.2988+1G>A) in both P1 and P4.

RT-PCR products in the exon 16-19 region (Figure 3.17D) revealed an additional higher band in P3 which, given its length (~25 extra nts, not shown), was hypothesised to correspond to an extended exon 17a due to the previously characterised 3272-26A>G (HGVS: c.3140-26A>G) mutation (Amaral et al., 2001), which is common in Portugal. Indeed, sequencing of IVS17a gDNA confirmed this mutation in P3.

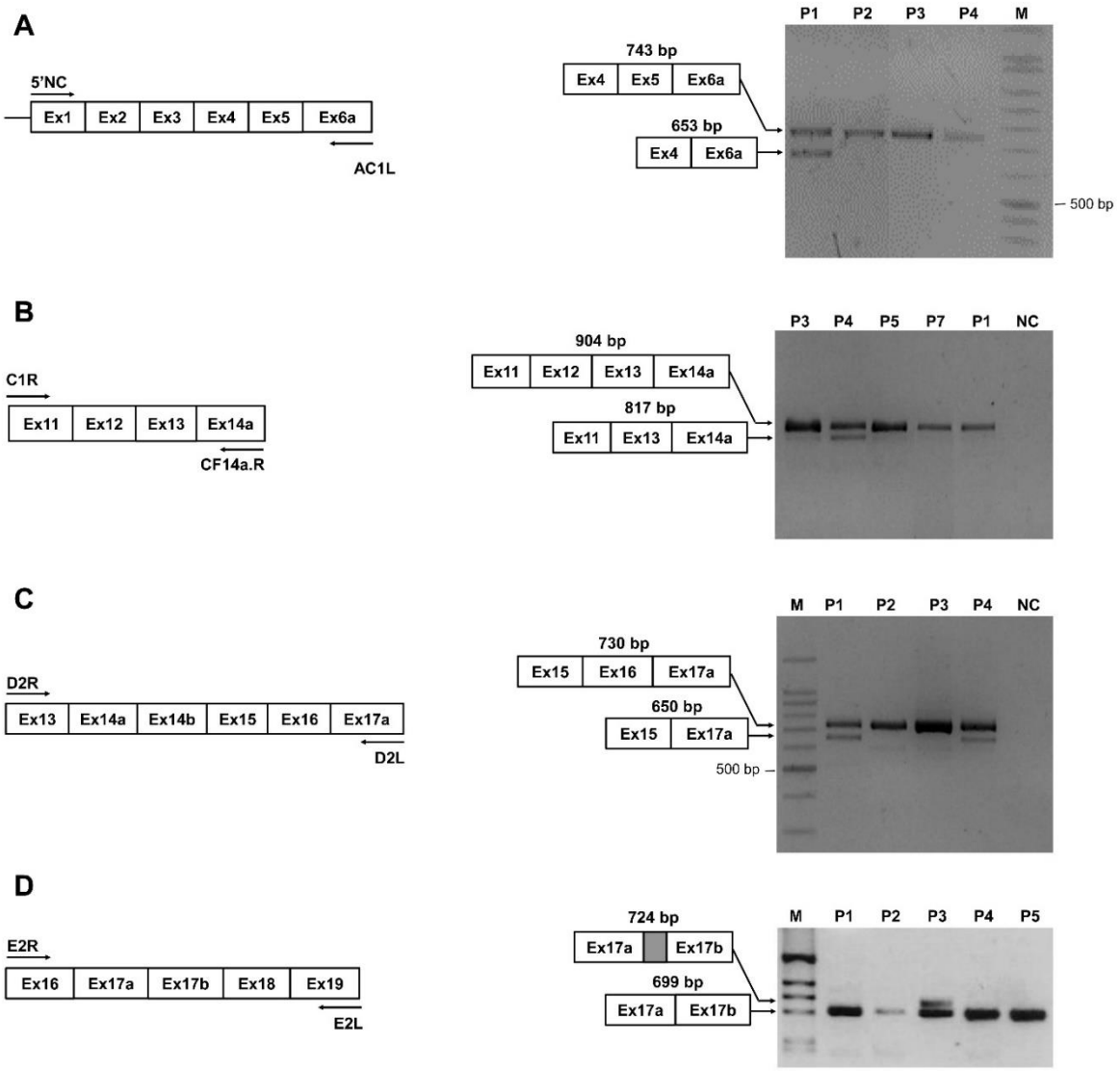



**Figure 3.17 - Application of the polymerase chain reaction (RT-PCR) strategy to successfully detect multiple splicing mutations.** Examples of four out of the nine RT-PCR reactions to analyse cystic fibrosis transmembrane conductance regulator (CFTR) transcripts in the regions of: (A) exons 1-6a; (B) exons 11-14a; (C) exons 13-17a; and (D) exons 16-19. Images show PCR strategy with primers used (left panels) and respective data for the transcripts of individuals under study (right panels). **(A)** Detection of 711+3A>T (HGVS: c.579+3A>T). RT-PCR products in the exon 1-6a region evidence an additional lower band in P1, not occurring in P2-P4. Based on size, this is probably due to a mutation causing IVS5 skipping which was confirmed by gDNA sequencing (IVS5) identifying 711+3A>T in P1. **(B)** Detection of 1812-1G>A (HGVS: c.1680-1G>A). RT-PCR products in the exons 11-14 a region show a lower band in all individuals analysed, corresponding to IVS12 skipping. InP4, this lower band is of higher intensity that in the others, probably due to a mutation enhancing IVS12 skipping, as confirmed by sequencing of gDNA, identifying 1812-1G>A in P4. **(C)** Detection of 3120+1G>A (HGVS: c.2988+1G>A). Analysis of CFTR mRNA in the exon 13-17a region shows an additional lower band in both P1 and P4, which is absent in P2 and P3. Based on size, this is probably due to a mutation causing IVS16 skipping, as confirmed by sequencing of gDNA, which identified 3120+1G>A in both P1 and P4. **(D)** Detection of 3272-26A>G mutation (HGVS: c.3140-26A>G). RT-PCR products in the exon 16-19 region show an additional higher band in P3 which is absent in the others. By sequencing this band, it was found to have~25 extra nts (not shown) and gDNA sequencing (IVS17a) identified 3272-26A>G in P3. M-DNA molecular mass markers 100bp (a, c) or 1kb (b).

## 4. Discussion

In the era of mutation-specific therapies for CF, it is critical to determine the patients' genotype for adequate clinical management. Given its large size and the very high number of gene variants reported there is a need to establish robust, yet inexpensive methods to identify CFTR gene mutations.

CFTR genetic analysis panels, which should achieve a mutation detection rate>95%, fail to do so in populations with lower p.F508del incidence e.g. in Southern Europe, and even more so in those with high genetic heterogeneity, namely in South-America, Asia or Africa, making this goal extremely challenging. Even with the advent of NGS, complete sequencing of large disease-associated genes and intronic regions is economically not feasible. Often these approaches still fail to identify intronic mutations because only the exons and the intronic flanking regions are sequenced. Due to the large size of introns, it is not cost effective to sequence them fully and thus it was estimated that 1–5% of CF alleles remain to be determined (Groman et al., 2002). Consequently, there is an unmet need to genotype CF patients by fast, comprehensive and cost-effective approaches.

The simple, inexpensive mRNA-based protocol reported here covers the complete CFTR coding region in just nine RT-PCR reactions which, despite requiring sequencing are considerably less expensive than the usual 28 gDNA PCR reactions (Bonini et al., 2015). This approach also directly identifies aberrant transcripts resulting from intronic splicing mutations and gene deletions/insertions. Although gDNA sequencing should follow to identify the causative mutation, it will be limited to one (or two) intron(s). Using just one additional RT-PCR p.F508del-ARMS reaction, this approach also detects the most frequent CF mutation. One disadvantage, however, is the requirement of specific tissue (e.g. nasal or rectal). Nevertheless, collection of nasal cells is a simple procedure (Beck et al., 1999) which is becoming increasingly used.

The robustness of the approach was shown here by its application to successfully detect multiple mutations in eight different individuals with a suspicion of a CF diagnosis, including several in the introns, including for example the c.1584+18672A>G mutation, associated with classical CF (Costantino et al., 2010) which causes an insertion of a cryptic 104 bp exon between IVS10 and 11. Because this mutation lies deep within IVS10, it would be very unlikely to be detected by conventional approaches.

Other mutations identified included: 711+3A>T (HGVS: c.579+3A>T) causing IVS5 skipping and (when in trans with R334W) associated milder CF disease and pancreatic sufficiency (Casals et al., 1997); 1812-1G>A (HGVS: c.1680-1G>A) promoting IVS12 skipping and originally described in Spain in association with severe CF, PI and lung colonisation

(Chillón et al., 1994); 3120+1G>A (HGVS: c.2988+1G>A) abrogating IVS16 splicing is reported to be common in African/Afro-descendant patients (9–14% of CF chromosomes) in association with a severe clinical phenotype (Carles et al., 1996) and is also common in Brazil (Cabello et al., 2001); 3272-26A>G (HGVS: c.3140-26A>G) was reported in Portugal creating abnormal splicing in IVS17a and is associated with milder CF (Amaral et al., 2001).

## 5. Concluding remarks

Based on the successful detection of mutations in this cohort of CF patients, we conclude that the rapid, easy-to-perform mRNA-based method described here is a robust approach to detect rare coding or intronic CFTR mutations. It can be easily performed either as a first screen or after the routine DNA analysis. It is also significantly less expensive than NGS approaches. Furthermore, it already shows the functional consequences of mutations in case they occur at the RNA level.

## 6. Supplementary Information

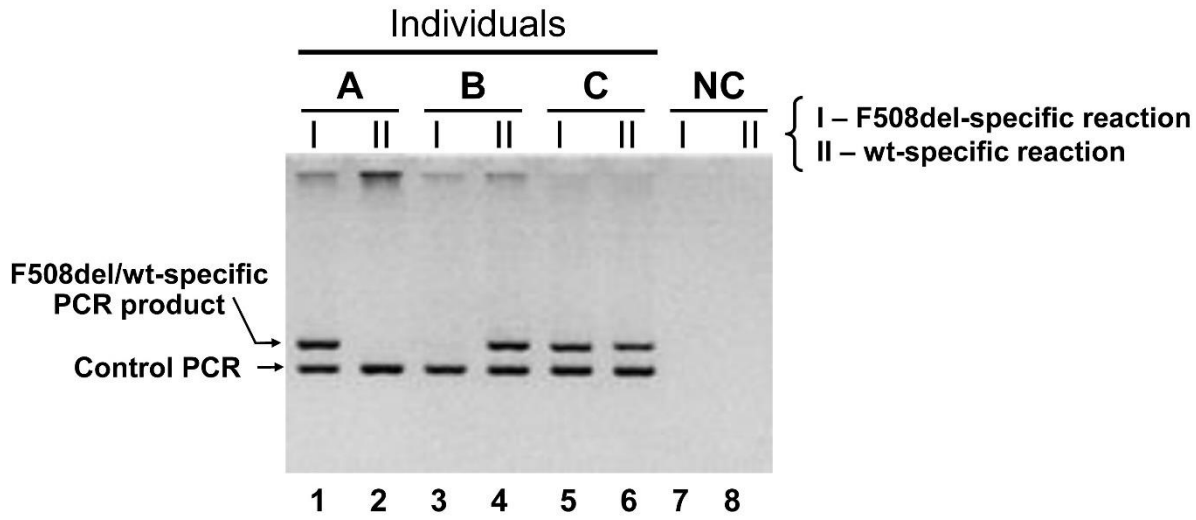



Figure S 3.1- **Representative results of the F508del-specific ARMS technique.** Analysis by 2% (w/v) agarose gel electrophoresis of F508del-specific ARMS products for 3 individuals with a suspicion of CF (A,B,C). For the cDNA of each individual, two PCR reactions were performed: I with the F508del-specific primer; II - with primer specific for the normal sequence at the F508del site (see also Table 2.4). In all reactions, an additional PCR product is generated ('Control PCR') to control for PCR amplification. Individual A (lanes 1,2) is homozygous for F508del mutation, since there was only specific amplification in reaction I (lane 1); individual B (lanes 3,4) does not have the F508del mutation, since there was only specific amplification in reaction II (lane 4); individual C (lanes 5,6) is heterozygous for F508del, since there was specific amplification in both reactions I (lane 5) and II (lane 6); NC (lanes 7,8) PCR negative control (water instead of cDNA).

# *Chapter 3 - A High-Throughput shRNA Screen to Identify Genes Rescuing CFTR Mutations Bearing Premature Termination Codons (PTCs).*

## 1. Abstract

**Background**: Nonsense mutations, which introduce premature stop codons (PTCs) account for about 8% of the ~2,000 CFTR gene variants reported in the CFTR Mutation Database. These mutations usually lead to extensive transcript degradation by nonsense-mediated mRNA decay (NMD) thus preventing protein production. PTC mutations are therefore usually associated with classical severe CF phenotypes. Several drugs promoting PTC suppression (or read-through) such as aminoglycoside antibiotics and Ataluren have shown some *in vitro* ability to suppress PTCs, albeit unspecifically, and show toxicity and low efficacy . Nevertheless, they have been under trial for CF, but were discontinued due to lack of demonstrated efficacy. There is thus an unmet need to treat patients with these mutations and novel therapeutic strategies should be explored.

**Aim:** Our aim is to identify novel genes which when knocked-down rescue either NMD of these transcripts or show PTC read-through (or both) and to validate them as novel drug targets.

**Methods:** To achieve this goal, we have used a CFTR mini-gene model (i.e., full-length CFTR cDNA with introns IVS14, 15 and 16; legacy name IVS13, 14a and 14b) harbouring the G542X mutation (at exon 12; legacy name 11), which recapitulates NMD process shown *in vivo*. The mini-gene features mCherry and eGFP tags at the CFTR N- and C-termini, respectively which allow the assessment of total CFTR protein levels (mCherry) as a readout of NMD and rescue of full-length protein (eGFP) as a readout of read-through. The construct also has a Flag-tag which becomes extracellular once the mCherry-Flag-G542X-CFTR-eGFP protein reaches the plasma membrane (which may not occur to all full-length proteins produced given the random insertion of amino acids in read-through). This construct was transfected into HEK293 Flp-In cells for stable expression, in parallel with a similar wt-CFTR construct. To validate this NMD reporter assay, we used the NMD inhibitor - SMG1 inhibitor ($Inh_{SMG1}$) as a positive control. The NMD correction was assessed by: *(i)* semi-quantitative RT-PCR; *(ii)* high-throughput (HT) microscopy analysis.

**Results and Discussion:** Our RT-PCR data show that levels of G542X-CFTR transcripts in non-treated and $Inh_{SMG1}$ treated cells were: 20% and 60%, respectively. By fluorescence microscopy we could not determine the presence of eGFP fluorescence in $Inh_{SMG1}$ treated cells because SMG1 inhibition only prevents NMD but has no read-through activity. However, we

observed a 5-fold increase in mCherry fluorescence, indicating a good dynamic range for detection of NMD suppression, thus validating the reporter assay for HT microscopy. Next, we used this CFTR-NMD reporter to identify novel NMD factors by screening a previously validated shRNA library – a subset of The RNAi Consortium (TRC) – which is enriched in shRNAs targeting genes with a previously described, or proposed involvement in transcript processing (425 genes). We selected 24 genes for additional validation: 11 genes with a NMD score ≥ 2 and more than 2 shRNAs with the same phenotype; 2 genes with a NMD score ≥ 2 and one shRNA that showed read-through activity; and 11 genes resulting from a HTS screen designed to identify CFTR splicing regulators (unpublished data). These splicing-related genes (S Igreja & MD Amaral, unpublished data) were also analysed because a relationship between NMD and splicing has been previously described. Our confirmation screen was inconclusive, since the NMD scores obtained were very low.

**Conclusion**: Some of the genes identified within this screen are potential drug targets due to their effects in inhibiting NMD, when knocked-down. To validate the 24 genes we are currently conducting validation assays including RT-PCR (qRT-PCR), and Western blot assays.

## 2. Introduction

Cystic fibrosis (CF) is the most common autosomal recessive disorder affecting approximately 85,000 individuals worldwide and around 34,000 individuals in Europe. Mutations in the CF Transmembrane conductance Regulator (CFTR) gene lead to absence or dysfunction of the CFTR protein, which is involved in the regulation of $Cl^-/HCO_3^-$ ion transport across the apical membrane at the surface of the epithelia. Mutated CFTR channels result in malfunctioning lungs, pancreas, and other organs. Indeed, defective CFTR protein leads to decreased airway surface liquid (ASL height) and dehydration, defective muco-ciliary clearance (MCC), mucus-clogged airways, recurrent bacterial infections, progressive loss of lung function with extensive lung damage, and eventually respiratory failure (Cutting, 2015; De Boeck and Amaral, 2016).

Until recently CF was just managed at the level of its symptoms, with patients being prescribed antibiotics, anti-inflammatory and mucolytics, to help alleviate infections, chronic inflammation and excessive production and thickness of mucus. However standard medical care is still far from providing appreciable levels of life quality for the patients.

Intensive drug discovery efforts have resulted in recent progress and approval of new CFTR-modulator therapies. These include CFTR correctors (lumacaftor and tezacaftor) that improve mutant protein folding and trafficking, and a potentiator (ivacaftor) that increases chloride channel activity. Approval of Kalydeko™ (ivacaftor) marked a milestone in CF therapeutics development, and with the advent of personalized medicine, these approaches could potentially revolutionize CF care and management (Amaral and Blach, 2015; Cutting, 2015; Schneider et al., 2017).

There are approximately 2,000 mutations described in the CFTR gene, comprising missense (39%), frameshift (16%), splicing (11%) and nonsense (8%) mutations; large (3%) and in-frame (2%) deletions or insertions; promotor mutations (1%); presumed no-pathological variants (13%) and unknown mutations (7%) (Cystic Fibrosis Mutation Database, 2017). According to their functional defects seven classes of mutations have been proposed (De Boeck and Amaral, 2016; Cystic Fibrosis Mutation Database, 2017). Most of these mutations cause CFTR protein defects by impairing protein translation, folding and stability, trafficking, and/or function.

However, many challenges remain in designing therapies for CF patients that present mutations in the *CFTR* gene which cannot be treated with the approved therapeutics (Fajac and Wainwright, 2017; Kerem, 2017; Marson, 2017). Among the ~2,000 CFTR gene variants reported, a significant fraction (~24%) generate PTCs, including: 8.4% nonsense and 15.7% frameshift mutations. PTC mutations lead to a significant reduction, or total absence, of normal

CFTR mRNA and protein expression, and are commonly associated with a severe clinical phenotype (Keeling et al., 2014a). Reduction in the levels of transcripts occurs via the mRNA surveillance mechanism termed Nonsense-Mediated mRNA Decay (NMD) which selectively degrades mRNAs harboring PTCs. To date, there are no CFTR modulator therapies to treat patients harboring PTC mutations, and although aminoglycoside antibiotics (used in the standard treatment of CF lung infections) and Ataluren/PTC-124 have shown some ability to suppress PTCs *in vitro,* they are unspecific, and show low efficacy. Nevertheless, they have been under clinical trials for CF, but were discontinued due to lack of demonstrated efficacy. As such, there is an unmet need to find novel therapeutic strategies to treat these patients.

NMD is an essential RNA quality control and gene regulatory mechanism conserved among eukaryotes. This surveillance mechanism detects and ensures the rapid degradation of PTC-harbouring mRNAs that could potentially result in the synthesis of C-terminally truncated proteins. Hence, the quality control function of NMD is centered on protecting the cell from expressing deleterious dominant-negative or gain-of-function (toxic) truncated proteins. Additionally, this regulatory mechanism also plays an important role in modulating the phenotypes resulting from diseases caused by PTCs (Linde et al., 2007; Kervestin and Jacobson, 2012; Peixeiro et al., 2012). The efficiency of the NMD pathway can vary by as much as 4-fold among the general population. Importantly, these differences in NMD efficiency may influence and modulate the clinical severity of numerous disorders, possibly due to changes in mRNA abundance, which influences the levels of truncated protein which may still retain residual function (or full-length protein produced by basal PTC read-through). (Khajavi et al., 2006; Keeling et al., 2014b). Furthermore, variable levels of NMD among patients can also influence the effectiveness of nonsense suppression therapies. Patients who express higher levels of PTC-containing mRNA, due to less efficient NMD, are expected to respond more robustly to suppression/read-through therapies, compared to patients who express lower levels of PTC-mRNA due to highly efficient NMD (Wilschanski et al., 2003; Linde et al., 2007).

Based on its role in regulating the abundance of PTC-containing transcripts, the NMD pathway represents a potential therapeutic target for diseases caused by PTCs. Numerous regulatory subunits have been implicated in the NMD pathway, including the core NMD factors UPF1, UPF2, and UPF3 as well as other factors that influence the phosphorylation status of UPF1, such as SMG-1, SMG-5, SMG-7, SMG-8, and SMG-9 (Kervestin and Jacobson, 2012; Schweingruber et al., 2013; Yamashita, 2013; Lykke-Andersen and Heick Jensen, 2015). Therefore, it seems plausible that increasing the level of PTC-containing mRNAs, by inhibiting NMD through the use of drugs that target NMD factors, would be a good therapeutic strategy. This, could potentially alleviate some disease phenotypes by increasing the expression of truncated polypeptides that could still retain some residual function, but especially in

combination with agents promoting read-through of the PTCs which would enable the production of "full-length" functional proteins (Keeling et al., 2014b; Mutyam et al., 2016; Liang et al., 2017).

Herein, our aim was to identify novel factors that modulate NMD of CFTR transcripts bearing PTCs. For this purpose, we have generated a novel triple-tagged CFTR NMD-PTC read-through reporter mini-gene (see section II – Material and Methods) which enables the detection of both NMD inhibition and PTC read-through. This mini-gene model was then used to screen a library of 425 genes, with a known or predicted involvement in transcript processing (Moura-Alves et al., 2011). The screening of these 425 genes yielded 11 hits. Using two different criteria, we have identified 11 top hits, i.e. genes which when suppressed inhibit NMD. Some of these NMD inhibitors were selected for further confirmation in an independent siRNA screen.

## 3. Results

### 3.1 Development and validation of stable cell lines for the HTS assay

To identify factors involved in NMD, which may be potential therapeutic targets, we have generated an NMD-PTC CFTR mini-gene model (i.e., full-length CFTR cDNA with mCherry at the C-terminus and eGFP at the N terminus plus introns IVS14, 15 and 16; legacy name IVS13, 14a and 14b), similar to a previously reported gene model (Igreja et al., 2016) but containing the NMD inducing G542X mutation in IVS12 (Figure 3.18). This mini-gene construct also bears a flag-tag at the CFTR 4$^{th}$ extracellular loop, which becomes extracellular once the mCherry-Flag-G542X-CFTR-eGFP protein reaches the plasma membrane (which may not occur to all full-length protein produced given the random insertion of amino acids in read-through). Due to the presence of G542X, this construct leads to a lower production of CFTR transcripts (Figure 3.19A).

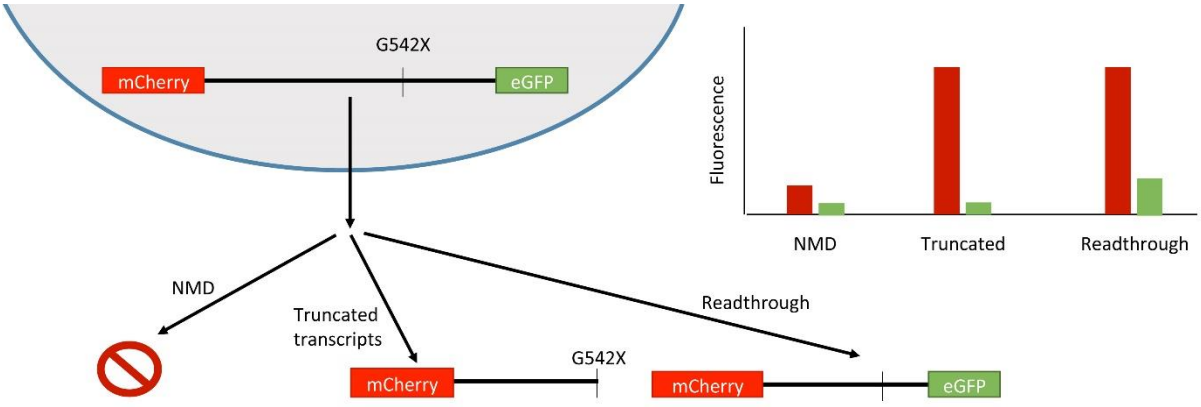



**Figure 3.18 - Schematic representation of three possible assay readouts**. Possible results obtained from the mCherry-GX-eGFP construct. Construct without any treatment leads to almost no fluorescence, since we have degradation by NMD. Inhibition of NMD leads to a truncated protein,

since we still have G542X mutation and the transcripts stop at this point. Full-length protein with mCherry and eGFP fluoresce once no NMD occurs and we have read-through the mutation.

To validate our NMD report, we first analysed the construct's performance by Western blot (WB) and semi-quantitative RT-PCR to ensure that the introduction of mCherry and eGFP with the three introns did not influence folding and traffic to the membrane (Figure 3.19) or lead to the production of a truncated protein in the wt construct.

For WB validation, HEK Flp-In cells stably transfected with the mini-gene were treated with DMSO for 48 h and protein was extracted. For semi-quantitative validation, cells stably transfected were treated with DMSO and SMG-1 inhibitor, a previously described NMD inhibitor (Yamashita, 2013) (kindly provided by M. Mense, CFFT labs) which was used as a positive control. SMG-1 is a member of the phosphoinositide 3-kinase-related kinase (PIKK) family that plays a critical role in NMD. SMG-1 directly phosphorylates UPF1 helicase, a key component of NMD, upon recognition of a PTC in the spliced mRNA, during the initial round of translation (Kervestin and Jacobson, 2012; Celik et al., 2015; Fatscher et al., 2015; Lykke-Andersen and Heick Jensen, 2015).

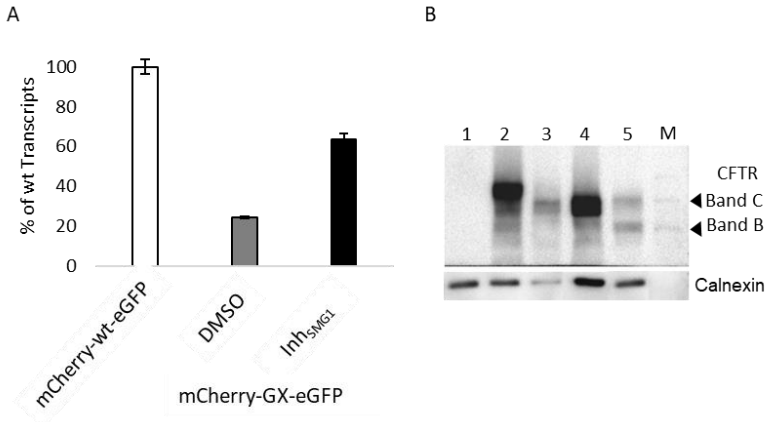



**Figure 3.19 - Representative characterization of the wt and mutant NMD constructs by semi-quantitative RT-PCR and Western blot. (A)** Quantification of CFTR transcripts by semi-quantitative RT-PCR for the wt-mCherry-wt-eGFP and mCherry-GX-eGFP constructs after cells treated with DMSO and Inh$_{SMG1}$. **(B)** WB shows the mCherry-GX-eGFP construct (lane 1) resulting in no CFTR protein production, as expected, due to the presence of the G542X mutation; mCherry-wt-eGFP (lane 2) showing both the core-glycosylated (band B) and the fully-glycosylated (band C) forms; and three different wt-CFTR constructs of smaller size: complete wt-CFTR-eGFP cDNA with eGFP in C-terminal (lane 3); complete cDNA mCherry-wt-CFTR with IVS's 14, 15 and 16 and m-Cherry in N-terminal (lane 4); and finally complete cDNA CFTR wt with intron 14, intron 15 and intron 16 (lane 5), all evidencing the presence of both band B and band C.

We can conclude that all wt mini-genes produce normal CFTR protein, i.e., with both fully glycosylated (band C) and core-glycosylated (band B) forms. Additionally, it was possible to confirm that no normal protein was produced from the G542X mutant, since no CFTR protein was detected.

After confirmation and validation of the constructs by RT-PCR and WB, we analysed the expression of mCherry and eGFP tags in both constructs under DMSO or $Inh_{SMG1}$ conditions by fluorescence microscopy to validate the assay (Figure 3.20). To this end, HEK Flp-In cells stably transfected with the construct were treated with DMSO and $Inh_{SMG1}$ and were imaged after Hoescht staining (see Materials and Methods, section 4.2).

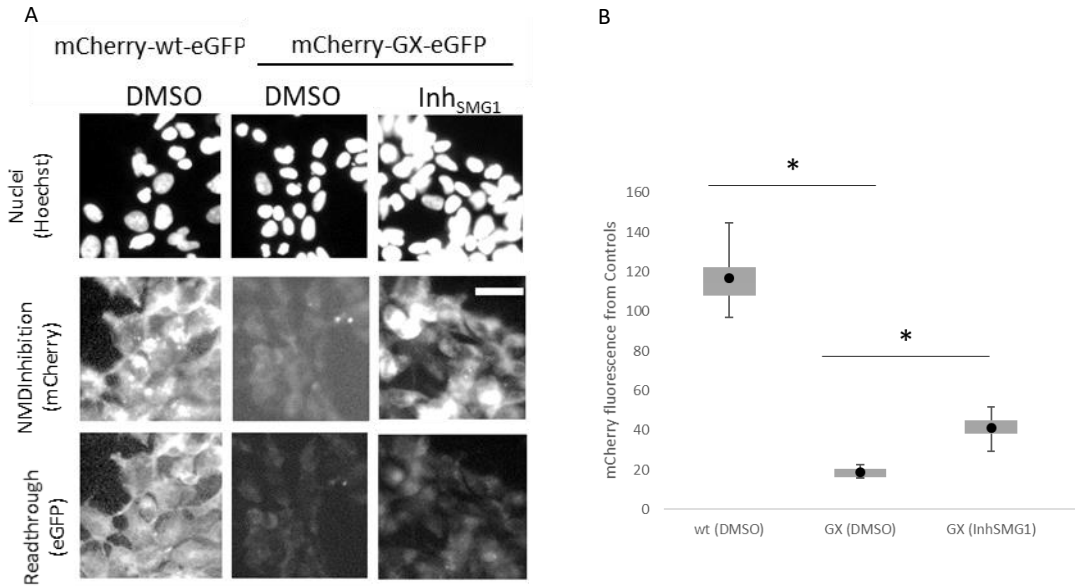



**Figure 3.20** - **Representative microscopy images of the NMD/PTC assay under control conditions. (A)** Microscopy images of the stable mCherry-wt-eGFP (left column) and the mCherry-GX-eGFP HEK cells (middle and right columns) treated with DMSO (left and middle columns) or with the $Inh_{SMG1}$ compound, an NMD inhibitor used as positive control (right column). Images show: Hoechst staining of the nuclei (top row), mCherry (middle row) and eGFP (bottom row). Scale bar=25µm. **(B)** Quantification of the data in A (n= 3). * means $p < 0.05$.

The expression of the mCherry-GX-eGFP NMD reporter was noticeable after 48 h of incubation, being significantly higher in cells treated with $Inh_{SMG1}$ vs control cells, i.e., under DMSO (Figure 3.20). These results show the suitability of our model reporter to study NMD and its sensitivity to detect the effects of the inhibition of factors that influence NMD. Quantification of the data shown in Figure 3.20B revealed a 119% increase in the average fluorescence intensity of the positive control versus DMSO, which was found to be statistically significant, thus indicating that this assay had an adequate dynamic range to be used in a HT screen.

### 3.2 HT shRNA Primary screen

To identify putative novel factors that inhibit NMD, which can be used to counteract the degradation of transcripts bearing PTCs, we used the above model in a high-throughput screen (HTS) of a shRNA library (NMD primary screen) containing factors with a known or predicted involvement in mRNA processing (Moura-Alves et al., 2011). The model was also

used to screen an siRNA library with 24 genes selected from the primary screen (NMD confirmation screen). Finally, the same cellular model was used to validate the most relevant targets using two different techniques: WB and qRT-PCR (Figure 3.21).

The primary screen was performed using a library of 425 genes (Table S 4.3) with an average of 5 shRNAs per gene in a total of 1967 shRNAs. To analyse this library, we used HT microscopy and image quantification, the latter was performed using CellProfiler. Statistical analysis was performed using shinyHTM (See Materials & Methods – section 4). After image acquisition we used CellProfiler and shinyHTM to apply a quality control (QC) to each well and each cell, in order to exclude out of focus images, images with high background fluorescence, and also to discard samples in which cells had too large or too small nuclei, or cellular damage, as well as apoptotic and mitotic cells (See Material & Methods – section 4.3). Each image was characterized by the median value of "NMD inhibition" and after averaging the NMD Inhibition score for all images relating to the same shRNA and which passed QC, the effect of each shRNA towards NMD Inhibition was expressed as a deviation score versus the appropriate negative control. The distribution of scores for each shRNA is shown in Figure 3.22. Of these, 75 shRNAs were shown to inhibit NMD with an NMD score ≥ 2 (Table S 3.1). By applying more stringent criteria, namely an NMD score ≥ 2 and at least 2 different shRNAs targeting the same gene (Figure 3.23 and Table S 3.1), as previously described (Moura-Alves et al., 2011), 11 genes were identified as top hits (Table 3.17). These were selected for further confirmation. Unexpectedly, some of the genes which when knocked-down inhibited NMD, also promoted read-through activity (Figure 3.24 and Table S 3.1). Therefore, we also selected two of these genes (that have an NMD score ≥ 2 but just 1 shRNA with the same phenotype) for further confirmation (Table 3.17). The complete list of genes and respective scores are shown in Table S 3.1.

After hit selection we performed a confirmation screen using siRNAs for the 13 genes selected and we also included 11 genes which were identified in a parallel screen for splicing-associated genes (Figure 3.21, Table 3.18 and refer to section 3.3). We included these genes because the splicing process is closely associated with NMD, as such we thought it would be relevant to determine their effect on NMD. Some of the splicing hits were also identified in the NMD screen, however these either failed to pass the QC analysis or did not meet the restricted criteria. In summary, we selected 24 genes to be validated through a confirmation siRNA screen. Notwithstanding, the results of this confirmation screen were inconclusive, since the NMD scores obtained showed very low values possibly due to low transfection efficiency. Nevertheless, we selected 4 genes with higher values (Table S 3.2). Since the confirmation screen was inconclusive, we are currently performing validation studies for the

24 genes using different approaches including qRT-PCR and WB (data not shown; ongoing work performed by S Igreja, L Clarke & MD Amaral).

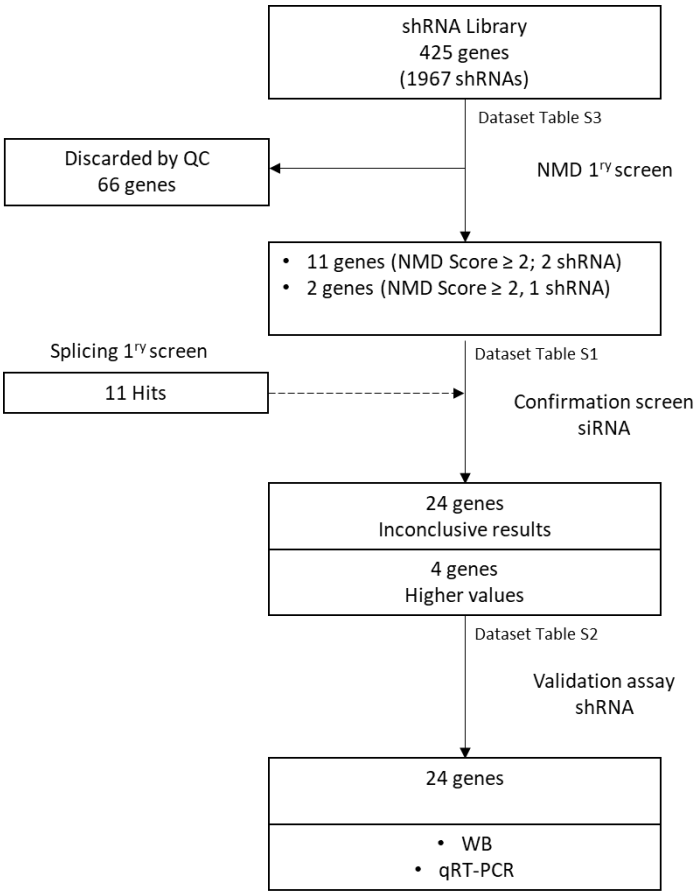



**Figure 3.21 – Workflow diagram of the study to identify novel genes which inhibit NMD or show PTC read-through.** Data from the NMD primary screen using a shRNA library, a siRNA screen i.e. confirmation screen using top hits from the NMD primary screen and from a splicing primary screen (performed by Igreja S. & Amaral MD). Ongoing validation studies using WB and qRT-PCR. See text for details.

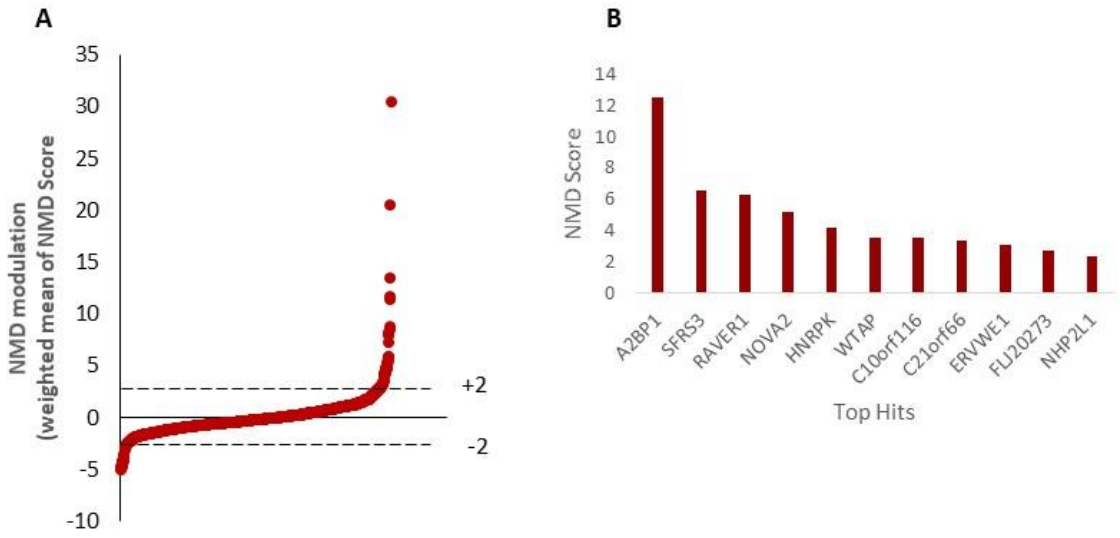


**Figure 3.22 - Distribution of the shRNA scores for HTS NMD assay. (A)** The distribution is determined according to the individual shRNA deviation from the negative control. NMD inhibitors are identified by having a 'NMD score' above +2, as described in Materials and Methods. **(B)** Distribution of top hits with a NMD score ≥ 2 and at least 2 shRNAs with the same phenotype.

The screen was performed using the HEK Flp-In cell line stably transfected with the G542X-mCherry-Flag-eGFP construct (abbreviated as mCherry-GX-eGFP cells). Analysis was performed using mCherry signal measured after 48h of incubation with the shRNAs. The N-terminal mCherry allows the quantification of the total CFTR protein being expressed in individual cells and its integrated fluorescence signal in each cell is inversely proportional to the extent of NMD and thus proportional to NMD inhibition. This is defined as the amount of CFTR transcripts that escape NMD and give rise to a polypeptide (independently of whether it is truncated or not). NMD inhibition can then be determined in individual cells using the mCherry fluorescence:

$$NMD\ Inhibition \propto \text{mCherry Integrated Fluorescence}$$

The $Inh_{SMG1}$ was used as a positive control and a 'scrambled' shRNA as negative control. After nuclei staining, image acquisition was performed in a Leica DMI 6000B widefield epifluorescence microscope. Fluorescence intensities were measured using CellProfiler, and statistical analysis was performed in shinyHTM (see Materials and Methods). Results are presented according to the NMD score:

$$NMD\ Score = \frac{\text{NMD Inhibition}_i - \text{NMD Inhibition}_{NegCtrl}}{SEM_{\text{NMD Inhibition},Neg\ Ctrl}}$$

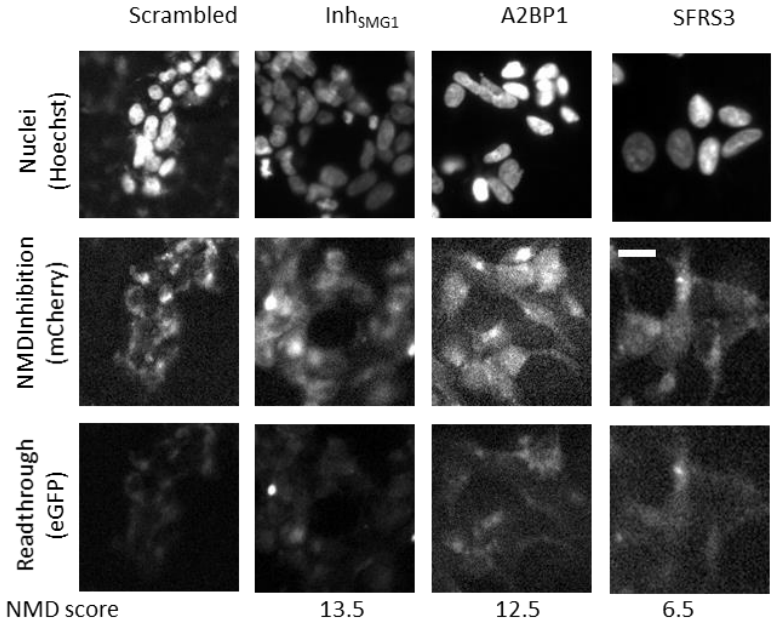



**Figure 3.23** - **Representative images for two top shRNA hits inhibiting NMD**. Representative microscopy images obtained from the stable mCherry-GX-eGFP HEK cells, 48h after being infected

with the library of shRNAs. Images show cells treated with: Scrambled shRNA (negative control, left column), $Inh_{SMG1}$ (positive control middle left column) and two hit shRNAs that inhibit NMD (middle right and right columns) as seen by the increased mCherry fluorescence (middle row). NMD score values for NMD are shown underneath. Scale bar = 20µm.

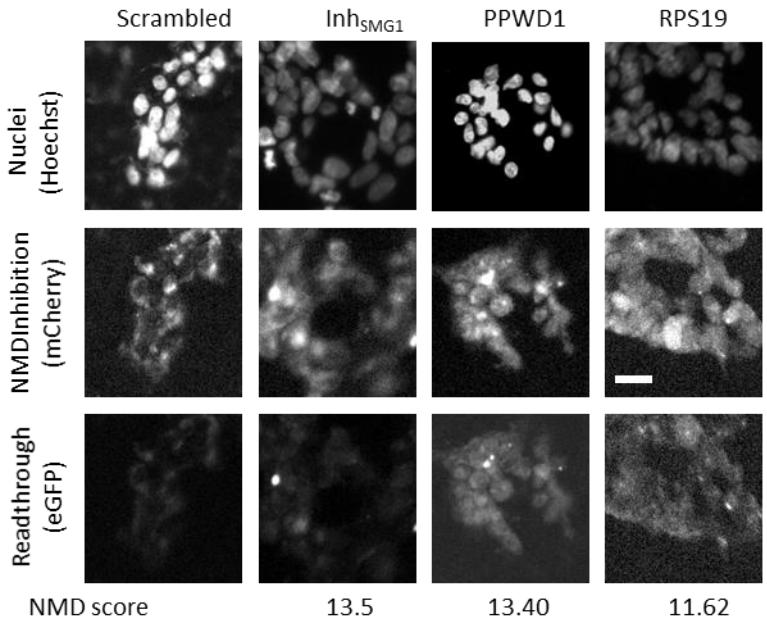



**Figure 3.24 - Representative images for two top shRNA hits exhibiting read-through activity.** Representative microscopy images obtained for the stable mCherry-GX-eGFP HEK cells 48h after being infected with the shRNA library. Images show cells treated with: Scrambled shRNA (negative control, left column), $Inh_{SMG1}$ (positive control, middle left column) and two shRNA hits both inhibiting NMD and enhancing read-through activity (middle right and right columns) as seen by the enhanced mCherry (middle row) and eGFP (bottom row) fluorescence. The respective NMD score values for genes that have read-through activity are shown underneath. Scale bar = 20µm.

**Table 3.17 - Top shRNA hits for NMD with NMD score ≥ 2 and targeted by two or more shRNAs***

| Gene | NMD Score | Nº of shRNA | Gene Name** | Description |
|---|---|---|---|---|
| FLJ20273 | 2.75 | 3 | RNA-binding protein, isoform CRA_d | Basic machinery for C to U RNA editing |
| A2BP1 | 12.5 | 3 | RNA binding protein fox-1 homolog 1 | RNA binding protein, alternative splicing events by binding to 5-UGCAUGU-3 elements |
| NOVA2 | 5.20 | 2 | RNA-binding protein Nova-2 | NOVA alternative splicing regulator 2 |
| C21orf66 | 3.40 | 2 | PAX3 and PAX7 binding protein 1 | Involvement in the regulation of transcription |
| NHP2L1 | 2.30 | 2 | NHP2-like protein 1 | SNU13 homolog, small nuclear ribonucleoprotein (U4/U6.U5) |
| WTAP | 3.60 | 2 | Pre-mRNA-splicing regulator WTAP | Efficiency of mRNA splicing, processing and mRNA stability |
| C10orf116 | 3.60 | 2 | Adipogenesis regulatory factor/ADIRF | promotes adipogenic differentiation and stimulates transcription initiation |
| ERVWE1 | 3.10 | 2 | Syncytin-1 | Endogenous retrovirus group W member 1 |
| HNRPK | 4.20 | 2 | Heterogeneous nuclear ribonucleoprotein K | One of the major pre-mRNA-binding proteins |
| RAVER1 | 6.30 | 2 | Ribonucleoprotein PTB-binding 1 | Cooperates with PTBP1 to modulate regulated alternative splicing events |
| SFRS3 | 6.60 | 2 | Serine/arginine-rich splicing factor 3 | Splicing factor that specifically promotes exon-inclusion during alternative splicing |
| PPWD1*** | 13.42 | 1 | Peptidylprolyl isomerase domain and WD repeat-containing protein 1 | May be involved in pre-mRNA splicing. |
| RPS19*** | 11.62 | 1 | 40S ribosomal protein S19 | This gene encodes a ribosomal protein that is a component of the 40S subunit. |

**Identification of 11 hits after application of both criteria*
*** Gene Name from UniProt Consortium* (UniProt, 2018)
**** Knocked-down genes that lead to read-through activity, (NMD score ≥ 2 but just 1 shRNA with the same phenotype)*

The top 11 genes that when knocked-down lead to NMD inhibition are listed in Table 3.17. Here we provide a brief description of each of them.

**FLJ20273/RBM47** (RNA-binding protein, isoform CRA_d) constitutes the basic machinery for C to U RNA editing (Fossat et al., 2014). RBM47 encodes a protein that binds molecules of messenger RNA.

**A2BP1/ RBFOX1** (RNA binding protein fox-1 homolog 1) is an RNA-binding protein that regulates alternative splicing events by binding to 5-UGCAUGU-3 elements. It was shown to regulate alternative splicing of tissue-specific exons and of differentially spliced exons during erythropoiesis (Kuroyanagi, 2009; Gene Cards, 2018).

**NOVA2** (NOVA Alternative Splicing Regulator 2) is a protein coding gene. Gene Ontology (GO)annotations related to this gene include nucleic acid binding and RNA binding (Gene Cards, 2018).

**C21orf66/ PAXBP1** (PAX3 and PAX7 Binding Protein 1) encodes a protein that may bind to GC-rich DNA sequences, which suggests its involvement in the regulation of transcription (Gene Cards, 2018).

**NHP2L1/SNU13** (Small Nuclear Ribonucleoprotein 13) is a protein coding gene (Gene Cards, 2018) that binds to the 5-stem-loop of U4 snRNA and may play a role in the late stage of spliceosome assembly. The protein undergoes a conformational change upon RNA-binding (UniProt, 2018).

**WTAP** (WT1 Associated Protein) is a protein coding gene that acts as an mRNA splicing regulator (Gene Cards, 2018).

**C10orf116/ADIRF** (Adipogenesis Regulatory Factor) is a protein coding gene that plays a role in fat cell development by promoting adipogenic differentiation and stimulates transcription initiation. The protein is an adipocyte lineage-specific nuclear factor which regulates master adipogenesis transcription factors during early differentiation. However, the precise functional properties of this gene have yet to be identified. According to its GO it seems to be involved in distinct biological processes, including: cell differentiation; cellular response to cisplatin and positive regulation of transcription from RNA polymerase II promoter .(UniProt, 2018).

**ERVWE1** (Endogenous Retrovirus Group W Member 1, Envelope) is a protein coding gene, which encodes an envelope glycoprotein (syncytin) with a likely involvement in trophoblast differentiation.

**HNRPK/HNRNPK** (Heterogeneous Nuclear Ribonucleoprotein K) is a protein coding gene and is one of the major pre-mRNA-binding proteins that binds tenaciously to poly(C) sequences (UniProt, 2018).

**RAVER1** (Ribonucleoprotein, PTB Binding 1) is a protein coding gene that cooperates with PTBP1 to modulate regulated alternative splicing events and also promotes exon skipping (Gene Cards, 2018).

**SFRS3** (Serine and Arginine Rich Splicing Factor 3) gene is a member of the serine/arginine (SR)-rich family of pre-mRNA splicing factors, which constitute part of the spliceosome (Gene Cards, 2018) and functions as a splicing factor that specifically promotes exon-inclusion during alternative splicing (UniProt, 2018).

Two of the hits that lead to PTC read-through (Figure 3.23) - PPWD1 (Peptidylprolyl isomerase domain and WD repeat-containing 1) and RPS19 (40S ribosomal protein S19) have been previously implicated in the NMD process.

**PPWD1** was found to be part of the catalytically competent form of the spliceosome C complex (Jurica and Moore, 2003). It has been associated with the mRNA splicing pathway and it was shown to interact with two major proteins in the NMD process (Gene Cards, 2018).

**RPS19** is related to the NMD pathway and is required for pre-rRNA processing and maturation of the 40S ribosomal subunits (Fabregat et al., 2017; Gene Cards, 2018; UniProt, 2018). According to Project (Fabregat et al., 2017), th3 RPS19 protein is related to NMD and is part of 6 complexes (i.e. SMG1:UPF1:EJC translated mRNP; UPF1:eRF3 Complex on translated mRNA) participating in 5 different reactions (i.e. Formation UPF1:eRF3 complex on mRNA with a premature termination codon). This protein is also related with MAGOH and eIF4A3, two important components of EJC (Figure 3.25).

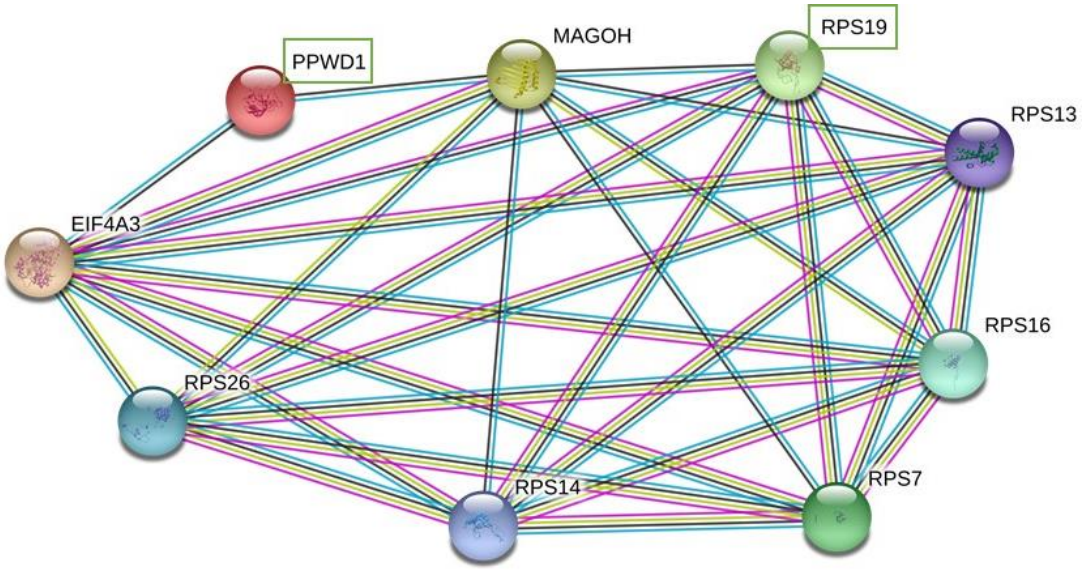



**Figure 3.25 - Representative image of the relationship between genes with read-through activity and NMD factors**. PPWD1 and RPS19 represent a relationship with two very important factors for NMD to occur, i.e., MAGOH and EIF4A3 (Da Costa et al., 2017; Gene Cards, 2018).

### 3.3 Hit confirmation by secondary screen - siRNA confirmation screen

The top hits from the NMD screen, i.e., those with an 'NMD Score' above +2 and targeted by at least 2 shRNAs with the same phenotype (Figure 3.23), were further validated using siRNAs targeting these genes. 13 hit genes from the primary screen were included in the confirmation screen (Table 3.17). In addition, we included siRNAs targeting 11 additional genes (Table 3.18) which were identified in a parallel HT screen for splicing associated genes (S Igreja & MD Amaral, unpublished data). We analysed these genes, because a relationship between splicing and NMD has been previously described with certain factors involved in splicing being also required for NMD to occur. Indeed, it is well known that many genes involved in the splicing process are also fundamental for NMD (Kashima et al., 2006; Moura-Alves et al., 2011; Da Costa et al., 2017; Iwatani-Yoshihara et al., 2017).

**Table 3.18 – Genes identified by the HTS shRNA splicing screen* that were assessed for their effects on NMD**

| Gene | NMD Score | Splicing Score | Gene Name** | Description |
|---|---|---|---|---|
| EIF4A3/DDX48 | -- | 3.8 | Eukaryotic initiation factor 4A-III | Involved in pre-mRNA splicing as component of the spliceosome |
| SREK1/SFRS12 | - 0.62 | 3.3 | Splicing regulatory glutamine/lysine-rich protein 1 | Participates in the regulation of alternative splicing |
| HNRPU | 0.45 | 3.3 | Heterogeneous nuclear ribonucleoprotein U-like protein 1 | Acts as a basic transcriptional regulator. |
| XRCC6 | -0.76 | 3.1 | X-ray repair cross-complementing protein 6 | Single-stranded DNA-dependent ATP-dependent helicase. |
| CCNA1 | 3.27 | 3.1 | Cyclin-A1 | Involved in the control of the cell cycle |
| RPL23A | -- | 3.0 | 60S ribosomal protein L23a | Component of the ribosome, binds a specific region on the 26S rRNA. |
| TPR | -0.33 | 2.9 | Nucleoprotein TPR | Component of the nuclear pore complex (NPC) |
| GCFC2/ C2orf3 | -- | 2.8 | GC-rich sequence DNA-binding factor 2 | Factor that represses transcription |
| S100A9 | 0.26 | 2.7 | Protein S100-A9 | S100A9 is a calcium- and zinc-binding protein |
| SAFB2 | -- | 2.6 | Scaffold attachment factor B2 | Binds to scaffold/matrix attachment region (S/MAR) DNA. |
| ZRSR2/U2AF1L2 | -0.91 | 2.5 | U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit-related protein 2 | Pre-mRNA-binding protein required for splicing of both U2- and U12-type introns. |

**Identification of 11 hits in an independent splicing screen (S Igreja & MD Amaral, unpublished data)*
*** Gene Name from UniProt Consortium* (UniProt, 2018)
*- - Genes discarded by QC in NMD Screen*

The genes included in the confirmation screen which were top hits in the splicing screen were also previously analysed in our NMD screen. However, they were excluded, either during the QC or because they did not meet the more restricted criteria used to identify the top hits. Here we provide a brief description of each of them.

**EIF4A3** (Eukaryotic initiation factor 4A-III) is an ATP-dependent RNA helicase, and a core component of the splicing-dependent multiprotein EJC deposited at splice junctions on mRNAs (Chan et al., 2004). Iwatani-Yoshihara and colleagues (Iwatani-Yoshihara et al., 2017) have shown in a recent study that eIF4A3 inhibitors supress NMD.

**SREK1** (Splicing regulatory glutamine/lysine-rich protein 1) encodes a member of a family of serine/arginine-rich (SR) splicing proteins containing RNA recognition motif (RRM) domains. The encoded protein interacts with other SR proteins to modulate splice site selection. It participates in the regulation of alternative splicing by modulating the activity of other splice factors (Gene Cards, 2018; UniProt, 2018).

**HNRNPUL1** (Heterogeneous Nuclear Ribonucleoprotein U Like 1) is a protein coding gene. An important paralog of this gene is HNRNPU. It acts as a basic transcriptional regulator, represses basic transcription driven by several virus and cellular promoters, and plays a role in mRNA processing and transport. This protein binds avidly to poly(G) and poly(C) RNA homopolymers in vitro.

**XRCC6** (X-Ray Repair Cross Complementing 6) is a DNA repair gene and together with XRCC7 forms an integral part of the double strand break repair (DSBR) pathway. The two genes are thought to play an important role in the repair of lethal double strand damage on DNA. The complex functions as a single-stranded DNA-dependent ATP-dependent helicase (Singh et al., 2018; UniProt, 2018).

**CCNA1** (Cyclin A1) the protein encoded by this gene belongs to the highly conserved cyclin family, whose members are characterized by a dramatic periodicity in protein abundance through the cell cycle. Cyclins function as regulators of CDK kinases. This cyclin binds both CDK2 and CDC2 kinases, which have two distinct kinase activities, one appearing in S phase, the other in G2, and thus regulate separate functions in cell cycle (Gene Cards, 2018; UniProt, 2018).

**RPL23A** (Ribosomal Protein L23a) this gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L23P family of ribosomal proteins. It is located in the cytoplasm. The protein may be one of the target molecules involved in mediating growth inhibition by interferon. In yeast, the corresponding protein binds to a specific site on the 26S rRNA. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome (Moin et al., 2017; Gene Cards, 2018; UniProt, 2018).

**TPR** (Translocated Promoter Region, Nuclear Basket Protein) this gene encodes a large coiled-coil protein that forms intranuclear filaments attached to the inner surface of nuclear pore complexes (NPCs). The protein directly interacts with several components of the NPC. It is required for the nuclear export of mRNAs and some proteins (Gene Cards, 2018; UniProt, 2018).

**GCFC2** (GC-Rich Sequence DNA-Binding Factor 2) is a factor that represses transcription. It binds to the GC-rich sequences (5-GCGGGGC-3) present in the epidermal growth factor receptor, beta-actin, and calcium-dependent protease promoters. It is involved in pre-mRNA splicing through the regulation of spliceosome C complex formation. It may also play a role during late-stage splicing events and turnover of excised introns (Gene Cards, 2018; UniProt, 2018).

**S100A9** (S100 calcium binding protein A9) is a calcium- and zinc-binding protein which plays a prominent role in the regulation of inflammatory processes and immune response. The protein is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins are localized in the cytoplasm and/or nucleus of a wide range of cells and are involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. S100 genes include at least 13 members which are located as a cluster on chromosome 1q21. This protein may function in the inhibition of casein kinase and altered expression of this protein is associated with cystic fibrosis (Kerkhoff et al., 1999; Gene Cards, 2018; UniProt, 2018).

**SAFB2** (Scaffold Attachment Factor B2) the protein encoded by this gene, along with its paralog (scaffold attachment factor B1), is a repressor of estrogen receptor alpha. The encoded protein binds scaffold/matrix attachment region (S/MAR) DNA and is involved in cell cycle regulation, apoptosis, differentiation, stress response, and regulation of immune genes. SAFB2 can function as an estrogen receptor corepressor and can also inhibit cell proliferation. Recently, the multifunctional capabilities of SAFB proteins have shown that they play crucial roles in DNA repair, mRNA processing and RNA regulation, as well as in interactions with chromatin-modifying complexes (Norman et al., 2016; UniProt, 2018).

**ZRSR2** (Zinc Finger CCCH-Type, RNA Binding Motif And Serine/Arginine Rich 2) or U2AF1L2 (U2(RNU2) Small Nuclear RNA Auxiliary Factor 1-Like 2) encodes an essential splicing factor. It interacts selectively with the 3-splice site of U2- and U12-type pre-mRNAs and promotes different steps in U2 and U12 intron splicing. It is recruited to U12 pre-mRNAs in an ATP-dependent manner and is required for assembly of the prespliceosome, a precursor to other spliceosomal complexes. For U2-type introns, it is selectively and specifically required for the second step of splicing (Gene Cards, 2018; UniProt, 2018).

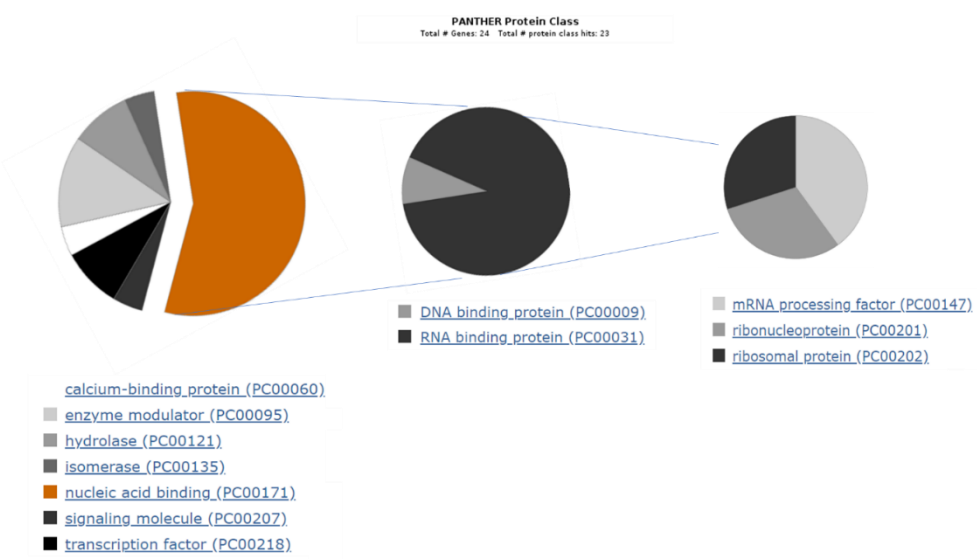
**Gene Ontology**

We have used Panther and DAVID programs to understand the interaction of these genes with NMD and to characterize them regarding protein class, molecular function, cellular components and biological process. Figure 3.26 shows the distribution of protein classes, gene ontology analysis (Biological Process (BP), Molecular Function (MF), Cellular components (CC)) and pathways for the 24 genes selected for validation. The top hits from the shRNA primary screen, as well as, the genes from the splicing screen were analysed by PANTHER gene ontology (GO) (Mi et al., 2013) and characterized by protein class, with these classes being mainly: RNA binding proteins (A2BP1; SREK1), mRNA splicing factors (RAVER1;NOVA2; HNRPK; SFRS3; RAVER1), ribosomal proteins (NHP2L1; RPS19; RPL23A), transcription factors (C21orf66; GCFC2), isomerase (PPWD1), DNA helicase

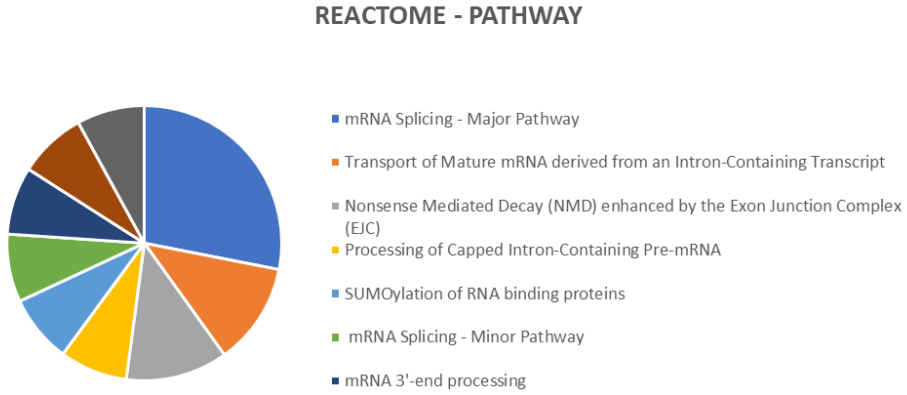
(XRCC6), Signalling molecule (S100A9), Ribonucleoprotein (ZRSR2), Kinase activator (CCNA1) and for 8 genes no function was defined(FLJ20273; WTAP; C10orf116; ERVWE1; eIF4A3; HNRNPUL1; SAFB2; TPR).

We separated our genes according to protein classes. The most evident class in the analysis of protein classes performed in PANTHER gene ontology (GO) is the Nuclei Acid binding class, with RNA binding protein as the most expressive in a subclass analysis. As for GO analysis, mRNA metabolic process has the highest percentage in BP, nucleic and binding is the most expressive in MF, and cell part, followed by nucleus, represent the locations with higher presence of the genes studied in this work.

**A**




**B**



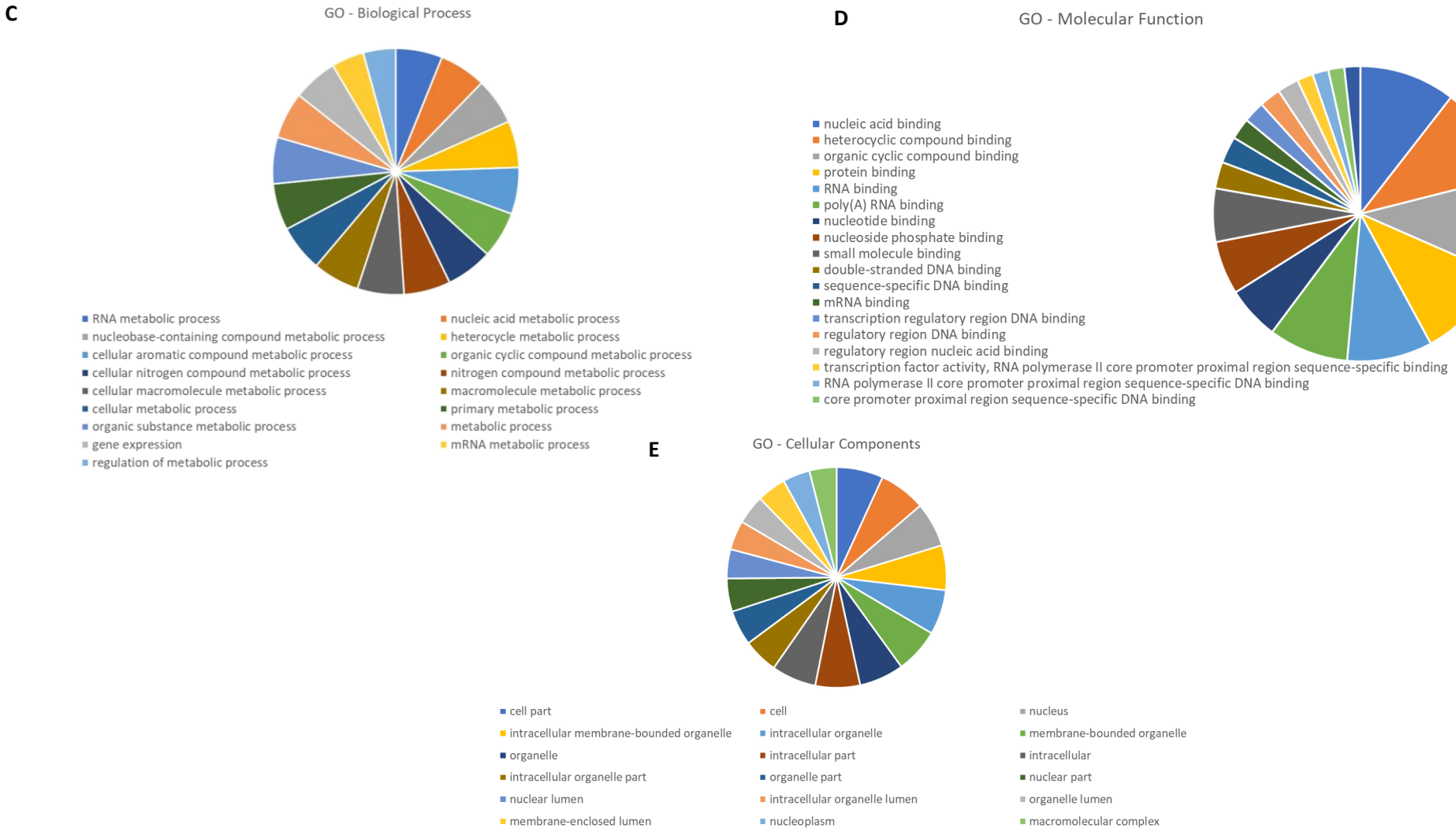



**Figure 3.26 – Schematic representation of protein classes, Biological Process, Molecular Function, Cellular components and pathway for 24 hits selected for validation. A)** Analysis of protein classes performed in PANTHER gene ontology (GO) (PANTHER, 2018) **B)** Pathway analysis performed using Reactome (Reactome, 2017). **C)** Biological Process **D)** Molecular Function and **E)** Cellular components identified in DAVID P<0.05

## Confirmation Screen

The siRNA confirmation screen consisted in reverse transfecting cells (see Materials and Methods) for 48h with siRNAs targeting the previously selected hit genes (hits from the NMD and splicing screens) as well as with UPF1 siRNA and SMG1 siRNA, which were used as positive controls, and 'scrambled' siRNA, which was used as a negative control. As a control for the transfection efficiency, an siRNA targeting CFTR was used (Figure 3.27).

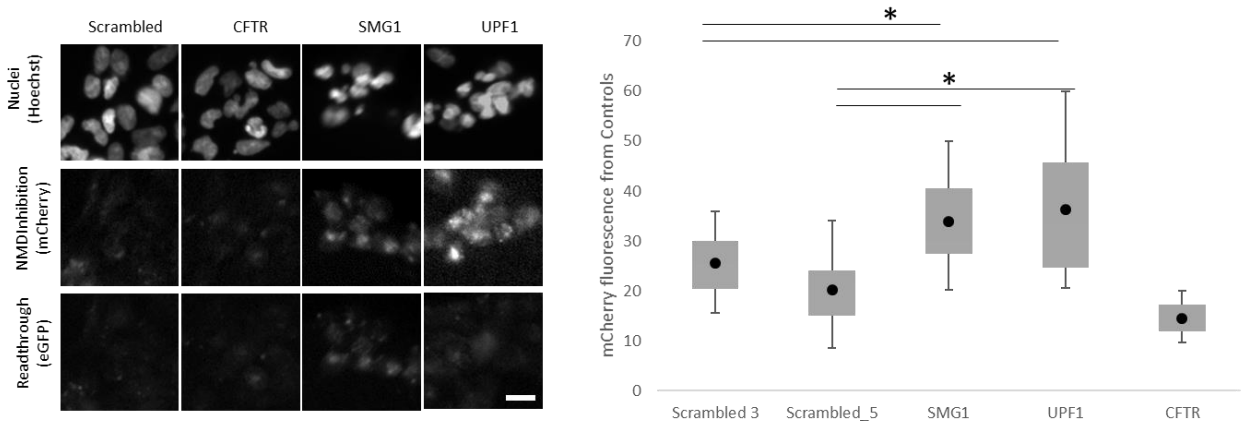



**Figure 3.27– Positive and negative siRNA controls for the NMD confirmation assay. (A)** Representative images obtained for the stable mCherry-GX-eGFP HEK cells 48h after being infected with the under 'scrambled' (left column), CFTR (middle left column), SMG1 (positive control, middle right column) and UPF1 siRNAs (positive control, right column). **(B)** Summary of the mCherry fluorescence quantification for different siRNA controls (n= 3). * means $p < 0.05$. Scale bar = 20µm.

In total, 24 genes were tested (plus controls) with each gene being targeted by two different Silencer Select™ siRNAs (Dharmacon). Image acquisition and image analysis were performed in the same conditions as described above for the shRNA primary screen. As shown by the overall siRNA distribution and respective NMD scores (

Figure ***3.28*** and Table S 3.2) in this assay we have lower levels of mCherry fluorescence and lower NMD scores when compared to the shRNA assay (Figure 3.22). This may be due to differences in transfection/transduction efficiency.

Herein, the results showed two siRNAs with a NMD score ≈ 1: UPF1 (used as a positive control) and eIF4A3, previously described as related to NMD (Iwatani-Yoshihara et al., 2017).

Figure ***3.28*** shows the average NMD scores from both siRNAs for each gene and Table S 3.2 shows the NMD scores for each individual siRNA. Analysing individual siRNAs we identified 4 genes with higher values such as eIF4A3, ADIRF, SREK1, and RPS19. However, when taking into account the NMD score and the images obtained in this confirmation screen (Figure 3.29) we conclude that the results from this study were inconclusive (Table S 3.2).

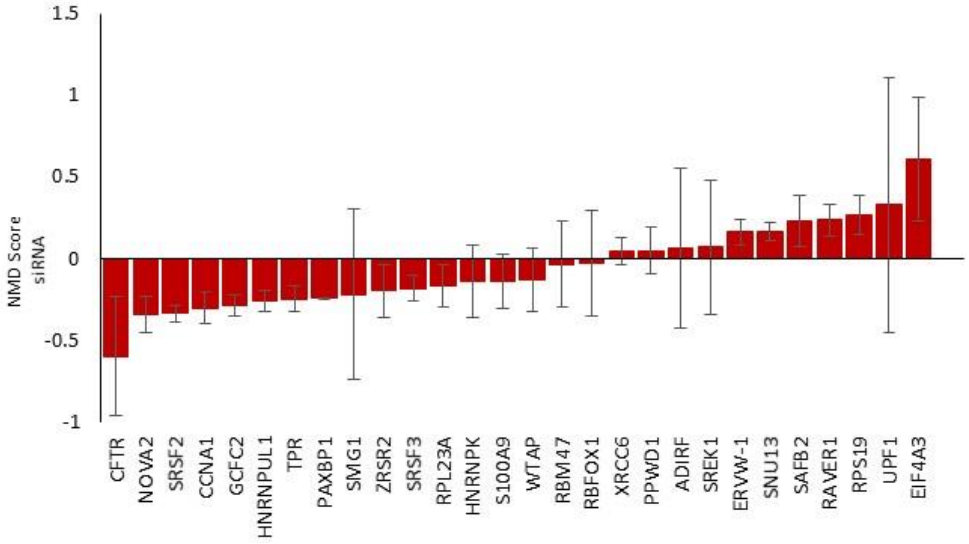



**Figure 3.28** - **Distribution of the siRNA scores for the NMD assay.** The distribution is performed according to the individual siRNA deviation from the negative control. Results presented as average of both siRNA for each gene and SD. NMD scores for each siRNA are presented in supplementary material in Table S 3.2.

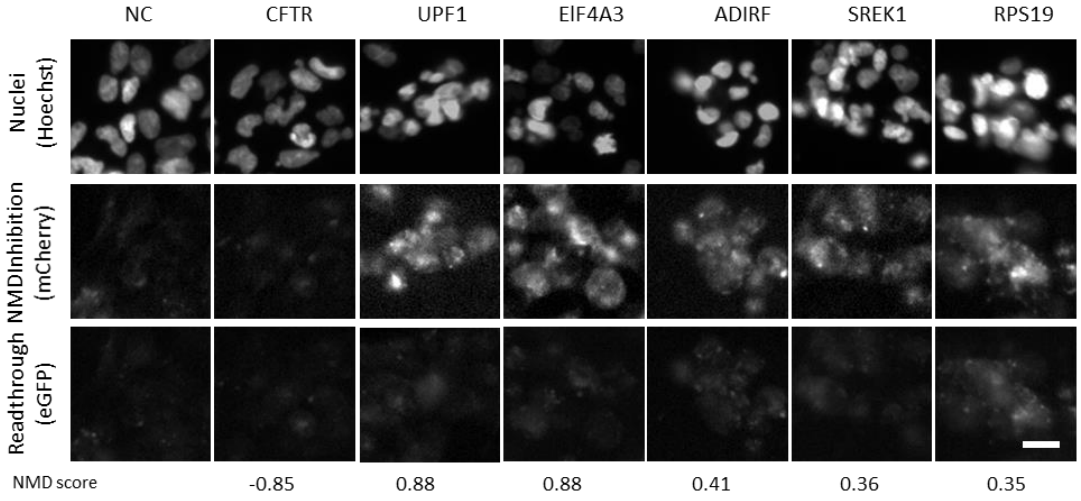



**Figure 3.29 – Representative images of top hit siRNAs that inhibit NMD in NMD confirmation assay.** Stable mCherry-GX-eGFP HEK cells were reverse transfected with the siRNAs for 48h. Widefield fluorescence images of non-permeabilized cells. Nuclei were stained with Hoechst 33342 (top row), mCherry fluorescence (middle row) is proportional to the total amount of expressed CFTR by inhibition of NMD and eGFP (bottom row) is proportional to the read-through activity. Fluorescent images were obtained under equivalent conditions for all cell lines. Scale bar = 20µm.

## 4. Discussion

The aim of this study was to identify putative NMD inhibitors that could be used as potential drug targets in CF patients bearing nonsense, or PTC mutations, with the goal of inhibiting the degradation of CFTR mRNA molecules containing a PTC. Inhibition of NMD in this case would allow an increased expression of a truncated CFTR protein, which in some cases may lead to

a partially functioning CFTR channel and therefore alleviate the CF symptoms. In cases where truncated CFTR is dysfunctional, this approach could be combined with a read-through drug.

To achieve our goal, we performed a NMD screen of ~425 genes involved in RNA processing (Moura-Alves et al., 2011). This screen was designed to identify the possible role of those genes in both NMD and in read-through. Our CFTR mini-gene construct has mCherry fused to the N-terminus of CFTR, which enables the quantification of truncated CFTR and the identification of NMD inhibitors, with the latter being detected by increased mCherry fluorescence. The same construct also has the ability to detect read-through activity from the eGFP fused to the C-terminus of CFTR. It was not expected that any of the genes that inhibit NMD would also display read-through activity. Nevertheless, we have also acquired eGFP images to make the most out of our construct and surprisingly some shRNAs had both NMD inhibition and read-through activity.

Analysis of the data from the primary screen led to the identification of 75 NMD inhibitors. However, using more stringent criteria for these data, namely an NMD score ≥ 2 and at least 2 shRNA with the same phenotype, led to the identification of 11 top hits which when knocked-down inhibit NMD. Furthermore, in this screen, we have identified 2 additional hits that not only inhibit NMD (NMD score ≥ 2 but only one shRNA) but also lead to PTC read-through (PPWD1 and RSP19) and thus these were also selected for further confirmation/ validation. In the subsequent confirmation screen, we considered the 11 top hits inhibiting NMD, the 2 hits with read-through activity, and included an additional 11 top hits from another HTS screen aimed at identifying CFTR splicing regulators (S Igreja & MD Amaral, unpublished data), since it is known that splicing is important for the NMD pathway.

In the confirmation screen, the mCherry fluorescence detected was much lower than that observed in the shRNA screen, as revealed by the lower NMD scores, with this assay being thus inconclusive. However, despite the low values of NMD scores (below +1), we selected the 4 genes with the highest values, as they may still be of relevance, since they were identified in the primary screen (or in the splicing assay) by more than two targeting shRNAs. These 4 genes include EIF4A3 (Eukaryotic initiation factor 4A-3), Adipogenesis regulatory factor (ADIRF), Splicing Regulatory Glutamic Acid And Lysine Rich Protein 1 (SREK1) and Ribosomal protein S19 (RPS19). As such, these genes may merit further validation studies.

Two annotation tools, Panther and DAVID, were used in order to identify a relationship between the 24 top hits and the NMD process as well as to characterize them regarding protein class, molecular function, cellular components and biological process. From these analyses we found that these genes were, mainly: RNA binding proteins (A2BP1; SREK1), mRNA splicing factors (RAVER1; NOVA2; HNRPK; SFRS3; RAVER1), ribosomal proteins (NHP2L1;

RPS19; RPL23A) and transcription factors (C21orf66; GCFC2). For 8 of them there was no function described (FLJ20273; WTAP; C10orf116; ERVWE1; eIF4A3; HNRNPUL1; SAFB2; TPR). From the 24 genes three have been previously associated with NMD, namely eIF4A3, RPS19 and RPL23A (Reactome, 2017), thus lending validity to our findings.

Further validation assays are being carried out (S Igreja and L Clarke) using the same cellular model and the 24 top hit shRNAs but different techniques including: i) mRNA studies using qRT-PCR to quantify CFTR transcript levels after gene knock-down (also to assess gene knock-down), and ii) CFTR protein expression studies by WB after gene knock-down to assess the levels and size of the CFTR protein produced from the PTC reporter. After gene knock-down we expect with both assays to detect an increase in CFTR expression, higher levels of truncated protein when NMD is inhibited and a full-length protein when there is read-through. Further functional assays in cell models as well as in patient-derived materials will also be performed.

In summary, some of the hits identified in this study are potential drug targets due to their effects in NMD inhibition as well as in read-though activity. Further characterization of the effects of these hits is being performed in order to clarify our findings. These data will provide important insights into the mechanisms of NMD and the establishment of novel therapeutic targets for CF as well as for other diseases caused by PTC mutations.

## 5. Supplementary information

**Table S 3.1 - NMD scores obtained for the shRNAs included in screen and genes selected for the confirmation screen.**

In green values ≥ 2; In red values ≤ -2; *Hits selected - NMD score ≥ 2 and more than 2 shRNA with the same phenotype; **Genes with 1 shRNA but showed read-through activity;

| Gene | shRNA1 | shRNA2 | shRNA3 | shRNA4 | shRNA5 |
|---|---|---|---|---|---|
| **NMD Inhibitors - NMD score ≥ 2 and at least 2 shRNA with same phenotype** | | | | | |
| *A2BP1 | -2.48433 | 3.281676 | 3.685467 | 30.49284 | |
| *C10orf116 | 0.899565 | 1.428727 | 2.754271 | 4.364878 | |
| *C21orf66 | -1.03826 | 4.55717 | 1.413343 | 1.691741 | 2.243449 |
| *ERVWE1 | -0.17222 | 0.056118 | 1.718877 | 2.782407 | 3.474885 |
| *FLJ20273 | -0.44888 | 1.52091 | 2.146793 | 2.911569 | 3.191739 |
| *HNRPK | -0.94397 | -0.74048 | -0.39985 | 2.962398 | 5.463193 |
| *NHP2L1 | -0.56273 | 0.01882 | 2.26331 | 2.378306 | |
| *NOVA2 | -1.55207 | -0.31615 | 0.010921 | 2.24212 | 8.072233 |
| *RAVER1 | -0.46702 | -0.07187 | 0.002271 | 4.400525 | 8.258601 |
| *SFRS3 | -3.908 | -0.65785 | -0.49501 | 5.202004 | 7.942901 |
| *WTAP | -0.81297 | 0.376567 | 2.385369 | 4.735453 | |
| **NMD Inhibitors - NMD score ≥ 2 and read-through activity** | | | | | |
| **RPS19 | -1.55477 | 0.08194 | 0.228747 | 0.769386 | 11.62153 |
| **PPWD1 | -0.34214 | 0.165852 | 0.679684 | 13.41716 | |
| **NMD Inhibitors - NMD score ≥ 2** | | | | | |
| BRUNOL5 | -0.87094 | -0.24261 | 0.875964 | 1.046141 | 3.050342 |
| BUB3 | 0.115852 | 0.46406 | 0.549187 | 1.779212 | 3.428402 |
| C21orf70 | | | 2.778576 | | |
| C22orf19 | 0.108738 | 1.458908 | 1.722982 | 2.730848 | |
| C9orf10 | 0.165911 | 0.543766 | 0.760546 | 1.152692 | 2.141336 |
| CCNA1 | | | | | 3.272812 |
| CDC5L | | | 2.20066 | | |
| CSN3 | -1.61898 | -1.1716 | 0.068931 | 1.831277 | 2.491595 |
| CWF19L1 | | | 2.873986 | | |
| DDX23 | -0.9204 | | -0.21927 | 2.817693 | |
| DHX15 | -1.10553 | -1.07588 | -1.03918 | -0.48396 | 4.317799 |
| DHX35 | -1.49195 | -0.82291 | -0.23592 | 4.745279 | |
| DHX9 | -4.76641 | -0.73264 | 2.830572 | | |
| EFTUD2 | -2.33143 | -1.8362 | -1.47057 | -1.249 | 2.062674 |
| eIF3S6 | -1.40078 | -0.82183 | -0.57031 | 0.248509 | 2.04996 |
| ERCC3 | -0.1574 | 0.811894 | 2.273763 | | |
| ET | -0.46807 | -0.09725 | 0.243443 | 0.836318 | 4.000796 |
| EXOSC9 | | -0.85383 | | | 2.343932 |
| FIP1L1 | -0.10109 | 0.224414 | 1.396537 | 1.854061 | 2.236685 |
| FKBP3 | -0.29883 | 0.446929 | 0.496339 | 3.13548 | |
| HNRNPG-T | | 2.931171 | | | |
| HNRPA3 | -1.16279 | -1.0297 | 0.059505 | 0.245841 | 3.055327 |
| HNRPAB | -0.84189 | -0.60706 | -0.27304 | -0.23085 | 2.907348 |
| HNRPDL | -0.32535 | 0.742018 | 1.127231 | 1.560656 | 3.095817 |
| HNRPM | -3.77971 | -0.77923 | -0.48722 | 1.288242 | 2.341067 |
| IMP-3 | -2.29055 | -1.69925 | 0.50371 | 1.047461 | 8.502965 |
| IQGAP1 | | -0.98545 | -0.95437 | -0.02013 | 2.581621 |
| LSM5 | -0.73322 | -0.42296 | 0.499236 | 0.954966 | 2.410485 |
| MGC2803 | -1.39815 | -1.31833 | 1.01208 | 2.681422 | |
| NCL | -0.41928 | 0.322678 | 1.987658 | 5.023905 | |
| NDUFA1 | -0.95005 | -0.70911 | -0.70724 | 4.758889 | |
| NIF3L1BP1 | -0.76069 | 0.172208 | 1.364969 | 1.728029 | 3.172771 |
| PHF5A | -0.68119 | -0.46046 | 0.320473 | 1.157849 | 4.975865 |
| PRPF3 | -0.77751 | -0.70118 | 0.793365 | 1.275076 | 3.204732 |
| PSIP1 | -1.20087 | -1.05298 | 0.792705 | 1.541428 | 3.121637 |
| RBM15 | -0.07104 | -0.0386 | 0.886392 | 1.479354 | 2.155021 |
| RBM5 | -2.00081 | -1.33978 | -0.95162 | -0.32308 | 8.833536 |
| RBM7 | 0.531051 | 0.541814 | 1.056061 | 1.952503 | 2.093601 |
| RBMS1 | -1.61086 | -1.27824 | -0.79916 | 0.460779 | 2.664278 |
| REXO1 | 0.040288 | 0.827607 | 1.090679 | 1.117308 | 2.396234 |
| RPL5 | | | | | 20.48134 |

| | | | | | |
|---|---|---|---|---|---|
| RPS10 | -0.52349 | 0.71434 | 1.059318 | 1.274539 | 2.175561 |
| RPS12 | -1.67846 | 0.013077 | 0.882671 | 1.281848 | 3.357818 |
| SAFB | | | | | 5.364237 |
| SF1 | -1.19981 | -0.7254 | 0.246876 | 0.824536 | 2.25106 |
| SF3B1 | -2.70417 | 0.408457 | 1.283139 | 1.432562 | 11.34466 |
| SFPQ | -1.97468 | 0.873791 | 1.71834 | 2.746347 | |
| SMC2L1 | -1.30606 | -0.59665 | -0.40809 | 0.037091 | 2.23222 |
| SNRPB | -0.11346 | 0.64251 | 0.859866 | 1.212524 | 4.204409 |
| SNRPD2 | -0.34496 | 0.801297 | 0.890274 | 2.712017 | |
| SNRPE | -0.46192 | -0.21467 | 0.524682 | 1.126053 | 5.855315 |
| SNRPF | -2.20959 | -1.49159 | -0.80101 | -0.39184 | 2.587853 |
| SNRPG | -1.30593 | -0.1645 | 0.352851 | 1.189919 | 2.668892 |
| SPPL3 | -2.62492 | -1.58643 | -1.3963 | -0.61715 | 2.403882 |
| SRP68 | | 4.738389 | | | |
| TAF15 | | | 0.161557 | | 4.183365 |
| TFIP11 | 0.559033 | 1.119317 | 3.777841 | | |
| TIA1 | 0.466446 | 0.698534 | 0.865948 | 1.078845 | 2.354701 |
| TXNL4A | -2.4977 | 0.205364 | 1.227198 | 4.905382 | |
| UNK | -0.71979 | -0.31114 | 0.378432 | 1.230809 | 2.634751 |
| USP39 | -0.25476 | -0.07764 | 1.061227 | 1.111202 | 2.896239 |
| WDR58 | -1.40926 | -0.54485 | -0.0752 | 4.192081 | |
| | | **NMD Score <2** | | | |
| ABT1 | | -0.61378 | -0.23544 | | |
| ACIN1 | -0.71246 | -0.38941 | -0.11456 | 0.251148 | 0.48194 |
| ADAR | -0.67243 | | -0.55892 | | |
| AKAP8 | | | -0.47482 | -0.4077 | 0.355917 |
| AQR | -0.58671 | -0.19846 | -0.03611 | 0.65892 | 1.373401 |
| ARS2 | 0.083601 | 0.382395 | 0.543988 | 0.824589 | 0.96553 |
| ASCC3L1 | -0.61299 | | | 0.209014 | |
| BAT1 | -0.8303 | -0.60656 | -0.42396 | -0.23013 | |
| BRUNOL4 | -1.8647 | -1.48681 | -1.34469 | 0.017142 | 0.05211 |
| BRUNOL6 | -0.76951 | 0.170021 | 0.395089 | 0.429616 | 0.848098 |
| C13orf10 | -0.55702 | 0.008379 | 0.066774 | 0.310598 | 0.625026 |
| C1orf55 | -2.06947 | -1.5765 | -1.44556 | -0.59893 | |
| C1orf60 | -1.23064 | -1.04992 | -1.04001 | -0.74569 | |
| C20orf14 | -1.89534 | -0.36183 | 0.338917 | 0.62767 | 1.163451 |
| C20orf23 | -0.77554 | -0.71085 | -0.5211 | -0.45038 | -0.0646 |
| CCDC12 | -1.17512 | | | | |
| CCNK | | | -0.05343 | -0.00334 | |
| CD2BP2 | -0.04163 | 0.035462 | | | |
| CDC2L2 | | | -0.18439 | | |
| CDC40 | -1.58318 | -1.24935 | 0.307701 | 1.186813 | |
| CIRBP | -0.29052 | -0.21638 | 0.26133 | | |
| CLK1 | -0.5118 | 0.817405 | | | |
| CLK2 | 0.774646 | 1.395776 | 1.586024 | | |
| CLK3 | 0.818235 | 1.165164 | 1.21771 | | |
| CLK4 | | | -0.19399 | | |
| CPSF1 | 0.05627 | 0.235202 | 0.4643 | 0.610743 | 0.681733 |
| CPSF4 | | | 0.563664 | 1.098619 | |
| CPSF6 | -1.81097 | -0.56576 | | | |
| CRK7 | | | -0.5864 | | |
| CRNKL1 | 0.021838 | 0.856865 | 0.862699 | 1.481834 | |
| CSDA | | | 0.044994 | | |
| CUGBP2 | -1.08636 | -0.14472 | 0.472744 | 1.166095 | 1.395211 |
| DDX17 | -1.80512 | -1.73157 | -0.56599 | -0.19811 | 1.040759 |
| DDX19B | -0.44461 | | | | |
| DDX21 | -0.81548 | -0.80802 | -0.64797 | | |
| DDX26 | -1.14444 | -0.21614 | 0.095638 | | |
| DDX39 | | | | | 0.505165 |
| DDX3X | -1.94888 | -1.35609 | -1.07705 | -0.47248 | |
| DDX41 | -1.89424 | -0.82122 | -0.11195 | 0.336407 | 1.285402 |
| DDX46 | -1.43344 | -0.98543 | 0.248099 | 0.839602 | 1.72313 |
| DDX49 | 0.582235 | | | | |
| DDX5 | -3.80564 | -0.64583 | 0.533204 | 0.551716 | 0.770214 |
| DDX54 | -0.36157 | | 0.546958 | | |
| DDX6 | -0.12974 | 0.186016 | 0.81049 | 1.431298 | |

| | | | | | |
|---|---|---|---|---|---|
| DEK | -1.4057 | | | 0.671538 | |
| DGCR14 | 0.30217 | 0.360572 | 0.482401 | 0.64592 | |
| DHX16 | -1.13009 | 0.940763 | | | |
| DHX38 | -0.36999 | 0.940187 | 1.180129 | 1.882155 | |
| DHX8 | -1.74942 | -1.11539 | -0.95593 | | |
| DIDO1 | -1.08082 | -0.93597 | -0.25924 | -0.16402 | -0.11968 |
| DKFZP434K1421 | -1.22908 | -1.13852 | -0.87208 | -0.68467 | |
| DNAJC6 | -0.71955 | -0.17647 | 0.358456 | | |
| EDG2 | | -0.72678 | | | |
| EEF1A1 | -0.6412 | -0.60086 | | | |
| EIF2S2 | -0.57715 | -0.20509 | 0.134062 | 0.354268 | 0.505729 |
| EIF3S10 | -1.92762 | -1.56785 | -0.38213 | 0.114027 | |
| EIF3S2 | -1.66545 | -0.86532 | -0.02587 | 1.031553 | |
| EIF3S6IP | -1.10051 | -0.92186 | -0.58984 | -0.52155 | |
| EIF4A2 | -0.84548 | -0.77853 | | | |
| ELAVL1 | | | -0.32145 | | 1.027211 |
| ELAVL2 | -1.10209 | -0.44847 | 0.481084 | | 0.937548 |
| ELAVL3 | -1.33655 | -1.16332 | 0.341588 | 0.767035 | 1.959175 |
| ELAVL4 | -0.98592 | | -0.67131 | -0.39731 | |
| EP400 | -1.10434 | | | | |
| EWSR1 | 0.275174 | 0.662949 | 1.31748 | | |
| EXOSC1 | -1.31488 | -0.55533 | | 0.21508 | |
| EXOSC10 | | | -0.7311 | -0.28571 | |
| EXOSC2 | | | -0.48346 | | |
| EXOSC4 | -0.50216 | -0.4134 | | | |
| EXOSC5 | -0.6683 | | | | |
| EXOSC8 | | | | 0.63773 | |
| FAU | -1.50898 | -0.04326 | | | |
| FLJ21827 | -0.29165 | -0.29165 | -0.1916 | 0.309771 | 1.136376 |
| FMR1 | -0.67246 | | | | |
| FUBP1 | | -0.41532 | 0.103434 | 0.120281 | |
| FUS | -3.52006 | | | | |
| FUSIP1 | -1.13224 | -0.34542 | 1.004971 | 1.611562 | |
| G10 | | -1.55843 | | | |
| GNB2L1 | -0.46131 | -0.06434 | 0.640207 | 1.791583 | |
| GPATC1 | | 1.798199 | | | |
| GTL3 | | | -0.4869 | -0.25107 | |
| HCFC1 | | | -0.23727 | | |
| HDAC2 | -2.12749 | | -1.35959 | | |
| HIST1H2BC | -2.26224 | -1.28114 | -0.86882 | -0.01647 | 0.442854 |
| HIST2H2AA | -1.63346 | -1.44005 | -1.1643 | -0.76187 | -0.19981 |
| HMGB1 | | | | -0.37239 | |
| HMGB3 | | | -0.01808 | 0.011483 | |
| HNRPA0 | | -0.41432 | | | |
| HNRPA1 | | -0.29395 | | | |
| HNRPA2B1 | -0.29588 | 1.588016 | | | |
| HNRPD | -1.19934 | -0.75512 | 0.479612 | 0.520578 | |
| HNRPF | -0.9636 | -0.76659 | -0.62359 | 0.183963 | 0.575126 |
| HNRPH1 | -1.11884 | -0.65265 | -0.02526 | | |
| HNRPH2 | -0.97081 | | | | |
| HNRPH3 | -0.11457 | 0.029616 | 0.101019 | 0.54148 | 1.220772 |
| HNRPL | | | -0.11369 | | |
| HNRPR | -2.84198 | -1.59648 | -0.40379 | | |
| HNRPU | | -0.86078 | -0.33176 | -0.14474 | 0.101461 |
| HNRPUL1 | -0.21944 | -0.07438 | 0.179665 | 0.272219 | 0.898631 |
| HSPA5 | -0.29033 | -0.25485 | 0.724269 | | |
| HSPA8 | | -1.46326 | -0.54999 | 0.141255 | 0.148461 |
| HTATSF1 | -0.88519 | | -0.60912 | | |
| IK | -2.2932 | -1.68004 | -1.51722 | -1.04771 | 1.298365 |
| ILF2 | | -0.88321 | | -0.25229 | |
| IMP-1 | | | 0.702151 | 0.759889 | |
| KIAA0773 | -0.6129 | -0.00437 | 0.328957 | 0.609862 | |
| KIAA0853 | | | | -0.06261 | |
| KIAA1008 | -0.52409 | -0.37303 | -0.33724 | | |
| KIAA1429 | -2.61533 | -0.49238 | -0.35047 | -0.15082 | |
| KIAA1542 | -1.88442 | -1.71671 | -1.26581 | -0.97816 | -0.29087 |
| KIAA1604 | -0.85192 | 0.135872 | 0.50995 | 0.782487 | 0.822156 |

| | | | | | |
|---|---|---|---|---|---|
| KIAA1967 | | | | 0.165934 | |
| KIN | -3.05644 | -0.56372 | | | |
| KPNA2 | -0.24622 | 0.460578 | 1.253501 | 1.341591 | |
| LOC138046 | -0.44538 | -0.40656 | -0.39908 | 1.360041 | |
| LSM1 | -1.19634 | | | | |
| LSM10 | -2.40446 | -0.20199 | 0.780275 | | |
| LSM11 | -1.63375 | -0.64428 | -0.61398 | -0.59691 | 0.527945 |
| LSM2 | -2.35605 | -1.8355 | -1.73923 | -1.22363 | -1.21195 |
| LSM3 | -0.98667 | 0.016132 | 0.317509 | 0.463717 | 1.166528 |
| LSM4 | -1.60266 | -0.62099 | 0.270629 | 0.396055 | 1.113183 |
| LSM6 | -1.67141 | 0.130041 | 0.685616 | 1.073289 | |
| LSM7 | -0.82031 | 1.006489 | | | |
| LSM8 | -2.5411 | -1.94123 | -1.66758 | -0.70788 | -0.55971 |
| LUC7L | -4.97381 | -1.02152 | -0.84295 | -0.75444 | 0.968934 |
| MAGOH | -0.89775 | | | | |
| MATR3 | -1.72636 | -0.22112 | -0.09041 | 0.168727 | 0.789431 |
| MFAP1 | -0.88387 | -0.57057 | -0.07283 | 0.150228 | 0.98512 |
| MGC13125 | -1.50523 | -0.70759 | -0.58983 | -0.02756 | 1.195198 |
| MGC14151 | -0.86764 | -0.31495 | -0.10386 | 0.1584 | |
| MKI67IP | -1.2479 | -0.80306 | -0.26182 | -0.25282 | -0.22686 |
| MORG1 | -0.9988 | -0.76125 | -0.6997 | -0.57526 | 1.270044 |
| MOV10 | -1.25625 | -1.1628 | | -0.18906 | |
| MSI1 | -1.60685 | -0.33479 | 0.339239 | | |
| MSI2 | -1.81976 | -1.49922 | -0.76011 | -0.48193 | |
| NCBP2 | | 0.061248 | 0.770882 | | |
| NONO | 0.006437 | 0.282285 | 0.939975 | 1.038253 | 1.185169 |
| NOSIP | -0.74271 | -0.13157 | | | |
| NOVA1 | -3.79062 | 0.278624 | 0.75934 | 1.409476 | |
| NUDT21 | -1.62603 | -0.10591 | 1.081401 | 1.46404 | 1.918313 |
| NUMA1 | -2.42219 | -1.78692 | -1.47363 | -0.99421 | 0.229912 |
| NXF1 | -0.67364 | -0.49479 | -0.44851 | | |
| PABPC1 | -1.28063 | -0.50429 | -0.25587 | -0.24268 | 0.188624 |
| PABPC4 | -1.48701 | -1.07798 | -0.6395 | 0.501127 | 0.542782 |
| PABPN1 | -1.37014 | -1.18002 | -0.13678 | 1.369013 | 1.718378 |
| PARP1 | | -0.31051 | | | |
| PCBP1 | -2.43444 | -1.80053 | -1.68888 | -1.67598 | -1.49922 |
| PCBP2 | -1.806 | -1.70275 | -1.17903 | -1.09677 | 0.790114 |
| PCBP3 | -1.61566 | -1.39276 | -0.88509 | -0.1402 | -0.10099 |
| PIAS1 | | -0.23237 | 0.01426 | | |
| PLRG1 | 0.174461 | | | | |
| PNN | -0.1808 | 0.973295 | 0.985036 | 1.026435 | 1.470678 |
| POLR2A | -0.38865 | | -0.07179 | | |
| PPIE | -1.03637 | | | | |
| PPIH | -0.91836 | -0.25443 | -0.03082 | 0.741957 | |
| PPIL3 | 0.700351 | | | | |
| PPM1G | -1.37321 | -1.24823 | 0.277897 | 0.998333 | |
| PRKRA | | 0.0065 | 1.131296 | | |
| PRPF18 | -0.31645 | -0.07822 | 0.6896 | 1.08649 | |
| PRPF19 | -2.029 | | | | |
| PRPF31 | -1.78071 | 0.131608 | 0.204972 | 0.721994 | |
| PRPF38A | -0.09205 | 0.754124 | 1.112775 | 1.81236 | |
| PRPF38B | -0.07734 | -0.05602 | 0.627529 | 0.671358 | 0.714326 |
| PRPF4 | -1.43291 | -1.16716 | -1.13933 | -1.00351 | -0.60039 |
| PRPF4B | -1.37479 | -1.33845 | -0.43069 | 0.199835 | 1.021436 |
| PRPF8 | -1.2787 | | | | |
| PSEN1 | -1.03759 | | | | |
| PTBP1 | -0.28394 | 1.586401 | | | |
| RALY | -1.09522 | -0.57356 | -0.32256 | | |
| RBM10 | -1.66432 | -1.56899 | -0.99154 | -0.80242 | 0.640903 |
| RBM12 | -1.66394 | -1.14578 | -0.43405 | -0.30187 | 0.186719 |
| RBM17 | -1.59927 | -0.91434 | 0.320339 | | |
| RBM22 | -0.66558 | | | | |
| RBM25 | -2.21218 | -1.73388 | -1.36174 | -1.10581 | -0.67894 |
| RBM8A | -0.75174 | | -0.05109 | | |
| RBM9 | -1.63934 | -0.23365 | -0.13663 | -0.06868 | 1.415805 |
| RBMX | -0.15429 | 0.110234 | 1.158987 | 1.315798 | 1.970589 |
| RBMX2 | -0.72475 | -0.58391 | -0.5247 | | |

| | | | | | |
|---|---|---|---|---|---|
| RBP7 | | | | -0.35094 | |
| REXO2 | -0.91284 | | | 0.495872 | |
| RKHD3 | -0.64815 | | 0.415879 | 1.046956 | |
| ROD1 | -0.44957 | | -0.0538 | | |
| RP13-297E16.1 | -1.11762 | -0.97467 | | | |
| RPL22 | -1.23437 | | | | |
| RPL31 | -0.44586 | | | 1.037017 | |
| RPS13 | -2.10266 | | -0.46776 | | |
| RPS15A | -1.24945 | -1.1949 | -0.66404 | -0.1808 | 1.417045 |
| RPS17 | -0.32535 | 0.482113 | 0.497335 | 1.141996 | 1.753097 |
| RPS18 | -0.80698 | -0.80475 | -0.55411 | -0.17773 | 0.121376 |
| RPS29 | -1.31445 | -0.66027 | -0.42381 | 0.02761 | |
| RPS3 | | | | -0.66627 | -0.01066 |
| RPS3A | -0.84277 | -0.5203 | -0.19465 | 0.265373 | |
| RPS4X | -0.35903 | -0.06459 | 0.019389 | 0.512935 | 1.255321 |
| RPS4Y1 | -0.30554 | 0.365854 | 0.741485 | 1.003683 | 1.081539 |
| RPS5 | -1.25426 | -0.82054 | -0.63599 | 0.329169 | |
| RPS7 | -1.63382 | -1.60644 | -1.59619 | -1.07395 | 0.501845 |
| RPS8 | -0.97177 | -0.5912 | -0.48387 | | |
| RPS9 | -1.53924 | -1.04837 | -1.03234 | -0.1248 | |
| RUVBL2 | -0.94739 | | -0.84314 | | |
| S100A9 | -0.49287 | -0.47835 | 0.260668 | | |
| SART1 | -1.49035 | 1.271895 | | | |
| SDCCAG10 | -1.58785 | -0.76988 | 0.094294 | 1.148426 | 1.643716 |
| SF3A1 | | | -0.77381 | -0.45307 | -0.36894 |
| SF3A2 | -1.55617 | -0.90107 | -0.52231 | 0.973411 | |
| SF3A3 | -1.57853 | -0.65432 | -0.22322 | | |
| SF3B3 | -1.43774 | -1.11662 | -1.00301 | 0.103146 | 1.228029 |
| SF3B4 | -1.72244 | -1.426 | -0.28744 | 0.068793 | |
| SF3B5 | -0.79033 | -0.78 | -0.66313 | 0.852647 | 1.624967 |
| SF4 | -1.06683 | -0.77865 | -0.21752 | 0.79052 | 0.845222 |
| SFRS1 | -1.81028 | -0.85835 | | | |
| SFRS10 | -1.34255 | -0.67637 | -0.22633 | | |
| SFRS11 | 0.184813 | 0.602882 | 1.355507 | 1.619744 | |
| SFRS12 | -0.69983 | -0.55148 | | | |
| SFRS2 | -0.71924 | 0.440807 | 1.699296 | | |
| SFRS5 | -0.88751 | -0.64129 | -0.47357 | | |
| SFRS6 | -0.85175 | | | -0.10527 | |
| SFRS7 | -2.12343 | -0.40277 | 0.405278 | | |
| SFRS9 | -1.1884 | -0.65387 | -0.61889 | -0.41046 | |
| SIAHBP1 | | | -0.42158 | | |
| SKIV2L | -1.12505 | | -0.59334 | | -0.3883 |
| SKIV2L2 | | | -0.16408 | | |
| SLU7 | -1.11624 | -0.89092 | 0.167275 | | |
| SMARCA5 | | | -0.43202 | -0.06949 | |
| SMC1L1 | -1.23743 | -0.92925 | -0.32772 | -0.0414 | 0.228354 |
| SMNDC1 | -3.73972 | -0.17346 | 0.080503 | 0.506675 | |
| SMU1 | -2.76375 | -1.69396 | -1.55819 | -1.46906 | -1.44699 |
| SNIP1 | | 0.218272 | | | |
| SNRP70 | -1.83624 | -0.63064 | -0.16231 | 1.865742 | |
| SNRPA | | -0.61593 | | 0.77354 | |
| SNRPA1 | -0.95232 | 0.380631 | 1.021338 | 1.123204 | 1.186756 |
| SNRPB2 | | -0.83164 | -0.76968 | -0.44288 | |
| SNRPC | -1.27354 | -1.05094 | -0.35512 | 0.218655 | 1.694308 |
| SNRPD1 | -1.48073 | | -0.82625 | -0.56287 | |
| SNRPD3 | 0.147244 | 0.279945 | 0.421399 | 0.989564 | |
| SNW1 | -4.18288 | -0.75606 | 0.129784 | 1.772247 | |
| SR140 | | | | 1.434132 | |
| SRP46 | -1.4936 | -1.13531 | -0.70138 | 0.600438 | 1.611693 |
| SRP9 | | | | | 1.756544 |
| SRPK1 | -0.8564 | 0.336096 | | | |
| SRRM1 | -2.02229 | -1.80726 | -1.53304 | | |
| SRRM2 | -0.50096 | | | | |
| SSB | | | | 1.248208 | |
| STK23 | | 1.47662 | | | |
| SYNCRIP | -1.70708 | | | | -0.13749 |
| TCERG1 | | -0.15322 | 0.141329 | 0.254216 | |

| | | | | | |
|---|---|---|---|---|---|
| TDRD3 | -1.91626 | -1.26384 | -1.12926 | -0.95558 | |
| THOC1 | -0.90493 | -0.73185 | 0.009393 | 0.968105 | 1.654829 |
| THOC3 | | | | -0.61596 | 0.116693 |
| THOC4 | -0.18794 | 0.490156 | 0.699093 | 1.173454 | |
| TIAL1 | | | -0.51921 | -0.43368 | -0.397 |
| TNPO1 | | 0.217808 | | | |
| TNRC4 | -0.77096 | -0.26485 | | | |
| TOP1MT | -0.6012 | | | 0.698441 | |
| TPR | | -0.33244 | | | |
| TPX2 | -0.76728 | -0.25713 | -0.03699 | 0.161808 | 0.716331 |
| TRA2A | -0.15151 | 0.105109 | 0.872833 | 1.167509 | 1.342977 |
| TTF2 | | | 0.508697 | 0.669166 | |
| TUBA1 | -1.86207 | -1.03705 | 0.638501 | 1.010412 | 1.697593 |
| TUBB | -0.63363 | -0.42303 | -0.14273 | -0.00473 | 0.073986 |
| U2AF1 | -0.67474 | 0.996379 | | | |
| U2AF1L2 | -1.88771 | -1.04494 | -0.6476 | -0.52445 | -0.47724 |
| U2AF1L3 | -0.09172 | -0.05549 | 0.041629 | 0.284019 | 0.80937 |
| U2AF2 | -0.95663 | -0.43314 | 0.205074 | 0.341237 | 0.866395 |
| VIM | -1.04606 | | | | |
| WBP11 | -1.61397 | -1.48021 | -1.47171 | -0.35932 | -0.28218 |
| WDR33 | -1.49698 | -0.92269 | -0.23108 | -0.07404 | 0.311411 |
| WDR57 | -0.87848 | -0.55487 | -0.49529 | 0.184237 | 0.226292 |
| XAB2 | 0.639141 | 0.759858 | 1.591767 | 1.700442 | |
| XRCC6 | -0.76363 | | | | |
| XRN2 | -0.89156 | -0.48137 | | | |
| ZFP36L1 | | -0.50731 | -0.40568 | | 0.092285 |
| ZFR | | -0.93326 | | -0.30983 | -0.04945 |
| ZMAT2 | -0.86797 | 0.054308 | 0.997064 | | |
| ZNF207 | -0.75476 | -0.45825 | | | |
| | **NMD Inhibitors - NMD score** ≤ 2 | | | | |
| DNAJC17 | -2.02215 | -2.01634 | -1.89278 | -1.72632 | -1.66236 |
| HIST1H2AC | -2.42689 | -2.11679 | -1.10578 | -0.9853 | -0.11778 |
| PPIL1 | -4.3203 | -4.21014 | -1.57098 | 1.703211 | 7.287349 |
| PPIL2 | -4.65285 | -2.44305 | 1.970461 | | |
| PTBP2 | -4.79378 | -4.34003 | -2.92043 | 1.17505 | |
| RBM3 | -2.61533 | -2.45619 | -2.16576 | -0.31195 | 2.98108 |
| RNPS1 | -4.9294 | -4.76843 | -4.21523 | -0.02207 | |
| RPS15 | -2.92649 | -2.12388 | -1.54915 | -1.47169 | -1.12428 |
| RPS25 | -2.30935 | -2.10215 | -1.90552 | -1.55616 | 0.828241 |
| RY1 | -2.59711 | -2.24658 | -2.07512 | -1.30488 | -0.80804 |

**Table S 3.2 - NMD scores obtained for the confirmation screen with siRNAs and top hits selected.**

In green genes selected.

| Gene | siRNA 1 | siRNA 2 |
|---|---|---|
| UPF1 | 0.880213 | -0.22287 |
| eIF4A3 | 0.879692 | 0.343308 |
| ADIRF | 0.412396 | -0.28273 |
| SREK1 | 0.360419 | -0.21822 |
| RPS19 | 0.353367 | 0.183417 |
| SAFB2 | 0.339279 | 0.120795 |
| RAVER1 | 0.305672 | 0.167189 |
| ERVW-1 | 0.222233 | 0.111169 |
| SNU13 | 0.206135 | 0.128915 |
| RBFOX1 | 0.201043 | -0.25191 |
| RBM47 | 0.152186 | -0.21565 |
| PPWD1 | 0.151744 | -0.0497 |
| SMG1 | 0.150184 | -0.58914 |
| XRCC6 | 0.109657 | -0.00828 |
| ZRSR2 | -0.08308 | -0.31214 |
| HNRNPUL1 | -0.21455 | -0.30723 |
| CCNA1 | -0.2308 | -0.37007 |
| SRSF3 | -0.23351 | -0.12701 |
| PAXBP1 | -0.2427 | -0.23637 |
| S100A9 | -0.25532 | -0.02364 |
| RPL23A | -0.26004 | -0.07297 |
| WTAP | -0.26535 | 0.010437 |
| HNRNPK | -0.29844 | 0.018695 |
| TPR | -0.29856 | -0.18991 |
| GCFC2 | -0.32888 | -0.24136 |
| SRSF2 | -0.37056 | -0.30224 |
| NOVA2 | -0.42081 | -0.26277 |
| CFTR | -0.85054 | -0.33973 |

# *Chapter 4 - A Small-Scale Screen to Identify Compounds that Correct CFTR Mutations Bearing PTCs.*

Experiments shown in this chapter were performed jointly with Susana Igreja (MD Amaral lab)

## 1. Abstract

**Background**: Among the ~ 2,000 CFTR gene variants reported, a significant fraction (~24%) generates a premature termination codon (PTC), including: 8.4% nonsense and 15.7% frameshift mutations. PTC mutations generally lead to a significant reduction, or total absence, of normal CFTR mRNA and protein expression, and are usually associated with a severe clinical phenotype. The reduction in transcript levels occurs through the mRNA degradation mechanism called NMD (Nonsense-Mediated Decay) that selectively degrades transcripts bearing PTCs. Patients expressing higher levels of mRNA containing PTC due to less efficient NMD are expected to respond more robustly to suppression / read-through therapies when compared to patients expressing lower levels of mRNA due to highly NMD efficiency. To date, there are no CFTR modulating therapies specifically targeted to treat patients with PTC mutations and, although the aminoglycoside antibiotics (some such as tobramycin used in the standard treatment of CF lung infections) and Ataluren / PTC-124 have demonstrated some ability to suppress PTCs *in vitro*, they are non-specific and show poor efficacy *in vivo*. Indeed, these compounds underwent clinical trials for CF but were discontinued because of lack of demonstrated efficacy. As such, there is a need to find new therapeutic strategies to treat these patients.

**Aim:** Our aim is to screen for novel compounds suppressing PTC mutations by automated microscopy using a cell-based model, as potential corrective therapies for CF.

**Methods:** To achieve this goal, we used a CFTR mini-gene reporter model harbouring the G542X mutation. The mini-gene features a mCherry and an eGFP tag at the CFTR N- and C-termini, which allow the assessment of both the extent of NMD (mCherry) and the full-length protein as a readout of read-through. The construct also has a Flag-tag which becomes extracellular once the mCherry-Flag-G542X-CFTR-eGFP protein is at the plasma membrane (which may not occur to all full-length proteins produced given the random insertion of amino acids in read-through). This construct was inserted into HEK293 Flp-In cells for stable expression, in parallel with a similar wt-CFTR construct. We used as a negative control DMSO and as a positive control G418.

**Results and Discussion:** Our data from this preliminary study show two compounds that stand out from the rest, in terms of their high scores. The compound with the highest read-through score is the G418 our positive control, and PTC124 did not demonstrate any effect.

**Conclusion**: To confirm the read-through efficiency, complementary experiments are required using other techniques. To validate the top hit compounds, it is necessary to test them in patient-derived materials and test them in combination with potentiators and correctors for maximal effect.

## 2. Introduction

Efforts to develop treatments for patients with CF with nonsense mutations have focused on strategies to promote suppression of PTCs (also known as translational read-through). Translational read-through is accomplished when an amino acid carried by a near- cognate aminoacyl transfer RNA is inserted into a polypeptide chain at the erroneous stop codon, allowing translation to continue, and partially restoring “full-length” functional protein (Keeling and Bedwell, 2011). Several pharmacologic approaches to induce read-through have been discovered, yet none has yielded an optimal combination of efficacy and safety.

Aminoglycoside antibiotics (gentamycin/G418, tobramycin) and the compound Ataluren (also known PTC-124 - (3-[5-(2-fluorophenyl)-[1,2,4] oxadiazol-3-yl]- benzoic acid), an orally bioavailable non-aminoglycoside compound) have been shown to have the ability to read-through PTCs, allowing translation of full-length proteins. Since aminoglycosides have severe side effects, such as nephrotoxicity and ototoxicity, when used at high concentrations and/or when used long-term, the search by High throughput screen (HTS) for non-aminoglycosides capable of promoting read-through, revealed the small molecule Ataluren, which evidenced some ability to carry out read-through PTCs without severe side effects. This small molecule has been tested in patients, albeit with controversial efficacy (Kerem et al., 2014). However, the clinical trials results showed that not all the patients responded to the treatment, in particular, the patients who are being treated with aminoglycoside antibiotics (to treat bacterial infections) were the ones showing worse results. So, a new Phase III trial excluding patients under tobramycin was performed, but since it did not meet the end points, Ataluren was discontinued. As such, there is an unmet need to find novel therapies to treat patients with PTC mutations.

The main purpose of this study is to restore functional protein production to PTC mutations using read-through compounds aimed at the treatment of CF patients. We thus assessed the effects of a small library of 24 compounds which are Ataluren/PTC124 analogues/derivatives as well as non-PTC124-like compounds chemically synthesised by a group of collaborators (Prof. Cristina Moiteiro, FCUL, Lisboa). The screen of the library was performed using HTS and using our NMD-PTC reporter (Chapter 3 and Figure 2.8). As mentioned previously (Chapter 3), this reporter is suitable to screen compounds by automated HT fluorescence microscopy. The construct has the G542X nonsense mutation, mCherry at the N-terminus and eGFP fused at the CFTR C-terminus. Because the G542X PTC mutation is present, this construct leads to very little expression of truncated CFTR protein and no expression of the full-length CFTR protein. Hence, it does not generate eGFP fluorescence. However, if a compound has read-through properties it will lead to the production of “full-length” CFTR,

which will be detected as an increase in green fluorescence (from eGFP). This reporter effectively allows the assessment of the read-through properties of the compounds and thus the identification of relevant or "hit" compounds in an automated, expedite way. Through the screen of the previously mentioned small library, we have identified new small-molecule compounds that induce PTC read-through.

### 3. Results and Discussion

Using the NMD/PTC reporter we evaluated the efficacy of 24 compounds which included PT124-like and non-PTC-like compounds, for their read-through activity. The experiments were performed using the stable Flp-In HEK cells expressing the mini-gene after incubation with the compounds for 48h, using DMSO (0.5% v/v) as a negative control and as positive controls G418 (a kwon read-through compound), PTC124 (Ataluren) and the SMG1 inhibitor ($Inh_{SMG1}$) (Figure 3.30). The assay was used in automated high-content microscopy in the 96-well plate format (see Materials and Methods). To avoid 'edge effects', controls were localized randomly throughout the plate. Image acquisition was performed with an automated Leica DMI 6000B widefield epifluorescence microscope. The fluorescence intensity was integrated and processed using CellProfiler (see Materials and Methods).

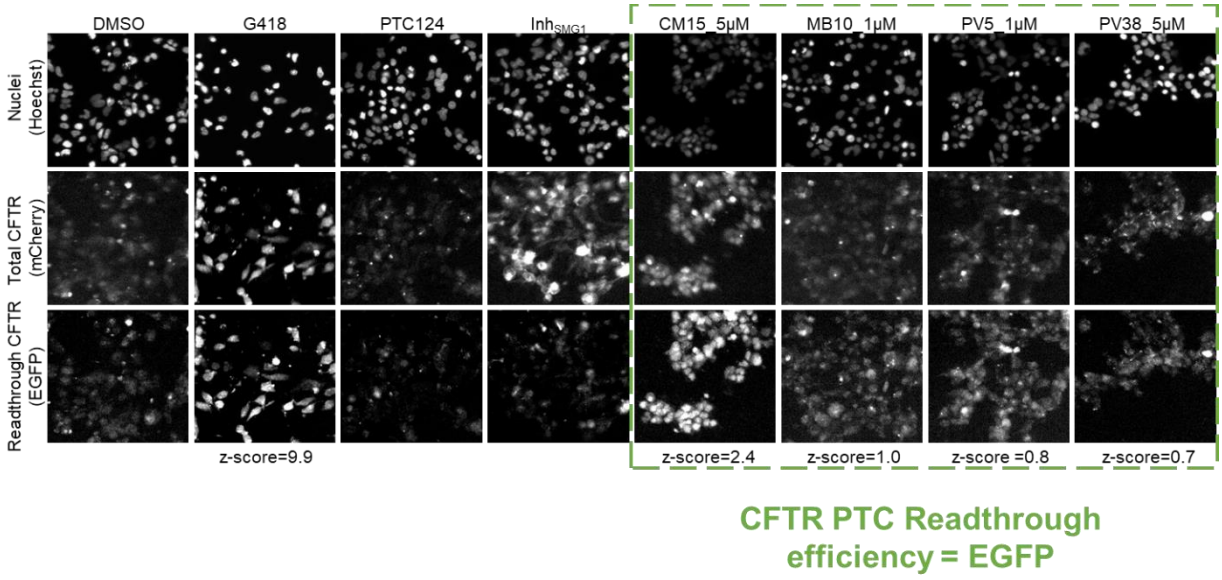



**Figure 3.30 – Representative images of hit compounds that promote PTC read-through.** Representative images of HEK 293 Flp-In cells grown for 48h under DMSO, G418, PTC124, SMG1 inhibitor and top hit compounds. Readouts are shown underneath.

The C-terminal eGFP allows the quantification of "full-length" protein being expressed in each individual cell and thus, the integrated fluorescence signal in each cell is proportional to the read-through activity of the compound. This is defined as the amount of CFTR transcripts that, by escaping NMD, originate full-length CFTR (independently of whether it is functional or not). For this analysis we used the eGFP fluorescence readout, however, it is expected that

mCherry fluorescence would also be increased, since transcripts that escape NMD have mCherry at N-terminal and if mRNAs are read-though they should also escape NMD. Read-through efficiency can then be determined in individual cells using a simple fluorescence readout:

$$PTC\ Readthrough\ Efficiency \propto eGFP\ Integrated\ Fluorescence$$

Figure 3.31 shows the distribution of the compound's PTC "Read-through Efficiency" (see Materials and Methods for calculation of this parameter). These results show that despite the relatively low values for PTC "Read-through Efficiency", some compounds stand out from the rest. Also, the highest concentrations do not always give the best results, possibly due to toxicity effects. Thus, the optimal concentration in terms of efficacy/toxicity differs between compounds.

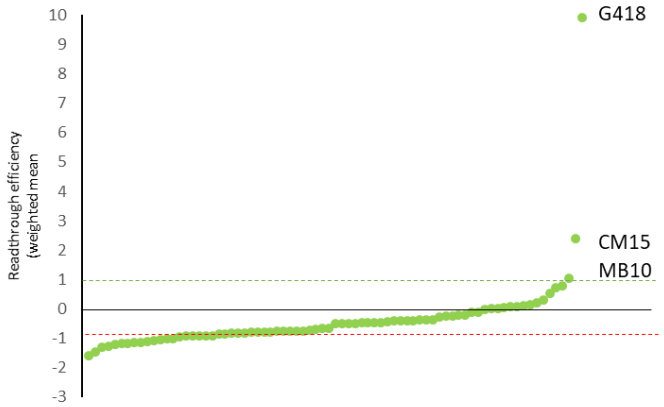



**Figure 3.31** – **Distribution of the PTC Read-through Efficiency of compound for the PTC read-through screening assay:** The distribution is performed according to the individual compound value from the negative control. Compounds with read-through efficiency are those having a PTC Read-through Efficiency above +1, as described in Materials and Methods.

We have identified in this preliminary study two compounds that stand out from the rest, in terms of their high scores. The compound with the highest read-through score is the G418, compound, used here as a positive control. It is also apparent that $Inh_{SMG1}$ only leads to an increase in mCherry fluorescence, which was also expected since this compound leads only to inhibition of NMD. Interestingly, the PTC124 compound that was in clinical trials for the correction of nonsense mutations and was previously reported to promote read-through, in our trial did not demonstrate any effect.

To further confirm the read-through efficiency of the tested compounds, complementary experiments are required using additional techniques, such as, qRT-PCR analysis of transcripts and Western blotting to confirm the presence of full-length CFTR. Finally, to

validate the top hit compounds, it is necessary to test them in patient-derived materials, including nasal primary cells for functional activity or intestinal organoids to determine the dose response and to test these compounds in combination with potentiators and correctors for maximal effect.

# *IV General Discussion and Perspectives*

## General Discussion

The growing understanding of human genetics, through the identification of disease associated variants, has contributed to our perception of human variability. With this genetic knowledge, the specificity of the clinical phenotype and the drug response of each individual are better understood. Genetics of CF has been widely studied. The observation that CF is a monogenic disease, caused by mutations in the CFTR gene, enables phenotype–genotype association studies (including drug response), which consider the wide clinical and laboratory spectrum dependent on the combined action of genotype, environment, and lifestyle. Despite the enormous advances in CF research since the identification of the CFTR gene in 1989, a cure cannot yet be offered to patients suffering from CF. The clinical manifestations of CF are complex, and even decades after the description of the disease and identification of the CFTR gene, not every spectrum of the disease is known or even understood, including the interaction between socioeconomic status and health outcomes (Oates and Schechter, 2016; Brown et al., 2017). For the vast majority of patients, the diagnosis of classic forms of CF is established early in life and suggested by one or more characteristic clinical features, including a history of CF in a sibling or a positive NBS result. Such diagnosis should ideally be supported by evidence of CFTR dysfunction through identification of two CF-disease causing mutations, two abnormal sweat-$Cl^-$ tests, and/or distinctive NPD measurements. The majority of treatments for CF aim to reduce its symptoms, such as mucolytics to reduce mucus viscosity, antibiotics to keep infection under control, anti-inflammatory drugs to suppress chronic inflammation, or digestive enzymes to compensate for pancreatic insufficiency. It is however, extremely important to treat the basic defect related to mutations in the *CFTR* gene so that these symptoms can be tackled globally.

The genetic part that characterizes the disease, determination of genes, their polymorphisms and mutations is essential to understand the individual phenotype. Furthermore, it is relevant for the study of new therapeutic models, such as conventional drug adjustment by individual patient dosage and drug type and response, development of new drugs, such as read-through, enhancer, stabilizer, and amplifier compounds.

The **first objective** of the present doctoral work was to identify CFTR gene mutations in patients with other chronic lung diseases. Since CF is a recessive disease, patients with just one *CFTR* gene mutation (i.e., carriers) are usually asymptomatic. However, several studies suggest the association of just one CFTR mutation with other respiratory diseases (Bombieri et al., 1998) such as DB (Pignatti et al., 1995), bronchopulmonary aspergillosis (Miller et al., 1996), sinusitis (Wang et al., 2000), COPD (Dahl and Nordestgaard, 2009) and bronchial asthma (Dahl et al., 1998b). With that purpose we analysed the 12 most common CFTR

mutations in Portuguese CF patients in non-CF patients with chronic lung diseases, such as DB (31 individuals), asthma (54 individuals) and COPD (51 individuals). Such CFTR mutations were detected by the ARMS, Tetra-ARMS and/or RFLP techniques.

Our data has shown that some patients with other respiratory diseases do carry CFTR mutations in one allele. We investigated the association of CFTR gene mutations with asthma and we have not found statistical significance, in contrast with other studies. We have found an increase in the frequency of *CFTR* gene mutations in patients with disseminated bronchiectasis, although it did not reach statistical significance, but these findings are in agreement with previous reports from other groups (Bombieri et al., 1998; Ngiam et al., 2006). However, we did find a significant increase in *CFTR* gene mutations in patients with COPD. In our study we focused only on the analysis of common *CFTR* gene mutations/variants. Although the F508del, the G542X and the G576A-R668C alleles were more frequent in the patient's group compared to the control group, the independent distribution of the tested CFTR mutations showed no statistical significance between COPD patients and controls. Our study included a small sample size resulting in a lower capability to detect minor to modest genotype disease association, future studies on larger cohorts should be performed.

In the COPD group we have identified one patient with the G542X mutation, a nonsense mutation that leads to the degradation of the transcripts and is associated with a severe phenotype. Another mutation identified was the complex-allele G576A-R668C, previously described as a polymorphism with varying effects on patient phenotypes. This patient has more than 10% of normal CFTR function and no CF phenotype, but from the mRNA analysis we conclude that G542X transcripts are degraded and that the G576A-R668C complex-allele leads to skipping of exon 13. Additional analysis to properly discriminate the effect of the G576A-R668C complex-allele in exon 13 skipping would require samples from a patient homozygous for this polymorphism.

The **second objective** of the current work was to analyse CFTR expression in native tissues (respiratory cells and colon) from patients to identify alterations in mRNA transcripts. Specifically, from Portuguese and Brazilian CF patients with unusual CFTR mutations potentially affecting gene expression through alternative splicing or NMD. Herein, we developed an RT-PCR analysis of CFTR transcripts for a select number of cases. These patients have sweat test values and activity of CFTR-mediated Cl- secretion in rectal biopsies that are suggestive of CF. RNA was extracted from either freshly-collected nasal epithelial cells (obtained by nasal scraping), colon biopsies (used for CF diagnosis) of CF individuals. In order to facilitate the identification of mutations, we have designed a mRNA-based protocol using just 9 RT-PCR reactions, whereby the complete CFTR coding region is analysed.

Moreover, one of the reactions for detection of the F508del mutation (exons 11-13) is based on the ARMS technique. With this 9 reaction technique we can detect any mutation occurring in the coding region as well as those that cause alternative splicing thus leading to the production of alternative transcripts.

This protocol is being applied to indirectly identify mutations in patients with total absence or residual CFTR-mediated $Cl^-$ secretion in rectal biopsies and with only 1 or no CFTR mutations identified by routine methods. With this technique we can detect for example a cryptic exon between exons 10-11 resulting from a mutation located deep into IVS 10 (c.1584+18672A>G), which is unlikely to be detected by CFTR exon sequencing. Other examples include the p.G576A, c.1717-1G>A, c.1812-1G>A, c.3272-26A>G, c.3120+1G>A mutations. We used this protocol in patients with a suspicion of CF and we were able to find 1 mutation that was already reported but had not been characterized, the 711+3A>T mutation, that we have shown causes skipping of IVS 5. We conclude that this is a rapid, robust and inexpensive method to detect rare mutations that can be easily used after a first screen. Furthermore, it can indicate the functional consequences of mutations in case they occur at the mRNA level.

The **third objective** of the current work was to identify key players in the NMD pathway by automated microscopy screens using a novel CFTR NMD-PTC mini-gene model. To identify factors involved in NMD, which may be potential drug targets, we developed an NMD CFTR mini-gene model. We have introduced in this mini-gene the G542X mutation in exon 12, which is the most common nonsense mutation in CF patients, and known to be related to a significantly lower production of CFTR transcripts. The mini-gene also features mCherry and eGFP tags at the CFTR N- and C-termini, respectively, which allow the assessment of the total CFTR levels (mCherry) and the “full-length” protein (eGFP). The construct also has a Flag-tag which becomes extracellular after G542X correction, originating mature mCherry-Flag-G542X-CFTR-eGFP. To induce NMD the nonsense mutation has to be recognized as a PTC. This in turn only occurs if an intron is downstream of the nonsense mutation, which will result as a consequence of the presence of proteins that remain at the EJC in the mRNA after the splicing process. For this reason, our cellular mini-gene NMD model contains three introns (IVS14, 15 and 16) downstream of the G542X mutation.

This model allowed us to use automated fluorescence microscopy to screen a shRNA library – a subset of The RNAi Consortium (TRC) – which is enriched in shRNAs targeting genes with a known or predicted involvement in transcript processing (425 genes), with the aim of identifying factors involved in the NMD pathway and detecting read-through compounds. This construct was stably transfected into HEK293 Flp-In cells, in parallel with a

similar wt-CFTR construct. After validation, the CFTR NMD-PTC reporter was used to screen the shRNA library using HT microscopy. Analysis of the data obtained from the primary screen led to the identification of 75 NMD inhibitors using the NMD score ≥ 2 criteria. When applying two different criteria (Z-Score ≥ 2 and at least 2 shRNA with the same phenotype) we obtained 11 top hits. Strikingly, in this screen for NMD inhibitors we also identified genes that show read-through activity. We identified 2 genes with read-through activity, PPWD1 and RSP19, and after literature search we selected them for further confirmation assays, since they are indirectly linked to NMD. For the confirmation screen (library of siRNAs), we included the 11 top hits form the primary screen, plus 2 genes from the screen that showed read-through activity and 11 genes from a parallel study related with alternative splicing. The results from our confirmation screen shown 4 genes with higher values of NMD score, however these values are much lower than expected. Thus, we are currently analysing this set of 24 genes using two different validation techniques: qRT-PCR and WB. Some of the hits identified within this study are potential drug targets due to their effects in NMD inhibition. Future data are expected to generate a valuable tool to shed light on the NMD pathway and on novel therapeutic targets for CF as well as for other diseases caused by PTC mutations.

The **fourth objective** of the current work was to restore functional protein production to PTC mutations using read-through compounds ultimately aimed at the treatment of CF patients. For this propose we analysed a library of 24 compounds which are PTC124 analogues/derivatives as well as non-PTC124-like compounds chemically synthesised by a group of collaborators. To perform HTS we used our NMD-PTC model (Figure 2.8) as before. If the presence of a compound with read-through properties leads to the production of “full-length” CFTR, this model enables the detection of the CFTR “full-length” protein by monitoring green fluorescence by fluorescence microscopy. This model, therefore, allows the identification of relevant (or “hit”) compounds in an automated and expedite way. Through the screen of the library, we identified new small-molecule compounds that induced PTC read-through.

To confirm the read-through efficiency more experiments are needed, using additional techniques, such as, WB and analysis of transcripts by semi-quantitative PCR and qRT-PCR. Finally, to validate the top hit compounds, it is necessary to test them in patient’s materials, including nasal primary cells for functional activity and intestinal organoids to determine a dose response and to test these compounds in combination with potentiators and correctors.

## *Concluding remarks and perspectives for future work.*

The studies presented in this dissertation had the overall aims of advancing the current knowledge of RNA-processing dysfunction, namely NMD, and of identifying novel compounds that modulate this process towards a "personalised medicine" approach. Previous work has demonstrated that a significant part (36%) of CF-causing mutations affects splicing efficiency and the stability of mRNA molecules. The results obtained in this thesis have indeed provided new knowledge pertaining to other respiratory diseases and the presence of CFTR mutations, new approaches to detect CFTR mutations in DNA and RNA and contributed to our understanding of the effect of NMD inhibition, as well as, read-through activity in the most common CFTR nonsense mutation. We have identified 3 relevant genes associated with splicing that lead to an increase in CFTR transcripts, overcoming the G542X mutation effect. The current results also opened new perspectives which merit further exploration.

Accordingly, future work should focus on the following aspects:

1) To further correlate the genotype-phenotype of patients with non-CF respiratory diseases, given that we have shown that there is an increased incidence of *CFTR* gene mutations among individuals with COPD. It thus becomes important to understand if individuals with one mutation in the *CFTR* gene have a greater predisposition for these diseases, constituting a particular subgroup of COPD disease which could benefit from approval CFTR modulators for CF.
2) To mechanistically validate the top hit genes identified as NMD inhibitors and test the possibility of using them as therapeutic targets for nonsense mutations in CF (and potentially in other diseases caused by PTC mutations).
3) More detailed studies need to be carried out to understand the general mechanism of action of compounds that show read-through activity to possibly use them in combination with NMD inhibitors as potential drugs for PTC mutations.

# *References*


Altamura N, Castaldo R, Finotti A, Breveglieri G, Salvatori F, Zuccato C, Gambari R, Panin GC, Borgatti M. 2013. Tobramycin is a suppressor of premature termination codons. J Cyst Fibros 806–811.

Amaral MD. 2014. Novel personalized therapies for cystic fibrosis: Treating the basic defect in all patients. J Intern Med 277:155–166.

Amaral MD, Blach WE. 2015. Hallmarks of Therapeutic Management of the Cystic Fibrosis Functional Landscape. J Clin Invest 14:687–699.

Amaral MD, Kunzelmann K. 2007. Molecular targeting of CFTR as a therapeutic approach to cystic fibrosis. Trends Pharmacol Sci 28:334–341.

Amaral MD, Pacheco P, Beck S, Farinha CM, Penque D, Nogueira P, Barreto C, Lopes B, Casals T, Dapena J, Gartner S, Vásquez C, et al. 2001. Cystic fibrosis patients with the 3272-26A>G splicing mutation have milder disease than F508del homozygotes: a large European study. J Med Genet 38:777–83.

Anagnostopoulou P, Riederer B, Duerr J, Michel S, Binia A, Agrawal R, Liu X, Kalitzki K, Xiao F, Chen M, Schatterny J, Hartmann D, et al. 2012. SLC26A9-mediated chloride secretion prevents mucus obstruction in airway inflammation. J Clin Invest 122:3629–3634.

Andersen DH. 1938. Cystic Fibrosis of the pancreas and its relations to celiac disease: A clinical and pathologic study. Am J Child 344–399.

Araújo D, Padrão E, Morais-Almeida M, Cardoso J, Pavão F, Leite RB, Caldas AC, Marques A. 2017. Asthma-chronic obstructive pulmonary disease overlap syndrome-Literature review and contributions towards a Portuguese consensus. Rev Port Pneumol (English Ed 23:90–99.

Arias-Palomo E, Yamashita A, Fernández IS, Núñez-Ramírez R, Bamba Y, Izumi N, Ohno S, Llorca O. 2011. The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. Genes Dev 25:153–164.

Baker JM, Hudson RP, Kanelis V, Choy W-Y, Thibodeau PH, Thomas PJ, Forman-Kay JD. 2007. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nat Struct Mol Biol 14:738–745.

Bárbara C, Rodrigues F, Dias H, Cardoso J, Almeida J, Matos MJ, Simão P, Santos M, Ferreira JR, Gaspar M, Gnatiuc L, Burney P. 2013. Chronic obstructive pulmonary disease prevalence in Lisbon, Portugal: The burden of obstructive lung disease study. Rev Port Pneumol 19:96–105.

Beck S, Penque D, Garcia S, Gomes A, Farinha C, Mata L, Gulbenkian S, Gil-Ferreira K, Duarte Â, Pacheco P, Barreto C, Lopes B, et al. 1999. Cystic fibrosis patients with the 3272-26A→G mutation have mild disease, leaky alternative mRNA splicing, and CFTR protein at the cell membrane. Hum Mutat 14:133–144.

Bergougnoux A, Viart V, Miro J, Bommart S, Molinari N, Georges M des, Claustres M, Chiron R, Taulan-Cadars M. 2015. Should diffuse bronchiectasis still be considered a CFTR-related disorder? J Cyst Fibros 14:646–653.

Bertrand CA, Zhang R, Pilewski JM, Frizzell RA. 2009. SLC26A9 is a constitutively active, CFTR-regulated anion conductance in human bronchial epithelia. J Gen Physiol 133:421–438.

Bienvenu T, Sermet-Gaudelus I, Burgel P-R, Hubert D, Crestani B, Bassinet L, Dusser D, Fajac I. 2010. Cystic fibrosis transmembrane conductance regulator channel dysfunction in non-cystic fibrosis bronchiectasis. Am J Respir Crit Care Med 181:1078–1084.

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

Boeck K De, Amaral MD. 2016. Progress in therapies for cystic fibrosis. Lancet Respir Med 4:662–674.

Boeck K De, Wilschanski M, Castellani C, Taylor C, Cuppens H, Dodge J, Sinaasappel M. 2006. Cystic fibrosis: terminology and diagnostic algorithms. Thorax 61:627–635.

Bombieri C, Benetazzo M, Saccomani A. 1998. Complete mutational screening of the CFTR gene in 120 patients with pulmonary disease. Hum Genet 103:718–722.

Bombieri C, Claustres M, Boeck K De, Derichs N, Dodge J, Girodon E, Sermet I, Schwarz M, Tzetis M, Wilschanski M, Bareil C, Bilton D, et al. 2011. Recommendations for the classification of diseases as CFTR-related disorders. J Cyst Fibros 10:86–102.

Bonini J, Varilh J, Raynal C, Thèze C, Beyne E, Audrezet M-P, Ferec C, Bienvenu T, Girodon E, Tuffery-Giraud S, Georges M Des, Claustres M, et al. 2015. Small-scale high-throughput sequencing–based identification of new therapeutic tools in cystic fibrosis. Genet Med 17:796–806.

Boucher RC. 2004. New concepts of the pathogenesis of cystic fibrosis lung disease. Eur Respir J 23:146–158.

Brogna S, Wen J. 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol 16:107–113.

Brown SD, White R, Tobin P. 2017. Keep them breathing: Cystic fibrosis pathophysiology, diagnosis, and treatment. J Am Acad Physician Assist 30:23–27.

Cabello GMK, Cabello PH, Jr JL, Fernandes O, Harris A. 2001. The 3120 + 1G → A Splicing Mutation in CFTR Is Common in Brazilian Cystic Fibrosis Patients. Hum Biol 73:403–409.

Cantin AM. 2016. Cystic Fibrosis Transmembrane Conductance Regulator. Implications in Cystic Fibrosis and Chronic Obstructive Pulmonary Disease. Ann Am Thorac Soc 13:S150–S155.

Cardoso J, Ferreira JR, Almeida J, Santos JM, Rodrigues F, Matos MJ, Gaspar M. 2013. Chronic obstructive pulmonary disease in portugal: Pneumobil (1995) and 2002 prevalence studies revisited. Rev Port Pneumol 19:88–95.

Carles S, Desgeorges M, Goldman A, Thiart R, Guittard C, Kitazos CA, Ravel TJ de, Westwood AT, Claustres M, Ramsay M. 1996. First report of CFTR mutations in black cystic fibrosis patients of southern African origin. J Med Genet 33:802–804.

Casals T, De-Gracia J, Gallego M, Dorca J, Rodríguez-Sanchón B, Ramos MD, Giménez J, Cisteró-Bahima A, Olveira C, Estivill X. 2004. Bronchiectasis in adult patients: An expression of heterozygosity for CFTR gene mutations? Clin Genet 65:490–495.

Casals T, Ramos MD, Larriba JGS, Nunes V, Estivill X. 1997. High heterogeneity for cystic fibrosis in Spanish families: 75 mutations account for 90% of chromosomes. Hum Genet 101:365–370.

Celik A, Kervestin S, Jacobson A. 2015. NMD: At the crossroads between translation termination and ribosome recycling. Biochimie 2–9.

CFTR2. 2018.

Chan CC, Dostie J, Diem MD, Feng W, Mann M, Rappsilber J, Dreyfuss G. 2004. eIF4A3 is a novel component of the exon junction complex. Rna 10:200–209.

Chillón M, Casals T, Giménez J, Nunes V, Estivill X. 1994. Analysis of the CFTR gene in the Spanish population: SSCP-screening for 60 known mutations and identification of four new mutations (Q30X, A120T, 1812-1 G→A, and 3667de14). Hum Mutat 3:223–230.

Chilvers ER. 2010. Advances in Neutrophil Biology : Clinical Implications. Respir Med 134:606–612.

Chiu S, Lejeune F, Ranganathan AC, Maquat LE. 2004. The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex. Genes Dev 18:745–754.

Cid R de, Chomel JC, Lazaro C, Sunyer J, Baudis M, Casals T, Le MN, Kitzis A, Feingold J, Anto J, Estivill X, Kauffmann F. 2001. CFTR and asthma in the French EGEA study. EurJHumGenet 9:67–69.

Clancy JP, Bebök Z, Ruiz F, King C, Jones J, Walker L, Greer H, Hong J, Wing L, Macaluso M, Lyrene R, Sorscher EJ, et al. 2001. Evidence that systemic gentamicin suppresses premature stop mutations in patients with cystic fibrosis. Am J Respir Crit Care Med 163:1683–1692.

Clancy JP, Rowe SM, Bebok Z, Aitken ML, Gibson R, Zeitlin P, Berclaz P, Moss R, Knowles MR, Oster RA, Mayer-Hamblett N, Ramsey B. 2007. No detectable improvements in cystic fibrosis transmembrane conductance regulator by nasal aminoglycosides in patients with cystic fibrosis with stop mutations. Am J Respir Cell Mol Biol 37:57–66.

Clarke LA, Botelho HM, Sousa L, Falcao AO, Amaral MD. 2015. Transcriptome meta-analysis reveals common differential and global gene expression profiles in cystic fibrosis and other respiratory disorders and identifies CFTR regulators. Genomics 106:268–277.

Clunes LA, Davies CM, Coakley RD, Aleksandrov AA, Henderson AG, Zeman KL, Worthington EN, Gentzsch M, Kreda SM, Cholon D, Bennett WD, Riordan JR, et al. 2012. Cigarette smoke exposure induces CFTR internalization and insolubility, leading to airway surface liquid dehydration. FASEB J 26:533–545.

Collins FS. 1992. Cystic fibrosis: molecular biology and therapeutic implications. Science (80- ) 256:774–779.

Costa PJ Da, Menezes J, Romão L. 2017. The role of alternative splicing coupled to nonsense-mediated mRNA decay in human disease. Int J Biochem Cell Biol 91(Pt B):168–175.

Costantino L, Claut L, Paracchini V, Coviello DA, Colombo C, Porcaro L, Capasso P, Zanardelli M, Pizzamiglio G, Degiorgio D, Seia M. 2010. A novel donor splice site characterized by CFTR mRNA analysis induces a new pseudo-exon in CF patients. J Cyst Fibros 9:411–418.

Crossley JR, Elliott RB, Smith PA. 1979. Dried-Blood Spot Screening for Cystic Fibrosis in the Newborn. Lancet 313:472–474.

Cutting GR. 2015. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat Rev Genet 16:45–56.

Cystic Fibrosis Mutation Database. 2017.

Dahl M, Nordestgaard B. 2009. Markers of early disease and prognosis in COPD. Int J Chron Obstruct Pulmon Dis 4:157–167.

Dahl M, Tybjaerg-Hansen A, Lange P, Nordestgaar B. 1998a. ΔF508 heterozygosity in cystic fibrosis and susceptibility to asthma. Lancet 351:1911–1913.

Dahl M, Tybjaerg-Hansen A, Lange P, Nordestgaar B. 1998b. ΔF508 heterozygosity in cystic fibrosis and susceptibility to asthma. Lancet 351:1911–1913.

Dahl M, Tybjærg-Hansen A, Lange P, Nordestgaard B. 2005. Asthma and COPD in cystic fibrosis intron-8 5T carriers. A population-based study. Respir Res 6:113.

Dang Y, Low WK, Xu J, Gehring NH, Dietz HC, Romo D, Liu JO. 2009. Inhibition of nonsense-mediated mRNA decay by the natural product pateamine a through eukaryotic initiation factor 4AIII. J Biol Chem 284:23613–23621.

Dekkers JF, Wiegerinck CL, Jonge HR de, Bronsveld I, Janssens HM, Winter-de Groot KM de, Brandsma AM, Jong NWM de, Bijvelds MJC, Scholte BJ, Nieuwenhuis EES, Brink S van den, et al. 2013. A functional CFTR assay using primary cystic fibrosis intestinal organoids. Nat Med 19:939–945.

Dixit P, Awasthi S, Maurya N, Agarwal S, Srinivasan M. 2013. CFTR Gene Mutations and Asthma in Indian Children: A Case–Control Study. Indian J Clin Biochem 30:35–42.

Douros K, Loukou I, Doudounakis S, Tzetis M, Priftis KN, Kanavakis E. 2008. Asthma and pulmonary function abnormalities in heterozygotes for cystic fibrosis transmembrane regulator gene mutations. Int J Clin Exp Med 1:345–9.

Dransfield MT, Wilhelm AM, Flanagan B, Courville C, Tidwell SL, Raju SV, Gaggar A, Steele C, Tang LP, Liu B, Rowe SM. 2013. Acquired cystic fibrosis transmembrane conductance regulator dysfunction in the lower airways in COPD. Chest 144:498–506.

Durand S, Cougot N, Mahuteau-Betzer F, Nguyen C-H, Grierson DS, Bertrand E, Tazi J, Lejeune F. 2007. Inhibition of nonsense-mediated mRNA decay (NMD) by a new chemical molecule reveals the dynamic of NMD factors in P-bodies. J Cell Biol 178:1145–1160.

Durante-Mangoni E, Grammatikos A, Utili R, Falagas ME. 2009. Do we still need the aminoglycosides? Int J Antimicrob Agents 33:201–205.

Eickmeier O, Huebner M, Herrmann E, Zissler U, Rosewich M, Baer PC, Buhl R, Schmitt-Grohé S, Zielen S, Schubert R. 2010. Sputum biomarker profiles in cystic fibrosis (CF) and chronic obstructive pulmonary disease (COPD) and association between pulmonary function. Cytokine 50:152–157.

El-Seedy A, Girodon E, Norez C, Pajaud J, Pasquet MC, Becdelièvre A de, Bienvenu T, Georges M des, Cabet F, Lalau G, Bieth E, Blayau M, et al. 2012. CFTR mutation combinations producing frequent complex alleles with different clinical and functional outcomes. Hum Mutat 33:1557–1565.

Elborn JS. 2016. Cystic fibrosis. Lancet 388:2519–2531.

Ellsworth RE, Jamison DC, Touchman JW, Chissoe SL, Braden Maduro V V, Bouffard GG, Dietrich NL, Beckstrom-Sternberg SM, Iyer LM, Weintraub L a, Cotton M, Courtney L, et al. 2000. Comparative genomic sequence analysis of the human and mouse cystic fibrosis transmembrane conductance regulator genes. Proc Natl Acad Sci U S A 97:1172–1177.

Erfle H, Neumann B, Liebel U, Rogers P, Held M, Walter T, Ellenberg J, Pepperkok R. 2007. Reverse transfection on cell arrays for high content screening microscopy. Nat Protoc 2:392–9.

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Duenas Roca C, et al. 2017. The Reactome Pathway Knowledgebase. Nucleic Acids Res 33:3461–3467.

Fajac I, Wainwright CE. 2017. New treatments targeting the basic defects in cystic fibrosis. Press Medicale 46:e165–e175.

Faria D, Schreiber R, Kunzelmann K. 2009. CFTR is activated through stimulation of purinergic P2Y2 receptors. Pflugers Arch 457:1373–1380.

Farrell PM, White TB, Derichs N, Castellani C, Rosenstein BJ. 2017a. Cystic Fibrosis Diagnostic Challenges over 4 Decades: Historical Perspectives and Lessons Learned. J Pediatr 181:S16–S26.

Farrell PM, White TB, Howenstine MS, Munck A, Parad RB, Rosenfeld M, Sommerburg O, Accurso FJ, Davies JC, Rock MJ, Sanders DB, Wilschanski M, et al. 2017b. Diagnosis of Cystic Fibrosis in Screened Populations. J Pediatr 181:S33–S44.

Farrell PM, White TB, Ren CL, Hempstead SE, Accurso F, Derichs N, Howenstine M, McColley SA, Rock M, Rosenfeld M, Sermet-Gaudelus I, Southern KW, et al. 2017c. Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation. J Pediatr 181:S4–S15.e1.

Farrell PMP, Rosenstein BJB, White TTB, Accurso FJ, Castellani C, Cutting GR, Durie PR, Legrys V a, Massie J, Parad RB, Rock MJ, Campbell PW, et al. 2008. Guidelines for Diagnosis of Cystic Fibrosis in Newborns through Older Adults: Cystic Fibrosis Foundation Consensus Report. J Pediatr 153:S4–S14.

Fatscher T, Boehm V, Gehring NH. 2015. Mechanism, factors, and physiological role of nonsense-mediated mRNA decay. Cell Mol Life Sci 72:4523–4544.

Feldmann D, Couderc R, Audrezet M-P, Ferec C, Bienvenu T, Desgeorges M, Claustres M, Mittre H, Blayau M, Bozon D, Malinge M-C, Monnier N, et al. 2003. CFTR genotypes in patients with normal or borderline sweat chloride levels. Hum Mutat 22:340.

Fossat N, Tourle K, Radziewic T, Barratt K, Liebhold D, Studdert JB, Power M, Jones V, Loebel DAF, Tam PPL. 2014. C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. EMBO Rep 15:903–910.

Fukuhara N, Ebert J, Unterholzner L, Lindner D, Izaurralde E, Conti E. 2005. SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Mol Cell 17:537–547.

Gadsby DC, Vergani P, Csanády L. 2006. The ABC protein turned chloride channel whose failure causes cystic fibrosis. Nature 440:477–483.

Gene Cards. 2018.

Gervais R, Lafitte J-J, Dumur V, Kesteloot M, Lalau G, Houdret N, Roussel P. 1993. Sweat chloride and ΔF508 mutation in chronic bronchitis or bronchiectasis. Lancet 342:997.

Ghosh A, Boucher RC, Tarran R. 2015. Airway Hydration and COPD Arunava. Cell Mol Life Sci 72:3637–3652.

Groman JD, Karczeski B, Sheridan M, Robinson TE, Fallin D, Cutting GR. 2005. PHENOTYPIC AND GENETIC CHARACTERIZATION OF PATIENTS WITH FEATURES OF “NONCLASSIC” FORMS OF CYSTIC FIBROSIS. J Pediatr 146:675–680.

Groman JD, Meyer ME, Wilmott RW, Zeitlin PL, Cutting GR. 2002. Variant Cystic Fibrosis Phenotypes in the Absence of CFTR Mutations. N Engl J Med 347:1892–1893.

Harman K, Dobra R, Davies JC. 2018. Disease-modifying drug therapy in cystic fibrosis. Paediatr Respir Rev 26:7–9.

Harris CM, Mendes F, Dragomir A, Doull IJM, Carvalho-Oliveira I, Bebok Z, Clancy JP, Eubanks V, Sorscher EJ, Roomans GM, Amaral MD, McPherson MA, et al. 2004.

Assessment of CFTR localisation in native airway epithelial cells obtained by nasal brushing. J Cyst Fibros 3:43–48.

Hassan F, Xu X, Nuovo G, Killilea DW, Tyrrell J, Tan C Da, Tarran R, Diaz P, Jee J, Knoell D, Boyaka PN, Cormet-Boyaka E. 2014. Accumulation of metals in GOLD4 COPD lungs is associated with decreased CFTR levels. Respir Res 15:1–9.

Hir H Le, Saulière J, Wang Z. 2016. The exon junction complex as a node of post-transcriptional networks. Nat Rev Mol Cell Biol 17:41–54.

https://github.com/hmbotelho/shinyHTM.

Hubert D, Chiron R, Camara B, Grenet D, Prévotat A, Bassinet L, Dominique S, Rault G, Macey J, Honoré I, Kanaan R, Leroy S, et al. 2017. Real-life initiation of lumacaftor/ivacaftor combination in adults with cystic fibrosis homozygous for the Phe508del CFTR mutation and severe lung disease. J Cyst Fibros 16:388–391.

Hug N, Longman D, Cáceres JF. 2015. Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Res 44:1483–1495.

Hwang L, Lou C-H, Chan W, Shum EY, Shao A, Stone E, Karam R, Song H, Wilkinson MF. 2011. RNA Homeostasis Governed by Cell Type-Specific and Branched Feedback Loops Acting on NMD. Mol Cell 43:950–961.

Hwang TC, Sheppard DN. 2009. Gating of the CFTR Cl − channel by ATP-driven nucleotide-binding domain dimerisation. J Physiol 587:2151–2161.

Igreja S, Clarke LA, Botelho HM, Marques L, Amaral MD. 2016. Correction of a Cystic Fibrosis Splicing Mutation by Antisense Oligonucleotides. Hum Mutat 37:209–215.

Inácio Â, Silva AL, Pinto J, Ji X, Morgado A, Almeida F, Faustino P, Lavinha J, Liebhaber SA, Romão L. 2004. Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. J Biol Chem 279:32170–32180.

Ishigaki Y, Li X, Serin G, Maquat LE. 2001. Evidence for a Pioneer Round of mRNA Translation. Cell 106:607–617.

Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JWB, Maquat LE. 2008. Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA Decay. Cell 133:314–327.

Iwatani-Yoshihara M, Ito M, Ishibashi Y, Oki H, Tanaka T, Morishita D, Ito T, Kimura H, Imaeda Y, Aparicio S, Nakanishi A, Kawamoto T. 2017. Discovery and Characterization of a Eukaryotic Initiation Factor 4A-3- Selective Inhibitor That Suppresses Nonsense-Mediated mRNA Decay. ACS Chem Biol 12:1760–1768.

Jarvi K, Zielenski J, Wilschanski M, Durie P, Buckspan M, Tullis E, Markiewicz D, Tsui LC. 1995. Cystic fibrosis transmembrane conductance regulator and obstructive azoospermia. Lancet 345:1578.

Jurica MS, Moore MJ. 2003. Pre-mRNA splicing: Awash in a sea of proteins. Mol Cell 12:5–14.

Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S. 2006. Nonsense-mediated mRNA decay (NMD) is a surveillance mechanism that degrades mRNA containing premature termination codons (PTCs). In mammalian cells, recognition of PTCs requires translation and depends on the presence on the mRNA with the splicing-depend. Genes Dev 355–367.

Keeling KM, Bedwell DM. 2011. Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases. Wiley Interdiscip Rev RNA 2:837–852.

Keeling KM, Du M, Bedwell DM. 2014a. Therapies of Nonsense-Associated Diseases. InMadame Curie Biosci Database.

Keeling KM, Xue X, Gunn G, Bedwell DM. 2014b. Therapeutics Based on Stop Codon Readthrough. Annu Re v Genomics Hum Genet Genomics Hum Genet 15:371–394.

Keenan K, Avolio J, Rueckes-Nilges C, Tullis E, Gonska T, Naehrlich L. 2015. Nasal potential difference: Best or average result for CFTR function as diagnostic criteria for cystic fibrosis? J Cyst Fibros 14:310–316.

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. 1989. Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080.

Kerem E. 2017. Cystic fibrosis: Priorities and progress for future therapies. Paediatr Respir Rev 24:14–16.

Kerem EMD, Konstan MW, Boeck K De, Accurso FJ, Bronsveld I, Fajac I, Malfroot A, Rosenbluth DB. 2014. A randomized placebo-controlled trial of ataluren for the treatment of nonsense mutation cystic fibrosis. Lancet Respir Med 2:539–547.

Kerkhoff C, Klempt M, Kaever V, Sorg C. 1999. The Two Calcium-binding Proteins, S100A8 and S100A9, Are Involved in the Metabolism of Arachidonic acid in Human Neutrophils. J Biol Chem 274:32672–32679.

Kervestin S, Jacobson A. 2012. NMD: a multifaceted response to premature translational termination Stephanie. Nat Rev Mol Cell Biol 13:700–712.

Khajavi M, Inoue K, Lupski JR. 2006. Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. Eur J Hum Genet 14:1074–1081.

Knowles MR, Clarke LL, Boucher RC. 1991. Activation by extracellular nucleotides of chloride secretion in the airway epithelia of patients with cystic fibrosis. N Engl J Med 325:533–538.

Kulyté A, Dryselius R, Karlsson J, Good L. 2005. Gene selective suppression of nonsense termination using antisense agents. Biochim Biophys Acta 1730:165–172.

Kurosaki T, Maquat LE. 2016. Nonsense-mediated mRNA decay in humans at a glance. J Cell Sci 129:461–467.

Kuroyanagi H. 2009. Fox-1 family of RNA-binding proteins. Cell Mol Life Sci 66:3895–3907.

Lérias J, Pinto M, Benedetto R, Schreiber R, Amaral M, Aureli M, Kunzelmann K. 2018. Compartmentalized crosstalk of CFTR and TMEM16A (ANO1) through EPAC1 and ADCY1. Cell Signal 44:10–19.

Levy H, Nugent M, Schneck K, Stachiw-Hietpas D, Laxova A, Lakser O, Rock M, Dahmer M, Biller J, Maker M, McColley S, Simpson P, et al. 2016. Refining the continuum of CFTR-associated disorders in the era of newborn screening. Clin Genet 89:539–549.

Liang F, Shang H, Jordan NJ, Wong E, Mercadante D, Saltz J, Mahiou J, Bihler HJ, Mense M. 2017. High-Throughput Screening for Readthrough Modulators of CFTR PTC Mutations. SLAS Technol 22:315–324.

Linde L, Boelz S, Nissim-Rafinia M, Oren YS, Wilschanski M, Yaacov Y, Virgilis D, Neu-yilik G, Kulozik AE, Kerem E, Kerem B. 2007. Nonsense-mediated mRNA decay affects nonsense transcript levels and governs response of cystic fibrosis patients to gentamicin. J Clin Invest 117:683–692.

López-Perrote A, Castaño R, Melero R, Zamarro T, Kurosawa H, Ohnishi T, Uchiyama A,

Aoyagi K, Buchwald G, Kataoka N, Yamashita A, Llorca O. 2015. Human nonsense-mediated mRNA decay factor UPF2 interacts directly with eRF3 and the SURF complex. Nucleic Acids Res 44:1909–1923.

Lowenfels AB, Maisonneuve P, Palys B, Schöni MH, Redemann B. 1998. Delta F508 heterozygosity and asthma. Lancet 352:985.

Lykke-Andersen S, Heick Jensen T. 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol 16:665–667.

Mall M, Wissner a, Seydewitz HH, Kuehr J, Brandis M, Greger R, Kunzelmann K. 2000. Defective cholinergic Cl(-) secretion and detection of K(+) secretion in rectal biopsies from cystic fibrosis patients. Am J Physiol Gastrointest Liver Physiol 278:G617–G624.

Marson FAL. 2017. Personalized or Precision Medicine? The Example of Cystic Fibrosis. Front Pharmacol 8:1–8.

Masvidal L, Igreja S, Ramos MD, Alvarez A, Gracia J de, Ramalho A, Amaral MD, Larriba S, Casals T. 2014. Assessing the residual CFTR gene expression in human nasal epithelium cells bearing CFTR splicing mutations causing cystic fibrosis. Eur J Hum Genet 22:784–791.

Mi H, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function analysis with the panther classification system. Nat Protoc 8:1551–1566.

Michl RK, Tabori H, Hentschel J, Beck JF, Mainz JG. 2016. Clinical approach to the diagnosis and treatment of cystic fibrosis and CFTR-related disorders. Expert Rev Respir Med 10:1177–1186.

Miller P, Hamosh A, Macek M, Greenberger P, MacLean J, Walden S, Salvin RG, Cutting GR. 1996. Cystic fibrosis transmembrane conductance regulator (CFTR) gene mutations in allergic bronchopulmonary aspergillosis. Am J Hum Genet 59:45–51.

Mishra BB, Moura-Alves P, Sonawane A, Hacohen N, Griffiths G, Moita LF, Anes E. 2010. Mycobacterium tuberculosis protein ESAT-6 is a potent activator of the NLRP3/ASC inflammasome. Cell Microbiol 12:1046–1063.

Moin M, Bakshi A, Madhav MS, Kirti PB. 2017. Expression Profiling of Ribosomal Protein Gene Family in Dehydration Stress Responses and Characterization of Transgenic Rice Plants Overexpressing RPL23A for Water-Use Efficiency and Tolerance to Drought and Salt Stresses. Front Chem 5:1–16.

Moniz S, Sousa M, Moraes BJ, Mendes AI, Palma M, Barreto C, Fragata JI, Amaral MD, Matos P. 2013. HGF Stimulation of Rac1 Signaling Enhances Pharmacological Correction of the Most Prevalent Cystic Fibrosis Mutant F508del- CFTR. ACS Chem Biol 8:.

Moura-Alves P, Neves-Costa A, Raquel H, Pacheco TR, D'Almeida B, Rodrigues R, Cadima-Couto I, Chora Â, Oliveira M, Gama-Carvalho M, Hacohen N, Moita LF. 2011. An shRNA-based screen of splicing regulators identifies SFRS3 as a negative regulator of IL-1β secretion. PLoS One 6:e19829.

Muthuswamy S, Agarwal S, Awasthi S, Singh S, Dixit P, Maurya N, Choudhuri G. 2014. Spectrum and distribution of CFTR gene mutations in asthma and chronic pancreatitis cases of North Indian population. Gene 539:125–131.

Mutyam V, Du M, Xue X, Keeling KM, White LE, Bostwick RoJ, Rasmussen L, Liu B, Mazur M, Hong J, Libby EF, Liang F, et al. 2016. Discovery of clinically approved agents that promote suppression of CFTR nonsense mutations. Am J Respir Crit Care Med 194:1092–1103.

Nagy E, Maquat LE. 1998. A rule for termination-codon position within intron-containing

genes: When nonsense affects RNA abundance. Trends Biochem Sci 23:198–199.

Neu-Yilik G, Amthor B, Gehring NH, Bahri S, Paidassi H, Hentze MW, Kulozik AE. 2011. Mechanism of escape from nonsense-mediated mRNA decay of human b-globin transcripts with nonsense mutations in the first exon. RNA 17:843–854.

Ngiam NSP, Chong SS, Shek LPC, Goh DLM, Ong KC, Chng SY, Yeo GH, Goh DYT. 2006. Cystic fibrosis transmembrane conductance regulator (CFTR) gene mutations in Asians with chronic pulmonary disease: a pilot study. J Cyst Fibros 5:159–64.

Nickless A, Bailis JM, You Z. 2017. Control of gene expression through the nonsense-mediated RNA decay pathway. Cell Biosci 7:26.

Nielsen AO, Qayum S, Bouchelouche PN, Laursen LC, Dahl R, Dahl M. 2016. Risk of asthma in heterozygous carriers for cystic fibrosis: A meta-analysis. J Cyst Fibros 15:563–567.

Norman M, Rivers C, Lee Y-B, Idris J, Uney J. 2016. The increasing diversity of functions attributed to the SAFB family of RNA-/DNA-binding proteins. Biochem J 473:4271–4288.

Oates GR, Schechter MS. 2016. Socioeconomic status and health outcomes: cystic fibrosis as a model. Expert Rev Respir Med 10:967–977.

Oberdoerffer S, Moita LF, Neems D, Freitas RP, Rao A. 2009. Regulation of CD45 Alternative Splicing by Heterogeneous Ribonucleoprotein, hnRNPLL. Science (80- ) 321:686–691.

Ooi CY, Dupuis A, Ellis L, Jarvi K, Martin S, Gonska T, Dorfman R, Kortan P, Solomon M, Tullis E, Durie PR. 2012. Comparing the American and European diagnostic guidelines for cystic fibrosis: same disease, different language? Thorax 67:618–624.

Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, Doudounakis S, Casals T, Baralle FE. 2003. New type of disease causing mutations: The example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet 12:1111–1120.

PANTHER. 2018. http://www.pantherdb.org/geneListAnalysis.do.

Paranjape SM, Zeitlin PL. 2008. Atypical cystic fibrosis and CFTR-related diseases. Clin Rev Allergy Immunol 35:116–123.

Peixeiro I, Inácio Â, Barbosa C, Silva AL, Liebhaber SA, Romão L. 2012. Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. Nucleic Acids Res 40:1160–1173.

Pignatti P., Bombieri C, Marigo C, Saccomani A. 1995. Increased incidence of cystic fibrosis gene mutations in adults with disseminated bronchiectasis. Hum Mol Genet 4:635–639.

Pignatti PF, Bombieri C, Benetazzo M, Casartelli A, Trabetti E, Gilè LS, Martinati LC, Boner a L, Luisetti M. 1996. CFTR gene variant IVS8-5T in disseminated bronchiectasis. Am J Hum Genet 58:889–892.

Popp MW, Maquat LE. 2017. Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. Cell 165:1319–1322.

PTC Therapeutics. 2017.

Puchelle E, Bajolet O, Abély M. 2002. Airway mucus in cystic fibrosis. Paediatr Respir Rev 3:115–119.

Quinton PM. 1983. Chloride impermeability in Cystic Fibrosis. Nature 301:421–422.

Raju S, Tate JH, Peacock SK, Fang P, Oster RA, Dransfield MT, Rowe SM. 2014. Impact of

heterozygote CFTR Mutations in COPD patients with Chronic Bronchitis. Respir Res 15:18.

Ramalho AS, Clarke LA, Amaral MD. 2011. Quantification of CFTR transcripts. Methods in Molecular Biology, p 115–135.

Ramalho AS, Lewandowska MA, Farinha CM, Mendes F, Gonçalves J, Barreto C, Harris A, Amaral MD. 2009. Deletion of CFTR translation start site reveals functional isoforms of the protein in CF patients. Cell Physiol Biochem 24:335–346.

Ratjen F, Döring G. 2003. Cystic fibrosis. Lancet 361:681–689.

Raxwal VK, Riha K. 2016. Nonsensemediated RNA decay and evolutionary capacitance. Biochim Biophys Acta 1859:1538–1543.

Reactome. 2017. https://reactome.org/.

Rich DP, Anderson MP, Gregory RJ, Cheng SH, Paul S, Jefferson DM, McCann JD, Klinger KW, Smith AE, Welsh MJ. 1990. Expression of cystic fibrosis transmembrane conductance regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells. Nature 347:358–363.

Riordan JR. 2005. Assembly of Functional CFTR Chloride Channels. Annu Rev Physiol 67:701–718.

Riordan JR, Rommens JM, Kerem B, Alon NO a, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou J, Drumm ML, Iannuzzi MC, et al. 1989. Identification the Cystic Fibrosis Gene : Cloning and Characterization of Complementary DNA. Science (80- ) 245:1066–1073.

Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, Zsiga M, Buchwald M, et al. 1989. Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping. Science (80- ) 245:1055–1065.

Rosenfeld M, Sontag MK, Ren CL. 2016. Cystic Fibrosis Diagnosis and Newborn Screening. Pediatr Clin North Am 63:599–615.

Rowe SM, Miller S, Sorscher EJ. 2005. Cystic fibrosis. N Engl J Med 352:1992–2001.

Roy B, Leszyk JD, Mangus DA, Jacobson A. 2015. Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. Proc Natl Acad Sci U S A 112:3038–3043.

Santos V, Cardoso AV, Lopes C, Azevedo P, Gamboa F, Amorim A. 2017. Cystic fibrosis – Comparison between patients in paediatric and adult age. Rev Port Pneumol 23:17–21.

Schneider EK, Reyes-Ortega F, Li J, Velkov T. 2017. Can Cystic Fibrosis Patients Finally Catch a Breath With Lumacaftor/Ivacaftor? Clin Pharmacol Ther 101:130–141.

Schroeder S, Gaughan D, Swift M. 1995. Protection against bronchial asthma by CFTR ΔF508 mutation: A heterozygote advantage in cystic fibrosis. Nat Med 1:703–705.

Schweingruber C, Rufener SC, Zünd D, Yamashita A, Mühlemann O. 2013. Nonsense-mediated mRNA decay - Mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochim Biophys Acta - Gene Regul Mech 1829:612–623.

Sehnert SS, Jiang L, Burdick JF, Risby TH. 2002. Breath biomarkers for detection of human liver diseases: preliminary study. Biomarkers 7:174–187.

Sheppard DN, Welsh MJ. 1999. Structure and function of the CFTR chloride channel. Physiol Rev 79:S23–S45.

Singh A, Singh N, Behera D, Sharma S. 2018. Role of polymorphic XRCC6 (Ku70)/XRCC7

(DNA-PKcs) genes towards susceptibility and prognosis of lung cancer patients undergoing platinum based doublet chemotherapy. Mol Biol Rep 45:253–261.

Solomon GM, Vamsee Raju S, Dransfield MT, Rowe SM. 2016. Therapeutic Approaches to Acquired Cystic Fibrosis Transmembrane Conductance Regulator Dysfunction in Chronic Bronchitis. Ann Am Thorac Soc 13:S169–S176.

Sosnay PR, Siklosi KR, Goor F Van, Kaniecki K, Yu H, Sharma N, Ramalho AS, Amaral MD, Dorfman R, Zielenski J, Masica DL, Karchin R, et al. 2013. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. Nat Genet 45:1160–1167.

Sousa M, Servidoni MF, Vinagre AM, Ramalho AS, Bonadia LC, Felício V, Ribeiro MA, Uliyakina I, Marson FA, Kmit A, Cardoso SR, Ribeiro JD, et al. 2012. Measurements of CFTR-Mediated Cl- Secretion in Human Rectal Biopsies Constitute a Robust Biomarker for Cystic Fibrosis Diagnosis and Prognosis. PLoS One 7:e47708.

Steiner B, Truninger K, Sanz J, Schaller A, Gallati S. 2004. The role of common single-nucleotide polymorphisms on exon 9 and exon 12 skipping in nonmutated CFTR alleles. Hum Mutat 24:120–129.

Taylor CJ, Hardcastle J, Southern KW. 2009. Physiological Measurements Confirming the Diagnosis of Cystic Fibrosis: The Sweat Test and Measurements of Transepithelial Potential Difference. Paediatr Respir Rev 10:220–226.

Team RC. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/].

The RNAi Consortium. 2017.

Trcek T, Sato H, Singer RH, Maquat LE. 2013. Temporal and spatial characterization of nonsense-mediated mRNA decay. Genes Dev 27:541–551.

Tzetis M, Efthymiadou A, Strofalis S, Psychou P, Dimakou A, Pouliou E, Doudounakis S, Kanavakis E. 2001. CFTR gene mutations--including three novel nucleotide substitutions--and haplotype background in patients with asthma, disseminated bronchiectasis and chronic obstructive pulmonary disease. Hum Genet 108:216–221.

UniProt. 2018.

Wang X, Moylan B, Leopold D, Kim J, Rubenstein R, Togias A, Proud D, Zeitlin PL, Cutting GR. 2000. Mutation in the gene responsible for cystic fibrosis and predisposition to chronic rhinosinusitis in the general population. JAMA 284:1814–1819.

Welch EM, Barton ER, Zhuo J, Tomizawa Y, Friesen WJ, Trifillis P, Paushkin S, Patel M, Trotta CR, Hwang S, Wilde RG, Karp G, et al. 2007. PTC124 targets genetic disorders caused by nonsense mutations. Nature 447:87–91.

Welsh M, Smith A. 1995. Cystic fibrosis. Sci Am 273:52–59.

WHO. 2004.

Wilschanski MMD, Yahav YMD, Yaacov YBS, Blau HMD, Bentur LMD, Rivlin JMD, Aviram MMD, Bdolah-Abram TMS, Bebok ZMD, Shushi LMS, Kerem BPD, Kerem EMD, et al. 2003. Gentamicin-Induced Correction of CFTR Function in Patients with Cystic Fibrosis and CFTR Stop Mutations. N Engl J Med 349:1433–1441.

Yamashita A. 2013. Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay. Genes to Cells 18:161–175.

Yamashita A, Izumi N, Kashima I, Ohnishi T, Saari B, Katsuhata Y, Muramatsu R, Morita T,

Iwamatsu A, Hachiya T, Kurata R, Hirano H, et al. 2009. SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. Genes Dev 23:1091–1105.

Zielenski J, Tsui LC. 1995. Cystic fibrosis: genotypic and phenotypic variations. Annu Rev Genet 29:777–807.

***Table S 4.3 – shRNA information***

| Gene Name | NM_Id | Entry | Gene Name | NM_Id | Entry |
|---|---|---|---|---|---|
| DDX3X | NM_001356 | O00571 | SMNDC1 | NM_005871 | O75940 |
| DHX15 | NM_001358 | O43143 | SFRS7 | NM_006276 | Q16629 |
| SNRP70 | NM_003089 | P08621 | SLU7 | NM_006425 | O95391 |
| PRPF18 | NM_003675 | Q99633 | USP39 | NM_006590 | Q53GS9 |
| DHX8 | NM_004941 | Q14562 | U2AF1 | NM_006758 | Q01081 |
| THOC1 | NM_005131 | Q96FV9 | U2AF2 | NM_007279 | P26368 |
| EWSR1 | NM_005243 | Q01844 | DHX38 | NM_014003 | Q92620 |
| SF3B4 | NM_005850 | Q15427 | DDX46 | NM_014829 | Q7L014 |
| RNPS1 | NM_006711 | Q15287 | PRPF31 | NM_015629 | Q8WWY3 |
| SF3A3 | NM_006802 | Q12874 | CDC40 | NM_015891 | O60508 |
| SF3A2 | NM_007165 | Q15428 | LUC7L | NM_018032 | Q9NQ29 |
| KIN | NM_012311 | O60870 | RBM17 | NM_032905 | Q96I25 |
| SF3B3 | NM_012426 | Q15393 | ZMAT2 | NM_144723 | Q96NC0 |
| SF3B1 | NM_012433 | O75533 | DHX9 | NM_001357 | Q08211 |
| SFRS2 | NM_003016 | Q01130 | PPM1G | NM_002707 | O15355 |
| C22orf19 | NM_003678 | Q13769 | SFRS3 | NM_003017 | P84103 |
| SFRS10 | NM_004593 | P62995 | SRPK1 | NM_003137 | Q96SB4 |
| PABPN1 | NM_004643 | Q86U42 | HNRPF | NM_004966 | P52597 |
| SFRS11 | NM_004768 | Q05519 | HNRPR | NM_005826 | O43390 |
| SRRM1 | NM_005839 | Q8IYB3 | HNRPM | NM_005968 | P52272 |
| SDCCAG10 | NM_005869 | Q6UX04 | RBM3 | NM_006743 | P98179 |
| SFRS5 | NM_006925 | Q13243 | DDX26 | NM_012141 | Q9UL03 |
| NUDT21 | NM_007006 | O43809 | DDX41 | NM_016222 | Q9UJV9 |
| CPSF6 | NM_007007 | Q16630 | RALY | NM_016732 | Q9UKM9 |
| CPSF1 | NM_013291 | Q10570 | DHX35 | NM_021931 | Q9H5Z1 |
| PPIL2 | NM_014337 | Q13356 | HNRPD | NM_031370 | Q14103 |
| PPWD1 | NM_015342 | Q96BP3 | HNRPU | NM_031844 | Q00839 |
| PPIL1 | NM_016059 | Q9Y3C6 | THOC3 | NM_032361 | Q96J01 |
| SRRM2 | NM_016333 | Q9UQ35 | SFRS12 | NM_139168 | Q8WXA9 |
| C21orf66 | NM_016631 | Q8N6E6 | CLK4 | NM_020666 | Q9HAZ1 |
| PPIL3 | NM_130906 | Q9H2H8 | CRK7 | NM_016507 | Q9NYV4 |
| THOC4 | NM_005782 | Q86V81 | DNAJC6 | NM_014787 | O75061 |
| PRPF4B | NM_003913 | Q13523 | RNGTT | NM_003800 | O60942 |
| CLK3 | NM_003992 | P49761 | PIAS1 | NM_016166 | O75925 |
| CLK2 | NM_003993 | P49760 | HDAC2 | NM_001527 | Q92769 |
| CLK1 | NM_004071 | P49759 | RPS3 | NM_001005 | P23396 |
| HSPA5 | NM_005347 | P11021 | SNRPD1 | NM_006938 | P62314 |
| HNRPA2B1 | NM_002137 | P22626 | RPS16 | NM_001020 | P62249 |
| PTBP1 | NM_002819 | P26599 | TNRC4 | NM_007185 | Q5SZQ8 |
| Pde4dip | NNM_031401 | | MGC5509 | NM_024093 | Q9BVC5 |
| SF1 | NM_004630 | Q15637 | EXOSC10 | NM_002685 | Q01780 |
| WTAP | NM_004906 | Q15007 | GNB2L1 | NM_006098 | P63244 |
| SFPQ | NM_005066 | P23246 | HNRPA1 | NM_002136 | P09651 |
| SART1 | NM_005146 | O43290 | HTATSF1 | NM_014500 | O43719 |
| NOVA1 | NM_006491 | P51513 | PRPF19 | NM_014502 | Q9UMS4 |
| SFRS1 | NM_006924 | Q07955 | SF3A1 | NM_005877 | Q15459 |
| SNW1 | NM_012245 | Q13573 | SNRPA | NM_004596 | P09012 |
| PTBP2 | NM_021190 | Q9UKA9 | MAGOH | NM_002370 | P61326 |

| Gene Name | NM_Id | Entry | Gene Name | NM_Id | Entry |
|---|---|---|---|---|---|
| CDC2L2 | NM_024011 | Q9UQ88 | POLR2B | NM_000938 | P30876 |
| STK23 | NM_014370 | Q9UPE1 | MYEF2 | NM_016132 | Q9P2K5 |
| RPS13 | NM_001017 | P62277 | RNPC2 | NM_004902 | Q14498 |
| FAU | NM_001997 | P35544 | ELAVL2 | NM_004432 | Q12926 |
| TOPORS | NM_005802 | Q9NS56 | RP13-297E16.1 | NM_005088 | Q02040 |
| NXF1 | NM_006362 | Q9UBU9 | ERCC3 | NM_000122 | P19447 |
| UBL5 | NM_024292 | Q9BZL1 | SAFB | NM_002967 | Q15424 |
| PARP1 | NM_001618 | P09874 | PLRG1 | NM_002669 | O43660 |
| YBX1 | NM_004559 | P67809 | TTF2 | NM_003594 | Q9UNY4 |
| EDG2 | NM_057159 | Q92633 | DNAJC8 | NM_014280 | O75937 |
| RBM22 | NM_018047 | Q9NW64 | NDUFA1 | NM_004541 | Q6IBB5 |
| HSPA1B | NM_005346 | P0DMV9 | VIM | NM_003380 | P08670 |
| ZNF207 | NM_003457 | O43670 | EEF1A1 | NM_001402 | P68104 |
| DEK | NM_003472 | P35659 | TRNT1 | NM_016000 | Q96Q11 |
| SMARCA5 | NM_003601 | O60264 | AKAP8 | NM_005858 | O43823 |
| CSDA | NM_003651 | P16989 | TNPO1 | NM_002270 | Q92973 |
| KHSRP | NM_003685 | Q92945 | ELAVL4 | NM_021952 | P26378 |
| FUBP1 | NM_003902 | Q96AE4 | MBD5 | NM_018328 | Q9P267 |
| G10 | NM_003910 | P41223 | RKHD3 | NM_032246 | Q6ZN04 |
| ZFP36L1 | NM_004926 | Q07352 | RBBP7 | NM_002893 | Q16576 |
| ILF2 | NM_004515 | Q12905 | RBM8A | NM_005105 | Q9Y5S9 |
| ILF3 | NM_004516 | Q12906 | XRCC6 | NM_001469 | P12956 |
| TCERG1 | NM_006706 | O14776 | LSM1 | NM_014462 | O15116 |
| C2orf3 | NM_003203 | P16383 | CCNK | NM_003858 | O75909 |
| DIDO1 | NM_022105 | Q9BTC0 | CCNA1 | NM_003914 | P78396 |
| QKI | NM_006775 | Q96PU8 | NOSIP | NM_015953 | Q9Y314 |
| GTL3 | NM_013242 | Q9Y6A4 | IQGAP1 | NM_003870 | P46940 |
| HCFC1 | NM_005334 | P51610 | PPIE | NM_006112 | Q9UNP9 |
| TAF6 | NM_005641 | P49848 | PPIH | NM_006347 | O43447 |
| ZFR | NM_016107 | Q96KR1 | TOP1MT | NM_052963 | Q969P6 |
| PCBP4 | NM_020418 | P57723 | KIAA1008 | NM_014953 | Q9Y2L1 |
| TIAL1 | NM_003252 | Q49AS9 | EXOSC2 | NM_014285 | Q13868 |
| SR140 | XM_031553 | O15042 | XRN2 | NM_012255 | Q9H0D6 |
| HNRPA0 | NM_006805 | Q13151 | REXO2 | NM_015523 | Q9Y3B8 |
| HNRPL | NM_001533 | P14866 | DDX54 | NM_024072 | Q8TDD1 |
| SIAHBP1 | NM_014281 | Q9UHX1 | MOV10 | NM_020963 | Q9HCE1 |
| SNRPB2 | NM_198220 |  | EP400 | NM_015409 | Q96L91 |
| SF3B14 | NM_016047 | Q9Y3B4 | EXOSC1 | NM_016046 | Q9Y3B2 |
| CIRBP | NM_001280 | Q14011 | EXOSC3 | NM_001002269 | Q9NQT5 |
| SYNCRIP | NM_006372 | O60506 | DDX1 | NM_004939 | Q92499 |
| ELAVL1 | NM_001419 | Q15717 | EXOSC5 | NM_020158 | Q9NQT4 |
| HSPA8 | NM_006597 | P11142 | DDX39 | NM_005804 | O00148 |
| CUGBP1 | NM_006560 | Q92879 | ADAR | NM_001111 | P55265 |
| ABT1 | NM_013375 | Q9ULW3 | EXOSC9 | NM_005033 | Q06265 |
| STRBP | NM_018387 | Q96SI9 | DDX49 | NM_019070 | Q9Y6V7 |
| KIAA0853 | NM_015070 | Q5T200 | DDX23 | NM_004818 | Q9BUQ8 |
| RUVBL1 | NM_003707 | Q9Y265 | EXOSC7 | NM_015004 | Q15024 |
| HMGB1 | NM_002128 | P09429 | DDX19B | NM_007242 | Q9UMR2 |
| HMGB3 | NM_005342 | O15347 | DDX21 | NM_004728 | Q9NR30 |
| GTF2I | NM_001518 | Q499G6 | EXOSC4 | NM_019037 | Q9NPD3 |

| Gene Name | NM_Id | Entry | Gene Name | NM_Id | Entry |
|---|---|---|---|---|---|
| SKIV2L | NM_006929 | Q15477 | KIAA1542 | NM_020901 | Q9P1Y6 |
| ASCC3L1 | NM_014014 | O75643 | UNK | NM_152302 | Q9C0B0 |
| eIF4A2 | NM_001967 | Q14240 | SNRPF | NM_003095 | P62306 |
| SKIV2L2 | NM_015360 | P42285 | TUBB | NM_178014 | Q5SU16 |
| PRKRA | NM_003690 | O75569 | LOC138046 | NM_173848 | Q86SE5 |
| POLDIP3 | NM_032311 | Q9BY77 | SF3B5 | NM_031287 | Q9BWJ5 |
| CHERP | NM_006387 | Q8IWX8 | FLJ21827 | NM_020153 | Q9NQC8 |
| KIAA1967 | NM_021174 | Q8N163 | C1orf55 | NM_152608 | Q6IQ49 |
| S100A8 | NM_002964 | P05109 | BAT1 | NM_004640 | Q13838 |
| S100A9 | NM_002965 | P06702 | DDX17 | NM_006386 | Q92841 |
| C1QBP | NM_001212 | Q07021 | C1orf60 | NM_023015 | Q68E01 |
| CD2BP2 | NM_006110 | O95400 | SNRPD2 | NM_004597 | P62316 |
| RBP7 | NM_052960 | Q96R05 | BRUNOL4 | NM_020180 | Q9BZC1 |
| NCBP1 | NM_002486 | Q09161 | XAB2 | NM_020196 | Q9HCS7 |
| FMR1 | NM_002024 | Q06787 | LSM10 | NM_032881 | Q969L4 |
| NCBP2 | NM_007362 | P52298 | FIP1L1 | NM_030917 | Q6UN15 |
| TPR | NM_003292 | P12270 | SNRPD3 | NM_004175 | P62318 |
| PSEN1 | NM_000021 | P49768 | BRUNOL5 | NM_021938 | Q8N6W0 |
| ROD1 | NM_005156 | O95758 | CRNKL1 | NM_016652 | Q9BZJ0 |
| DDX48 | NM_014740 | P38919 | LSM11 | NM_173491 | P83369 |
| SSB | NM_003142 | P05455 | REXO1 | NM_020695 | Q8N1G1 |
| ERVWE1 | NM_014590 | Q9UQF0 | BRUNOL6 | NM_052840 | Q96J87 |
| NCL | NM_005381 | P19338 | LSM5 | NM_012322 | Q9Y4Y9 |
| HNRPK | NM_002140 | P61978 | BUB3 | NM_004725 | O43684 |
| SMC2L1 | NM_006444 | O95347 | TIA1 | NM_022173 | P31483 |
| SMC1L1 | NM_006306 | Q14683 | C10orf116 | NM_006829 | Q15847 |
| MSI2 | NM_138962 | Q96DH6 | WBP11 | NM_016312 | Q9Y2W2 |
| FKBP3 | NM_002013 | Q00688 | MGC14151 | NM_032356 | Q9BRA0 |
| ELAVL3 | NM_001420 | Q14576 | MGC2803 | NM_024038 | Q9BQ61 |
| DGCR14 | NM_022719 | Q96DF8 | C20orf23 | NM_024704 | Q96L93 |
| MSI1 | NM_002442 | O43347 | MORG1 | NM_032332 | Q9BRX9 |
| TXNL4A | NM_006701 | P83876 | MKI67IP | NM_032390 | Q9BYG3 |
| KPNA2 | NM_002266 | P52292 | TFIP11 | NM_001008697 | Q9UBB9 |
| SNRPE | NM_003094 | P62304 | HNRPA3 | NM_194247 | P51991 |
| PNN | NM_002687 | Q9H307 | KIAA1604 | NM_020943 | Q9HCG8 |
| SNRPC | NM_003093 | Q5TAL4 | HNRPAB | NM_004499 | Q99729 |
| LSM4 | NM_012321 | Q9Y4Z0 | C13orf10 | NM_022118 | Q5T8P6 |
| LSM3 | NM_014463 | P62310 | RBM7 | NM_016090 | Q9Y580 |
| TUBA1 | NM_006000 | P68366 | TPX2 | NM_012112 | Q643R0 |
| SNRPB | NM_198216 | P14678 | NIF3L1BP1 | NM_025075 | Q6I9Y2 |
| SNRPA1 | NM_003090 | P09661 | RBM9 | NM_014309 | O43251 |
| RBMX | NM_002139 | P38159 | TRA2A | NM_013293 | Q13595 |
| A2BP1 | NM_145891 | Q9NWB1 | HNRPDL | NM_005463 | O14979 |
| TDRD3 | NM_030794 | Q9H7E2 | NONO | NM_007363 | Q15233 |
| SPPL3 | NM_139015 | Q8TCT6 | FLJ20273 | NM_019027 | A0AV96 |
| NUMA1 | NM_006185 | Q14980 | RY1 | NM_006857 | Q8WVK2 |
| MFAP1 | NM_005926 | P55081 | HIST1H2AC | NM_003512 | Q93077 |
| IK | NM_006083 | Q13123 | CSN3 | NM_005212 | P07498 |
| KIAA0773 | NM_014690 | Q86XD5 | RBMX2 | NM_016024 | Q9Y388 |
| SRP46 | NM_032102 | Q9BRL6 | HIST2H2AA | NM_003516 | Q6FI13 |

| Gene Name | NM_Id | Entry | Gene Name | NM_Id | Entry |
|---|---|---|---|---|---|
| HIST1H2BC | NM_003526 | P62807 | RPS15A | NM_001019 | P62244 |
| WDR57 | NM_004814 | Q96DI7 | AQR | NM_014691 | O60306 |
| CUGBP2 | NM_006561 | O95319 | RBM5 | NM_005778 | P52756 |
| HNRPH3 | NM_012207 | P31942 | PHF5A | NM_032758 | Q7RTV0 |
| RAVER1 | NM_133452 | Q8IY67 | RPS17 | NM_001021 | P08708 |
| U2AF1L2 | NM_005089 | Q15696 | RBM12 | NM_006047 | Q9NTZ6 |
| C20orf14 | NM_012469 | O94906 | MGC13125 | NM_032725 | Q9BRD0 |
| PABPC1 | NM_002568 | P11940 | RPS18 | NM_022551 | P62269 |
| ARS2 | NM_015908 | Q9BXP5 | MATR3 | NM_018834 | A0A0R4J2E8 |
| U2AF1L3 | NM_144987 | Q8WU68 | PRPF38A | NM_032284 | Q8NAV1 |
| EFTUD2 | NM_004247 | Q15029 | RPS19 | NM_001022 | P39019 |
| PABPC4 | NM_003819 | Q13310 | C9orf10 | NM_014612 | Q9NZB2 |
| DNAJC17 | NM_018163 | Q9NVM6 | RPS25 | NM_001028 | P62851 |
| PCBP1 | NM_006196 | Q15365 | ET | NM_024311 | O43934 |
| IMP-3 | NM_006547 | O00425 | RBMS1 | NM_016836 | P29558 |
| LSM2 | NM_021177 | Q9Y333 | DKFZP434K1421 | NM_032141 | Q9H0G5 |
| PCBP2 | NM_005016 | Q15366 | RPS29 | NM_001032 | P62273 |
| ACIN1 | NM_014977 | Q9UKV3 | BCAS2 | NM_005872 | O75934 |
| DDX6 | NM_004397 | P26196 | CPSF4 | NM_006693 | O95639 |
| PCBP3 | NM_020528 | P57721 | RDBP | NM_002904 | P18615 |
| RBM15 | NM_022768 | Q96T37 | SYF2 | NM_015484 | O95926 |
| RPS3A | NM_001006 | P61247 | FAM32A | NM_014077 | Q9Y421 |
| RPS4X | NM_001007 | P62701 | CCDC12 | NM_144716 | J3KR35 |
| LSM6 | NM_007080 | P62312 | RPS11 | NM_001015 | P62280 |
| KIAA1429 | NM_015496 | Q69YN4 | CPSF3 | NM_016207 | Q9UKF6 |
| HNRPUL1 | NM_007040 | Q9BUJ2 | C21orf70 | NM_058190 | Q9NSI2 |
| RBM10 | NM_005676 | P98175 | RPL5 | NM_000969 | P46777 |
| RPS4Y1 | NM_001008 | P22090 | GPATC1 | NM_018025 | Q9BRR8 |
| LSM7 | NM_016199 | Q9UK45 | CPSF2 | NM_017437 | Q9P2I0 |
| RBM25 | NM_021239 | P49756 | FRG1 | NM_004477 | Q14331 |
| RPS5 | NM_001009 | P46782 | RPL22 | NM_000983 | P35268 |
| LSM8 | NM_016200 | O95777 | CDC5L | NM_001253 | Q99459 |
| RPS7 | NM_001011 | P62081 | CXXC6 | NM_030625 | Q8NFU7 |
| PRPF4 | NM_004697 | O43172 | RPL23A | NM_000984 | P62750 |
| SF4 | NM_021164 | Q8IWZ8 | CSTF3 | NM_001326 | Q12996 |
| RPS8 | NM_001012 | Q5JR94 | RPL31 | NM_000993 | P62899 |
| PRPF3 | NM_004698 | O43395 | GRSF1 | NM_002092 | Q12849 |
| EIF2S2 | NM_003908 | P20042 | SRP9 | NM_003133 | P49458 |
| RPS9 | NM_001013 | P46781 | SAFB2 | NM_014649 | Q14151 |
| NHP2L1 | NM_001003796 | P55769 | HSPC148 | NM_016403 | Q9P013 |
| EIF3S10 | NM_003750 | Q14152 | SRP19 | NM_003135 | P09132 |
| RPS10 | NM_001014 | P46783 | HNRNPG-T | NM_014469 | O75526 |
| EIF3S2 | NM_003757 | Q5U0F4 | PRCC | NM_005973 | Q92733 |
| PSIP1 | NM_021144 | O75475 | SRP68 | NM_014230 | Q9UHB9 |
| RPS12 | NM_001016 | P25398 | HNRPCL1 | NM_001013631 | O60812 |
| EIF3S6 | NM_001568 | P60228 | SFRS16 | NM_007056 | Q8N2M8 |
| FUSIP1 | NM_006625 | O75494 | HNRPH2 | NM_019597 | P55795 |
| WDR33 | NM_001006622 | Q9C0J8 | HNRPLL | NM_138394 | Q8WVV9 |
| EIF3S6IP | NM_016091 | Q9Y262 | CTNNBL1 | NM_030877 | Q8WYA6 |
| WDR58 | NM_024339 | Q86W42 | PRPF8 | NM_006445 | Q6P2Q9 |

| Gene Name | NM_Id | Entry |
|---|---|---|
| KIAA1160 | NM_020701 | Q9ULR0 |
| SNRPN | NM_003097 | P63162 |
| HSPC117 | NM_014306 | Q9Y3I0 |
| SNIP1 | NM_024700 | Q8TAD8 |
| IMP-1 | NM_006546 | Q9NZI8 |
| HYPC | NM_012272 | Q6NWY9 |
| ZCCHC8 | NM_017612 | Q6NZY4 |
| SPEN | NM_015001 | Q96T58 |
| NHN1 | NM_144604 | Q86VM9 |
| DDB1 | NM_001923 | Q16531 |
| TCERG1 | NM_006706 | O14776 |
| HNRPC | NM_031314 | P07910 |
| POLR2A | NM_000937 | A0A0C4DGZ0 |
| EXOSC8 | NM_181503 | Q96B26 |
| SNRPG | NM_003096 | P62308 |
| NOVA2 | NM_002516 | Q9UNW9 |
| SMU1 | NM_018225 | Q2TAY7 |
| C14orf166 | NM_016039 | Q9Y224 |
| FUS | NM_004960 | P35637 |
| SFRS9 | NM_003769 | Q13242 |
| TAF15 | NM_003487 | Q92804 |
| RUVBL2 | NM_006666 | Q9Y230 |
| PRPF38B | NM_018061 | Q5VTL8 |
| HNRPH1 | NM_005520 | P31943 |
| RPS15 | NM_001018 | P62841 |
| CWF19L1 | NM_018294 | Q69YN2 |
| DDX5 | NM_004396 | P17844 |
| SFRS4 | NM_005626 | Q08170 |
| DHX16 | NM_003587 | O60231 |
| SFRS6 | NM_006275 | Q13247 |