Mutation profiling of the F508del CFTR allele using haplotype-resolved long-read next generation sequencing

Dario Dilernia[1,2], Pooneh Amin[2], Julie Flores[3], Arlene Stecenko[3], Eric Sorscher[3]

[1] Department of Pathology, School of Medicine, Emory University.

[2] Emory Vaccine Center, Emory University

[3] Department of Pediatrics, School of Medicine, Emory University.

Source of funding: DILERN17G0 (Cystic Fibrosis Foundation)

Corresponding author:
Dario A. Dilernia, Ph.D.
Assistant Professor,
amfAR Mathilde-Krim Fellow,
Emory Vaccine Center,
954 Gatewood Road
Atlanta, GA 30329
Phone: 404-727-5733
Fax: 404-727-9316
ddilern@emory.edu

ABSTRACT

Current approaches to characterize the mutational profile of CFTR are based on targeted mutation analysis (TMA) or whole gene studies derived from short-read next generation sequencing (NGS). However, these methods lack phasing capability which, in certain scenarios, can provide clinically valuable information. In the present work, we performed near-full length CFTR using Single-Molecule Real-Time Sequencing to produce haplotype resolved data from F508del homozygous and F508del compound heterozygous individuals. This approach utilizes target enrichment of the CFTR gene using biotinylated probes, facilitates multiplexing samples in the same sequencing run, and utilizes fully-automated bioinformatics pipelines for error correction and variant calling. We show a remarkable conservation of F508del haplotype, consistent with the single gene founder effect, as well as diverse mutational profiles in non-F508del alleles. By the same method, 105 single nucleotide polymorphisms exhibiting invariant linkage to F508del CFTR (which better define the founder haplotype) were identified. High level homology between F508del sequences derived from heterozygotes, and those obtained from homozygous individuals, demonstrate accuracy of this method to produce haplotype resolved sequencing. The studies provide a new diagnostic technology for detailed analysis of complex CFTR alleles linked to disease severity.

Keywords: CFTR, long-read NGS, haplotype

INTRODUCTION

Cystic fibrosis (CF) is an autosomal recessive genetic disease caused by mutations in the CF transmembrane conductance regulator (*CFTR*) gene that impair functionality of the CFTR protein(Welsh, Denning, Ostedgaard, & Anderson, 1993; Welsh & Smith, 1993). While CF is a

multi-organ system disorder, progressive lung disease is the major cause of morbidity and mortality. Defective CFTR leads to aberrant chloride transport in epithelial tissues, which in turn depletes airway surface liquid and causes defective mucociliary clearance, resulting in increased susceptibility to respiratory infection and inflammation(Strug, Stephenson, Panjwani, & Harris, 2018). There are over 2,000 mutations described in the *CFTR* gene, and more than 300 are known to be pathogenic ("Cyst. Fibros. Found., Johns Hopkins Univ. Hosp. Sick Child. CFTR2 variant list history," ; Sosnay et al., 2013). The most common mutation is F508del, which is present in more than 80% of patients diagnosed with CF(*Cystic Fibrosis Foundation Patient Registry. 2018 Annual Data Report*, 2019). This mutation leads to misfolding of the protein and prevents the gene product from reaching the cell membrane(Cheng et al., 1990; Jensen et al., 1995; Qu, Strickland, & Thomas, 1997; Ward, Omura, & Kopito, 1995). Previous studies have shown a remarkable conservation of *CFTR* alleles that harbor F508del, supporting the notion of a single origin for this disease-causing haplotype(Vecchio-Pagan et al., 2016). The current approach to resolving complex CFTR alleles in the clinic, mainly for diagnostic purposes, is to performing either Targeted mutation analysis (TMA) or whole gene sequencing based on short-read Next Generation Sequencing (NGS)(Bergougnoux et al., 2019; Bonini et al., 2015; Vecchio-Pagan et al., 2016). The limitation of these methods is that phasing information (i.e., linking different polymorphisms on the same allele) is lost. Based on the high likelihood that CFTR polymorphisms on an F508del background – either alone or in combination – can impact overall disease severity and therapeutic response, an urgent and practical need exists for better means of phasing CFTR DNA so that polymorphisms and their effects on clinical course, response to modulator treatment, protein folding, etc. can be evaluated and understood. The challenge of properly phasing (i.e., assigning DNA variation to a specific allele) has severely limited progress in this area for cystic fibrosis and other inherited diseases(Bansal, Tewhey, Topol, & Schork, 2011; Tewhey, Bansal, Torkamani, Topol, & Schork, 2011). The technology of choice for such a task is long-read NGS, which could allow fully-phased sequencing of CFTR alleles, and may provide a level of resolution with important clinical implications. In the present study, we describe a novel CFTR sequencing approach that implements Single-Molecule Real-Time sequencing (Pacific Biosciences, CA) to assess the potential of long-read CFTR analysis.

MATERIALS AND METHODS

Editorial Policies and Ethical Considerations. The present study involves samples from the Emory + Children's Center for Cystic Fibrosis and Airways Disease Research Biospecimen

Registry (CFBR). All patients willingly gave consent and were enrolled utilizing the informed consent process in accordance with the policies and regulations set forth by the Emory University Institutional Review Board(IRB00042577).

Sample processing. Whole-blood samples were collected from patients enrolled in the CF Cohort at the Emory University Center for Cystic Fibrosis Research through an IRB-approved protocol. DNA was isolated from 1-5ml of whole-blood using the Blood and Cell Culture DNA midi kit (Qiagen, 13343). Genomic DNA was quantified using a Qubit 4 Fluorometer (Invitrogen). Fragmentation of genomic DNA was performed using g-Tubes (Covaris, 520079), and 20-30ug of DNA. The g-tubes were centrifuged at 8000rpm for 2 minutes, followed by an inversion spin for 2 additional minutes. Sheared gDNA collected from g-tubes was concentrated using AMpure PB beads (PacBio, 100-265-900) in 50ul of H_2O. Libraries were prepared starting from 6-8 µg sheared genomic DNA following the PacBio multiplex genomic DNA target capture protocol with IDT Xgen Lockdown Probes. Briefly, sheared gDNA samples were end repaired and A-tailed using the KAPA hyper prep kit (Kapa Biosystems, 07962371001), and ligated to linear barcoded adapters, using up to 2-3ug DNA as input per ligation reaction. Barcoded gDNA was size selected using 0.75% BluePippin gel cassettes (Sage Science, BLF7510) in the BluePippin system (Sage Science PB01825). Size-selected barcoded gDNA was PCR amplified using 100uM PacBio universal primer set and the Takara LA Taq HS enzyme (Takara, RR042A). PCR-amplified, size-selected barcoded gDNA was subjected to target enrichment using xGen Lockdown Probes designed to span the CFTR gene, together with the XGen hybridization and wash kit (IDT, 1080584). Post-capture PCR amplification of enriched gDNA was performed under the same conditions as pre-capture PCR. CFTR enriched gDNA was used as input material to build the PacBio library prep using the SMRTbell template prep kit 1.0 (PacBio, 100-259-100) and barcoded adapters. Library preps were pooled after normalizing for concentration, and sequenced in a Sequel Instrument, using a 10-hour movie.

Variant calling. Raw data was demultiplexed using Lima (https://github.com/PacificBiosciences/barcoding). On-target sequencing reads were separated from off-target reads by mapping against the Grch38 human genome reference using MashMap (https://github.com/marbl/MashMap). Sequencing reads that mapped against the CFTR genomic coordinates were selected using in-house developed scripts. On-target sequencing reads were input by a pipeline previously developed for error correction and variant calling(Dilernia et al., 2015), with a sliding window approach to produce haplotype resolved error corrected sequences across the CFTR gene.

Phylogenetic analysis. CFTR haplotype sequences were aligned using MAFFT(Katoh, Asimenos, & Toh, 2009; Katoh, Misawa, Kuma, & Miyata, 2002; Katoh & Toh, 2010). Alignment segments were gap-stripped, segments with missing sequences removed, and Bayesian phylogenetic trees constructed using MrBayes(Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) under the Nst=6 and invgamma model for substitution variation among sites, using 1 million generations, four chain, and a burn-in of 25%. Phylogenetic trees were visualized with FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS

In the present study, haplotype-resolved near-full length sequencing of the CFTR gene was performed in 14 de-identified samples collected from CF patients, eight of which were previously typed as homozygotes for the F508del mutation, and six previously typed as heterozygous for this variant (Table 1). In each case, genomic DNA was extracted from whole-blood, sheared to an average size of 6-8Kb, and enriched in CFTR DNA using biotinylated probes that span most of the CFTR gene. During library preparation, samples were barcoded and then sequenced in a 4- or 5-plex approach using a Sequel I instrument (Pacific Biosciences). Raw sequencing reads were mapped against the human genome reference Grch38 to select on-target data, which were then mapped against a CFTR reference sequence.

After error correction and variant calling, polymorphic positions were defined as those that exhibit mutations present in at least two independent overlapping sets of sequencing reads. Five of the eight F508del homozygous samples showed no evidence of polymorphic positions across the entire sequenced gene region, suggesting homozygosity extending throughout the entire CFTR DNA interval. By considering unsupported nucleotide variation as variant calling errors, the error rate was estimated by comparing each individual sequence from these five patients against the consensus sequence produced by their analysis. A total 73 errors were found across the 847,269 nucleotide positions sequenced across the five subjects, indicating a global error rate of 0.008%. Error rate for substitutions (4 total mutations - 0.0005%) was significantly lower than error rate for INDELs (25 insertions and 44 deletions – 0.008%). All except two insertions, were located either in homopolymer tracks (N=37) or in short repeat sequences (N=34).

Using the F508del global consensus sequence derived from the five homozygous patients that showed no evidence of nucleotide variability from the reference sequence for F508del

haplotype, we assessed diversity among the remaining three F508del homozygous subjects exhibiting some level of variation across the CFTR gene. This new F508del reference sequence extended from 5,813 nucleotides upstream of the first position in the 5'UTR through 3,718 nucleotides downstream from the end of exon 24. There were two segments within this region that could not be sequenced due to lack of sufficient coverage: a region of 2,498 nucleotides within intron 3, spanning positions 7,508 to 10006 downstream from the end of exon 3. The second region with missing data comprised a 413 base/interval located between nucleotide 69 of exon 20 and nucleotide 255 of intron 20. As shown in Figure 1, a total of seven polymorphisms were identified across the three homozygous samples showing evidence of diversity. Due to the distance between polymorphisms and the high homology between alleles, polymorphisms in sample CF06 could not be unequivocally assigned (i.e., phased) to one or the other allele. In samples CF07 and CF08, only one polymorphism was observed in each study subject.

In six F508del heterozygous patients with CF, levels of nucleotide diversity facilitated long-range haplotype phasing across the three extended regions of the CF gene. There were two smaller segments in which sequencing coverage was not sufficient to perform variant calling, thereby limiting haplotype linkage of the phased regions described above. However, there was 100% homology between one allele of each subject for each of the three regions and the corresponding region of the F508del allele sequence obtained from homozygous samples. Accordingly, it was possible to infer linkage between haplotypes from three large CFTR segments based on the most parsimonious solution. In this manner, we observe an identical homology between F508del alleles obtained from heterozygous patients and consensus sequence obtained from homozygous individuals. In addition, based on results presented in Figure 2, extensive sequence divergence between F508del and non-F508del alleles among the six heterozygotes was demonstrated. In particular, samples CF13 and CF11 share identical F508del and non-F508del alleles. The evolutionary relationship between CFTR sequences derived from these patients was further investigated by reconstructing a Bayesian phylogenetic tree. As shown in Figure 3, while all alleles harboring the F508del mutation clustered together in a monophyletic clade with limited variability, CFTR alleles not harboring the F508del mutation appear very distant from each other in the phylogenetic analysis.

Collectively, sequence information was produced from 21 F508del alleles (16 alleles among the 8 homozygotes and 5 alleles among the 5 heterozygotes), and 105 single nucleotide polymorphisms were consistently observed across all 21 F508del alleles (compared to the

published CFTR consensus sequence) in addition to the F508del mutation. All the polymorphisms found in addition to F508del were located in intronic CFTR regions. Most of the mutations observed in F508del were common variants across multiple populations based on frequencies of these polymorphisms retrieved from gnomAD Database (Figure 4A). F508del linked polymorphisms were present at frequencies higher than 5% and only the F508del mutation (p.phe508del) and three polymorphisms located in a small cluster in intron 3 are infrequently observed across 9 populations with data available from gnomAD (i.e., at less than 2%) (Figure 4B).

DISCUSSION

In the present study, we performed detailed characterization of the mutational profile CFTR among both homozygous and heterozygous patients carrying the F508del mutation. Implementation of long-read NGS facilitated resolution not only of the F508del haplotype, but also of a number of non-F508del CFTR alleles, allowing a direct comparison that highlights the conserved nature of the F508del genetic background in contrast to significant variability of the overall CFTR locus.

Our results show that F508del alleles share a similar sequence, with only seven mutations found across the 21 CFTR alleles sequenced from homozygous and heterozygous individuals with CF. These findings are consistent with the expected highly-conserved mutation profile of the F508del CFTR allele, while also establishing that even in this limited cohort of patients, well supported mutations can be readily identified. The number of intronic polymorphisms identified even in the small sample size studied here has significance. Genome-wide association and other studies in cystic fibrosis directed towards understanding variability in overall CF prognosis and response to emerging modulator treatments have indicated significant levels of “missing heritability,” and suggested the importance of polymorphisms within CFTR, itself, as responsible for phenotypic heterogeneity. The finding that three of eight homozygous individuals for F508del exhibited at least one genetic difference in CFTR compared to the consensus F508del allele establishes the importance of further studies to investigate impact of polymorphisms such as those shown in Figure 2 for their effects on mRNA expression, splicing, utilization, etc., including clinical correlation with clinical pulmonary status and therapeutic responsiveness to emerging CFTR modulators. Such studies could indicate new contributors to “missing heritability” in the disease, and illustrate importance of properly phasing complex CFTR alleles (i.e., encoding multiple polymorphisms) to allow more complete interpretation for studies of this

type. The degree of variability among the F508del alleles observed in our study group also suggests new polymorphisms have emerged after the event that generated the original F508del mutation. This is supported, for example, by phylogenetic analysis showing the monophyletic nature of the entire F508del cluster (Figure 3).

The possibility of unequivocally identifying every mutation linked to F508del based on our ability to produce fully phased CFTR sequences can provide insight regarding the origin of this mutation. Examining the frequency distribution of the 105 SNPs identified in F508del allele across different populations is of interest in this regard. Frequency of F508del linked polymorphisms was found to be highly variable across populations, ranging from 0.1% to more than 50% depending on the subgroup. The presence of rare polymorphisms in F508del allele implies the CFTR allele that exhibits all these 105 SNPs is unusual and that the event that lead to the generation of the F508 deletion occurred in a rather rare CFTR allele. Intriguingly, the three polymorphisms that were found to be as infrequent as the F508del mutation itself cluster together in a small region of 706 nucleotides in intron 3. Whether this cluster plays a role that influences functional effects mediated by F508del is unknown, and this is the first study to our knowledge to report the cluster. Identifying and studying the phylogenetic history of the original F508del allele will provide important new information regarding the origin of CF, and would be possible by implementing the approach described in the present study for a larger CF population.

Finally, the present study suggests a means to better understanding disease outcome in patients with CF. While most F508del alleles share a very similar nucleotide sequence, patients diagnosed with CF disease who harbor F508del show significant variability in disease severity and overall phenotype. This observation applies to multiple organ systems affected by the disease, including respiratory failure(Castellani & Assael, 2017). Several studies have demonstrated high degrees of heritability for specific CF phenotypes(Blackman et al., 2006; Bradley, Blackman, Watson, Doshi, & Cutting, 2012; Mekus et al., 2000; Vanscoy et al., 2007), and indicate more than 50 percent of variations in disease outcome not linked to known CFTR mutations(O'Neal & Knowles, 2018). However, previous studies have been able to explain only 5% of such variability, and most of clinical heterogeneity remains unexplained(Corvol et al., 2015). It is possible that a large component of the missing heritability might be accounted for by the genetic makeup of the F508del allele (in addition to non-CFTR disease modifiers affecting CF clinical outcome), especially in F508del heterozygotes. Proper characterization of the F508del allele has been challenging since: 1) the extent of non-F508del DNA variation of the

F508del haplotype is poorly defined, and 2) methods have not been previously applied (or available) to properly phase complex polymorphisms on the F508del background. In this context, a method to perform *de novo* characterization of the mutational profile of the F508del allele, as presented here, will help advance understanding of the influence of CFTR haplotypes on disease outcome. The new approach to perform detailed mutation profiling of CFTR alleles therefore provides a means to better explain origins of CF and variability in disease phenotype, and possible also new insight relevant to predicting heterogeneity of clinical response to novel treatments.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

FIGURE LEGENDS

Figure 1. Highlight plot of the mutation profile of F508del homozygous patients with cystic fibrosis. Five of the eight F508del homozygous individuals studied in the present study showed 100% sequence identity across the entire targeted region. In the figure, location of the variants identified among the three F508del homozygous patients that exhibit mutations not present in the F508del consensus allele sequence are highlighted.

Figure 2. Highlight plot of mutational profile for F508del compound heterozygous individuals. All six F508del heterozygous patients shared a very similar F508del allelic sequence, showing 100% homology to the consensus F508del sequence produced from homozygous patients with CF. The figure describes location for polymorphisms in the CFTR sequence versus the GRCh38 human reference sequence.

Figure 3. Phylogenetic analysis of CFTR allelic sequences. A Bayesian phylogenetic tree was constructed using MrBayes. All F508del haplotype sequences cluster together, irrespective of whether they were obtained from homozygote or heterozygote patients. * denotes CFTR alleles harboring F508del mutation with non-consensus mutations as shown in Figure 1.

Figure 4. Population frequencies of dF508 SNPs. Frequency data was retrieved from the gnomAD Database for each of 105 polymorphisms observed in the dF508 CFTR allele consensus sequence. **Panel A.** Frequency of each polymorphism across the 9 population groups. **Panel B.** Count of the number of populations in which each polymorphism was present at a frequency lower than 0.02. Note that only F508del and a cluster of five mutations in intron 3 are rare variants across most populations.

REFERENCES

Bansal, V., Tewhey, R., Topol, E. J., & Schork, N. J. (2011). The next phase in human genetics. *Nat Biotechnol*, *29*(1), 38-39. doi:10.1038/nbt.1757

Bergougnoux, A., Deletang, K., Pommier, A., Varilh, J., Houriez, F., Altieri, J. P., . . . Taulan-Cadars, M. (2019). Functional characterization and phenotypic spectrum of three recurrent disease-causing deep intronic variants of the CFTR gene. *J Cyst Fibros*, *18*(4), 468-475. doi:10.1016/j.jcf.2018.10.012

Blackman, S. M., Deering-Brose, R., McWilliams, R., Naughton, K., Coleman, B., Lai, T., . . . Cutting, G. R. (2006). Relative contribution of genetic and nongenetic modifiers to intestinal obstruction in cystic fibrosis. *Gastroenterology*, *131*(4), 1030-1039. doi:10.1053/j.gastro.2006.07.016

Bonini, J., Varilh, J., Raynal, C., Theze, C., Beyne, E., Audrezet, M. P., . . . Taulan-Cadars, M. (2015). Small-scale high-throughput sequencing-based identification of new therapeutic tools in cystic fibrosis. *Genet Med*, *17*(10), 796-806. doi:10.1038/gim.2014.194

Bradley, G. M., Blackman, S. M., Watson, C. P., Doshi, V. K., & Cutting, G. R. (2012). Genetic modifiers of nutritional status in cystic fibrosis. *Am J Clin Nutr*, *96*(6), 1299-1308. doi:10.3945/ajcn.112.043406

Castellani, C., & Assael, B. M. (2017). Cystic fibrosis: a clinical view. *Cell Mol Life Sci*, *74*(1), 129-140. doi:10.1007/s00018-016-2393-9

Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., . . . Smith, A. E. (1990). Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell*, *63*(4), 827-834. doi:10.1016/0092-8674(90)90148-8

Corvol, H., Blackman, S. M., Boelle, P. Y., Gallins, P. J., Pace, R. G., Stonebraker, J. R., . . . Knowles, M. R. (2015). Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nat Commun*, *6*, 8382. doi:10.1038/ncomms9382

Cyst. Fibros. Found., Johns Hopkins Univ. Hosp. Sick Child. CFTR2 variant list history. Retrieved from https://cftr2.org/mutations_history

Cystic Fibrosis Foundation Patient Registry. 2018 Annual Data Report. (2019). Retrieved from Bethesda, Mayrland: https://www.cff.org/Research/Researcher-Resources/Patient-Registry/2018-Patient-Registry-Annual-Data-Report.pdf

Dilernia, D. A., Chien, J. T., Monaco, D. C., Brown, M. P., Ende, Z., Deymier, M. J., . . . Hunter, E. (2015). Multiplexed highly-accurate DNA sequencing of closely-related HIV-1 variants using continuous long reads from single molecule, real-time sequencing. *Nucleic Acids Res*, *43*(20), e129. doi:10.1093/nar/gkv630

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754-755. doi:10.1093/bioinformatics/17.8.754

Jensen, T. J., Loo, M. A., Pind, S., Williams, D. B., Goldberg, A. L., & Riordan, J. R. (1995). Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. *Cell*, *83*(1), 129-135. doi:10.1016/0092-8674(95)90241-4

Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol*, *537*, 39-64. doi:10.1007/978-1-59745-251-9_3

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*, *30*(14), 3059-3066. doi:10.1093/nar/gkf436

Katoh, K., & Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, *26*(15), 1899-1900. doi:10.1093/bioinformatics/btq224

Mekus, F., Ballmann, M., Bronsveld, I., Bijman, J., Veeze, H., & Tummler, B. (2000). Categories of deltaF508 homozygous cystic fibrosis twin and sibling pairs with distinct phenotypic characteristics. *Twin Res*, *3*(4), 277-293. doi:10.1375/136905200320565256

O'Neal, W. K., & Knowles, M. R. (2018). Cystic Fibrosis Disease Modifiers: Complex Genetics Defines the Phenotypic Diversity in a Monogenic Disease. *Annu Rev Genomics Hum Genet*, *19*, 201-222. doi:10.1146/annurev-genom-083117-021329

Qu, B. H., Strickland, E., & Thomas, P. J. (1997). Cystic fibrosis: a disease of altered protein folding. *J Bioenerg Biomembr*, *29*(5), 483-490. doi:10.1023/a:1022439108101

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572-1574. doi:10.1093/bioinformatics/btg180

Sosnay, P. R., Siklosi, K. R., Van Goor, F., Kaniecki, K., Yu, H., Sharma, N., . . . Cutting, G. R. (2013). Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet*, *45*(10), 1160-1167. doi:10.1038/ng.2745

Strug, L. J., Stephenson, A. L., Panjwani, N., & Harris, A. (2018). Recent advances in developing therapeutics for cystic fibrosis. *Hum Mol Genet*, *27*(R2), R173-R186. doi:10.1093/hmg/ddy188

Tewhey, R., Bansal, V., Torkamani, A., Topol, E. J., & Schork, N. J. (2011). The importance of phase information for human genomics. *Nat Rev Genet*, *12*(3), 215-223. doi:10.1038/nrg2950

Vanscoy, L. L., Blackman, S. M., Collaco, J. M., Bowers, A., Lai, T., Naughton, K., . . . Cutting, G. R. (2007). Heritability of lung disease severity in cystic fibrosis. *Am J Respir Crit Care Med*, *175*(10), 1036-1043. doi:10.1164/rccm.200608-1164OC

Vecchio-Pagan, B., Blackman, S. M., Lee, M., Atalar, M., Pellicore, M. J., Pace, R. G., . . . Cutting, G. R. (2016). Deep resequencing of CFTR in 762 F508del homozygotes reveals clusters of non-coding variants associated with cystic fibrosis disease traits. *Hum Genome Var*, *3*, 16038. doi:10.1038/hgv.2016.38

Ward, C. L., Omura, S., & Kopito, R. R. (1995). Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell*, *83*(1), 121-127. doi:10.1016/0092-8674(95)90240-6

Welsh, M. J., Denning, G. M., Ostedgaard, L. S., & Anderson, M. P. (1993). Dysfunction of CFTR bearing the delta F508 mutation. *J Cell Sci Suppl*, *17*, 235-239.

Welsh, M. J., & Smith, A. E. (1993). Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell*, *73*(7), 1251-1254. doi:10.1016/0092-8674(93)90353-r