BIF 701: Database Searching for Cystic Fibrosis Associated Genes and Proteins

By Christopher Eeles

Introduction

Cystic Fibrosis (CF)—formerly know as Cystic Fibrosis of the pancreas—is an autosomal recessive genetic disorder.[2] Manifestation of the disorder affects exocrine function of the pancreas, intestinal glands, biliary tree, bronchial glands, and sweat glands as well as inducing infertility.[2] Based on a 2012 study by Feuchtbaum *et al.* birth prevalence in California for CF was approximately 20 per 100 000 births across all races/ethnicities.[2] The results of this study are comparable to the estimated prevalence from Cystic Fibrosis Canada of 1 in 3600, which works out to about 28 per 100 000 births.[1] Incidence of CF have been decreasing slowly over time, likely due to increased use of prenatal screening and advances in genetic testing which lower cost.[2] Associated molecular genetics and manifestation of pathology in CF will be discusses in subsequent sections.

Results

Initial search was conducted for "Cystic Fibrosis" in OMIM database yield an overview of the disorder. A link to the NCBI therein yielded genetic information and gene sequence data. Links were also followed to UniProt to retrieve information about the proteomics and molecular mechanism of action in CF pathology.

Gene information

Mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene are associated with CF phenotypes. Considering that the gene is autosomal recessive, CF phenotypes only arise with homozygous or compound heterozygous mutations (*i.e.*, mutations on both chromosomes).[3] The cytogenetic location is 7q31.2 meaning it is located on the small arm of chromosome 7 in region 3, band 1, sub-band 2.[3] The gene ID in NCBI is 1080 and 60241 in OMIM.[4]

FASTA Sequence: Gene[4a]

```
>NG_016465.4:19180-207882 Homo sapiens cystic fibrosis transmembrane
conductance regulator (CFTR), RefSeqGene (LRG_663) on chromosome 7
AATTGGAAGCAAATGACATCACAGCAGGTCAGAGAAAAAGGGTTGAGCGGCAGGCACCCAGAGTAGTAGG
TCTTTGGCATTAGGAGCTTGAGCCCAGACGGCCCTAGCAGGGACCCCAGCGCCCGAGAGACCATGCAGAG
GTCGCCTCTGGAAAAGGCCAGCGTTGTCTCCAAACTTTTTTTCAGGTGAGAAGGTGGCCAACCGAGCTTC
GGAAAGACACGTGCCCACGAAAGAGGAGGGCGTGTGTATGGGTTGGGTTTGGGGTAAAGGAATAAGCAGT
TTTTAAAAAGATGCGCTATCATTCATTGTTTTGAAAGAAAATGTGGGTATTGTAGAATAAAACAGAAAGC
…
ATTGACTTTTTATGGCACTAGTATTTCTATGAAATATTATGTTAAAACTGGGACAGGGGAGAACCTAGGG
TGATATTAACCAGGGGCCATGAATCACCTTTTGGTCTGGAGGGAAGCCTTGGGGCTGATGCAGTTGTTGC
CCACAGCTGTATGATTCCCAGCCAGCACAGCCTCTTAGATGCAGTTCTGAAGAAGATGGTACCACCAGTC
TGACTGTTTCCATCAAGGGTACACTGCCTTCTCAACTCCAAACTGACTCTTAAGAAGACTGCATTATATT
TATTACTGTAAGAAAATATCACTTGTCAATAAAATCCATACATTTGTGTGAAA
```

For full sequence please see the link provided in reference 4a.

Protein Information

The protein product of the CFTR gene is an epithelial ion channel present in the various body tissues affected by CF.[5] The transporter is a member of the ATP-binding cassette family and functions as a low conductance Cl(-) channel gated by ATP binding and hydrolysis.[3,5] It also plays a role in regulation of other

ion channels and transporters.[5] Overall the proteins plays a role in fluid homeostasis as well as contributing to pH and ion content of body surface fluids.[5] The implications for generating the different aspects of CF pathology will be addressed in the discussion section.

FASTA Sequence: Protein[5a]

```
>sp|P13569|CFTR_HUMAN Cystic fibrosis transmembrane conductance regulator
OS=Homo sapiens OX=9606 GN=CFTR PE=1 SV=3
MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDRE
LASKKNPKLINALRRCFFWRFMFYGIFLYLGEVTKAVQPLLLGRIIASYDPDNKEERSIA
IYLGIGLCLLFIVRTLLLHPAIFGLHHIGMQMRIAMFSLIYKKTLKLSSRVLDKISIGQL
VSLLSNNLNKFDEGLALAHFVWIAPLQVALLMGLIWELLQASAFCGLGFLIVLALFQAGL
GRMMMKYRDQRAGKISERLVITSEMIENIQSVKAYCWEEAMEKMIENLRQTELKLTRKAA
YVRYFNSSAFFFSGFFVVFLSVLPYALIKGIILRKIFTTISFCIVLRMAVTRQFPWAVQT
WYDSLGAINKIQDFLQKQEYKTLEYNLTTTEVVMENVTAFWEEGFGELFEKAKQNNNNRK
TSNGDDSLFFSNFSLLGTPVLKDINFKIERGQLLAVAGSTGAGKTSLLMVIMGELEPSEG
KIKHSGRISFCSQFSWIMPGTIKENIIFGVSYDEYRYRSVIKACQLEEDISKFAEKDNIV
LGEGGITLSGGQRARISLARAVYKDADLYLLDSPFGYLDVLTEKEIFESCVCKLMANKTR
ILVTSKMEHLKKADKILILHEGSSYFYGTFSELQNLQPDFSSKLMGCDSFDQFSAERRNS
ILTETLHRFSLEGDAPVSWTETKKQSFKQTGEFGEKRKNSILNPINSIRKFSIVQKTPLQ
MNGIEEDSDEPLERRLSLVPDSEQGEAILPRISVISTGPTLQARRRQSVLNLMTHSVNQG
QNIHRKTTASTRKVSLAPQANLTELDIYSRRLSQETGLEISEEINEEDLKECFFDDMESI
PAVTTWNTYLRYITVHKSLIFVLIWCLVIFLAEVAASLVVLWLLGNTPLQDKGNSTHSRN
NSYAVIITSTSSYYVFYIYVGVADTLLAMGFFRGLPLVHTLITVSKILHHKMLHSVLQAP
MSTLNTLKAGGILNRFSKDIAILDDLLPLTIFDFIQLLLIVIGAIAVVAVLQPYIFVATV
PVIVAFIMLRAYFLQTSQQLKQLESEGRSPIFTHLVTSLKGLWTLRAFGRQPYFETLFHK
ALNLHTANWFLYLSTLRWFQMRIEMIFVIFFIAVTFISILTTGEGEGRVGIILTLAMNIM
STLQWAVNSSIDVDSLMRSVSRVFKFIDMPTEGKPTKSTKPYKNGQLSKVMIIENSHVKK
DDIWPSGGQMTVKDLTAKYTEGGNAILENISFSISPGQRVGLLGRTGSGKSTLLSAFLRL
LNTEGEIQIDGVSWDSITLQQWRKAFGVIPQKVFIFSGTFRKNLDPYEQWSDQEIWKVAD
EVGLRSVIEQFPGKLDFVLVDGGCVLSHGHKQLMCLARSVLSKAKILLLDEPSAHLDPVT
YQIIRRTLKQAFADCTVILCEHRIEAMLECQQFLVIEENKVRQYDSIQKLLNERSLFRQA
ISPSDRVKLFPHRNSSKCKSKPQIAALKEETEEEVQDTRL
```

Discussion and Conclusion

Cystic Fibrosis (CF) is an exocrine gland disorder which results in inadequate clearance of secretions on various body surfaces.[5] Defective regulation of ion channels leads to inappropriate gating, and therefore closure of ion transporters.[3] The result is organ specific impairment of secretion and/or reabsorption of salt and fluids.[3] In the pancreas, intestine and biliary tree this results in blockage of exocrine outflow and therefore insufficient secretion of enzymes, fluids and small molecules necessary for normal tissue function.[3] Pancreatic insufficiency can the lead to malabsorption and associated growth retardation.[3] Respiratory disruptions result in accumulation of dehydrated mucus in the bronchi and bronchioles; this, combined with decrease pH of lung surface secretions, leads to decrease immune function and increased risk of respiratory infection.[3] Defects in chloride and pH regulation are also the major cause of infertility in both men and women due to impaired formation and development of gametes.[3] Sweat gland pathology affects reabsorption of salt and fluid, yielding electrolyte depletion and the salty tasting sweat characteristic of cystic fibrosis.[3]

Such wide-ranging manifestations of mutations in one gene illustrate both the complexity of physiological systems and the necessity of understanding them in order to improve human health and wellbeing. Cystic Fibrosis is the leading genetic disorder in Caucasians and current life expectancies sit

around 30 years; the disorder causes morbidity, and eventual mortality, in ~4100 Canadians at the time of this report.[1] Thus the importance of understanding the molecular biology, genomics and proteomics underlying the physiological manifestation is apparent. Application of bioinformatics to this and other genetic disorders can improve the rate and efficiency of research and therefore ultimately reduce the economic, societal and personal costs of genetic mutations and their resultant pathologies.

References

1. Cystic Fibrosis Canada [CFC] (2014). The Canadian Cystic Fibrosis Registry: 2014 Annual Report. Retrieved from https://cysticfibrosis.uberflip.com/i/705240-cysticfibrosis-canada-registry.

2. McKusick, V.A., Hamosh, A. (2018). Cystic Fibrosis; CF. In *OMIM Entry #219700*. Retrieved September 28, 2018 from http://omim.org/entry/219700?search=Cystic%20Fibrosis&highlight=fibrosi%20cystic.

3. McKusick, V.A., Hamosh, A. (2018). Cystic Fibrosis Transmembrane Conductance Regulator; CFTR. In *OMIM Entry *60241*. Retrieved September 28, 2018 from http://omim.org/entry/602421?search=Cystic%20Fibrosis&highlight=fibrosi%20cystic.

4. NCBI. (2018). CFTR cystic fibrosis transmembrane conductance regulator [*Homo Sapiens* (humans)]. Retrieved from https://www.ncbi.nlm.nih.gov/gene/1080.
    a. FASTA Sequence. Retrieved September 30, 2018 from https://www.ncbi.nlm.nih.gov/nuccore/NG_016465.4?report=fasta&from=19180&to=207882.

5. Uniprot. (2018). UniProtKB – P13560 (CFTR_HUMAN). Retrieved September 28, 2018 from https://www.uniprot.org/uniprot/P13569.
    a. FASTA Sequence. Retrieved September 28, 2018 from https://www.uniprot.org/uniprot/P13569.fasta.