cftbarcodes Documentation

Release 1.0.0

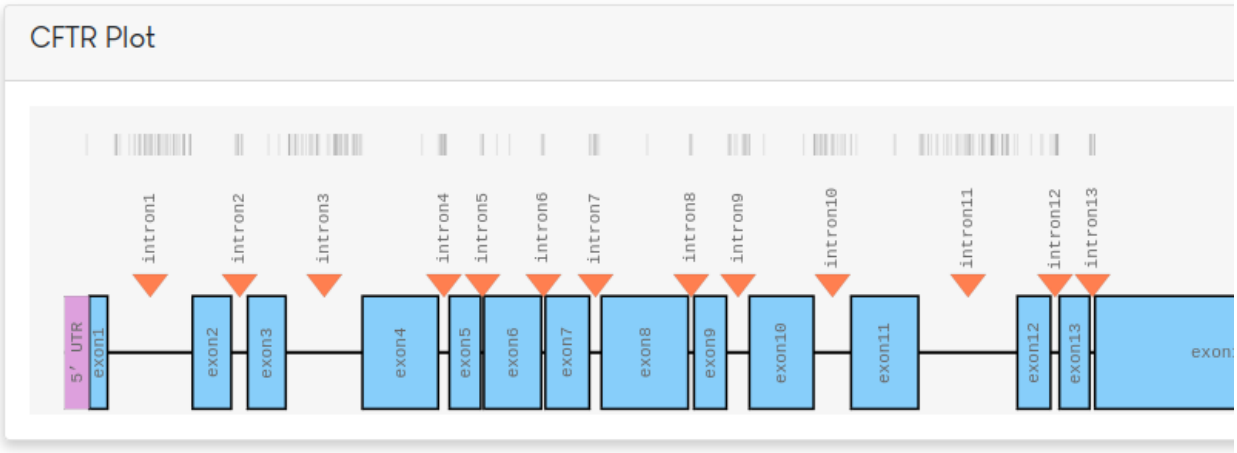
Scott Mastromatteo

Sep 01, 2022

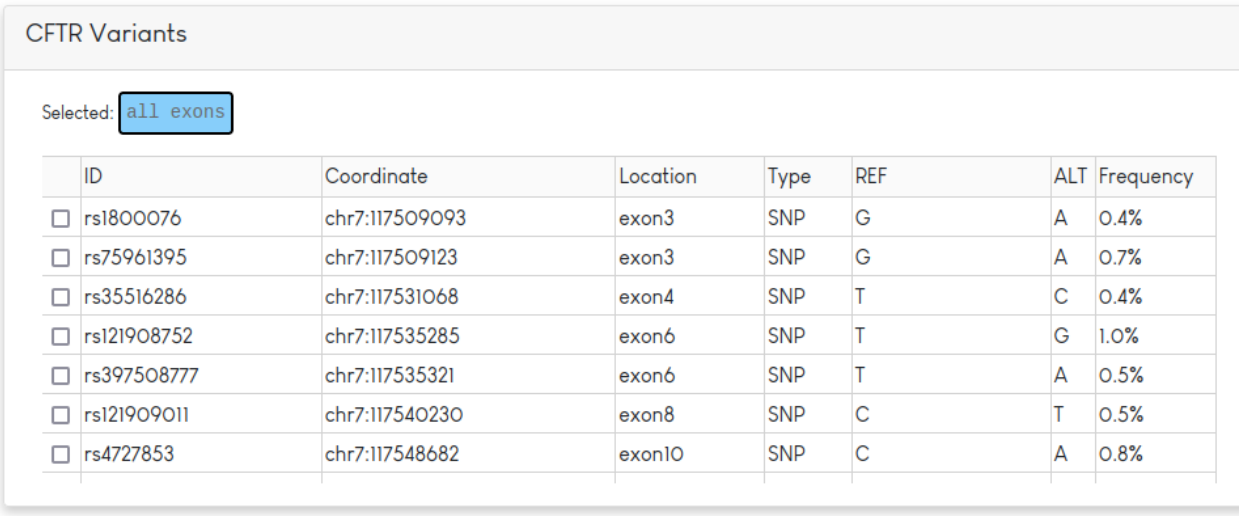
CONTENTS

A web application for understanding CFTR haplotypes. The software is available here

The first panel of the software shows a plot of the *CFTR* gene.



Along the top of the figure, known variants are plotted as small vertical black lines. Different parts of the gene can be selected for further exploration. By default, all exons are selected. Based on the selection, a table of variants will be shown in a panel below.

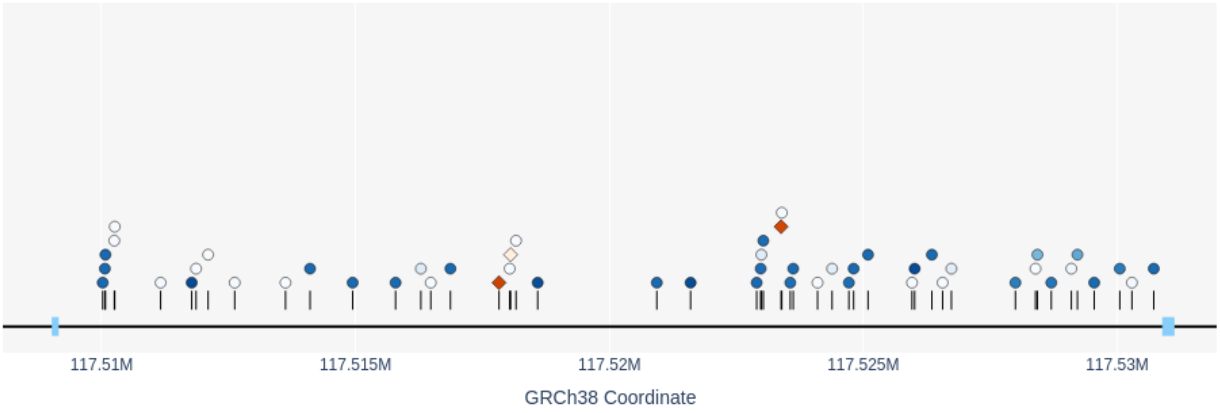


	ID	Coordinate	Location	Type	REF	ALT	Frequency
☐	rs1800076	chr7:117509093	exon3	SNP	G	A	0.4%
☐	rs75961395	chr7:117509123	exon3	SNP	G	A	0.7%
☐	rs35516286	chr7:117531068	exon4	SNP	T	C	0.4%
☐	rs121908752	chr7:117535285	exon6	SNP	T	G	1.0%
☐	rs397508777	chr7:117535321	exon6	SNP	T	A	0.5%
☐	rs121909011	chr7:117540230	exon8	SNP	C	T	0.5%
☐	rs4727853	chr7:117548682	exon10	SNP	C	A	0.8%

This table lists information regarding the rsID, position, REF/ALT and frequency of the variant in the dataset. In addition to the table, the variants from the table plotted by coordinate. SNPs are drawn as blue circles and indels appear as orange diamonds. The allele frequency of each variant is indicated by the darkness of the colour (darker = more common).

Variants can be selected by either the checkbox on the table view or by clicking the shape in the plotted view. Selected variants appear in the below panel.

Variant information is displayed in the table. Variants can be tagged with one of a set of predefined symbols (shown in last column). After clicking on the “Redraw Plot” button, variants tagged with a symbol will be shown in a haplotype table:



Selected Variants

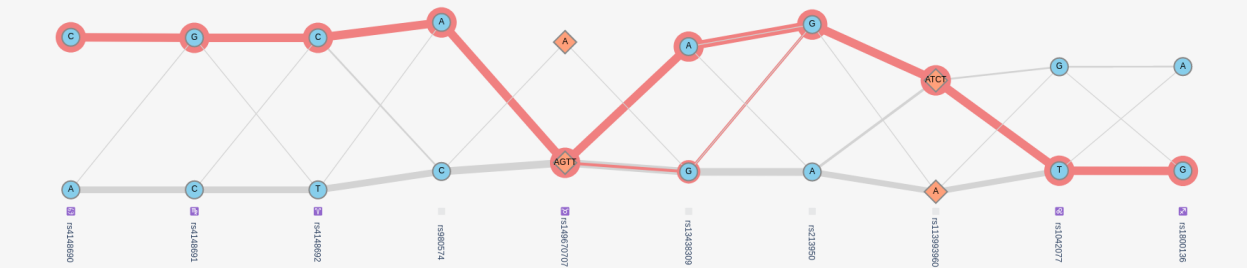
	ID	Coordinate	Location	Type	REF	ALT	Frequency	Symbol
×	rs213950	chr7:117559479	exon11	SNP	G	A	89.5%	
×	rs113993960	chr7:117559590	exon11	DEL	ATCT	A	68.4%	
×	rs1896887	chr7:117560427	intron11	SNP	G	C	78.1%	
×	rs1896886	chr7:117560456	intron11	SNP	G	T	78.0%	
×	rs34410639	chr7:117561309	intron11	DEL	CT	C	11.5%	
×	rs6957018	chr7:117570314	intron11	SNP	C	T	11.7%	

Haplotype Graph Redraw plot

	# Haps	Frequency	♋	♉	♍	♌
○	759	67.3%	1	1	1	1
○	124	11.0%	1	0	0	0
○	118	10.5%	0	0	0	0
○	116	10.3%	1	0	1	1
○	8	0.7%	1	1	0	0

2 rare haplotypes hidden

The haplotype table shows a count and frequency for each unique haplotype observed in the dataset. Rare haplotypes are hidden to protect privacy. The "Redraw Plot" button also generates a textile plot:



In this plot, variants are shown in columns, lines connect alleles that are found on the same haplotype. The thicker the line, the more frequent a pair of alleles is found on the same haplotype. The plot optimized the vertical position of each allele such that alleles more often found on the same haplotype are aligned horizontally. More details can be found in the textile plot paper.

Selecting a haplotype in the haplotype table highlights the corresponding path on the plot in red. Hovering over each variant provides frequency information.

The code is freely available on GitHub