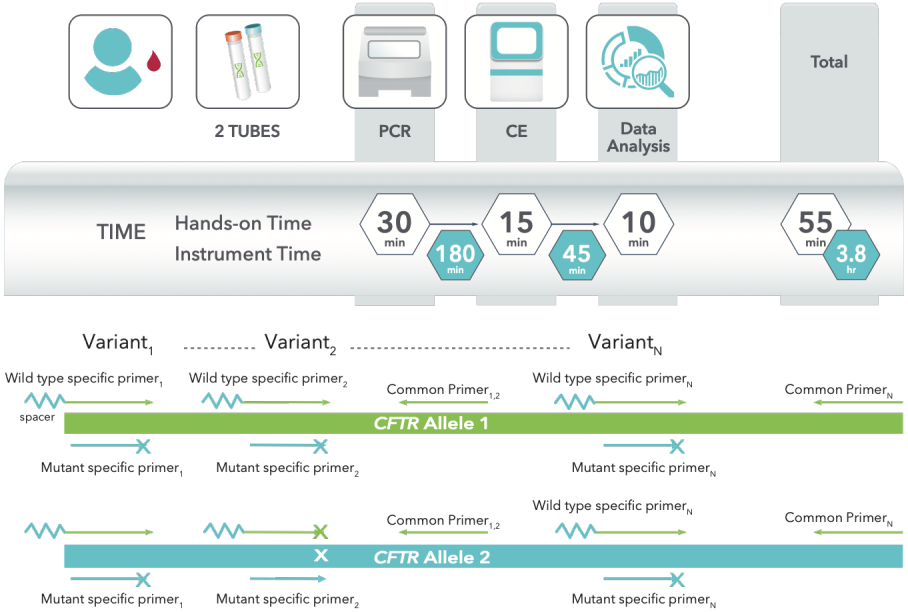
Automated Deep Learning Software for PCR/Capillary Electrophoresis Fragment Analysis Enables Efficient Pan-Ethnic *CFTR* Testing at Scale

Elliot Hallmark, Jacob Ashton, Ryan Routsong, Brian Haynes, Gary J Latham, Bradley Hall and John N Milligan
Asuragen, Inc., Austin, TX

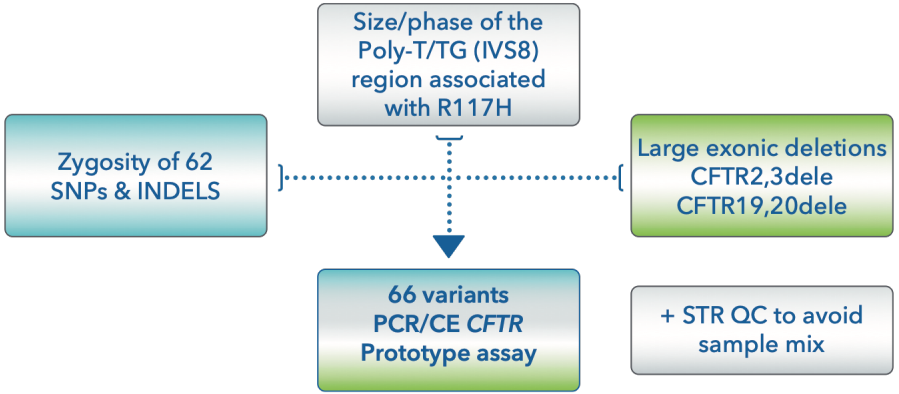
Poster # P17.017.D



The *CFTR* gene is known to harbor >350 CF-causing variants although just 66 cover >93% of diverse US and global populations[1]. The mutations covered by the AmplideX® PCR/CE *CFTR* Prototype is responsive to both European (ECFS guidelines) & US recommendations (ACMG/ACOG guidelines) for *CFTR* mutation detection.

The AmplideX® *CFTR* PCR/CE assay* requires two allele-specific PCR (AS-PCR) reactions per sample. This generates two CE output files representing separate wildtype and mutant peaks for each of the 62 SNP and INDEL variants covered by the assay, as well as peaks for the Poly T/TG region (IVS8) and fingerprinting STRs.

The analysis software automatically pairs the CE output files for analysis based on output file name, and fingerprint STRs are checked by the software to flag mislabeled files.



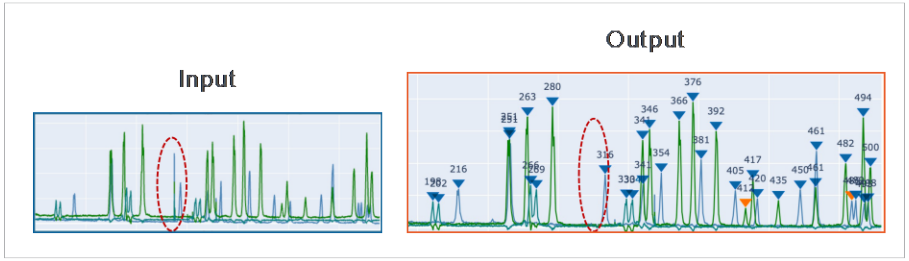
[1]Beauchamp K. et al. (2019) Genet in Med 21:1948-1957
Karczewski et al. (2020) Nature 581: 434-443

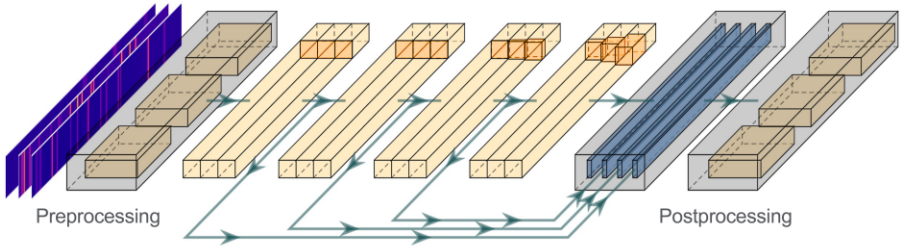
*Product in development. Specifications not finalized.
Presented at ESHG 2021

Asuragen
a bio-techne brand

A Sequence-to-Sequence, Multiscale Convolutional Neural Network Translates Raw Multi-Channel Traces into Genotype Results

Elliot Hallmark, Jacob Ashton, Ryan Routsong, Brian Haynes, Gary J Latham, Bradley Hall and John N Milligan
Asuragen, Inc., Austin, TX





Over 3600 electropherograms (traces) were generated by Asuragen's prototype AmplideX® PCR/CE *CFTR* assay using cell lines, blood samples, and synthetic constructs across six CE instrument configurations and different extraction and assay conditions.

1083 multichannel traces were selected for manual peak annotation, resulting in 46120 true peak annotations. 543 traces from the PCR mix B primer pool were used to train the neural network and 543 from mix A were used for evaluation.

Genotype accuracy was 0.9991 for mix A variants and 0.9997 for mix B, demonstrating that model over fitting was not an issue.

Analysis Pipeline performs:

- Signal Preprocessing
  - » Normalization
  - » Particle removal (shown in red circle)
  - » TAMRA dye pull-up correction
- Peak calling
  - » Sequence to sequence convolutional neural network
  - » Multi-scale due to telescoping receptive field (shown in orange)
- Post-processing logic
  - » Base pair sizing with ROX ladder
  - » Cross talk correction
  - » Genotyping logic
- QC checks happen throughout process

*Product in development. Specifications not finalized.
Presented at ESHG 2021

Asuragen
a bio-techne brand

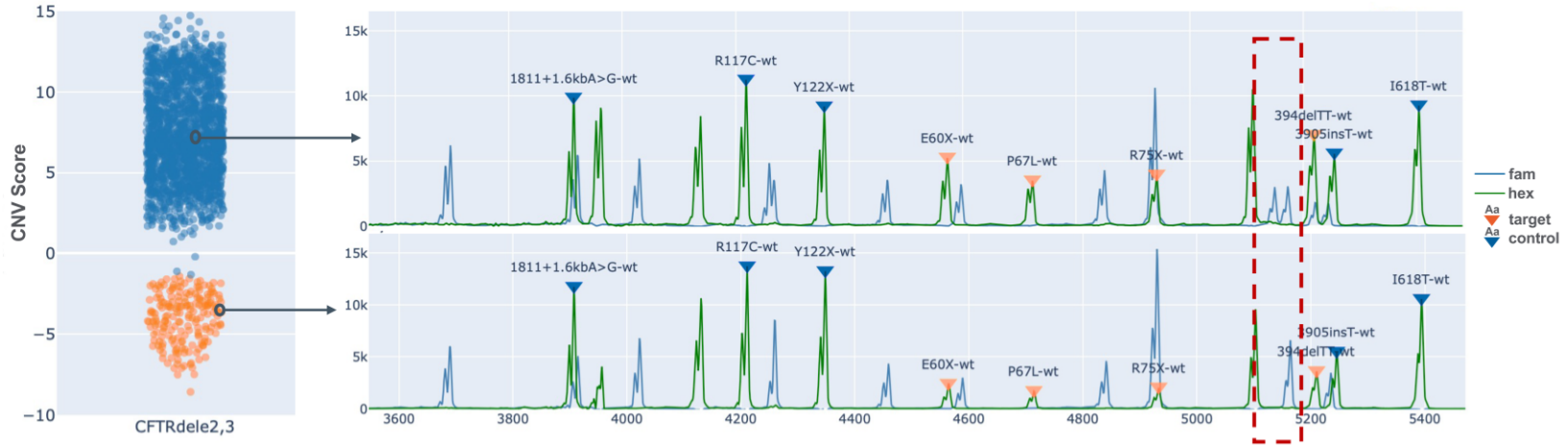
Analysis Utilizes Genotypic Interdependence of Peaks for Exon Deletion Detection and Dynamic Peak Interpretation

Elliot Hallmark, Jacob Ashton, Ryan Routsong, Brian Haynes, Gary J Latham, Bradley Hall and John N Milligan
Asuragen, Inc., Austin, TX

Because the assay contains variants in many exons, relative peak heights can be used to compute exon copy number. The two most common exon deletions, exons 3 and 20, are implemented in the software analysis. For threshold setting, relative peak heights for deletions and non-deletion samples were collected and used to inform a distribution. Training data was drawn from this distribution. Shown below are examples of a CFTRdele2,3 WT (above) and HET (below) samples.

The peak heights of the exon 3 alleles (target peaks) are noticeably lower than control peaks in more stable regions.

The red box shows the interdependence of 1677delTA on deletion F508del (not pictured). The software must take into account F508del's zygosity in order to interpret the peaks associated with 1677delTA.



*Product in development. Specifications not finalized.
Presented at ESHG 2021

Asuragen
a bio-techne brand

Performance Breakdown Over 1846 Samples Demonstrates 98-100% Accuracy Across Variants

Elliot Hallmark, Jacob Ashton, Ryan Routsong, Brian Haynes, Gary J Latham, Bradley Hall and John N Milligan
Asuragen, Inc., Austin, TX

3692 electropherograms

- The "External" instrument consists of PCR plates run at Asuragen and CE injected at various other sites on 3500, 3730 and SeqStudio™ instruments
- Truth for SNP and PolyT/TG determined by Sanger sequencing at Asuragen
- CFTRdele2,3 Exon deletion truth was reported by vendor and confirmed by Sanger
- CFTRdele20 Exon deletion truth was determined by partner using MLPA
- QC truth for QC accuracy was determined by trained operators or experimental design

Instrument	# Samples	# Variants	SNP PPA	SNP PPV	SNP Accuracy	Poly T/TG Accuracy	Exon Deletion Accuracy	QC Accuracy	Sample QC Pass Rate
3500	368	550	1	0.993	1	0.997	0.996	0.986	96.2%
3730	**628**	**997**	**1**	**0.979**	**0.999**	**0.997**	**0.996**	**0.964**	**92.8%**
3130	450	621	0.982	0.973	0.999	0.979	0.995	0.960	89.1%
SeqStudio™	**238**	**312**	**1**	**0.990**	**1**	**1**	**0.998**	**0.978**	**95.0%**
External	162	332	1	0.976	0.999	0.977	0.980	0.916	89.5%
Total	**1846**	**2812**	**0.996**	**0.981**	**0.999**	**0.992**	**0.995**	**0.965**	**92.6%**

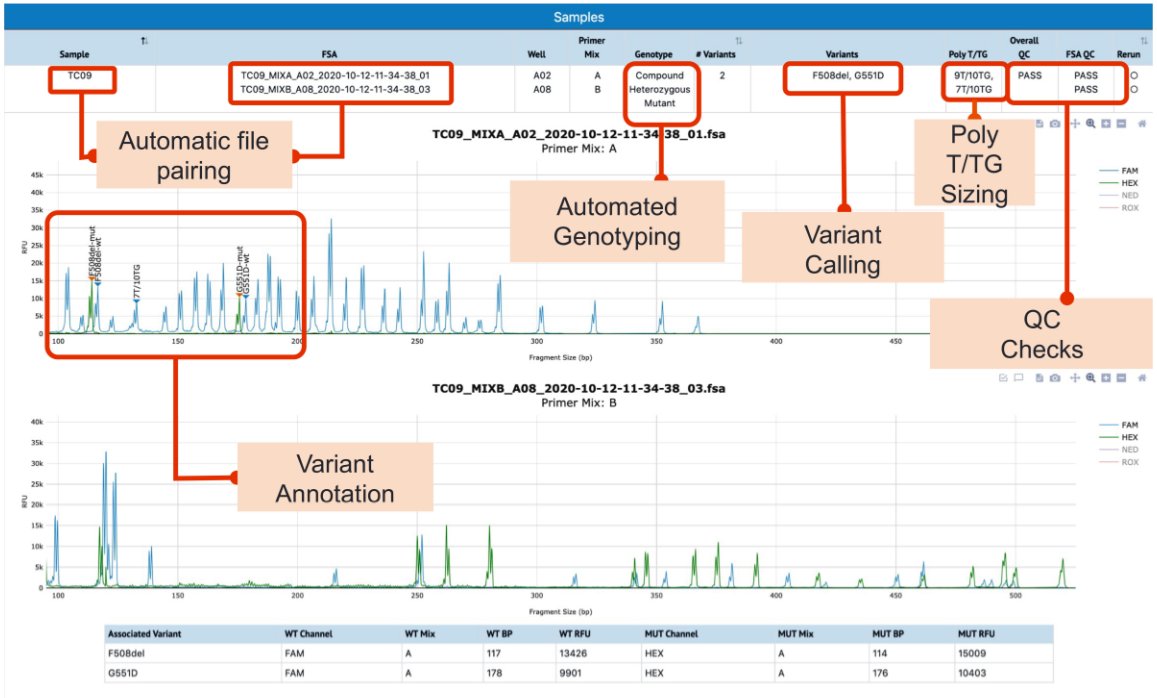
*Product in development. Specifications not finalized.
Presented at ESHG 2021

Asuragen
a bio-techne brand

AmplideX® Software Provides Genotyping and QC with Reduced Analysis Time

Elliot Hallmark, Jacob Ashton, Ryan Routsong, Brian Haynes, Gary J Latham, Bradley Hall and John N Milligan
Asuragen, Inc., Austin, TX

Input Files to Answer in Under 5 Minutes For 48 Sample (96 well) Plate



*Product in development. Specifications not finalized.
Presented at ESHG 2021

Asuragen
a bio-techne brand