PB3189

Detection of *CFTR* polyT and TG variations by an NGS-based method

Qian Zeng, Yuanyu Cao, Neil Russell, Binbin Huang, Jean Smith, Patty Okamoto, Stan Letovsky, Hui Zhu, Natalia Leach and Angela Kenyon
Center of Excellence for Data Sciences, AI and Bioinformatics, Labcorp, Westborough, MA

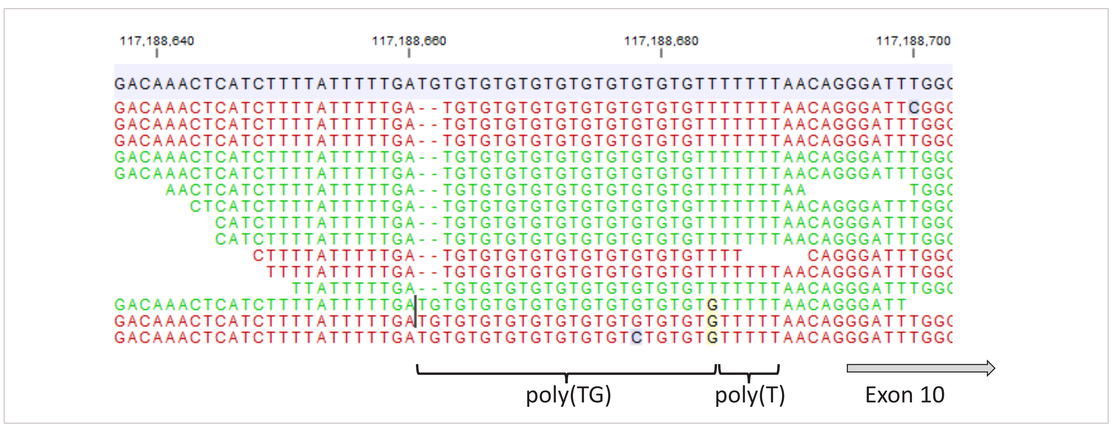
Introduction

Pathogenic variants in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene[1] have significant impact on the structure and functions of *CFTR*, resulting in cystic fibrosis (CF) and a range of other conditions.

The *CFTR* polyT and TG tract (**Figure 1**):

- Located in intron 9 just six bases upstream of *CFTR* exon 10
- The polyT and TG length can vary between the two alleles
- c.1210-34TG[11]T[7] is the most common haplotype and is benign
- Shorter polyT tracts (<=5T) have been linked to *CFTR*-related disorders
- Number of TG repeats can modulate effect of 5T on *CFTR* expression
- Sanger sequencing is the gold standard method for polyT and TG analysis
- Determination of polyT and TG lengths is challenging for next generation sequencing (NGS) based methods[2]
- Need to develop a more robust computational strategy for detecting *CFTR* polyT and TG lengths by NGS

Figure 1. *CFTR* polyT and TG region on Chr 7



Methods (Figure 2)

Generatation of simulated NGS samples[3]

- For the *CFTR* polyT and TG region plus 500 basepairs (bp) on each side
- Illumina paired reads: 150bp read length, 300bp fragment size, 100X coverage
- PolyT length range: T[1-13]
- PolyTG length range: TG[7-15]
- Simulated allele count: 2 (in each simulated sample)
- Total simulated samples: 13,689

Creation of a lookup table

- Run simulated samples through NGS analysis pipeline
- Link input polyT and TG truths with NGS variant fingerprint profiles

Analysis of clinical specimens

- Run clinical samples through NGS analysis pipeline
- Report variant fingerprint profile
- Use the lookup table to find the corresponding polyT and TG lengths
- If needed, update lookup table to include new variant fingerprint

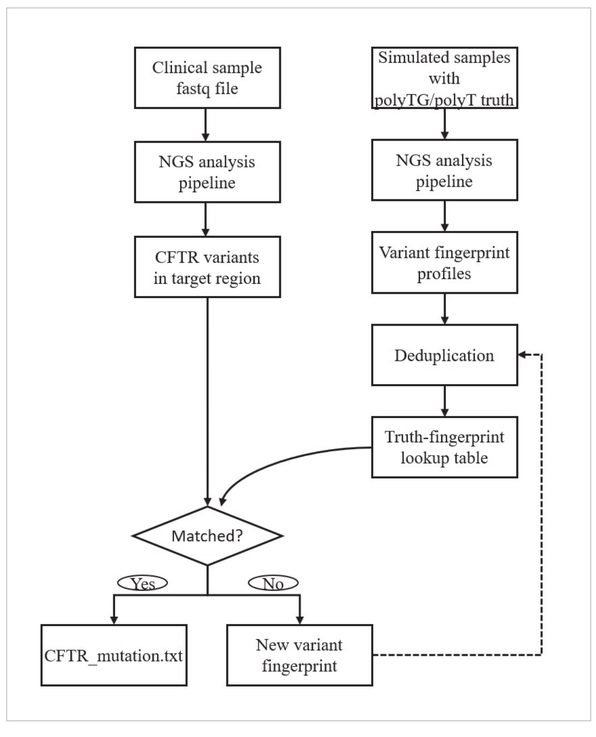


Figure 2. *CFTR* polyT and TG analysis for clinical specimens

Results

Table 1. NGS and Sanger are 100% concordant for *CFTR* polyT and TG length analysis

Sample ID	Haplotype 1	Haplotype 2	Agree with SANGER
Sample001	TG[9]T[9]	TG[11]T[7]	Y
Sample002	TG[10]T[7]	TG[11]T[9]	Y
Sample003	TG[10]T[7]	TG[11]T[5]	Y
Sample004	TG[10]T[7]	TG[12]T[7]	Y
Sample005	TG[10]T[7]	TG[11]T[5]	Y
Sample006	TG[10]T[7]	TG[10]T[9]	Y
Sample007	TG[10]T[7]	TG[12]T[5]	Y
Sample008	TG[10]T[7]	TG[11]T[5]	Y
Sample009	TG[10]T[7]	TG[12]T[7]	Y
Sample010	TG[10]T[7]	TG[12]T[5]	Y
Sample011	TG[10]T[7]	TG[11]T[9]	Y
Sample012	TG[10]T[7]	TG[11]T[5]	Y
Sample013	TG[10]T[7]	TG[11]T[7]	Y
Sample014	TG[10]T[7]	TG[12]T[7]	Y
Sample015	TG[10]T[9]	TG[11]T[5]	Y
Sample016	TG[10]T[9]	TG[11]T[9]	Y
Sample017	TG[10]T[9]	TG[12]T[7]	Y
Sample018	TG[10]T[9]	TG[11]T[5]	Y
Sample019	TG[10]T[9]	TG[11]T[9]	Y
Sample020	TG[11]T[5]	TG[11]T[7]	Y
Sample021	TG[11]T[5]	TG[11]T[7]	Y
Sample022	TG[11]T[7]	TG[11]T[9]	Y
Sample023	TG[11]T[7]	TG[12]T[7]	Y
Sample024	TG[11]T[7]	TG[11]T[5]	Y
Sample025	TG[11]T[9]	TG[11]T[9]	Y
Sample026	TG[12]T[7]	TG[13]T[5]	Y
Sample027	TG[12]T[7]	TG[12]T[7]	Y

Example: "Sample010"

Sanger analysis

- Result: "TG[10]T[7]; TG[12]T[5]"
- Analog signal (see Sanger trace in **Figure 3**)
- No phasing information among variants on the two alleles

NGS analysis

- Results: "TG[10]T[7]; TG[12]T[5]"
- Digital signal with high resolution
- Variants on the two alleles are automatically phased
- PolyT and TG read count (**Figure 4**) is consistent with the NGS and Sanger results

Figure 3. Sanger sequencing trace

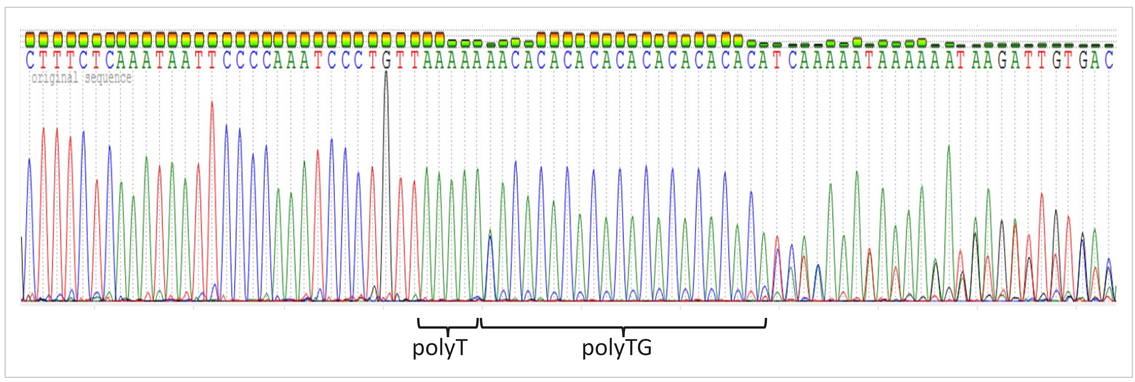
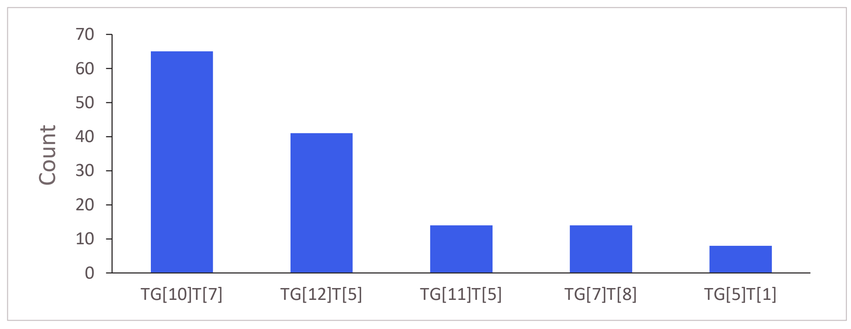


Figure 4. NGS polyT and TG read count profiles



Conclusions

- Sanger is the current gold standard for *CFTR* polyT and TG length determination, but its analog read out presents a challenge in complex cases, especially when indels are present in either of the two alleles.
- NGS performance agrees with the gold standard method, and has better resolution in complex cases.

References

1. Riordan J.R. et al., *Science*, 1989, 245, 1066-1073.
2. Lincoln S.E. et al., *Genet. Med*., 2021, 23, 1673-1680.
3. Huang W. et al., *Bioinformatics*, 2012, 28, 593-594.

ASHG Annual Meeting, November 1-5, 2023, Washington, DC