A NOVEL DDPCR METHOD FOR COPY NUMBER VARIATION ANALYSIS IN THE SEGMENTAL DUPLICATION REGIONS OF NEBULIN AND TITIN

Lydia Sagath[1], Vilma-Lotta Lehtokari[1], Carina Wallgren-Pettersson[1], Katarina Pelin[2], Kirsi Kiiski[1]

[1] The Folkhälsan Research Center, Finland
[2] Department of Biosciences, University of Helsinki, Finland

1 INTRO

Nebulin and titin (encoded by the genes *NEB* and *TTN*) are sarcomeric giant proteins, and both harbour large segmental duplication (SD) regions known to be expressed in muscle. In *NEB*, this region contains eight exons repeated three times. In *TTN*, nine exons are repeated threefold, after which the two first exons of the block reappear alone a fourth time. Mutations in *NEB* are known to cause nemaline myopathy and copy number variations (CNV) of the *NEB* repetitive region are known to be pathogenic (Kiiski *et al.* 2016).

We are able to detect CNVs and determine copy number (CN) in the *NEB* SD using our custom Comparative Genomic Hybridisation arrays (aCGH; Sagath *et al.* 2018, Kiiski *et al.* 2013). The SD region of *TTN*, however, is shorter, making unique probe design more difficult.

Droplet digital PCR (ddPCR) allows us to quantify the CN of the SD regions of *NEB* and *TTN* by using sample partitioning, PCR and hydrolysis probes. By targeting unique sequences within the SDs we are able to determine the CN with greater precision than using other methods.

2 OBJECTIVES

The objectives of the project are to:

- Develop a ddPCR assay for the CNV analysis of *NEB* and *TTN*
- Determine normal copy number for the*TTN* SD region
- Create estimates for the level of variation in *TTN* SD
- Find the putative pathogenic threshold for the *TTN* SD region

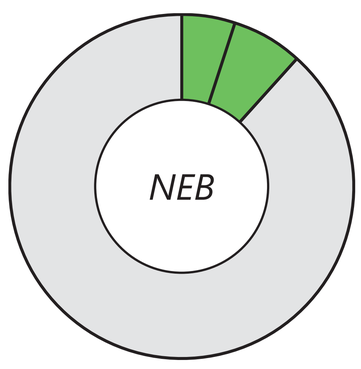

Figure 1. Copy number variation of the SD regions of *NEB* and *TTN*. In *NEB*, ~12% of samples run harbor variation in the SD region (in green). Of these, 5% are deletions and 7% are duplications. In *TTN*, variation in the SD region in seen in ~27% of the samples, and there is an equal distribution between duplication and deletions (in green).

3 METHODS

We have designed custom assays to target different parts of the segmental duplication regions of *NEB* and *TTN* using Primer3Plus. The assays target the exons 4 and 8 of the repeated *NEB* region and exon-intron regions 1-2 and 7-8 of the *TTN* SD. The ddPCR was run with Bio-Rads QX200 ddPCR system and analysed with `twoddpcr`. Subsequent statistical analyses were performed in R.

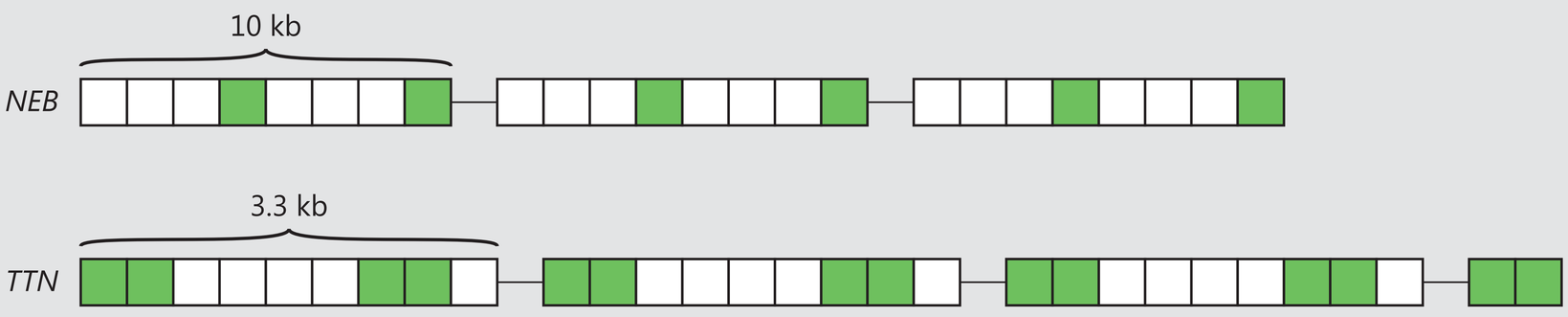

Figure 1. The segmental duplication regions of nebulin and titin. In nebulin, 8 exons are repeated three times over a sequence of around 30 kb. In titin, 9 exons are repeated three times over a sequence of around 10 kb, after which the two first exons of the block are repeated a fourth time. Our assays are designed for the nebulin exons 4 and 8, and the titin exon-intron regions 1-2 and 7-8. Each square represents an exon, and the targeted exons are presented in green.

Figure 2. A ddPCR schematic. Samples are partitioned into 20 000 individual droplets, and the PCR reaction happens individually in each of them. In duplex reactions droplets can be categorised into four categories: target-positive/reference-positive, target-positive/reference-negative, target-negative/reference-positive and target-negative/reference/negative.

4 RESULTS

We have optimised the ddPCR protocol for the custom assays, and have compared the *NEB* SD assays against data acquired with our custom CGH-arrays. Simultaneously, we are gathering *TTN* SD data. As to date, we have run around 80 samples. Of these over half are healthy controls, while the rest are patient and family samples with or without pathological variation in the SD regions.

We are expecting to be able to extrapolate the information gathered from the *NEB* SD ddPCR assays to the *TTN* SD ddPCR assays, as to be able to determine the copy number of the *TTN* SD with more precision than the NMD-CGH-array.

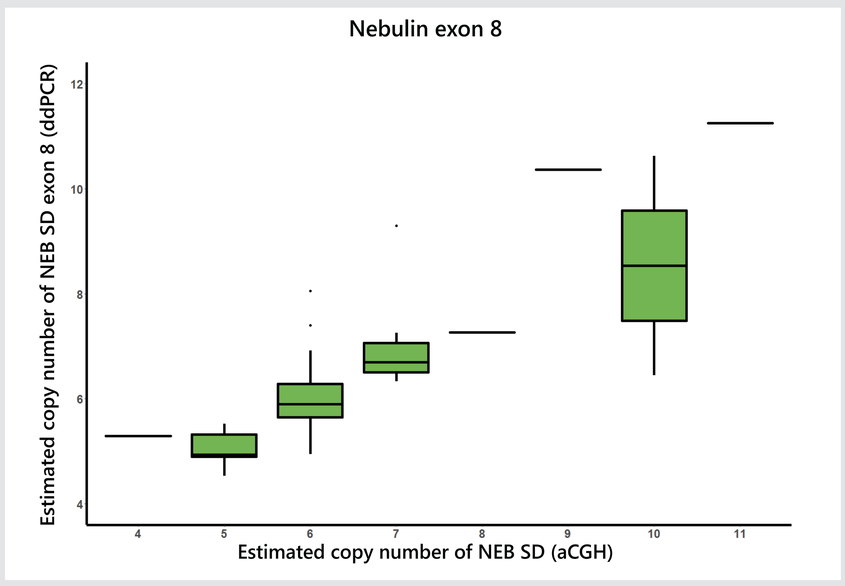

Figure 3. Validation data of the assay for nebulin SD exon 8. The y-axis shows ratio of target sequence to diploid reference. The "normal" samples (n=39) showed 6 SD copies on the CGH-arrays. The statistical significance between the 5, 6 and 7 copy clusters is $p<0.0005$. No statistical significance could be calculated for the other groups.

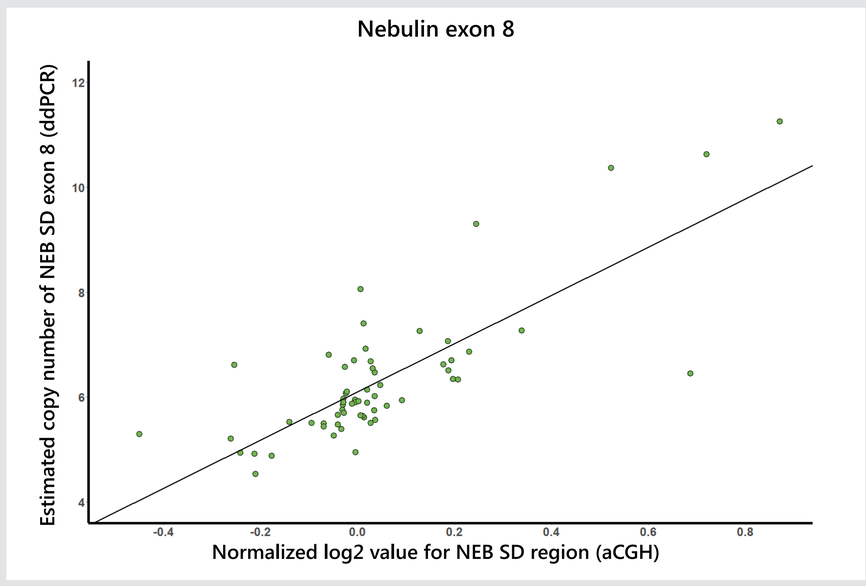

Figure 4. Validation data of the assay for nebulin SD exon 8. The y axis shows the ratio of target sequence to diploid refernce doubled. The x-axis shows the normalized $\log_2$ value for the complete nebulin SD region as recorded from the NM- or NMD-CGH-array data. The R squared value of the linear regression model is 0.6094 and the p-value 4.441e-14.

5 DISCUSSION

The proposed methods brings a novel segmental duplication region analysis method to the table with a short turnaround time, low costs and high preciseness.

The method is however still under development. We are rigorously aquiring data and developing the analysis pipeline to achieve a diagnostically robust status.

We are currently also evaluating possible error sources. Sample quality seems to systematically affect the custom assays. We are investigating the possible ways of circumventing different sample quality issues.

The ddPCR approach could be adapted to suit other similar SD regions in other genes, and thus paves the way for diagnostic uses of ddPCR in CNV analysis.

CONTACT

Lydia Sagath, MSc
PhD Student

lydia.sagath@helsinki.fi
+358-294125078

lydiasagath.com
nemalinehelsinki.fi

FUNDING

This work is funded by Muscular Dystrophy UK, the Finska Läkaresällskapet, the Medicinska understödsföreningen Liv och Hälsa, l'Association Française contre les Myopathies, and the Folkhälsan Research Center.

REFERENCES

Chiu A, Ayub M, Dive C, Brady G, Miller C (2017). "twoddpcr: An R/Bioconductor package and Shiny app for Droplet Digital PCR analysis." Bioinformatics.

Kiiski K, Laari L, Lehtokari VL, Lunkka-Hytönen M, Angelini C, Petty R, Hackman P, Wallgren-Pettersson C, Pelin K (2013). "Targeted array comparative genomic hybridization--a new diagnostic tool for the detection of large copy number variations in nemaline myopathy-causing genes". Neuromuscul Disord.

Kiiski K, Lehtokari VL, Löytynoja A, Ahlstén L, Laitila J, Wallgren-Petterson C, Pelin K (2016). "A recurrent copy number variation of the NEB triplicate region: only revealed by the targeted nemaline myopathy CGH array". Eur J Hum Genet.

Sagath L, Lehtokari VL, Välipakka S, Udd B, Wallgren-Pettersson C, Pelin K, Kiiski K (2018). "An extended targeted copy number variation detection array including 187 genes for the diagnostics of neuromuscular disorders". J Neuromuscul Dis.