NGS goes diagnostics: Multiplex analysis of genes for limb-girdle muscular dystrophies

J. E. Bach, S. Rost, A. Gehrig, W. Kress, C. R. Müller

Department of Human Genetics, University of Wuerzburg, Germany

INTRODUCTION

A main focus of our diagnostic laboratory lies in the identification of mutations causing muscular diseases. Besides classical Sanger sequencing we are now establishing next generation sequencing (NGS) applications for single large genes and groups of genes e.g. those causing limb-girdle muscular dystrophies (LGMDs).

Clinically, LGMDs can be classified into juvenile and adult onset and association with high and low creatine kinase (CK) levels, respectively. Inheritance is autosomal recessive or dominant and more than 25 genes are known to date as causative for the various subtypes of LGMDs.

The autosomal recessive LGMD types 2A (OMIM #253600), 2B (OMIM #253601) and 2L (OMIM #611307) have a similar clinical presentation and are associated with very high CK levels. Since they are at the same time the most prevalent subtypes in Central Europe, we have established a NGS "mini-panel" for the three genes *CAPN3*, *DYSF* and *ANO5*.

Causative mutations in these genes are mostly point mutations and therefore can easily be detected by sequencing analysis. For the detection of the much rarer bigger rearrangements like deletions of several exons other methods such as MLPA can be used.

METHODS AND RESULTS

Target enrichment of the LGMD panel has been performed with the Access Array™ System of Fluidigm which allows parallel amplification of 48 target regions for up to 48 samples in one single PCR setup (Figure 1). The assay comprises 98 primer pairs covering the coding regions of the three genes *CAPN3*, *DYSF* and *ANO5*, i.e. 24, 55 and 22 exons, respectively. The resulting libraries were sequenced using the GS Junior System of Roche.

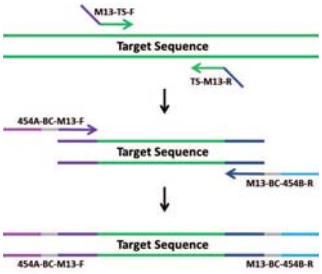

Figure 1:
Overview of the PCR: Region-specific primers tagged with M13 sequence can be combined with second primers with M13, barcode and sequencer-specific tag.

TS = target-specific sequence
BC = barcode sequence
454A = 454 sequencer-specific forward sequence
454B = 454 sequencer-specific reverse sequence

In a first step, *CAPN3*, *DYSF* and *ANO5* have been resequenced in a total of 18 patients with known mutations and SNPs in at least one of the three genes. Table 1 gives an overview of the type of the variants. 58 of the 59 known variants could be retrieved, one was in an exon not covered in the NGS data. Apart from that this provides a sensitivity estimate of more than 98 % and a specificity of more than 97 % of this NGS panel.

type of variant		No.
substitutions		34
deletions (1-16 bp)		11
duplications (1-7 bp)		12
indels	del4ins3	1
	del5ins4	1

Table 1:
Overview of the type of the 59 known variants of 18 patients in the genes *CAPN3*, *DYSF* and *ANO5*, respectively, 58 of the 59 variants could be retrieved. One variant (a duplication of 7 nucleotides) could not be seen as the affected exon was not covered in NGS data.

However, as expected, short indels or duplications of single nucleotides in homopolymeric regions posed problems and cannot reliably be identified in the NGS data. Therefore, they have to be resequenced in parallel with Sanger sequencing.

See Figure 2 and 3 for a comparison of a badly detectable single nucleotide duplication and a well visible single nucleotide deletion in a homopolymer of 7 and 2 A, respectively.

Figure 2:
Visualization of c.191dup in a homopolymer of 7 "A" in the *ANO5* gene leading to LGMD 2L. Insertions are shown by white lines between two nucleotides. This mutation can hardly be detected by standard analysis of the NGS data as the duplication is shown on either sides of the homopolymer with frequencies of less than 20% each.

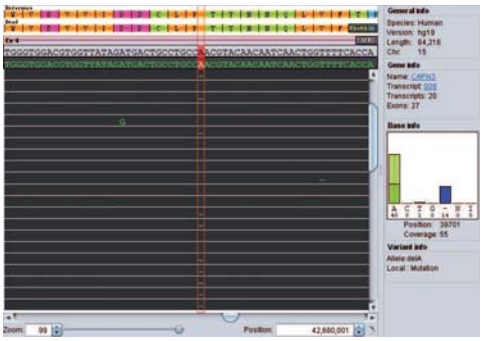

Figure 3:
Visualization of c.550del in the *CAPN3* gene leading to LGMD 2A. This mutation is easily detectable as it is in a short homopolymer of only 2 "A".

Since then, more than 57 LGMD patients have been analysed by this panel. The number of patients analysed per run varies between 8 and 12. With a minimum average of 70,000 reads per run on a GS Junior, this leads to a calculated coverage of 60x - 90x per amplicon. In fact, coverage was not homogeneous across all amplicons, but in most cases at least 15-fold, i.e. suitable for reliable analyses. Each variant suspected to be causal was confirmed by Sanger sequencing. Up to now, we identified causative mutations in about 47 % of LGMD patients using this NGS panel with no false-positive results (upon verification by Sanger sequencing).

CONCLUSION

The combination of Access Array System and GS Junior sequencing has proven to be a reliable and practical NGS method for diagnostic analyses of the three LGMD genes *CAPN3*, *DYSF* and *ANO5*. Compared to classical PCR and Sanger sequencing, the NGS application has proven to be a time and cost efficient diagnostic approach. However, mutations in homopolymers are not reliably detectable and when indicated need to be resequenced by Sanger sequencing in parallel. Further NGS panels for other groups of muscle disease genes like those associated with myofibrillar myopathies are under development.