Acta Myologica • 2012; XXXI: p. 196-200

Next generation sequencing (NGS) strategies for the genetic testing of myopathies

Vincenzo Nigro and Giulio Piluso

Telethon Institute of Genetics and Medicine, Naples, Italy; Dipartimento di Biochimica, Biofisica e Patologia Generale, Seconda Università di Napoli, Italy

Next generation sequencing (NGS) technologies offer the possibility to map entire genomes at affordable costs. This brings the genetic testing procedure to a higher level of complexity. The positive aspect is the ease to cope with the complex diagnosis of genetically heterogeneous disorders and to identify novel disease genes. Worries arise from the management of too many DNA variations with unpredictable meaning and incidental findings that can cause ethical and clinical dilemmas. The technology of enrichment makes possible to focus the sequencing to the exome or to a more specific DNA target. This is being used to provide insights into the genetics underlying Mendelian traits involved in myopathies and to set up cost-effective diagnostic tests. This huge potential of the NGS applications makes likely that these will soon become the first approach in genetic diagnostic laboratories.

Key words: Next generation sequencing, NGS, neuromuscular disorders

The power of NGS

The field of DNA sequencing is quickly moving from uses within genetic labs towards a huge number of other applications for human health. Until now, sequencing based on the Sanger method (1) showed a limited throughput, due to the necessity to obtain a single homogeneous DNA fragment per each sequencing reaction. With the next generation sequencing (NGS) (2) procedures, the ability to confine each individual clone of identical molecules on a physical support can provide the opportunity for parallel reactions. Using the ultimate platforms such systems are able to perform billions of sequencing reactions with a read length of 150-250 nt. This corresponds to the ability to sequence many times individual human genomes to have a complete picture of all the clinically relevant variations. However, although these possibilities are fascinating and receive the attention of politics and investments, they are well beyond the scope of a diagnosis and on the other side contain some risks and difficulties. In particular, there is the dilemma of the incidental findings of mutations with high clinical importance, i.e. those in cancer-susceptibility genes that are detected at a considerable frequency (3).

Enrichment procedures

The "enrichment" techniques provide the ability to focus this huge NGS power towards a more specific DNA target (4). The ability to produce custom libraries composed of hundred thousands different oligonucleotides allows a very large number of alternative options. In addition to different systems for high-throughput multiplex PCR (Fluidigm (5), Raindance (6), AmpliSeq), there are two basic approaches for selecting out regions of the genome for sample preparations: the first is by hybridization only (Agilent SureSelect and NimbleGen/Roche SeqCap EZ) and the second is by hybridization followed by an extension step (Halogenomics by Agilent and TruSeq by Illumina). Different targeting options have been used (Table 1).

In the first case, the enrichment procedure selects as sequencing target a chromosome-specific region corresponding to a previously restricted linkage interval, with the aim to discover a new causative gene (7, 8).

Second, the sequencing target is a collection DNA fragments of many genes, located on different chromosomes, that are all possibly involved in genetically heterogeneous disorders, such as ataxias (9), Usher syndrome (10), or inherited retinal disorders (11, 12). In this second option, the mutations usually occur in known causative genes and genetic testing can be also provided in isolated cases.

Address for correspondence: Vincenzo Nigro - Telethon Institute of Genetics and Medicine, Napoli, Italy; Dipartimento di Biochimica, Biofisica e Patologia Generale, Seconda Università di Napoli, via Luigi De Crecchio 7, 80138 Napoli, Italy. E-mail: vincenzo.nigro@unina2.it

Table 1. The term NGS defines very different sizes of analysis, depending on the target size.

NGS analysis	Target size	Coverage required	Targetting method	Sequenced bases /sample
Whole genome	3.2 Gb	40x	No enrichment	128 Gb
Whole exome	50 Mb	80x	Affinity capture	4 Gb
A linkage interval of ~10 cM	8.5 Mb	100x	Affinity capture	0.85 Gb
1200 exons (~100 genes)	0.4 Mb	200x	Affinity capture or HT-PCR	0.08 Gb

Third, the target is the entire set of the human exons (exome) that may also include untranslated sequences. In this case, the computational analysis of exome is made by comparing the shared variations in a significant number of affected individuals (13, 14). An alternative use of exome sequencing is to search for *de novo* DNA mutations (15) in case-parent trios, composed by a single affected child and both unaffected parents. An important study on the Proteus syndrome sequenced the exomes in biopsy samples obtained from affected tissues compared with those of unaffected tissues (16).

Muscle disease gene discovery by NGS

A number of genes were recently associated to myopathies by the use of targeted NGS or exome sequencing (Table 2). The following papers report the identification of novel genes using a combination of linkage analysis, next-generation sequencing, and validation by Sanger sequencing. In addition, when no other mutation outside the original family was found, functional studies as well as modelling in the animals were performed.

The first example is the targeted NGS of 640 exons from a chromosomal region located on chromosome 5q23, identified by phased haplotype analysis that was used to discover the cause of EMARDD, a disease characterized by early onset myopathy, areflexia, respiratory distress and dysphagia (17). These infantile myopathies with diaphragmatic paralysis are genetically heterogeneous and clinical symptoms do not assist in differentiating between them. EMARDD is inherited as an autosomal recessive disorder. Affected member of a consanguineous family from Pakistan showed a homozygous 10-bp duplication (c.2288_2297dup) in the coding sequence of exon 19 of MEGF10 (multiple epidermal growth factor-like domains-10 protein). Other independent families were homozygous or compound heterozygous for other loss-of-function mutations in MEGF10, thus proving proof of the causative role for this gene. MEGF10 is a regulator of satellite cell myogenesis, highly expressed in activated satellite cells, that regulates their proliferation, differentiation, and fusion into multinucleated myofibers, which are greatly reduced in muscle.

A second example is the identification of the cause of a form of congenital myopathy with prominent internal nuclei and atypical cores (18). Congenital myopathies are well suited for whole exome NGS since they are clinically and genetically heterogeneous diseases. In this case the Authors performed a SNP linkage analysis on ten individuals (including five affected members) of a family with autosomal dominant inheritance characterized by distal weakness and core-like areas and increased internalized nuclei at biopsy. The top LOD score was only 1.87 on chromosome 16. The DNA from the index case alone was analyzed by whole-exome sequencing using the NimbleGen exome capture and NGS. Among many unique variants, the disease was linked to a heterozygous C>T change at c.68-1 of CCDC78, an uncharacterized coiled-coiled domain-containing gene located on 16p13 and expressed in skeletal muscle. This change alters the splicing of exon 2. The mutation was confirmed in the original family and tested in the zebrafish using a morpholino- mediated splice-site alteration. The CCDC78 alteration in zebrafish resulted in altered motor function and abnormal muscle ultrastructure.

A third example is the use of whole-exome NGS or traditional positional cloning by two different groups to reveal the causative gene in an autosomal dominant limb-girdle muscular dystrophy (LGMD1D). LGMD1D is characterized by skeletal muscle vacuoles, previously mapped to chromosome 7q36. Sarparanta et al. performed the characterization of LGMD1D in Finnish families and refined the locus to a 3.4-Mb region containing 12 genes. Sanger Sequencing of the positional candidates RNF32, UBE3C, DNAJB6 and PTPRN2 identified a c.279C>G (p.Phe93Leu) change in exon 5 of DNAJB6 in all affected individuals in the Finnish families. Another group used whole exome analysis in 3 affected individuals from another LGMD1D family and identified novel candidate mutations in 22 genes, but further linkage analysis excluded all variants except the Phe93Leu mutation of the DNAJB6 gene. Sequencing data from other independent pedigrees with dominant myopathy identified a second G/F domain mutation (Pro96Arg) in DNAJB6 (19). DNAJB6 is a member of the HSP40/DNAJ family of molecular co-chaperones tasked with protect-

Table 2. Muscle disease gene discovery by NGS.

Disease	OMIM	Inheritance	Strategy	Gene	Locus	Reference
Early-onset myopathy, areflexia, respiratory distress, and dysphagia (EMARDD)	614399	AR	Haplotype analysis + Targeted NGS	MEGF10	5q23.2	Logan et al. (17)
Centronuclear myopathy-4 (CNM4)	614807	AD	Linkage + Exome	CCDC78	16p13.3	Majczenko et al. (18)
Limb-girdle muscular dystrophy type 1E (LGMD1E)	603511	AD	Exome + Functional analysis in Zebrafish	DNAJB6	7q36.3	Harms et al. (19); Sarparanta et al. (20)
Hereditary myopathy with early respiratory failure (HMERF)	603689	AD	Linkage + Exome	TTN	2q31.2	Pfeffer et al. (21)
Facioscapulohumeral muscular dystrophy type 2 (FSHD2)	158901	Digenic	Exome	SMCHD1	18p11.32	Lemmers et al. (24)
Congenital muscular dystrophy-dystroglycanopathy with brain and eye anomalies, type A7 (MDDGA7)	614643	AR	On-Cell complementation assay + Homozygosity mapping + Targeted NGS	ISPD	7p21.2	Cirak et al. (25); Willer et al. (26)

ing client proteins from irreversible aggregation during protein synthesis or during times of cellular stress (20). LGMD1D muscle showed early disruption of Z-disks and autophagic pathology.

A fourth example of a possible use of the NGS is that related to detect mutations in apparently unrelated families that share clinical finding. This is the case of autosomal dominant hereditary myopathy with early respiratory failure. All patients shared adult onset muscle weakness in the pelvic girdle, neck flexors, respiratory and trunk muscles, with the majority showing calf hypertrophy. They also shared myofibrillar lesions with marked Z-disc alterations. Single nucleotide polymorphism arrays mapped a shared 6.99 Mb-haplotype to chromosome 2q31, suggesting a common ancestry. Whole exome sequencing in four individuals from the same family revealed a heterozygous missense mutation, g.274375T>C; p.Cys30071Arg, in the titin gene (TTN). The mutation segregated with the disease in all three families (21). A parallel study was carried out on 31 mutation carriers followed for 31 years. Muscle weakness was earlier onset and more severe in the lower extremities in nearly all patients, while other patients also had axial muscle weakness. A combination of genome-wide linkage and whole exome sequencing also revealed the variant in the titin (TTN) gene (g.274375T>C; p.Cys30071Arg) (22).

A completely new approach was used to identify a role for digenic inheritance and an epigenetic modifier in facioscapulohumeral muscular dystrophy type 2 (FSHD2). Facioscapulohumeral dystrophy type 1 (FSHD1) and FSHD2 are phenotypically indistinguishable, with the difference that in type 2 a normal-sized D4Z4 array on a chromosome 4 may be found. In FSHD2 there is a focal region of extreme demethylation within a 5' domain, which was named DR1 (23). Whole exome sequencing was performed in 14 individuals from 7 unrelated families with FSHD2. In 79% of families out-of-frame deletions, heterozygous splice-site mutations or heterozygous missense mutations were identified in the SMCHD1 gene that encodes a Structural Maintenance of Chromosomes flexible Hinge Domain containing 1 (24). SMCHD1 mutant alleles may modify the epigenetic repression mutations and could also modify the penetrance of FSHD1.

Very recently, a large collaborative study, headed by Francesco Muntoni (25), mutations in the gene isoprenoid synthase domain containing (ISPD) have been also associated to dystroglycanopathy (26) phenotypes ranging from congenital muscular dystrophy to limb-girdle muscular dystrophy. The Authors identified by Illumina-based exome sequencing allelic ISPD variants in nine cases belonging to seven families. The same ISPD gene and the TMEM5 gene have been previously identified as the genetic causes of the Cobblestone lissencephaly (27).

Diagnosis of genetic muscle disorders by NGS

The use of targeted NGS for clinical diagnostics should be considered as a cost-effective alternative, when the total number of PCR fragments of the candidate genes exceeds the 96 wells of a PCR plate used to perform Sanger sequencing reactions. All the different disorders of muscle present diagnostic challenges due to phenotypic variability, and difficulties with muscle immunohistochemical studies (28-31).

The dystrophin gene has been the first challenge for targeted NGS by our group (32) and recently, by others (33). This is because the Dystrophin (DMD) gene is large and the spectrum of point mutations is unpredictable. However, the muscle cDNA Sanger sequencing in

DMD cases remains the more convenient option, because any targeting method can miss some sequences.

Valencia et al. (34) used NGS to identify mutations in 321 exons representing 12 different genes involved with congenital muscular dystrophies. Two different enrichment technologies were used, solution-based hybridization and microdroplet-based PCR target enrichment. NGS results were analyzed and compared with Sanger sequencing. Both enrichment technologies produced suitable data for clinical laboratories. In a recent study, 267 neuromuscular disease genes were targeted by affinity capture for enrichment and eight patients were studied (35). With this protocol more than 97% of the targeted exons were fully covered.

However, clinical labs are fast moving towards the routine use of whole exome sequencing. The exome sequencing has been also applied to the diagnosis of known conditions, such as a case of autosomal recessive Emery-Dreifuss muscular dystrophy caused by a novel homozygous mutation (R225Q) in the lamin A/C gene (36). Another successful use of whole exome sequencing was the identification of the cause of autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy (ARVC) in a Swedish family. A heterozygous mutation was identified that replaces p.Pro419Ser in the desmin gene on chromosome 2q35 (37). Whole exome sequencing was also used to diagnose a LGMD2A, erroneously assigned as non-4q FSHD (FSHD2). This also shows how wrong diagnoses can be corrected by NGS (38).

Perspectives

This huge potential of the next generation sequencing applications makes likely that these will soon become the first approach in clinical laboratories. There is a growing number of patients studied and of novel muscle disease genes so far identified. This may improve genetic counselling in myopathic patients and will favour inclusion into novel therapeutic trials that require a prior knowledge of the mutation type.

References

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977;74(12):5463-7.
2. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2009;11:31-46.
3. Johnston JJ, Rubinstein WS, Facio FM, et al. Secondary variants in individuals undergoing exome sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. Am J Hum Genet 2012;91:97-108.
4. Mamanova L, Coffey AJ, Scott CE, et al.. Target-enrichment strategies for next-generation sequencing. Nat Methods 2010;7:111-8.
5. Schlipf NA, Schule R, Klimpe S, et al. Amplicon-based high-throughput pooled sequencing identifies mutations in CYP7B1 and SPG7 in sporadic spastic paraplegia patients. Clin Genet 2011;80:148-60.
6. Brouzes E, Medkova M, Savenelli N, et al. Droplet microfluidic technology for single-cell high-throughput screening. Proc Natl Acad Sci USA 2009;106:14195-200.
7. Nikopoulos K, Gilissen C, Hoischen A, et al. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. Am J Hum Genet 2010;86:240-7.
8. Rehman AU, Morell RJ, Belyantseva IA, et al. Targeted capture and next-generation sequencing identifies C9orf75, encoding taperin, as the mutated gene in nonsyndromic deafness DFNB79. Am J Hum Genet 2010;86:378-88.
9. Hoischen A, Gilissen C, Arts P, et al. Massively parallel sequencing of ataxia genes after array-based enrichment. Hum Mutat 2010;31:494-9.
10. Licastro D, Mutarelli M, Peluso I, et al. Molecular diagnosis of Usher syndrome: application of two different next generation sequencing-based procedures. PLoS One 2012;7:e43799.
11. Audo I, Bujakowska KM, Leveillard T, et al. Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases. Orphanet J Rare Dis 2012;7:8.
12. Neveling K, Collin RW, Gilissen C, et al.: Next-generation genetic testing for retinitis pigmentosa. Hum Mutat 2012;33:963-72.
13. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2009;42:30-5.
14. Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010;42:790-3.
15. Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nat Rev Genet 2012;13:565-75.
16. Lindhurst MJ, Sapp JC, Teer JK, et al. A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N Engl J Med 2011;365:611-9.
17. Logan CV, Lucke B, Pottinger C, et al. Mutations in MEGF10, a regulator of satellite cell myogenesis, cause early onset myopathy, areflexia, respiratory distress and dysphagia (EMARDD). Nat Genet 2011;43:1189-92.
18. Majczenko K, Davidson AE, Camelo-Piragua S, et al. Dominant mutation of CCDC78 in a unique congenital myopathy with prominent internal nuclei and atypical cores. Am J Hum Genet 2012;91:365-71.
19. Harms MB, Sommerville RB, Allred P, et al. Exome sequencing reveals DNAJB6 mutations in dominantly-inherited myopathy. Ann Neurol 2012;71:407-16.
20. Sarparanta J, Jonson PH, Golzio C, et al. Mutations affecting the cytoplasmic functions of the co-chaperone DNAJB6 cause limb-girdle muscular dystrophy. Nat Genet 2012;44:450-5, S1-2.
21. Pfeffer G, Elliott HR, Griffin H, et al. Titin mutation segregates with hereditary myopathy with early respiratory failure. Brain 2012;135(Pt 6):1695-713.
22. Ohlsson M, Hedberg C, Bradvik B, et al. Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin. Brain 2012;135(Pt 6):1682-94.
23. Hartweck LM, Anderson LJ, Lemmers RJ, et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. Neurology 2013 Jan 2. [Epub ahead of print].
24. Lemmers RJ, Tawil R, Petek LM, et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet 2012;44:1370-4.

25. Cirak S, Foley AR, Herrmann R, et al. ISPD gene mutations are a common cause of congenital and limb-girdle muscular dystrophies. Brain 2013 Jan 3. [Epub ahead of print].
26. Willer T, Lee H, Lommel M, et al. ISPD loss-of-function mutations disrupt dystroglycan O-mannosylation and cause Walker-Warburg syndrome. Nat Genet 2012;44:575-80.
27. Vuillaumier-Barrot S, Bouchet-Seraphin C, Chelbi M, et al. Identification of mutations in TMEM5 and ISPD as a cause of severe cobblestone lissencephaly. Am J Hum Genet 2012;91:1135-43.
28. Mercuri E, Muntoni F. The ever-expanding spectrum of congenital muscular dystrophies. Ann Neurol 2012;72:9-17.
29. Nigro V, Aurino S, Piluso G. Limb girdle muscular dystrophies: update on genetic diagnosis and therapeutic approaches. Curr Opin Neurol 2011;24:429-36.
30. Laing NG. Genetics of neuromuscular disorders. Crit Rev Clin Lab Sci 2012;49:33-48.
31. Nance JR, Dowling JJ, Gibbs EM, et al. Congenital myopathies: an update. Curr Neurol Neurosci Rep 2012;12:165-74.
32. Bonnal RJ, Severgnini M, Castaldi A, et al. Reliable resequencing of the human dystrophin locus by universal long polymerase chain reaction and massive pyrosequencing. Anal Biochem 2010;406:176-84.
33. Xie S, Lan Z, Qu N, et al. Detection of truncated dystrophin lacking the C-terminal domain in a Chinese pedigree by next-generation sequencing. Gene 2012;499:139-42.
34. Valencia CA, Rhodenizer D, Bhide S, et al. Assessment of target enrichment platforms using massively parallel sequencing for the mutation detection for congenital muscular dystrophy. J Mol Diagn 2012;14:233-46.
35. Vasli N, Bohm J, Le Gras S, et al. Next generation sequencing for molecular diagnosis of neuromuscular diseases. Acta Neuropathol 2012;124:273-83.
36. Jimenez-Escrig A, Gobernado I, Garcia-Villanueva M, et al. Autosomal recessive Emery-Dreifuss muscular dystrophy caused by a novel mutation (R225Q) in the lamin A/C gene identified by exome sequencing. Muscle Nerve 2012;45:605-10.
37. Hedberg C, Melberg A, Kuhl A, et al. Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy 7 is caused by a DES mutation. Eur J Hum Genet 2012;20:984-5.
38. Leidenroth A, Sorte HS, Gilfillan G, et al. Diagnosis by sequencing: correction of misdiagnosis from FSHD2 to LGMD2A by whole-exome analysis. Eur J Hum Genet 2012;20:999-1003.