

Clinical application of massively parallel sequencing in the molecular diagnosis of glycogen storage diseases of genetically heterogeneous origin

Jing Wang, MD¹, Hong Cui, PhD¹, Ni-Chung Lee, MD², Wuh-Liang Hwu, MD², Yin-Hsiu Chien, MD², William J. Craigen, MD PhD¹, Lee-Jun Wong, PhD¹, Victor Wei Zhang, MD, PhD¹

Purpose: Glycogen storage diseases are a group of inborn errors of glycogen synthesis or catabolism. The outcome for untreated patients can be devastating. Given the genetic heterogeneity and the limited availability of enzyme study data, the definitive diagnosis of glycogen storage diseases is made on the basis of sequence analysis of selected potentially causative genes.

Methods: A massively parallel sequencing test was developed for simultaneous sequencing of 16 genes known to cause muscle and liver forms of glycogen storage diseases: *GYS2*, *GYS1*, *G6PC*, *SLC37A4*, *GAA*, *AGL*, *GBE1*, *PYGM*, *PYGL*, *PFKM*, *PHKA2*, *PHKB*, *PHKG2*, *PHKA1*, *PGAM2*, and *PGM1*. All the nucleotides in the coding regions of these 16 genes have been enriched with sufficient coverage in an unbiased manner.

Results: Massively parallel sequencing demonstrated 100% sensitivity and specificity as compared with Sanger sequencing.

Massively parallel sequencing correctly identified all types of mutations, including single-nucleotide substitutions, small deletions and duplications, and large deletions involving one or more exons. In addition, we have confirmed the molecular diagnosis in 11 of 17 patients in whom glycogen storage diseases were suspected.

Conclusion: This report demonstrates the clinical utility of massively parallel sequencing technology in the diagnostic testing of a group of clinically and genetically heterogeneous disorders such as glycogen storage diseases, in a cost- and time-efficient manner.

Genet Med 2013;15(2):106–114

Key Words: glycogen storage disease; massively parallel sequencing; next-generation sequencing; target gene enrichment

INTRODUCTION

Glycogen storage diseases (GSDs) are a group of inborn errors of glycogen metabolism that primarily affect the liver and/or muscle, the main sites of glycogen storage.^{1,2} The overall incidence of all forms of glycogen storage disease in the population is ~1 in 10,000.³ The prevalence rates of some forms of GSD have been underestimated because of their mild clinical presentation and rare occurrence, and the lack of efficient molecular diagnostic methods.

GSDs are classified on the basis of specific enzyme defects in glycogen metabolic pathways (Figure 1a,b). Typical presentations of the liver forms of GSDs are hepatomegaly and hypoglycemia due to the accumulation of glycogen and impaired mobilization of glucose for release into the blood during fasting.⁴ Muscle forms of GSDs result from a similar inability to degrade glycogen (e.g., GSD type V (GSD V)) or a block in glycolysis (e.g., GSD VII) during exercise, leading to exercise intolerance, muscle weakness, and muscle cramps.⁵ The clinical phenotypes depend on the specific defective enzyme and the type of mutation. Some types of GSD, such as GSD II (Pompe disease), due to deficiency in acid maltase (α -glucosidase), can affect multiple organ systems, including the heart, kidneys, and central nervous system.⁶ Some

GSDs may have prenatal or neonatal onset, leading to fetal demise or death within the first year of life, whereas other GSDs may exhibit only mild exercise intolerance or no clinical symptoms until adulthood (e.g., adult polyglucosan body disease).⁷ Early diagnosis is critical for prompt and proper patient management to minimize organ damage and maximize the life span of the patient.

The diagnosis of GSD used to depend largely on invasive liver or muscle biopsies and biochemical assays.² Enzyme study data are limited, and the more broadly available DNA-based testing allows accurate diagnosis when enzymological results are ambiguous or unavailable. Nevertheless, given the genetic heterogeneity in the population, serial testing of each gene is expensive and time-consuming, and often results in a delayed diagnosis that slows the implementation of appropriate care.

Massively parallel sequencing (MPS), also known as next-generation sequencing, has been shown to be an efficient, accurate, and cost-effective method of identifying disease genes.^{8–10} For clinically and genetically heterogeneous diseases caused by a group of genes involving a common metabolic pathway, MPS can also be used for simultaneous sequencing of the group of candidate genes.^{11–13} To facilitate

¹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; ²Department of Pediatrics and Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan. Correspondence: Lee-Jun Wong (ljwong@bcm.edu)

Submitted 4 June 2012; accepted 16 July 2012; advance online publication 16 August 2012. doi:10.1038/gim.2012.104

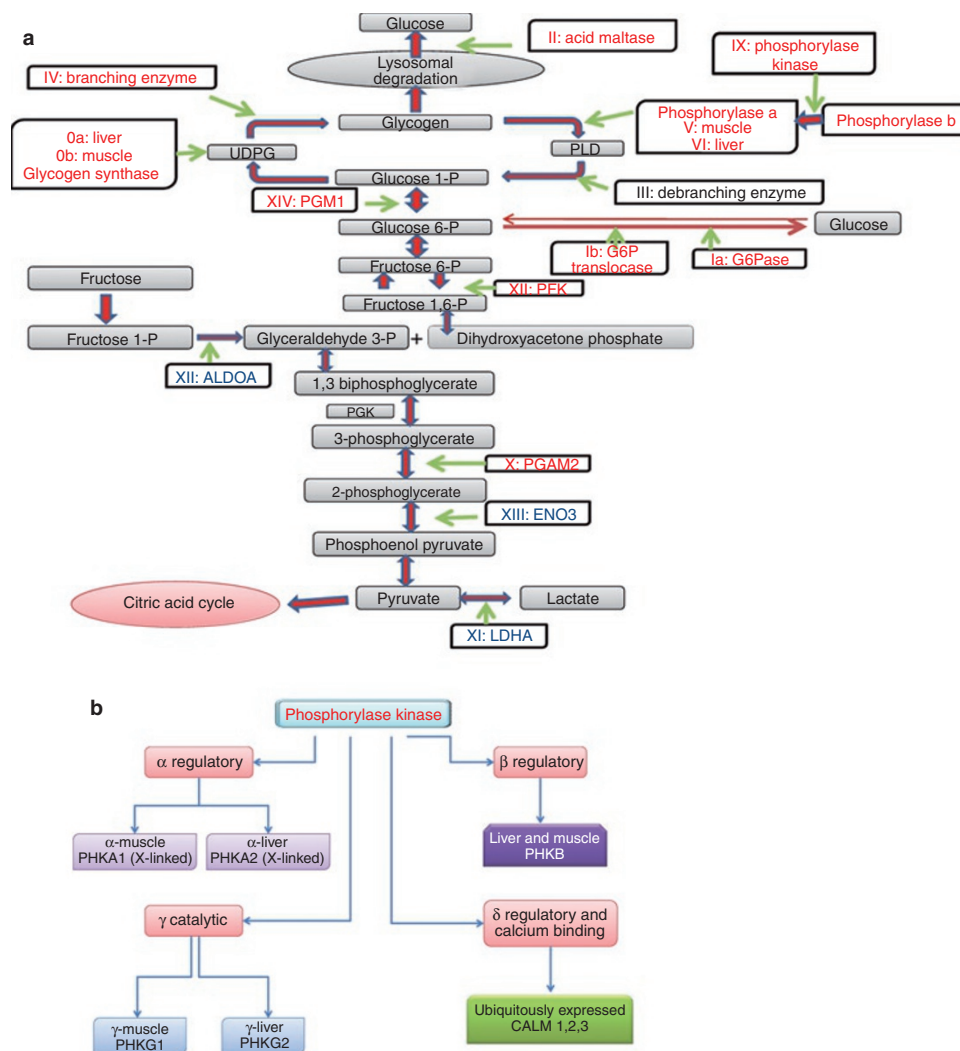


Figure 1 Glycogen metabolism and glycolysis pathway. (a) Enzymes involved in glycogen metabolism and the glycolysis pathway are listed using Roman numerals. They are categorized into glycogen storage diseases (GSD) types 0 and I–XIV, presented as liver form, muscle form, or both. (b) Phosphorylase kinase deficiency causes GSD IX. The enzyme comprises four copies of each of four subunits: α , β , γ , and δ . The schematic figure shows subunit function and the genes that cause liver or muscle forms of GSD IX.

the molecular diagnosis of patients with suspected GSD, we have developed strategies for using MPS to analyze a panel of genes responsible for this group of diseases (**Supplementary Table S1** online). This approach eliminates the hurdle of having to prioritize multiple candidate genes for step-wise sequencing.

MATERIALS AND METHODS

Patients

Patient samples were submitted to the medical genetics laboratories at Baylor College of Medicine for sequence analysis of the gene(s) causing GSDs. Samples that were negative for mutations in the gene(s) requested to be tested were subjected to MPS analysis of a panel of 16 genes potentially responsible for GSDs. The analyses were performed in accordance with the protocols approved for human subjects by the institutional review board at the Baylor College of Medicine.

Validation using control samples containing previously identified mutations

DNA samples from seven patients with known mutations in GSD genes (as identified by Sanger sequencing) were selected as positive controls for MPS validation (patients (Ps)1–7 in **Table 1**). This included two mutations each in *G6PC* and *PHKA2*, and one each in *SLC37A4*, *GBE1*, and *PYGM*. The types of mutation were single-nucleotide substitutions, small deletions/insertions, and large exonic deletions.

Detection of mutations in patients with no previous molecular diagnoses

Seventeen samples (P9–24 in **Table 1**) from patients with clinical, histochemical, and/or enzymatic findings of a GSD, but no molecular diagnoses, were analyzed using MPS in an attempt to identify causative mutations.

Table 1 GSD mutations confirmed or detected by MPS-GSD panel

Patient	Age	Gender	Clinical indication	Previous lab tests	Gene	Allele 1	Allele 2
P1	6 months	M	Hypoglycemia and mild hepatosplenomegaly	<i>G6PC</i> sequencing negative	<i>G6PC</i>	E3–E5 del	E3–E5 del
P2	7 years	M	Hypoglycemia and hepatomegaly	<i>G6PC</i> sequencing negative	<i>G6PC</i>	c.379_380dupTA (p.Tyr128Thrfs*3)	c.533C>T (p.Pro178Leu)
P3	38 years	M	FmHx, high cholesterol, gout, liver adenoma, and low WBC	<i>G6PC</i> sequencing negative	<i>SLC37A4</i>	c.467C>T (p.Ala156Val) c.572C>G (p.Pro191Arg)	c.1024T>C (p.Ser342Pro)
P4	3 weeks	M	Neonatal demise and GBE deficiency	<i>G6PC</i> sequencing negative	<i>GBE1</i>	E16 del	E16 del
P5	14 years	F	Recurrent rhabdomyolysis	<i>G6PC</i> sequencing negative	<i>PYGM</i>	c.21_28dup8 (p.Lys10Thrfs*9)	WT
P6	1 year	M	Hypoglycemia, hepatomegaly, and elevated lactate	<i>G6PC</i> sequencing negative	<i>PHKA2</i>	c.2812delG (p.Glu938Argfs*6)	NA
P7	3 years	F	Hepatomegaly, elevated liver enzymes, and low PhK activity	<i>G6PC</i> sequencing negative	<i>PHKA2</i>	c.3614C>T (p.Pro1205Leu)	WT
P8	10 months	F	Hypoglycemia and hepatomegaly	<i>G6PC</i> sequencing negative	<i>SLC37A4</i>	c.785-3_786del5	c.785-3_786del5
P9	1.5 years	F	Hyperlipidemia, hyperlactatemia, failure to thrive, and hepatomegaly	<i>G6PC</i> sequencing negative	<i>SLC37A4</i>	c.817G>A (p.Gly273Ser)	c.1042_1043delCT (p.Leu348Valfs*53)
P10 ^a	13 years	F	Clinically diagnosed as GSD Ia? Received liver transplant at age of 6 years. Psychomotor retardation and urinary infection	<i>G6PC</i> sequencing negative	<i>SLC37A4</i>	c.595delC (p.Leu199Trpfs*13)	c.1043T>C (p.Leu348Pro)
P11	2 years	M	Hepatomegaly	<i>G6PC</i> sequencing negative	<i>AGL</i>	c.256C>T (p.Gln86*)	c.2723T>G (p.Leu908Arg)
P12	4 months	M	Hypoglycemia and hepatomegaly	<i>G6PC</i> sequencing negative	NA	Negative	WT
P13	13 years	F	Fatigue, encephalopathy, and abnormal liver function	<i>G6PC</i> sequencing negative	NA	Negative	WT
P14	3 years	M	Hypoglycemia, hyperuricemia, recurrent infections, and bone fractures	<i>G6PC</i> sequencing negative	NA	Negative	WT
P15	3 months	F	Pulmonary hypertension, large liver, and elevated lipids/uric acid/lactate	<i>G6PC</i> sequencing negative	NA	Negative	WT
P16 ^a	10 years	M	Hepatomegaly and GSD III?	Debranching enzyme: 0 (nl: 0.31 ± 0.1), but sample was too small to confirm the diagnosis	<i>PYGL</i>	c.698G>A (p.Pro233Asp)	c.2467C>T (p.Gln823*)
P17 ^a	10 years	M	GSD III	Debranching enzyme: 0 (nl: 0.31 ± 0.1)	<i>AGL</i>	c.658C>T (p.His220Tyr)	c.1735+1G>T
P18 ^a	9 years	F	GSD III	Debranching enzyme: 0 (nl: 0.31 ± 0.1)	<i>AGL</i>	c.1735+1G>T	c.1735+1G>T
P19 ^a	12 years	M	GSD IX	GSD IX enzyme assay: highly suggestive of GSD IX	<i>PHKA2</i>	c.2609delC (p.Pro870Glnfs*44)	NA
P20 ^b	4 years	M	Hepatomegaly, abnormal liver function, and GSD VI or IX?	Phosphorylase enzyme: 0.003 (10.3 ± 1.7%). GSDVI sequencing negative	<i>PHKA2</i>	Exons 27–30 del	NA
P21 ^b	3 years	M	Hepatomegaly, abnormal liver function, and EM and LM suggesting GSD IX	NA	<i>PHKA2</i>	c.133C>T (p.Arg45Trp)	NA
P22	1 year	M	GSD Ia	Hypoglycemia, lactic acidosis, and hepatomegaly	<i>G6PC</i>	c.379_380dupTA (p.Tyr128Thrfs*3)	c.379_380dupTA (p.Tyr128Thrfs*3)
P23	9 years	F	Hypoglycemia/GSD 0? No hepatosplenomegaly, and no lactic acidemia	NA	<i>PHKG2</i>	c.986G>A (p.Arg329Gln)	WT
P24	19 years	M	Lipid myopathy	NA	<i>PGAM2</i>	c.244G>A (p.Met82Val)	WT

Bold type represents novel unclassified variants. EM, electron microscope; GSD, glycogen storage disease; LM, light microscope; MPS, massively parallel sequencing; NA, *PHKA2* is X-linked gene, P6, P19, and P20 are hemizygous for the detected mutations; P, patient; PhK, phosphorylase kinase; WBC, white blood cells; WT, wild type.

^aParental testing confirmed that the proband is compound heterozygous or homozygous for the mutation. ^bMother is heterozygous for the mutation.

Quality control

In addition to probes directed at GSD genes, the custom library contains probes to capture 14 unique single-nucleotide polymorphism regions located in various chromosomes. These 14 single-nucleotide polymorphisms of each sample are genotyped either by Sanger sequencing or by TaqMan assay. At the conclusion of MPS and sequencing analyses, these 14 single-nucleotide polymorphisms are verified against the original genotyping results for assurance of sample identity. Detailed procedures for the inclusion of quality control single-nucleotide polymorphisms have been described elsewhere.¹⁴

Sample preparation for MPS

An in-solution-based capture library was custom-designed to enrich the coding regions of target GSD genes in accordance with the SeqCap EZ Choice Library User's Guide (Roche NimbleGen, Madison, WI). The sample preparation was carried out in accordance with the manufacturer's recommendation. Equal molar ratios of 10 indexed samples were pooled to be loaded to each lane of the flow cell for sequencing on a HiSeq2000 (Illumina, San Diego, CA) with 75 cycle single-end reads.

MPS data analysis

Raw data in base call files (.bcl format) were converted to qseq files before demultiplexing using CASAVA v1.7 (Illumina). Demultiplexed data were further processed by NextGENE software for alignment (SoftGenetics, State College, PA). All mutation calls were further reviewed and confirmed by Sanger sequencing. An in-house bioinformatics pipeline was used for the variant annotation.

Confirmation of Sanger sequencing data

All mutations and novel variants detected using MPS were confirmed by Sanger sequencing. Sequence-specific oligonucleotide primers linked to M13 universal primers were designed to amplify all coding exons and the flanking intronic regions. Sequencing reactions were performed using the BigDye Terminator cycle sequencing kit (version 3.1) (Life Technologies, Grand Island, NY), analyzed on an ABI3730XL automated DNA sequencer with Sequencing Analysis Software version 5.1 (Applied Biosystems Carlsbad, CA) and Mutation Surveyor version 3.97 (SoftGenetics, State College, PA). The GenBank accession numbers for each gene are listed in **Supplementary Table S1** online.

RESULTS

Characteristics of target gene capture and depth of coverage

A total of 16 genes (*GYS2*, *GYS1*, *G6PC*, *SLC37A4*, *GAA*, *AGL*, *GBE1*, *PYGM*, *PYGL*, *PFKM*, *PHKA2*, *PHKB*, *PHKG2*, *PHKA1*, *PGAM2*, and *PGM1*) were included in this panel. All 294 coding regions, with a total 50062 bases for the 16 genes were enriched in an unbiased fashion, with sufficient coverage. Every base of each of the coding regions and at least 20 bp of flanking intronic regions were covered, the mean coverage being 758X. The average total reads per 100bp was 1,025. The minimum coverage per base was 300X per coding exon (**Supplementary Figure S1 and Table S2** online). Both the sensitivity and the specificity for 50,062 base calls per sample were 100% for the comparison with the results of Sanger sequencing of the same set of samples (**Table 2**). Besides the highly reproducible nature of the MPS, one particularly beneficial feature of this capture/sequencing is the deep coverage of every coding exon of every gene. As previously reported by others, the depth of coverage is such that complementary Sanger sequencing is not required for coding regions of low or no coverage.¹¹

Verification of known mutations by MPS

Mutations previously identified in seven patients with GSDs were correctly detected by MPS (**Table 1**).

Patient 1 was an infant with mild hepatosplenomegaly. Previous PCR analysis had failed to amplify exons 3–5 of the *G6PC* gene. Subsequent targeted array comparative genome hybridization confirmed a homozygous deletion of exons 3–5. MPS (**Figure 2a**) failed to detect exons 3–5, whereas exons 1 and 2 in this patient, and all *G6PC* exons in control samples in the same run had an average coverage of >600X per exon (**Figure 2a**), suggesting a homozygous deletion of exons 3–5 in this patient. Similarly, a homozygous deletion of a single exon of the *GBE1* gene was detected in P4, a male infant whose parents were second cousins. The infant was hypotonic and had respiratory difficulty immediately after delivery. An echocardiogram identified a large patent ductus arteriosus. Histopathologic studies of a muscle biopsy suggested GSD IV, given the presence of abnormal glycogen. Follow-up enzyme studies confirmed glycogen branching enzyme deficiency.¹⁵ PCR amplification of coding exons for Sanger analysis failed to amplify exon 16, the last exon of the *GBE1* gene, and no mutations were detected in

Table 2 Sensitivity and specificity of GSD panel calculated by three phase I validation samples

Sample ID	Tissue type	MPS						Sanger
		TP	FN	TN	FP	% Sensitivity	% Specificity	Positives
V1	Blood	35	0	50027	0	100	100	35
V2	Blood	36	0	50026	0	100	100	36
V3	Fibroblasts	27	0	50035	0	100	100	27
Sum		98	0	150088	0	100	100	

Region of interest: coding exons with 20bp flanking regions. The total number of base call is 50062 base per sample.

FN, false negatives; FP, false positives; GSD, glycogen storage disease; MPS, massively parallel sequencing; TN, true negatives, which are all the reads matching with reference sequences; TP, true positives, which include all mutation and variant calls in the region of interest.

Sanger positives: total number of mutations and variants.

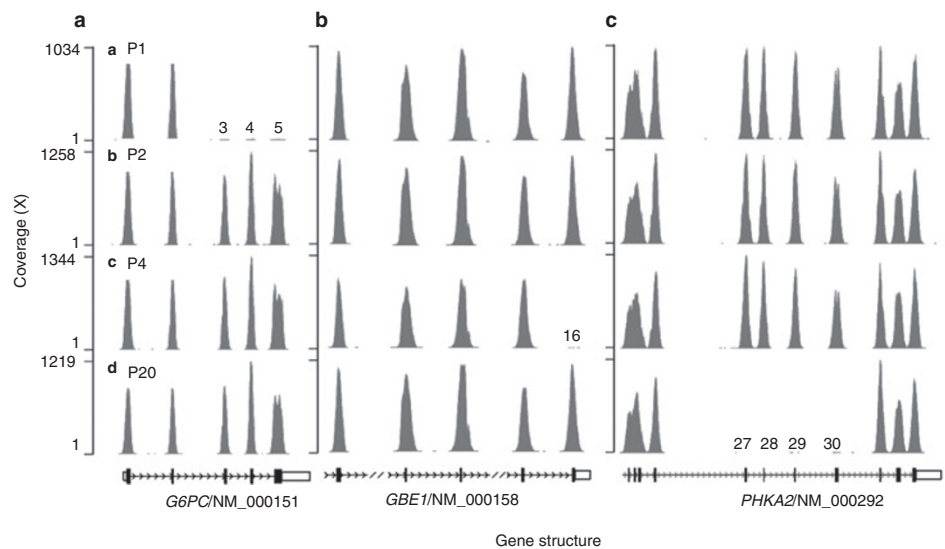


Figure 2 Large exonic homozygous and hemizygous deletions can be detected by massively parallel sequencing. The x-axis denotes the gene structure and corresponding exons, and the y-axis denotes the coverage scale. Four samples, patient (P1, P2, P4, and P20), are presented. P2 is the normal control in the same run. (a) The exons 3–5 in the *G6PC* gene in P1 were not covered by any reads (top row), whereas the same exons in samples taken from P2, P4, and P20 were uniformly well covered (lower three rows). This result is consistent with the homozygous deletion of exons 3–5 previously detected in P1.³⁰ (b) Exon 16 in the *GBE1* gene showed no coverage in P4. (c) A hemizygous deletion of exons 27–30 was detected in the *PHKA2* gene in P20.

the other exons of the *GBE1* gene. The MPS sequencing did not detect exon 16 (Figure 2b). Other exons in the *GBE1* gene of this patient, and all exons in control samples, had full coverage of greater than 600X. These results indicate that deep coverage of all coding exons allows the detection of homozygous large exonic deletions. This patient died of GSD IV-related complications within 1 month after birth.

Patient 2 presented with hypoglycemia, hepatomegaly, elevated triglycerides, and lactic acidemia, suggestive of GSD Ia. A heterozygous c.379_380dupTA (p.Tyr128Thrfs*3) mutation and a novel heterozygous c.533C>T (p.Pro178Leu) variant were detected in the *G6PC* gene by Sanger sequencing (Table 1). The c.379_380dupTA (p.Tyr128Thrfs*3) is a common mutation in patients with GSD Ia, whereas the c.533C>T (p.Pro178Leu) is a novel change. Other amino acid substitutions at the same position, namely, p.Pro178Ala and p.Pro178Ser, have been reported in patients with GSD Ia.^{16,17} Targeted MPS successfully detected these two mutations with a coverage of 730X. No other deleterious mutations were detected in other GSD genes in the panel. However, samples from the parents were not available for the determination of the phase of these two mutations.

P3 harbored three heterozygous novel variants, c.467C>T (p.Ala156Val), c.572C>G (p.Pro191Arg), and c.1024T>C (p.Ser342Pro) in the *SLC37A4* gene. All three missense variants were confirmed by our targeted gene capture MPS method (Table 1). The c.467C>T (p.Ala156Val) and c.572C>G (p.Pro191Arg) are relatively close to each other and always appear in the same reads in MPS data; therefore, they appear to be in *cis* configuration. Unfortunately, samples from the parents were not available for testing to determine whether the c.1024T>C (p.Ser342Pro) variant is in *trans* configuration with the other two variants.

The details of P4 have been described earlier in the text, together with those of P1.

P5 was a teenage patient with recurrent rhabdomyolysis, suggestive of GSD V. Sanger sequencing analysis identified a heterozygous frameshift mutation, c.21_28dup8 (p.Lys10Thrfs*19), in the *PYGM* gene. The MPS analysis correctly detected this heterozygous 8bp duplication but found no mutations in other GSD genes (Figure 3a). Given that GSD V is an autosomal recessive disorder, the identification of a single heterozygous mutation does not confirm the diagnosis. Recurrent rhabdomyolysis may be caused by mutations in other genes, including other muscle forms of GSDs, as well as disorders of fatty acid oxidation and mitochondrial respiratory chain disorders.¹⁸ DNA from this patient was analyzed using MPS for a group of 24 genes responsible for metabolic myopathy. Of note, a heterozygous c.1784delC (p.Ala596Glnfs*2) frameshift mutation and a heterozygous c.122C>T (p.Pro41Leu) novel missense variant in the *CPT2* gene were detected. It was thereby shown that the rhabdomyolysis in this patient was caused by CPTII deficiency.

P6 and P7 had mutations in the glycogen phosphorylase kinase alpha subunit gene, *PHKA2* (GSD IXa, X-linked). Using MPS, a hemizygous frameshift mutation, c.3648_3649delAA (p.Arg1217Serfs*26), was confirmed in the male patient (P6) and a heterozygous missense mutation, c.3614C>T (p.Pro1205Leu) was confirmed in the female patient (P7) (Table 1). No other mutations were identified in other GSD genes in either of these patients.

Detection of mutations in previously undiagnosed patients

A total of 17 DNA samples from unrelated individuals with clinical and/or enzymatic findings suggestive of GSD but without identified mutations were analyzed using the

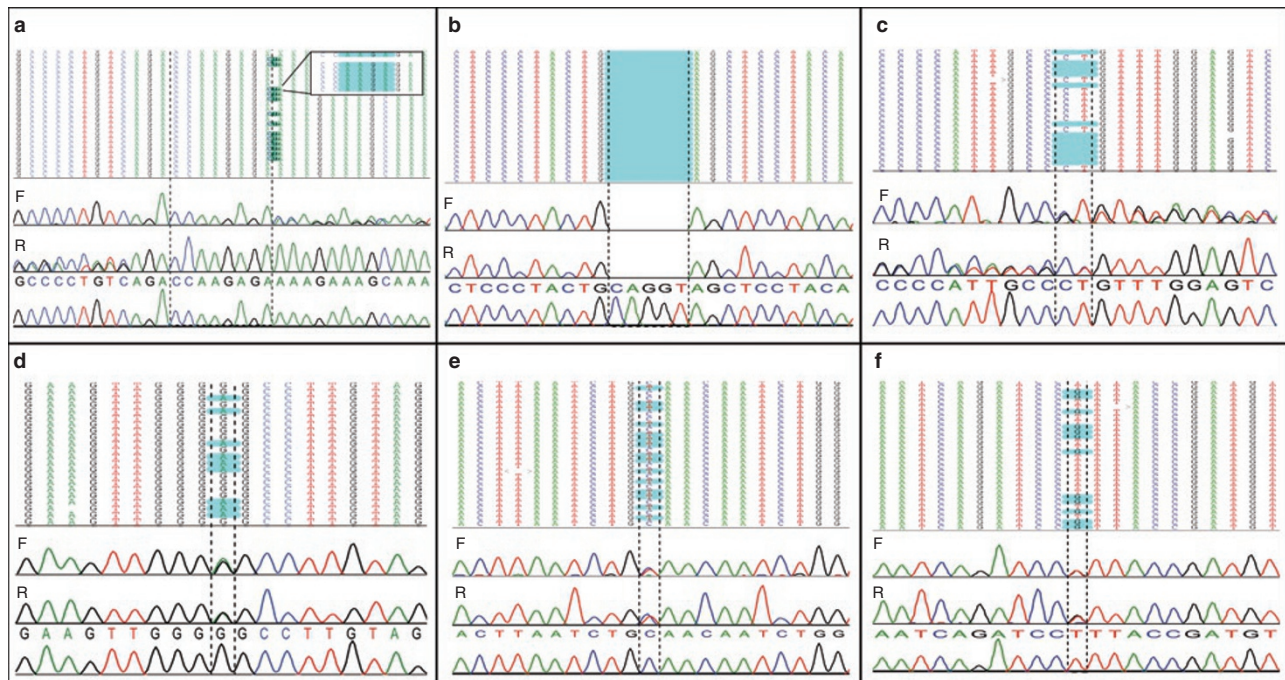


Figure 3 Examples of different types of mutations were successfully captured in phase II and phase III validation samples and confirmed by Sanger sequencing. (a) A heterozygous frameshift mutation, c.21_28dup8 (p.Lys10Thrfs*19), was detected in the *PYGM* gene in patient (P)5. (b) A novel homozygous c.785-3_786del5 (p.?) deletion was detected in the *SLC37A4* gene in P8. (c,d) A heterozygous frameshift mutation, c.1042_1043delCT (p.Leu348Valfs*53), and a heterozygous novel missense variant, c.817G>A (p.Gly273Ser), were detected in the *SLC37A4* gene in P9. (e,f) A heterozygous mutation, c.256C>T (p.Gln86*), and a heterozygous novel missense variant, c.2723T>G (p.Leu908Arg), were detected in the *AGL* gene in P11.

capture/sequencing approach. Deleterious mutations or novel variants in various GSD genes were detected in 11 of the patients (11/17 = 65%). All mutations or novel variants detected by MPS were confirmed by Sanger sequencing.

P8–P15 had clinical symptoms suggestive of a liver-related form of GSD. Earlier Sanger sequencing analysis of the *G6PC* gene (GSD Ia) in these patients did not detect deleterious mutations. MPS analysis of the GSD-related genes identified mutations in four of the patients (Table 1). A novel homozygous c.785-3_786del5 (p.?) deletion was detected in the *SLC37A4* gene (G6P translocase, GSD Ib) in P8. This deletion disrupts the invariant splice acceptor site of intron 5 of *SLC37A4* and is therefore classified as a deleterious mutation¹⁹ (Figure 3b). P9 and P10 both harbored a heterozygous frameshift mutation and a novel missense variant (Table 1, Figure 3b,c). In patient 11, a heterozygous c.256C>T (p.Q86*) mutation (Figure 3e) and a heterozygous novel missense variant, c.2723T>G (p.Leu908Arg) (Figure 3f) in the *AGL* gene (debranching enzyme, GSD III) were detected.

P16–P20 were deficient in one of the enzymes of glycogen metabolism, but the diagnoses had not been confirmed by molecular testing. MPS analysis detected mutations in the genes corresponding to the deficient enzymes in P17–P19. Two heterozygous mutations were detected in the *AGL* gene in P17 and P18, both of whom had debranching enzyme deficiency (Table 1). A hemizygous frameshift mutation, c.2609delC (p.Pro870Glnfs*44), was detected in the *PHKA2* gene in P19; this patient's enzyme study result was highly suggestive of GSD IX. However, in the other two patients, P16 and P20, the enzyme

study results were inconsistent with molecular findings. P16 had debranching enzyme deficiency, whereas no mutations were detected in the *AGL* gene. Instead, a heterozygous missense mutation, c.698G>A (p.Pro233Asp),²⁰ and a heterozygous nonsense mutation, c.2467C>T (p.Gln823*), were detected in the liver isoform of glycogen phosphorylase, the *PYGL* gene (GSD VI) (Table 1). P20 had phosphorylase enzyme deficiency suggestive of GSD VI; however, previous Sanger sequencing analysis of the *PYGL* gene had not detected any mutations. The MPS analysis detected a hemizygous large deletion encompassing exons 27–30 of the *PHKA2* gene. The mother of this patient was confirmed to be carrying the same deletion.

P21 and P22 had clinical and histochemical findings strongly suggestive of GSD. Enzymatic analysis of the liver biopsy was not performed in these patients. MPS analysis confirmed the diagnosis of GSD by identifying a hemizygous mutation, c.133C>T (p.Arg45Trp) in the *PHKA2* gene in P21 and a homozygous frameshift mutation, c.379_380dupTA (p.Tyr128Thrfs*3), in the *G6PC* gene in P22 (Table 1).

The diagnosis of GSD could not be confirmed in P12–P15, and P23 and P24. Either no deleterious mutations were detected in any of the GSD genes analyzed, or only one heterozygous unclassified variant was detected in an autosomal recessive gene.

DISCUSSION

The ability to sequence a group of candidate genes simultaneously makes the MPS technology an ideal approach to the molecular diagnosis of a genetically heterogeneous and clinically difficult

to distinguish genetic disorder such as GSD. Sixteen genes that are known to cause either liver or muscle forms of GSDs were included in this panel test (**Supplementary Table S1** online). The unbiased capture and deep coverage of each coding exon and adjacent intronic region of all genes in this panel ensures the accuracy of mutation detection. Currently, the reported MPS-based analysis of target genes relies on multiplex PCR enrichment of the coding sequences, requiring routine confirmation with additional Sanger sequencing for genes harboring low- or no-coverage exons.^{11,21–23} The MPS-based sequence analysis of the genes involved in glycogen metabolism presented here relies on in-solution probe hybridization for the capture of target sequences under one uniform condition, followed by MPS with deep coverage. This strategy provides an average base-by-base coverage of >600X in all target regions. Therefore it does not require additional Sanger procedures to cover any problematic exons; it provides reliable results in a single step, with reduced turnaround time in reporting the results.

Clinical utility of MPS-based analysis of GSD genes

GSD is a group of genetically heterogeneous inborn metabolic disorders. Current molecular diagnoses rely on step-wise Sanger sequencing of individual genes. This conventional approach is expensive and time consuming. The overall yield is low because of the limited number of candidate genes that can be sequenced; therefore a molecular diagnosis cannot be established for many persons in whom GSD may be suspected. We have developed MPS-based analyses of the genes that are known to be implicated in GSD and validated them for clinical testing (**Supplementary Table S1** online). Depending on the predominant clinical features, GSD-related genes can be divided into two groups: those giving rise to the liver forms of GSD and those giving rise to the muscle forms of GSD; some GSDs (types II, III, and IXb) can affect both liver and muscle. In patients who have clinical and histochemical indications of a GSD but in whom evidence of organ involvement is ambiguous, the molecular testing of all GSD genes is likely to confirm a diagnosis. The testing of a panel of relevant genes simultaneously can greatly shorten the time required to reach a confirmed diagnosis, thereby facilitating appropriate patient care and genetic counseling.

There is no doubt that whole-exome sequencing is the future of molecular diagnosis. This is especially so for complex disorders that have less specific clinical findings and therefore present difficulties in identifying the group of genes to be analyzed.⁹ However, in defined diseases or syndromes such as GSD, biochemical markers or clinical features can easily direct the investigation of a specific pathway or a group of genes responsible for the disease condition. In such cases, MPS-based target gene analysis has been proven to be an efficient and cost-effective approach.^{11,24,25} This study demonstrated that 65% (11/17) of the patients with no previous molecular diagnosis of GSD did indeed carry mutations in one of the GSD-related genes, thereby confirming the presumptive clinical diagnosis (**Table 1**). Of course, the detection rate is highly dependent on an accurate clinical evaluation.

Of a total of eight patients in whom GSD Ia was suspected, three did not have mutations in *G6PC*; rather, they had mutations in *SLC37A4* (GSD Ib) (**Table 1**). Theoretically, GSD Ia can be distinguished from GSD Ib by a clinical evaluation. GSD Ib is associated with impaired neutrophil and monocyte function and chronic neutropenia, resulting in recurrent bacterial infections and oral and intestinal mucosal ulcers.²⁶ However, these clinical features may not occur in the first few years of life. P10 and P11 were infants, and neutropenia was not part of their clinical presentation. Therefore the absence of neutropenia from the clinical picture may not be sufficient *per se* to suggest the presence of GSD Ia rather than GSD Ib in young patients. GSD Ia is the more common of the two types, and molecular testing as well as biochemical assays are readily available to identify it. Biochemical evaluation for GSD Ia involves assaying the catalytic activity of glucose-6-phosphatase, which can be readily measured in snap-frozen liver biopsy samples. In contrast, GSD Ib involves measuring the activity of glucose-6-phosphate translocase (transporter), which is difficult to carry out in frozen liver samples. A fresh liver specimen is often needed to accurately assay the enzyme activity. Consequently, most clinical diagnostic laboratories do not offer this enzyme activity assay.³ The confirmation of the diagnosis of GSD Ib therefore relies mostly on DNA analysis of the *SLC37A4* gene. Mutations in *SLC37A4* are estimated to account for ~20% of GSD I.²⁷ However, this mutation frequency may be an underestimate, given the limited availability of molecular and biochemical testing for GSD Ib. Although the sample size in our study is small, our results suggest that close to 40% (3/8) of patients with no GSD Ia mutations may nevertheless have GSD Ib. With the availability of MPS-based analysis of GSD, the estimated frequency of GSD Ib may need to be revised.

GSD III is caused by defects in the glycogen debranching enzyme. It is characterized by variable levels of involvement of the liver, cardiac muscle, and skeletal muscle. Liver involvement is typically present in infancy, whereas hypertrophic cardiomyopathy develops during childhood. Skeletal muscle weakness is usually not evident in childhood but slowly progresses and becomes prominent in the third to fourth decade of life.²⁸ P11 was a 2-year-old boy who showed only liver involvement, suggestive of GSD Ia. Previous Sanger sequencing analysis of *G6PC* did not detect any mutations in this patient. An analysis of the genes responsible for the liver form of GSD revealed a heterozygous c.256C>T (p.Gln86*) mutation and a heterozygous novel missense variant, c.2723T>G (p.Leu908Arg) (**Table 1**) in the *AGL* gene in this patient, suggesting a diagnosis of GSD III. These examples show that patients may not have developed the full spectrum of symptoms at the time of their clinical evaluation. With limited availability of enzymatic analyses and ambiguous histochemical findings, making a specific clinical diagnosis of GSD in young patients may be difficult.

P16 had debranching enzyme deficiency suggestive of GSD III; however, no mutations were detected in the *AGL* gene. Instead, two heterozygous mutations in the *PYGL* gene were detected. Debranching enzyme activity can be measured in

biopsy samples from liver or muscle. However, improper handling of samples or an inaccurate enzyme assay can lead to false-positive or false-negative results. P20 was a boy who had liver glycogen phosphorylase enzyme deficiency consistent with GSD VI. Previous Sanger sequencing of the hepatic phosphorylase gene *PYGL* did not reveal any deleterious mutations. The MPS-based analysis detected a hemizygous large deletion encompassing exons 27–30 in the X-linked phosphorylase kinase, *PHKA2* gene (Table 1). The *PHKA2* gene encodes the α subunit of phosphorylase kinase (PhK) in the liver. The enzyme PhK activates the inactive form of glycogen phosphorylase b to the active form, phosphorylase a, and the total phosphorylase activity is regulated by PhK.²⁹ Therefore, a deficiency in glycogen phosphorylase may be the result of impairment of the phosphorylase itself or of its regulatory protein, the glycogen phosphorylase kinase, PhK. Furthermore, there are multiple tissue-specific forms of PhK subunits that may be responsible for the apparent phosphorylase deficiency. These two examples indicate that, although enzyme assays are important in identifying GSD, the ultimate diagnosis must rely on molecular confirmation. This approach would also reduce the need for tissue biopsies.

In this study, the diagnosis of GSD could not be confirmed in 6 of 17 (35%) patients. These were patients who either carried no identifiable deleterious mutations or harbored only one heterozygous unclassified variant in one of the autosomal recessive genes. There are several possible reasons for these results. First, these six patients may not have GSD. Some of the clinical features, such as hepatomegaly and muscle weakness are non-specific and difficult to distinguish from those caused by fatty acid oxidation disorders, mitochondrial diseases, and other metabolic storage disorders. Indeed, none of the six patients had clear clinical indications or laboratory findings supporting the diagnosis of GSD. Second, although this test has 100% sensitivity and 100% specificity as compared with the Sanger method, it is limited to the targeted coding exons and the adjacent 20 bp of the intronic regions. Mutations in nontargeted deep intronic and regulatory regions will not be detected by this assay. Finally, the mutations could have been in genes that were not targeted in the probe design.

Large deletions and other types of mutations detected using MPS

The uniform deep coverage of all coding exons and the flanking intron regions allows the detection not only of single-nucleotide substitutions and small deletions and insertions (Figure 3a–f) but also of large deletions involving single or multiple exons. Homozygous (Figure 2a,b) or hemizygous (Figure 2c) deletions can be easily identified by the absence of coverage in the targeted regions. However, detection of heterozygous exonic deletions will be more challenging. Our data show that there is uniform coverage of the same coding regions across different samples (Figure 2a–c). Computational programs similar to the ones used in the detection of heterozygous copy-number changes in the oligonucleotide microarray comparative genomic hybridization analysis can be designed

for the analysis of the sequence read coverage to detect a heterozygous exonic deletion. All large deletions detected by MPS should be further confirmed through a second methodology such as targeted array comparative genome hybridization,³⁰ Multiplex Ligation-dependent Probe Amplification (MLPA), or quantitative PCR. Because the MPS targeted enrichment focuses only on the coding regions, this GSD panel test cannot identify a deletion breakpoint if it falls within deep intronic regions, but it is able to map the exact breakpoint if the deletion is within a coding region.¹⁴

Advantages and limitations of MPS testing for molecular diagnosis

In comparison with the most commonly used multiplex PCR enrichment method for MPS sequencing,³¹ the custom-designed in-solution capture is better suited for clinical diagnostics because it can be easily scaled up and automated with robotic liquid handling. For a clinical test, all novel variants with possible clinical significance must be verified. This verification step is necessary so as to remove incorrect calls caused by experimental error and to confirm the variant calls. The confirmation of a large number of novel variant calls is time-consuming, resulting in a long turnaround time. Therefore, to adapt MPS-based testing to a clinical diagnostic setting, it is recommended that all primers to be used for confirmation of any positive findings be validated.

The turnaround time for this MPS-based panel testing, including confirmation of the positive calls through Sanger sequencing, is ~8 weeks. This is a very reasonable time span for a test of such high complexity, as compared to the 3–6 weeks taken for a simple single-gene analysis using the Sanger method. The simultaneous analysis of 16 genes reduces the average time and cost involved when compared to the traditional step-wise approach. With the improvement of sequencing chemistries, computational algorithms, bioinformatics analytical tools, interpretation of variants,³² and shortened turnaround time, reliable and fully validated MPS-based clinical tests will eventually become the mainstay of molecular diagnoses.

In summary, this study is the first to demonstrate the utility of a “clinical grade” MPS approach with 100% sensitivity and specificity of sequence analysis. Our recent review of the up-to-date publications revealed that there are issues related to clinical testing validation when adapting MPS technology for the molecular diagnosis of genetic disease in a clinical setting.³³ Sanger sequencing can detect the variants at any position in the regions of interest, while the low coverage of exons by MPS may produce substantial false-negative results. The targeted enrichment, sequencing chemistry, and computational analysis can also lead to false-positive results, as has been demonstrated in these publications. Our data underscore the importance and clinical utility of MPS-based analysis in the molecular diagnosis of a defined disorder that may be caused by defects in multiple genes. The high-throughput MPS strategy for simultaneous analysis of all the genes responsible for liver and muscle forms of GSD greatly helps in the diagnosis of GSD in a cost- and time-efficient manner.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

- Scriver CR. *The metabolic and molecular bases of inherited disease*, 7th ed. McGraw-Hill: New York, NY, 1995.
- Hicks J, Wartchow E, Mierau G. Glycogen storage diseases: a brief review and update on clinical features, genetic abnormalities, pathologic features, and treatment. *Ultrastruct Pathol* 2011;35:183–196.
- Priya S, Kishnani DK, Yuan-Tsong C. Glycogen storage diseases. *Online Metabolic and Molecular Bases of Inherited Disease*. McGraw-Hill: New York, NY, 2001; 2009. Chapter 17. http://www.ommbid.com/OMMBID/a/c.html/carbohydrates/glycogen_storage_diseases.
- Chou JY, Matern D, Mansfield BC, Chen YT. Type I glycogen storage diseases: disorders of the glucose-6-phosphatase complex. *Curr Mol Med* 2002;2:121–143.
- Tsujino S, Shanske S, Nonaka I, DiMauro S. The molecular genetic basis of myophosphorylase deficiency (McArdle's disease). *Muscle Nerve* 1995; 3:S23–S27.
- Desnuelle C, Salviati L. Challenges in diagnosis and treatment of late-onset Pompe disease. *Curr Opin Neurol* 2011;24:443–448.
- Bruno C, Servidei S, Shanske S, et al. Glycogen branching enzyme deficiency in adult polyglucosan body disease. *Ann Neurol* 1993;33:88–93.
- Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;12:745–755.
- Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;42:30–35.
- O'Roak BJ, Deriziotis P, Lee C, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 2011;43:585–589.
- Jones MA, Bhide S, Chin E, et al. Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation. *Genet Med* 2011;13:921–932.
- Neveling K, Collin RW, Gilissen C, et al. Next-generation genetic testing for retinitis pigmentosa. *Hum Mutat* 2012;33:963–972.
- Coppieters F, De Wilde B, Lefever S, et al. Massively parallel sequencing for early molecular diagnosis in Leber congenital amaurosis. *Genet Med* 2012; 14:576–585.
- Cui H, Zhang W, Wong L-JC. Comprehensive molecular analyses of mitochondrial genome by next-generation sequencing. *12th International Congress of Human Genetics/61st Annual Meeting of The American Society of Human Genetics*. Montreal, Canada, 2011.
- Ward TL, Valberg SJ, Adelson DL, Abbey CA, Binns MM, Mickelson JR. Glycogen branching enzyme (GBE1) mutation causing equine glycogen storage disease IV. *Mamm Genome* 2004;15:570–577.
- Ki CS, Han SH, Kim HJ, et al. Mutation spectrum of the glucose-6-phosphatase gene and its implication in molecular diagnosis of Korean patients with glycogen storage disease type Ia. *Clin Genet* 2004;65:487–489.
- Lei KJ, Chen YT, Chen H, et al. Genetic basis of glycogen storage disease type 1a: prevalent mutations at the glucose-6-phosphatase locus. *Am J Hum Genet* 1995;57:766–771.
- Berardo A, DiMauro S, Hirano M. A diagnostic algorithm for metabolic myopathies. *Curr Neurol Neurosci Rep* 2010;10:118–126.
- Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet Med* 2008;10:294–300.
- Tang NL, Hui J, Young E, et al. A novel mutation (G233D) in the glycogen phosphorylase gene in a patient with hepatic glycogen storage disease and residual enzyme activity. *Mol Genet Metab* 2003;79:142–145.
- Hu H, Wrogemann K, Kalscheuer V, et al. Mutation screening in 86 known X-linked mental retardation genes by droplet-based multiplex PCR and massive parallel sequencing. *Hugo J* 2009;3:41–49.
- Gowrisankar S, Lerner-Ellis JP, Cox S, et al. Evaluation of second-generation sequencing of 19 dilated cardiomyopathy genes for clinical applications. *J Mol Diagn* 2010;12:818–827.
- Simpson DA, Clark GR, Alexander S, Silvestri G, Willoughby CE. Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa. *J Med Genet* 2011;48:145–151.
- Calvo SE, Compton AG, Hershman SG, et al. Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci Transl Med* 2012;4:118ra110.
- 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;86:378–388.
- Visser G, Rake JP, Labrune P, et al. Granulocyte colony-stimulating factor in glycogen storage disease type 1b. Results of the European Study on Glycogen Storage Disease Type 1. *Eur J Pediatr* 2002;161 Suppl 1:S83–S87.
- Janecke AR, Mayatepek E, Utermann G. Molecular genetics of type 1 glycogen storage disease. *Mol Genet Metab* 2001;73:117–125.
- Lucchiari S, Santoro D, Pagliarini S, Comi GP. Clinical, biochemical and genetic features of glycogen debranching enzyme deficiency. *Acta Myol* 2007;26: 72–74.
- Hendrickx J, Dams E, Coucke P, Lee P, Fernandes J, Willems PJ. X-linked liver glycogenosis type II (XLG II) is caused by mutations in PHKA2, the gene encoding the liver alpha subunit of phosphorylase kinase. *Hum Mol Genet* 1996;5:649–652.
- Wang J, Zhan H, Li FY, Pursley AN, Schmitt ES, Wong LJ. Targeted array CGH as a valuable molecular diagnostic approach: Experience in the diagnosis of mitochondrial and metabolic disorders. *Mol Genet Metab* 2012. 106:221–30.
- Clark MJ, Chen R, Lam HY, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* 2011;29:908–914.
- Zhang VW, Wang J. Determination of the clinical significance of an unclassified variant. *Methods Mol Biol* 2012;837:337–348.
- Zhang W, Cui H, Wong LJ. Application of next generation sequencing to molecular diagnosis of inherited diseases. *Top Curr Chem* 2012; e-pub ahead of print 11 May 2012.