Top Curr Chem (2014) 336: 19–46
DOI: 10.1007/128_2012_325

Published online: 11 May 2012

Application of Next Generation Sequencing to Molecular Diagnosis of Inherited Diseases

Wei Zhang, Hong Cui, and Lee-Jun C. Wong

Abstract Recent development of high throughput, massively parallel sequencing (MPS or next generation sequencing, NGS) technology has revolutionized the molecular diagnosis of human genetic disease. The ability to generate enormous amount of sequence data in a short time at an affordable cost makes this approach ideal for a wide range of applications from sequencing a group of candidate genes, all coding regions (known as exome sequencing) to the entire human genome. The technology brings about an unprecedented application to the identification of the molecular basis of hard-to-diagnose genetic disorders. This chapter reviews the up-to-date published application of next generation sequencing in clinical molecular diagnostic laboratories. We also emphasize the various target gene enrichment methods and their advantages and shortcomings. Obstacles to compliance with regulatory authorities like CLIA/CAP in clinical settings are also briefly discussed.

Keywords Human genetic diseases, Massively parallel sequencing (MPS), Molecular diagnosis, Next generation sequencing (NGS), Target gene enrichment

Contents

W. Zhang, H. Cui and L.-J.C. Wong (✉)
Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, NAB 2015, Houston TX 77030, USA
e-mail: ljwong@bcm.edu

1 Introduction

Over the past 20–30 years, DNA sequencing by the Sanger method has been regarded as the gold standard for the identification of aberrant DNA sequence changes (mutations, in genetic terms) to support a genetic disease diagnosis [1]. For single gene disorders with clear clinical/biochemical indication and/or known mutation hot spots, Sanger sequencing of the target region is an accurate and cost effective way to obtain a definitive molecular diagnosis. However, the selection of candidate gene(s) for sequence analysis is extremely difficult, as most inherited disorders exhibit genetic and clinical heterogeneity. Thus, it often takes a long time to obtain the right molecular diagnosis by stepwise sequencing the probable causative genes with consequent high cost and anxiety for the patient's family. An example of an extreme situation is the diagnosis of mitochondrial disorders, for which there are significant clinical overlaps and over 1,300 genes are responsible [2–5]. In recent years, owing to the fast development of a variety of sequencing technologies in the post human genome project era, large scale sequencing such as (1) a group of target genes, (2) all of protein coding regions of the human genome, and (3) even the whole human genome have become a reality [6–12]. Next Generation Sequencing (NGS) or Massively Parallel Sequencing (MPS) offers a way to detect mutations in many different genes in a cost and time efficient manner through their ability to generate deep coverage of the target sequences [13–19].

Various sequencing platforms using different sequencing chemistry have been developed [20]. The three major currently commercially available platforms that produce gigabases of output are Illumina's sequencing by DNA synthesis, Roche's 454 by pyro-sequencing, and ABI SOLiD by oligonucleotide ligation [21–24]. There are also platforms for smaller scale sequencers, such as Ion Torrent, which is based on semiconductors to detect protons generated by polymerase reactions [25], single molecule sequencing by Helicos Helioscope [26], and Pacific Bioscience's single molecular real time (SMRT) instrument [27, 28].

Regardless of the chemistry and platforms, these methods share one common principle, which is to sequence numerous spatially separated genomic regions in a massively parallel manner [13, 20, 29]. The detailed chemistry, sequencing principles, their hardware, and their advantages and pitfalls are beyond the scope of this chapter but are discussed in chapter 1 of this book.

The MPS approaches have been successfully applied to many different research studies to identify new disease genes [30], mutations in non-coding regions [31–34], and epigenetic changes in the whole genome [35, 36]. Similar approaches have been applied to molecular diagnosis of inherited disorders, particularly complex diseases with heterogeneous clinical phenotype and multiple underlying genetic causes [37–40]. However, it is still not economically feasible and is technically demanding to sequence the 3×10^9 base pairs of the whole human genome. Therefore, in the context of clinical applications, it is often desirable to capture or enrich a group of genes known to be responsible for a certain type of clinical phenotypes, followed by MPS. Good examples are genes causing cardiomyopathy or mitochondrial respiratory chain disorders [38, 40]. Therefore, depending on the amount of sequence to be analyzed, the purpose of the analysis, and the available sequencing platforms, the method of target gene enrichment may vary greatly. In addition, in order to use newly developed MPS tests for clinical diagnosis, stringent validation and quality control procedures need to be instituted in compliance with regulating authorities, such as CAP or CLIA to assure quality of service [41–43]. It is likely that all laboratory-developed tests (LDTs) will be subjected to regulation in the near future in many developed countries [44].

In this chapter we will review the reported clinical applications of MPS to the molecular diagnosis of inherited disorders, various gene enrichment methods, potential pitfalls, necessary quality control procedures, and remaining obstacles for comprehensive genomic analysis using MPS.

2 Current Clinical Application of MPS

A review of the recent publications on the introduction of MPS technology into clinical practice has revealed the broad utility of this novel technology in the molecular diagnosis of a variety of human genetic disorders (Table 1). The clinical applications vary from single gene disorders such as neurofibromatosis Type 1 (NF1), Marfan syndrome (MFS), and spastic paraplegia [45–47] to diseases caused by a group of related genes such as hypertrophic cardiomyopathy and congenital disorders of glycosylation (CDG) [38, 39, 48]. MPS has also been applied to multi-gene disorders including X-linked intellectual disability (XLID) and retinitis pigmentosa [37, 49] as well as defined disorders without identified genetic causes [9, 52, 53]. Various gene enrichment methods were employed (Table 2) depending on the purpose of application.

2.1 Target Gene(s) Sequencing in the Clinical Setting

MPS has been applied to assist with molecular diagnosis of well-defined disorders caused by a group of genes. In order to sequence specifically the regions of interest, these target genes need to be enriched tens of thousands of times or more to avoid the interference of the remaining bulk of the genome. Depending on the amount of target sequences, the enrichment method may vary.

Table 1 Summary of the recent publications on the introduction of MPS technology to clinical practice

	Gene/disease	Enrichment method	MPS Sequencing platform/ chemistry	Sample	Average coverage	% reads mapped to target	Analysis software	References
1	*NF1*	NimbleGen oligo array capture	GS-FLX (SBS, pyro-sequencing)	2 known	>30X	~52% 59% of 52% mapped to ch 17 *NF1*	NextGENe	[45]
2	Marfan+LDS *FBN1 +TGFBR1+ TGFBR2*	Multiplex PCR	GS-FLX (SBS, pyro-sequencing)	5 known 87 unknown	~174X	55–85%	In house variant interpret pipeline	[46]
3	*CYP7B1* (SPG5) *+SPG7*	Fluidigm (FD)	GS-FLX (SBS, pyro-sequencing)	187 patients	72X for run1 25X for run2	80% of target exons have more than 20X coverage	SeqNext v3.4.1 Build 504 (JSI German)	[47]
4	DCM 19 genes	Pooled PCR amplicons	GAII (SBS)	5 known	~50X	59–69%	NextGENe	[38]
5	CDG 24 gene	Fluidigm (FD) RainDance (RD)	SOLiD version 3/50 base read, SE (SBL)	12 known	616X (FD) 455X (RD)	48%	NextGENe	[48]
6	RP 45 genes	NimbleGen oligo array capture	GAII/32 base read, SE (SBS)	2 known 3 unknown	486X (1 sample per lane) 98X (4 samples per lane)	~35%	Genomic Workbench	[49]
7	XLMR 86 genes	RainDance (RD)	GAII (SBS)	3 known 21 unknown	Coverage per base ranging from 92X to 445X	67.9%	SOAP2.20	[37]
8	mtDNA	2 overlapping PCR fragments	GAII (SBS)	2 known	~1,785x	N/A	DNAStar & NextGENe	[50]
9	mtDNA and 362 nuclear genes	Agilent array-based capture	GAII/36 base read, SE (SBS)	2 patients 1 normal	37X-51X for nuclear genes, 3,000–5,000X for mtDNA	17–35% mapped to nuclear genome, 20–37% mapped to mtDNA	MAQ	[40]

10	Carrier test of 437 genes	Agilent solution-based capture RainDance (RD)	GAII/50 base read (SBS) SOLiD3/50 base read (SBL) HiSeq/130 base read, PE (SBS)	104 known	Varies between different platforms	Varies from 13.7% to 31.7%	GSNAP (for SBS data) BioScope v1.2 (for SBL data)	[51]
11	Miller syndrome WES	Agilent array-based capture	GAII/76 base read, SE (SBS)	3 kindreds	40X	97%	MAQ	[52]
12	Kabuki syndrome WES	Agilent array-based capture	GAII/SE or PE (SBS)	10 unknown	40X	N/A	MAQ, Phaster v1.100122a	[53]
13	Inflammatory bowel disease (IBD) WES	NimbleGen exome array-based capture	GS-FLX (SBS, pyro-sequencing)	1 patient	34X	76.2%	gsMapper (Roche 454)	[54]
14	Charcot–Marie–Tooth (CMT) WGS	Direct genomic DNA	SOLiD (SBL)	Family members	30X	N/A	In house variant interpret pipeline	[9]
15	Dopa-responsive dystonia (DRD) WGS	Direct genomic DNA	SOLiD4 (SBL)	Twins and family members	30X	N/A	In house variant interpret pipeline	[11]
16	Prenatal diagnosis	Direct plasma DNA	SOLiD3 (SBL)	15 unknown	N/A	21%	SOLiD alignment tool	[55]

SE single-end, *PE* paired-end, *SBS* sequencing-by-synthesis, *SBL* sequencing-by-ligation, *WES* whole exome sequencing, *WGS* whole genome sequencing

Table 2 Comparison of different gene enrichment methods

Method	Target size	Throughput	Automation adaptability	Advantage	Disadvantage
PCR-based enrichment (commercial platform)					
Regular PCR/ multiplex	Small	Low	Yes	Simple, straight forward Flexible in optimizing conditions	Cost-inefficient Limited target size
Array PCR (Fluidigm)	Small to medium	Low	No	Cost-efficient reactions Easy to perform Uniform PCR condition	High cost in synthesizing primers Varied amplification efficiency Fixed format
Microdroplet PCR (RainDance)	Medium	Low	Yes	Up to 20,000 amplicons Relatively equal amplification efficiency	High cost in synthesizing primers Varied amplification efficiency Need special equipment Less flexibility
Capture-based enrichment					
Oligonucleotide array-based capture	Medium to high	Low	No	Target size is expandable Low cost per gene	Need one extra step to elute the captured DNA Need special equipment Missing GC rich regions
Oligonucleotide Solution-based capture	Medium to high	High	Yes	Target size is expandable Low cost per gene Easy for automation	Cost-inefficient for small target region Missing GC rich regions

2.1.1 Examples of Single Gene Disorders and the Target Enrichment Approach

Neurofibromatosis Type 1 (NF1)

Neurofibromatosis type 1 is an autosomal dominant disorder caused by mutations in the *NF1* (17q11.2) gene. It is one of the most commonly inherited genetic disorders with an estimated prevalence of 1 in 3,000 individuals [56]. Molecular defects in the

NF1 gene include a variety of mutation types from point mutations and small indels to large complex rearrangements [57]. In addition, the presence of pseudogene complicates the sequencing analysis of the target gene. A new molecular analysis was carried out with a microarray based target gene capture and enrichment method, followed by MPS [45]. In this report, DNA samples from two patients with a definitive diagnosis of neurofibromatosis type 1 and known *NF1* mutations were analyzed. The 282 kb *NF1* gene region was fragmented and captured on an array containing specific oligonucleotide probes followed by sequencing using Roche (454) Genome Sequencer FLX system. About 59% of mapped reads were localized to the targeted *NF1* loci on chromosome 17 in both cases with a depth of coverage >30X. After filtering out the false positive calls introduced by low sequencing coverage, and manual removal of pseudogene, an Alu sequence insertion and a frame-shift single base deletion were identified in each of the two patients respectively. However, only 3 out of 17 and 3 out of 9 identified variants in each of these two samples respectively were confirmed by Sanger sequencing [45]. It indicated that sequence variants identified by MPS might have a high percentage of false positives.

Marfan Syndrome

MPS has also been applied to the molecular diagnosis of MFS [46]. MFS is a common (1 in 5,000) autosomal dominant connective tissue disorder, characterized by tall stature with long limbs, dislocated lens, and severe cardiovascular manifestation including aorta aneurysms [58, 59]. The majority of mutations causing MFS are missense, nonsense, and small insertion/deletions in the *FBN1* gene encoding the fibrillin-1 protein. Due to the large number of exons, sequencing by PCR/Sanger is very time consuming and costly. Baetens et al. developed a multiplexed PCR based-MPS approach to sequence simultaneously all exons in 87 samples [46]. In addition to *FBN1*, transforming growth factor beta receptor genes *TGFBR1* and *TGFBR2* are also included. A total of 117 amplicons were amplified in 17 multiplex PCR sets, each containing 4–11 amplicons and sequencing was performed on the Roche (454) Genome Sequencer FLX system. Five samples from known patients diagnosed with MFS or LDS (Loeys–Dietz Syndrome) were selected for a pilot study, which confirmed 23 out of the total 25 calls made by Sanger sequencing, leaving 2 variants unidentified due to low coverage. Additionally, 20 false positive variants were called, all residing in homopolymeric regions. Consequently, to rule out false positives, 13 homopolymer regions were sequenced by the Sanger method separately. Subsequent analysis of 87 patients with typical MFS identified 75 *FBN1* mutations by MPS, equivalent to an 86% (75/87) detection rate.

Hereditary Spastic Paralegias

Hereditary spastic paralegias (HSPs) are a group of inherited neurodegenerative disorders characterized by weakness and spasticity in the lower limbs. They are

genetically heterogeneous and at least 46 genetic loci have been identified so far, showing autosomal dominant, autosomal recessive, or X-linked form of inheritance [60]. In the study carried out by Schlipf and coworkers, 2 genes, *CYP7B1* (*SPG5*) and *SPG7*, that are involved in the autosomal recessive form of HSP, were evaluated in 187 patients using MPS [47]. Target gene regions were enriched by an array based microfluidic chip amplification method (Fluidigm) followed by next-generation pyro-sequencing using the Roche (454) Genome Sequencer FLX system. Results from two independent runs showed that 80% of the exons were covered at more than 20X. The remaining poorly covered regions were analyzed by Sanger sequencing separately. Mutations were identified in a total of ten patients: three with *SPG5* mutations and seven with *SPG7* mutations.

2.1.2 Panels of Target Genes for a Disease that Could Be Caused by Mutations in One of Many Potentially Responsible Genes

Dilated Cardiomyopathy

Dilated cardiomyopathy (DCM) is a group of genetically heterogeneous disorders with more than 30 genes identified so far. For three decades, Sanger sequencing of the genes one-by-one has been used, but this is expensive, and time consuming. Resequencing arrays has been considered as an alternative method, but it is well known that they cannot reliably detect insertions and deletions or mutations in high GC rich regions. Additionally, if there were a need for revision in the original design, the cost of making a small change is not economical for most laboratories. Thus, it is impractical to expand quickly the array content as new genes are discovered [61, 62]. These obstacles are likely to be solved by the newly developed MPS technologies that allow simultaneous sequencing of a large number of genes. Gowrisankar and coworkers applied this method to analyze a subset of 19 genes (*ABCC9, ACTC, ACTN2, CSRP3, CTF1, DES, EMD, LDB3, LMNA, MYBPC3, MYH7, PLN, SGCD, TAZ, TCAP, TNNI3, TNNT2, TPM1,* and *VCL*) known to cause DCM. The target coding exons were amplified by the PCR method and their products were pooled for library construction [38]. Five samples with known mutations were validated by this method on the Illumina Genome Analyzer II (GAII), a first generation MPS platform. Results were compared to those of Sanger sequencing and an array-based resequencing chip [38].

An estimated false positive rate of 10.8% and false negative rate of 3.4% were observed. Exons with poor coverage were clearly noticeable, which was the primary reason for missing calls (false negative). The importance of confirmation by a second method to remove the false positive variant calls was emphasized. About 9.3% (23 out of 246) of poorly covered amplicons required Sanger sequencing to minimize the percentage of missed calls. The authors concluded that even though the quality of MPS based testing for the set of 19 genes related to DCM is better compared to the array based re-sequencing method, but the cost associated with test setup and the turnaround time need to be improved before implementing as a

routine clinical testing [38]. At the time of writing this review, clinical MPS DCM testing has been already available from a few clinical laboratories.

Congenital Disorders of Glycosylation

The congenital disorders of glycosylation (CDG) are a group of diseases caused by molecular defects in more than 30 genes involved in the N-linked glycosylation pathway [63]. New underlying genetic causes for CDG are continuously being discovered [64, 65]. The prevalence of CDG is estimated to be 1 in 20,000. The disease is usually devastating [63, 66]. Approximately 40% of CDG patients do not have a definitive molecular diagnosis, which is required for the prenatal diagnosis for the affected families. A newly developed MPS-based method was designed to analyze a panel of 24 genes (*ALG2, ALG3, ALG6, ALG8, ALG9, ALG12, ATP6V0A2, B4GALT1, COG1, COG7, COG8, DOLK, DPAGT1, DPM1, GNE, MGAT2, MOGS, MPDU1, MPI, PMM2, RFT1, SLC35A1, SLC35C1,* and *TUSC3)* known to cause CDGs [48].

A total of 215 coding exons were enriched by two different PCR methods: the microdroplet-based PCR (RainDance, RD) with custom designed primers and multiplex PCR by the microfluidic chip (Fluidigm, FD) with in-house designed primers. PCR products were pooled and subjected to ABI SOLiD library preparation and sequencing by ligation [24].

Twelve samples from patients with a known diagnosis of CDG were sequenced for validation using the methods described above. The results showed that about 26–32% of the total variant calls appear to depend on the target enrichment methods. A total of 455 and 616 variants were called by RD and FD methods, respectively. After filtering the data to eliminate the low coverage and low quality calls, 85% and 94% of the unique variants called by the RD and FD, respectively, were likely false positives. In contrast, only 27% of the variants called by both methods were likely not real. About 28 exons (ca. >10% library failure rate) had low or no coverage, which required Sanger sequencing to fill these gaps. The false-negative rate could not be determined due to the lack of DNA [48].

This 24 CDG gene analysis demonstrated the cost-effectiveness of highly multiplexed analysis vs the conventional step-wise single gene approach. The authors also emphasized the importance of verification of all clinically significant variants by Sanger sequencing [48].

Retinitis Pigmentosa

Similar to DCM and CDG, retinitis pigmentosa (RP) is also a group of genetically diverse diseases caused by mutations in any one of more than 40 known genes. Diagnosis of RP is complicated by the lack of clear clinical indication for a specific genetic defect. This disease can be inherited as autosomal dominant, autosomal recessive, or X-linked. About 50% of all patients with RP are sporadic simplex

cases and it is clinically difficult to pinpoint the causal genes for molecular diagnosis [67]. The recent development of MPS overcomes these obstacles.

Oligonucleotide DNA probes for a total of 681 coding exons encompassing 359 kb regions of a panel of 45 RP genes were printed onto a custom-designed solid-phase capture array [49]. Target sequences of DNA samples from five patients with RP were enriched by array capture followed by MPS using Genome Analyzer II with 40 bp single end reads [49].

This study identified one patient with compound heterozygous missense mutations in the *CRB1* gene, a second patient with a known homozygous missense mutation in the *PDE6B* gene, and a third patient with a homozygous novel missense variant in the *CNGB1* gene that is predicted to be deleterious. Although predicted deleterious variants were found in the remaining two patients, due to non-segregation of these variants with the disease the molecular diagnosis could not be confirmed.

In this study, if one sample was analyzed in one lane of the flow cell, the mean coverage was 486X with 99% regions covered at >20X, compared to the mean coverage of 98X and 88% regions with >20X coverage by pooling four samples. A total of 582 variants were detected, 150 of them novel. Only the possibly pathogenic variants were verified by Sanger sequencing. The data were not sufficient to establish false positive and false negative rates. The authors emphasized the importance of achieving the requirement of >15X coverage in at least 90% of target regions as a clinical standard and also suggested outsourcing as an alternative to lower the cost [49].

X-Linked Intellectual Disability

As much as 2–3% of the general population meets clinical criteria of intellectual disability. XLID is very heterogeneous. More than 100 XLID genes have been identified, which account for only a portion of families with unambiguous XLID [68]. Using MPS, analysis of a panel of 86 XLID genes has been offered as a clinical molecular test in order to improve the detection rate.

The target regions of the 86 XLID genes of 24 unrelated patients, each from a large XLID family, were enriched by microdroplet-based multiplex PCR (RainDance) with custom designed primer sets for 1,912 amplicons; MPS was performed on Illumina GA II [37].

This amplicon-based approach revealed that 88.5% of targeted regions were covered with a median coverage of 249X. Between 87.9% and 99.5% of targeted regions were covered by sequence reads with a base quality score ≥30 at a minimal coverage of 2X [37]. It was reported that 131 out of 1,912 amplicons cannot be unambiguously mapped, and 34 amplicons failed to amplify, corresponding to an 8.6% failure rate.

The most noticeable is the large variations (from 12.2% to 57.4%) in the percentage of reads mapped to the targeted regions. The authors attributed the variation to different amplification efficiencies in different samples. In addition to three known positive controls in the 24 samples, deleterious mutations were

identified in 7 additional patients, with a detection rate of 33% (7/21). Nevertheless, the authors considered this method to be robust and usable as a screening test for XLID [37].

2.1.3 Sequencing the Mitochondrial Genome and Autosomal Genes Coding for Mitochondrial Proteins

In addition to 23 pairs of nuclear chromosomes, each human cell contains hundreds to thousands of copies of the mitochondrial genome. The human mitochondrial genome is a 16.6-kb circular double stranded DNA encoding 37 genes: 22 tRNA, 2 rRNA, and 13 mRNAs for 13 respiratory chain complex protein subunits. Mutations in mitochondrial DNA (mtDNA) cause a broad clinical spectrum of respiratory chain disorders with the tissues of high energy demand such as brain and muscle preferentially affected [69]. Disease severity depends on the degree of mutation heteroplasmy in the affected tissues [69]. Therefore, diagnosis of mtDNA disorders must not only identify the disease causing mutations but also quantify the mutation heteroplasmy. Since accurate quantification of mutation heteroplasmy is important in disease prognosis and outcome, MPS needs to provide sufficiently deep coverage for the measurement of heteroplasmy at every single nucleotide position.

A pilot study using MPS technology has been conducted to investigate the possibility and feasibility of mtDNA sequencing using the MPS approach [50, 70]. The entire 16.6-kb mtDNA genome was amplified with two overlapping fragments followed by sequencing with the Illumina Genome Analyzer GAII. The authors demonstrated the ability of MPS to detect reliably mtDNA heteroplasmy down to 5% level and with high sensitivity but poor specificity for variant detection. This method had the advantages of cost effectiveness and fast turnaround time, but cannot detect large mtDNA deletions or mtDNA copy number variation and may include incorrect mtDNA heteroplasmy at certain nucleotide positions. Cui et al. have recently developed a one-step comprehensive analysis of mtDNA by MPS that provides accurate detection of point mutations at every single nucleotide position of the entire mitochondrial genome and determination of large deletions with deletion heteroplasmy and exact breakpoints [71].

Mitochondrial Disorders: Diseases of the Two Genomes

Mitochondrial disorders are a group of complex diseases that can be caused by mutations in both nuclear and mitochondrial genomes. More than 99% of the proteins involved in mitochondrial biogenesis and functions are encoded by an estimated 1,300 nuclear genes. These features of heterogeneous but overlapping clinical presentations and thousands of possible underlying genetic loci make the diagnosis of mitochondrial disorders the most difficult among all inherited disorders [69, 72–75]. The ability of MPS technology to sequence thousands of

different genes with deep coverage is an attractive solution to the diagnosis of complex mitochondrial disorders.

A proof-of-concept study was carried out by using a microarray based capture method to enrich the mtDNA genome and 362 nuclear genes related to mitochondrial disorders, followed by sequence analysis on an Illumina Genome Analyzer GAII [40]. Their results showed that more than 94% of the targeted regions were sequenced at a coverage of about 45X for the nuclear genes. Two known mutations in the positive controls were identified correctly. The coverage depth is approximately 4,000X for mtDNA [40].

In this study about 5% of the target regions were not covered, and about 6–10% of the variants identified were novel, which required Sanger confirmation. The authors commented that the sample preparation step was tedious, but the entire procedure, after careful validation and improvement, was likely to be readily adapted for clinical application.

2.1.4 Carrier Testing in Family Members

Of the approximately 7,000 Mendelian genetic disorders, 16% are classified as autosomal recessive disorders. Most autosomal recessive disorders manifest in early childhood with severe disease course and a poor outcome [76, 77]. Preconception screening, coupled with genetic counseling of carriers has helped tremendously in the reduction of the incidence of severe recessive diseases. Thus, an affordable comprehensive preconception screening would be desirable for general population if the assay includes actionable, highly penetrant, and severe recessive mutations. The recently developed target gene capture and MPS approach has made this type of screening assay possible. Bell et al. reported their work on a preconception carrier screening test for 448 severe recessive childhood diseases [51]. A total of 7,717 regions from 437 target genes were enriched by either SureSelect RNA-based in solution capture (Agilent) or microdroplet-based multiplex PCR (RainDance), followed by sequence analysis with Illumina Genome Analyzer GAII or ABI SOLiD.

The average coverage depth was 160X and 93% of regions had >20X coverage. The results demonstrated a specificity of 99.6% and a sensitivity of 94.9%. A study of 104 unrelated individuals revealed an average carrier burden of 2.8 severe pediatric recessive mutations per individual, and the distribution of mutations appeared to be random and pan-ethnic. The authors also commented that about 27% (122 of 460) of the literature-documented mutations are actually common polymorphisms or have been mis-annotated, underscoring the need for better mutation databases.

The authors believed that the combination of two target enrichment methods could provide better detection sensitivity, although at a higher cost. Additionally they commented that sequencing cost could be further decreased with higher test volume. An important issue in the discussion was the availability of informatics support and result interpretation. This remains a major challenge for laboratories

looking to implement MPS-based population carrier screening tests in clinical settings in which regulatory guidelines set by CLIA and CAP are required to be followed for certified clinical diagnostic laboratory [51].

2.2 Whole Exome Sequencing

The common procedures for analyzing exome sequencing results include primary analysis to convert image data to base-calling, secondary analysis to align and map sequence reads, and tertiary analysis to annotate the variants. Examples of bioinformatics tools needed for the processing of sequence data can be found at http://seqanswers.com/wiki/How-to/exome_analysis. The bulk of variants detected by MPS are routinely filtered using publicly available SNP databases, the population allele frequency, computational algorithms for function prediction, and disease databases. Comparison of the variants among family members or unrelated patients with similarly defined clinical syndrome is very helpful in narrowing down the genes for further confirmation and investigation.

As described in the examples of target gene sequencing, a large number of patients with clinical diagnosis were left without an identified molecular cause despite sequence analysis of a group of candidate genes responsible for the clinical condition. These results suggested that the disease causing mutations might not be in the genes that were targeted for sequencing or that mutations were not in the coding regions of the targeted genes. Complete exome sequencing (WES) addresses the first problem, but sequencing of the whole human genome, including the non-coding region (WGS), would be needed to solve the second problem.

It has been shown that whole exome sequencing (WES) of a small number of affected, unrelated individuals could potentially be used to identify a causal gene underlying a rare Mendelian monogenic disorder [52]. In a recent report, DNA samples from four affected individuals in three unrelated kindreds with Miller syndrome were subjected to exome capture followed by single-end, 76 bp cycle sequencing [30]. About 40X coverage of 26.6 Mb mappable exome sequence was obtained. After filtering through dbSNP129 and eight HapMap exomes, a single candidate gene, dihydroorotate dehydrogenase (*DHODH*), a key enzyme in the pyrimidine de novo synthesis pathway, was identified. Sanger sequencing confirmed the presence of *DHODH* mutations in three additional families with Miller syndrome [30]. Using this approach, the same group successfully analyzed exomes from ten patients with autosomal dominant Kabuki syndrome. After filtering against SNP databases, candidate genes containing novel variants in all affected individuals were not found. With less stringent filtering criteria, allowing for the presence of modest genetic heterogeneity, multiple candidate genes were found. After phenotypic and genotypic stratification, a gene *MLL2* encoding the trithorax-group histone methyltransferase was identified. Follow-up sequencing detected *MLL2* mutations in 2 of the remaining 3 patients and 26 of 43 additional cases with Kabuki syndrome [52, 53].

The two examples described above are characteristic Mendelian disorders that occurred in multiple unrelated families. The use of exome sequencing identified new disease genes. Recently, this approach enabled researchers to make a clinical diagnosis of a previously undefined disease that altered the treatment in a single child with a life threatening inflammatory bowel disease (IBD) [54]. The male child presented at 15 months with perianal abscesses and proctitis, suggesting an immune defect. However, comprehensive clinical evaluation could not reach a definitive diagnosis. Exome sequencing performed on this individual identified 16,124 variants. After analysis, a novel hemizygous missense alteration that changed a highly conserved cysteine residue to tyrosine in the X-linked inhibitor of apoptosis (*XIAP*) gene was identified as the causative mutation [54]. Functional studies using peripheral blood mononuclear cells (PBMCs) from the patient and normal controls demonstrated defective responsiveness to NOD2 ligands and enhanced apoptosis in the patient's PBMCs, suggesting that the mutation leads to the loss of XIAP activity. Based on the diagnosis, an allogenic transplant of hematopoietic progenitor cells was performed and the child was able to ingest food without recurrence of the gastrointestinal disease. This report demonstrates the power of exome sequencing in molecular diagnosis of a novel disease and the utility of exome sequencing method in a clinical laboratory [54].

2.3 Whole Genome Sequencing

Before WES became commonly used for clinical diagnosis, whole-genome sequencing (WGS) was also proven to be a useful research tool for new gene discovery [9, 78, 79]. Using the ABI SOLiD platform, WGS was performed on the proband of a family with a recessive form of Charcot–Marie–Tooth disease. Compound heterozygous mutations in the *SH3TC2* gene were identified and the mutation were found to segregate with disease in four affected siblings [9].

Similarly, using the ABI SOLiD 4 platform, the disease gene was discovered in a twin brother and sister initially diagnosed with cerebral palsy in early childhood. The WGS revealed an average coverage of 30X and 2,500,000 variants. Extensive filtering resulted in 70 variants. Among them, only three genes were considered to be candidates for an autosomal recessive disorder. One of them was the *SPR* gene, encoding a sepiapterin reductase, responsible for the synthesis of tetrahydrobiopterin, an important cofactor for the biosynthesis of neurotransmitters. Based on the finding, a new therapeutic regimen was administered that led to the improvement of clinical symptoms [11]. This discovery of the gene responsible for DOPA responsive dystonia demonstrated that, similar to WES, the WGS approach can not only provide a molecular diagnosis for patients with rare genetic disorders but also guide effective treatments that would otherwise not have been considered [11, 54].

2.4 Application of MPS to Prenatal Diagnosis

Current molecular prenatal diagnosis requires invasive procedures for amniocentesis or chorionic villus sampling. In addition to cost, these procedures have an approximately 0.5% miscarriage risk. Thus, it is desirable to develop a non-invasive procedure for prenatal diagnosis to avoid the risk of fetal loss.

Recently there has been rapid progress in applying MPS techniques to the detection of fetal chromosomal aneuploidies using maternal plasma DNA. These developments have built upon the pioneering work of Denis Lo and his coworkers at The Chinese University of Hong Kong [80], who demonstrated that at the end of first trimester more than 10% of cell-free DNA is from fetal genome. The detection of trisomy 21 aneuploidy from maternal plasma DNA by MPS has been shown to be more reliable than the detection of other aneuploidies such as trisomy 18 and 13 [55]. Subsequently, three large-scale studies involving multiple centers have further confirmed the clinical validity of this approach [81–83]. These studies established that non-invasive detection of fetal trisomy 21 could be carried out at nearly 100% sensitivity and 98% specificity by multiplexed MPS of maternal plasma DNA. Thus, the implementation of the MPS-based detection method will undoubtedly decrease the risk of fetal loss associated with the screening of high risk pregnancies. However, practical considerations, such as turnaround time, cannot be overlooked. It is likely that an invasive procedure will still be required for follow-up.

As for the application of MPS to the prenatal mutation detection of monogenic disorders, differentiating germline from somatic mutations appears to be challenging. A pilot study to detect fetal β-thalassemia mutations using maternal blood and massively parallel sequencing illustrates the possibility of investigating specific genetic disease loci [84].

3 Target Gene Enrichment Methods

3.1 Multiplex PCR

3.1.1 Microfluidic Chip Multiplex PCR

As the most common method for Sanger sequencing, the PCR-based enrichment method has the potential to be tailored for MPS based tests. Simplex PCR is unlikely to be applicable since it is quite labor intensive if the target region is large and thousands of PCR are needed. To generate high throughput target amplicons by PCR, multiple methods have been investigated including multiplex PCR and Fluidigm (South San Francisco, CA) Access Array [85, 86]. In a multiplex PCR, several pairs of primers are mixed together in order to amplify multiple regions of the genome in a single PCR tube under the same cycling condition. The Fluidigm

access array utilizes a microfluidic chip, which contains separated minute chambers and thus solves the problems of encountered non-specific amplification in regular multiplex PCR. Simplex PCR is performed in each physically partitioned space without cross interference with the others. Under this setting, up to 48 samples can be mixed individually with up to 10 unique primer pairs. Each amplicon or set of amplicons is maintained separately after amplification. With the fixed capacity, microfluidic chip multiplex PCR can be tailored for a small target gene region for enrichment.

3.1.2 Microdroplet-Based Multiplex PCR

Another solution for performing multiple PCR reactions simultaneously with comparable amplification efficiency is using microdroplet-based multiplex PCR. In this setup, each reaction droplet is a reaction well in isolation with a lipid capsule and contains a small amount of template DNA along with one primer pair and the PCR reagents (polymerase, dNTPs, and buffers) [87]. To generate such reaction droplets, template/reagents droplets and primer droplets are first formed separately as individual libraries and then merged together on a microfluidic chip, which is currently commercialized by RainDance Technologies (Lexington, MA). During the PCR reactions, amplifications are carried out independently in each droplet, mimicking a simplex PCR. Therefore, more uniform enrichment of the target regions can be achieved as compared to the conventional multiplex PCR method.

Nevertheless, due to the same intrinsic characteristics, microdroplet PCR technology also faces several challenges shared with other PCR-based assays, including limited availability of primer design at certain genomic regions and relatively small target size. Other enrichment methods, such as capture-based enrichment methods, may be considered as an alternative when large chromosomal regions are targeted.

3.2 Oligonucleotide Probe Based Capture on a Solid Phase

In the array-based capture assays, DNA probes are synthesized on microarrays, ranging from 60 bases to 90 bases in length, for example in NimbleGen probes [88, 89]. Hybridization is performed directly on solid phase array chips to capture the target regions, followed by extensive washing to remove non-specifically bound DNA. Captured DNA is then eluted from the arrays for MPS library construction. Currently, different commercial companies have extended their array capacity with the size of the targeted regions to ~30 Mb. Although it is more time- and cost-efficient than PCR based enrichment, array based capture has its drawbacks: low through-put and difficulty to scale up.

3.3 Oligonucleotide Probe Based Capture in Solution Phase

Solution capture shares most features of array-based capture, except that it does not require dedicated hybridization instrumentation since the capture hybridization is carried out on a regular thermocycler at either 42 °C (for NimbleGen DNA probe based capture) or 65 °C (for Agilent SureSelect RNA probe based capture). In addition, the solution capture processes can easily be scaled up and automated with robotic liquid handling for simultaneous capture of 96 samples in a 96-well plate set-up. Multiple studies have been performed to compare the performance of DNA probe- and RNA probe-based in-solution capture. The results demonstrated similar capture capability with slightly better uniformity achieved by the DNA probe-based method [90–93]. Nevertheless, one common hurdle shared by all capture-based enrichment is the co-capturing of targeted loci/genes with corresponding pseudogenes. At times it is difficult to retrieve only the target sequence if the regions of interest contain long stretches of highly (>90%) homologous sequences such as the case of pseudogene. This is a problem that may not be resolved by the modification of the probe design and experimental procedures. However, pseudogene sequences may sometimes be filtered out by stringent alignment in post-sequencing data management.

4 Cautionary Tale of Current MPS Tests Offered for Clinical Application

As reviewed above, different gene enrichment methods may have certain limitations and these limitations directly affect the subsequent MPS analysis [94, 95]. In addition, read coverage, sensitivity, specificity, and turnaround time may vary depending on the size of the target genes and the chemistry of the different MPS platforms. To bring this new research technology (a series of new technology, in fact) to the standard required for medical application needs stringent validation. In this section we will review the potential shortcomings of the current MPS methodologies and how they will impact on the application to clinical settings.

4.1 PCR-Based Target Amplification Has Intrinsic Drawbacks

This PCR-based amplification followed by MPS encounters some potential drawbacks, which usually appear to produce uneven coverage of the target regions. Poor amplification of some targets may be due to high GC content or particular DNA structures. The presence of SNPs at the primer sites may also cause allelic dropout. In addition, it is difficult to detect large deletion/insertion events using PCR-based enrichment methods.

4.2 Lack of Sufficient Clinical Validation, Specificity, Sensitivity

Unlike the Sanger method, MPS analyses involve more technical steps, including sample preparation, multiplexing, sequencing, and image detection, setting parameters for variant detection, filtering algorithms and cut off values, as well as evaluation of detection sensitivity and specificity. Each of these steps requires quality control and clinical validation. Depending on the enrichment methods, sequencing chemistry and platforms, as well as analytical tools, individual laboratories may institute different standards and procedures. Therefore, some kind of consensus on standard setting is required.

Review of the published articles revealed that the majority of laboratories used a limited number of pre-selected, known positive samples as controls for validation. For example, 5 control samples were from MFS [46] and only 12 known positive controls were used for CDG validation [48]. This level of validation underestimates the false negative rate and thus these methods may have lower assay sensitivity. The validation procedures should be designed in such a way that the full spectrum of the target regions is examined with a large sample size in two phases; phase I with blinded samples and phase II with known positive controls, in order to fulfil the requirement of a full clinical validation.

4.3 High False Negative Rate Due to Low Coverage

Regardless of the methods used for target gene enrichment, low or no coverage for certain exons is a common problem. This is more common in GC rich regions or regions with particular DNA sequence structures that are susceptible to DNA fragmentation. In addition, different chemistry of sequencing platforms, and the different computational algorithms, may miss particular regions of target sequences. For example, indels in the homopolymer regions are easily missed by pyro-sequencing-based technology, and short tandem repeats (STRs) may be missed by ligation based short reads. The low and no coverage regions are the major cause of false negative results, which in a clinical setting are the least acceptable analytical errors. For a panel containing a small number of genes, these low or no coverage regions may be easily filled by conventional PCR/Sanger sequencing. However, even if the missed regions are known, it may be impossible to fill these gaps by Sanger sequencing if there are a large numbers of low coverage regions. As described in the published papers, 28 out of the 215 coding regions of the CDG study [48] and 20% of the exons in the HSP study were not well covered. Moreover, 34 out of 1,912 amplicons in the XLID study failed amplification [37]. Although there is no standard regarding the minimum coverage required for MPS-based clinical tests, a preliminary cut-off of at least 40 reads from both directions was suggested (TECHGENE forum 2011, Leuven, Belgium). At this cut-off, on average, about 5–10% of the target regions are considered not sufficiently covered. That would mean that about ~9,000–18,000 coding exons are not well covered in the whole exome sequence analysis. A recent

analysis on the coverage depth with regard to the detection sensitivity and sequencing errors indicated that a few hundred-fold coverage may be needed for accurate diagnosis in some clinical situations [96].

4.4 Confirmation of Detected Variants to Rule Out False Positives

False positive rate is high in MPS experiments. Typically, about 10–20 variants per gene are detected [97]. This number can be greatly reduced after filtering, depending on the filtering criteria. On average, there may be up to two missense variants per gene. As such, if a panel of 20 genes is analyzed, there will be approximately 20 novel variants that require confirmation by a second method, usually by Sanger sequencing. This verification step is necessary for two reasons – to remove incorrect calls due to experimental error, and to confirm the diagnosis. The confirmation may become burdensome or impossible when the number of genes analyzed increases to 1,000 or 20,000 for the whole exome. For a clinical test, all novel variants with possible clinical significance must be verified before reporting. The confirmation of a large number of novel variants is time-consuming, resulting in a long turnaround time, which is impractical for clinical diagnosis.

4.5 Interpretation of Novel Variants: The Most Challenging Task

The human genome reference sequence has been updated constantly (Homo sapiens GRCh37.2 (hg19) is being used now). Although the majority of variants detected by MPS have been observed previously, a normal healthy individual may carry a number of sequence changes, of which, majority are benign [98, 99].

The population based variant database dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) has documented allele frequencies, but some ethnic specific variants may be underrepresented; such alleles are often confused for disease causing mutations. Bioinformatics tools have been used extensively to help with the interpretation of novel variants [100, 101]. However, none of these tools have been validated for clinical use and most of the commonly used algorithms have very high false-positive and false-negative rates [102]. Another issue is that a high number of sequencing errors may be mistaken as de novo events. Family-based sequencing has therefore been proposed to aid in the gene discovery [98, 103, 104]. In chapter 4 of this book, another approach using protein structure information is reviewed for its potential use in MPS.

5 Discussion

In the field of clinical medical genetics, the advancement of molecular diagnosis for detecting inherited genetic defects usually accompanies the development of new technologies. These novel technologies often lead to the rapid discovery of new genes and new diseases, leading to further improvement in molecular diagnosis, patient care, and management [20, 105–107].

Since the development of the Sanger dideoxynucleotide-terminator sequencing method, it gradually became, and has remained, the gold standard for clinical molecular diagnostics, due to its accuracy in detecting small genetic variants. The diagnosis of monogenic genetic disorders using Sanger sequencing usually has a reasonable turnaround time of approximately 6–8 weeks depending on the size and complexity of the gene interrogated. However, the laborious workflow, error-prone nature, and high operational cost make it unattractive in the new genomic era, especially when sequencing large genomic regions are frequently involved.

The sequencing of James Watson's genome in 2008 was the first "personal genome" accomplished by using MPS. The era of genomic medicine has arrived and MPS will be a key feature, which will evolve with reducing price and increasing throughput [7]. The methodology developed at the human genome sequencing center (HGSC) at the Baylor College of Medicine has since paved the road to personalized genome re-sequencing for disease gene discovery, risk factor assessment, and MPS-based molecular diagnoses in a clinical setting.

However, there are growing pains as we move forward with these novel technologies. The biggest challenge is the interpretation of the enormous amount of genomic data and the establishment of genotype and phenotype correlations. There are about 3,000,000 single nucleotide polymorphisms in each individual, and everyone is a carrier of some recessive diseases [108]. The complexity posed by these technologies is illustrated by the recent comparative genomic analysis of an individual using different platforms, only about 60% of the variants called are likely to be real [109]. It is clear that, without careful validation, MPS at its present shape cannot be reliably and widely used in clinical setting for diagnosis of genetic diseases.

Another aspect of the complexity is establishing the physiological and pathological significance of a genetic variant in an individual's genome. For example, a recent newborn screening study revealed that the pathogenic status of the previously identified c.1436C>T (p.P479L) mutation of the *CPT1A* gene is debatable due to its high frequency and low penetrance in the Inuit population [110, 111].

The sequencing errors generated during an MPS run may not be a big problem for the purpose of gene discovery and other research applications, as these studies often involve multiple family members or multiple unrelated pedigrees with the same disease, which could be used for internal validation [52, 53]. However, when clinical molecular diagnosis is to be made for a sample of single patient with uncertain disease diagnosis, the accuracy of the sequencing results becomes very

critical and crucial [18, 54, 112, 113]. The relatively high false positive findings demonstrated by different groups in this review need to be resolved before the MPS approach can make widespread impact on patient care.

The use of MPS technology is not a perfect solution to all genetic problems. Similar to other molecular technologies, there are limitations such as the inability to detect large, complex genomic structural rearrangements [95]. It also has difficulty in distinguishing active gene from pseudogene or highly homologous genomic regions. Other technologies, including aCGH and Sanger sequencing, are presently required to complement these shortcomings in order to detect the full spectrum of mutations responsible for genetic disorders. In light of these practical issues in the routine application of MPS to clinical diagnosis and patient care MPS will likely require further improvements in both accuracy and standardization [18].

6 Conclusion

The studies reviewed here demonstrate promising results in the development of MPS technologies for clinical applications [20, 114, 115]. The adaptation of MPS approaches to clinical diagnostics has been an on-going active pursuit. Continual improvement of the accuracy of sequencing chemistries, computational algorithms for alignment, bioinformatics analytical tools, and the improved methods for the interpretation of variants will make MPS a primary tool of clinical laboratories. Although finding the causal change for a simplex case is in fact a difficult undertaking, it will become routine [106]. Many hurdles are expected to be resolved in the near future [18, 94, 104, 116]. There is every reason to expect that reduction in cost and improvement in accuracy and speed will be seen in the coming years.

Acknowledgement This work is partially supported by a grant from Muscular Dystrophy Association (Research #175369) to Dr. Lee-Jun Wong.

References

1. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74(12):5463–5467
2. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong S-E et al (2008) A mitochondrial protein compendium elucidates complex I disease biology. Cell 134(1):112–123
3. Calvo S, Jain M, Xie X, Sheth SA, Chang B, Goldberger OA et al (2006) Systematic identification of human mitochondrial disease genes through integrative genomics. Nat Genet 38(5):576–582
4. Thorburn D (2004) Mitochondrial disorders: prevalence, myths and advances. J Inherit Metab Dis 27(3):349–362

5. Scharfe C, Lu HH-S, Neuenburg JK, Allen EA, Li G-C, Klopstock T et al (2009) Mapping gene associations in human mitochondria using clinical disease phenotypes. PLoS Comput Biol 5(4):e1000374
6. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP et al (2007) The diploid genome sequence of an individual human. PLoS Biol 5(10):e254
7. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A et al (2008) The complete genome of an individual by massively parallel DNA sequencing. Nature 452 (7189):872–876
8. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE et al (2010) Clinical assessment incorporating a personal genome. Lancet 375(9725):1525–1535
9. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DCY, Nazareth L et al (2010) Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med 362(13):1181–1191
10. Dewey FE, Chen R, Cordero SP, Ormond KE, Caleshu C, Karczewski KJ et al (2011) Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. PLoS Genet 7(9):e1002280
11. Bainbridge MN, Wiszniewski W, Murdock DR, Friedman J, Gonzaga-Jauregui C, Newsham I et al (2011) Whole-genome sequencing for optimized patient management. Sci Transl Med 3(87):87re3
12. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467(7319):1061–1073
13. Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9(1):387–402
14. Metzker ML (2010) Sequencing technologies – the next generation. Nat Rev Genet 11 (1):31–46
15. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26 (10):1135–1145
16. Gibbs RA (2011) Bringing genomics and genetics back together. Science 331(6017):548
17. Collins FS (2011) Faces of the genome. Science 331(6017):546
18. Green ED, Guyer MS (2011) Charting a course for genomic medicine from base pairs to bedside. Nature 470(7333):204–213
19. Lander ES (2011) Initial impact of the sequencing of the human genome. Nature 470 (7333):187–197
20. Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55(4):641–658
21. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437 (7057):376–380
22. Bennett S (2004) Solexa Ltd. Pharmacogenomics 5(4):433–438
23. Bennett ST, Barnes C, Cox A, Davies L, Brown C (2005) Toward the $1000 human genome. Pharmacogenomics 6(4):373–382
24. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM et al (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309 (5741):1728–1732
25. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475 (7356):348–352
26. Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I et al (2008) Single-molecule DNA sequencing of a viral genome. Science 320(5872):106–109
27. Korlach J, Bjornson KP, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ et al (2010) Chapter 20: Real-time DNA sequencing from single polymerase molecules. Methods in Enzymology. Academic, pp 431–455

28. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA et al (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7(6):461–465
29. Turner EH, Ng SB, Nickerson DA, Shendure J (2009) Methods for genomic partitioning. Annu Rev Genomics Hum Genet 10(1):263–284
30. Ng SB, Nickerson DA, Bamshad MJ, Shendure J (2010) Massively parallel sequencing and rare disease. Hum Mol Genet 19(R2):R119–R124
31. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ (2005) Elucidation of the small RNA component of the transcriptome. Science 309(5740):1567–1569
32. Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) A two-hit trigger for siRNA biogenesis in plants. Cell 127(3):565–577
33. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu A-L et al (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18(4):610–621
34. Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12 (10):671–682
35. Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. Nat Rev Genet 11(7):476–486
36. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462 (7271):315–322
37. Hu H, Wrogemann K, Kalscheuer V, Tzschach A, Richard H, Haas SA et al (2009) Mutation screening in 86 known X-linked mental retardation genes by droplet-based multiplex PCR and massive parallel sequencing. Hugo J 3(1–4):41–49
38. Gowrisankar S, Lerner-Ellis JP, Cox S, White ET, Manion M, LeVan K et al (2010) Evaluation of second-generation sequencing of 19 dilated cardiomyopathy genes for clinical applications. J Mol Diagn 12(6):818–827
39. Voelkerding KV, Dames S, Durtschi JD (2010) Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. J Mol Diagn 12(5):539–551
40. Vasta V, Ng S, Turner E, Shendure J, Hahn SH (2009) Next generation sequence analysis for mitochondrial disorders. Genome Med 1(10):100
41. Lubin IM, Caggana M, Constantin C, Gross SJ, Lyon E, Pagon RA et al (2008) Ordering molecular genetic tests and reporting results: practices in laboratory and clinical settings. J Mol Diagn 10(5):459–468
42. Maddalena A, Bale S, Das S, Grody W, Richards S, the ALQAC (2005) Technical standards and guidelines: molecular genetic testing for ultra-rare disorders. Genet Med 7(8):571–583
43. Chen B, Gagnon M, Shahangian S, Anderson NL, Howerton DA, Boone JD (2009) Good laboratory practices for molecular genetic testing for heritable diseases and conditions. MMWR Recomm Rep 58(RR-6):1–37, quiz CE-1-4
44. Vance GH (2011) College of American pathologists proposal for the oversight of laboratory-developed tests. Arch Pathol Lab Med 135(11):1432–1435
45. Chou L-S, Liu CSJ, Boese B, Zhang X, Mao R (2010) DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model. Clin Chem 56(1):62–72
46. Baetens M, Van Laer L, De Leeneer K, Hellemans J, De Schrijver J, Van De Voorde H et al (2011) Applying massive parallel sequencing to molecular diagnosis of Marfan and Loeys-Dietz syndromes. Hum Mutat 32(9):1053–1062
47. Schlipf NA, Schüle R, Klimpe S, Karle KN, Synofzik M, Schicks J et al (2011) Amplicon-based high-throughput pooled sequencing identifies mutations in CYP7B1 and SPG7 in sporadic spastic paraplegia patients. Clin Genet 80(2):148–160

48. Jones MA, Bhide S, Chin E, Ng BG, Rhodenizer D, Zhang VW et al (2011) Targeted polymerase chain reaction-based enrichment and next generation sequencing for diagnostic testing of congenital disorders of glycosylation. Genet Med 13(11):921–932
49. Simpson DA, Clark GR, Alexander S, Silvestri G, Willoughby CE (2011) Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa. J Med Genet 48(3):145–151
50. Tang S, Huang T (2010) Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. Biotechniques 48(4):287–296
51. Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J et al (2011) Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med 3(65):65ra4
52. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM et al (2009) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42(1):30–35
53. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI et al (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 42(9):790–793
54. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B et al (2011) Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med 13(3):255–262
55. Chiu RWK, Sun H, Akolekar R, Clouser C, Lee C, McKernan K et al (2010) Maternal plasma DNA analysis with massively parallel sequencing by ligation for noninvasive prenatal diagnosis of trisomy 21. Clin Chem 56(3):459–463
56. Rasmussen SA, Friedman JM (2000) NF1 gene and neurofibromatosis 1. Am J Epidemiol 151 (1):33–40
57. Messiaen LM, Callens T, Mortier G, Beysen D, Vandenbroucke I, Van Roy N et al (2000) Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. Hum Mutat 15(6):541–555
58. Judge DP, Dietz HC (2005) Marfan's syndrome. Lancet 366(9501):1965–1976
59. Robinson PN, Arteaga-Solis E, Baldock C, Collod-Baroud G, Booms P, De Paepe A et al (2006) The molecular genetics of Marfan syndrome and related disorders. J Med Genet 43 (10):769–787
60. Salinas S, Proukakis C, Crosby A, Warner TT (2008) Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms. Lancet Neurol 7(12):1127–1138
61. Zimmerman RS, Cox S, Lakdawala NK, Cirino A, Mancini-DiNardo D, Clark E et al (2010) A novel custom resequencing array for dilated cardiomyopathy. Genet Med 12(5):268–278. doi:10.1097/GIM.0b013e3181d6f7c0
62. Sakai H, Suzuki S, Mizuguchi T, Imoto K, Yamashita Y, Doi H et al (2012) Rapid detection of gene mutations responsible for non-syndromic aortic aneurysm and dissection using two different methods: resequencing microarray technology and next-generation sequencing. Hum Genet 131(4):591–599
63. Jaeken J, Matthijs G (2007) Congenital disorders of glycosylation: a rapidly expanding disease family. Annu Rev Genomics Hum Genet 8(1):261–278
64. Cantagrel V, Lefeber DJ, Ng BG, Guan Z, Silhavy JL, Bielas SL et al (2010) SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder. Cell 142(2):203–217
65. Ng BG, Sharma V, Sun L, Loh E, Hong W, Tay SKH et al (2011) Identification of the first COG-CDG patient of Indian origin. Mol Genet Metab 102(3):364–367
66. Matthijs G, Schollen E, Bjursell C, Erlandson A, Freeze H, Imtiaz F et al (2000) Mutations in PMM2 that cause congenital disorders of glycosylation, type Ia (CDG-Ia). Hum Mutat 16 (5):386–394
67. Fishman GA (1978) Retinitis pigmentosa: visual loss. Arch Ophthalmol 96(7):1185–1188
68. Ropers HH (2008) Genetics of intellectual disability. Curr Opin Genet Dev 18(3):241–250

69. Wong L-JC (2010) Molecular genetics of mitochondrial disorders. Dev Disabil Res Rev 16 (2):154–162
70. Huang T (2010) Next generation sequencing to characterize mitochondrial genomic DNA heteroplasmy. Current Protocols in Human Genetics. Wiley
71. Cui H, Zhang W, Wong L-JC (2011) 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; 2011; Montreal, Canada, 2011
72. Haas RH, Parikh S, Falk MJ, Saneto RP, Wolf NI, Darin N et al (2008) The in-depth evaluation of suspected mitochondrial disease. Mol Genet Metab 94(1):16–37
73. Wong L-JC, Scaglia F, Graham BH, Craigen WJ (2010) Current molecular diagnostic algorithm for mitochondrial disorders. Mol Genet Metab 100(2):111–117
74. Berardo A, DiMauro S, Hirano M (2010) A diagnostic algorithm for metabolic myopathies. Curr Neurol Neurosci Rep 10(2):118–126
75. DiMauro S, Schon EA (2008) Mitochondrial disorders in the nervous system. Annu Rev Neurosci 31(1):91–123
76. Costa T, Scriver CR, Childs B (1985) The effect of Mendelian disease on human health: a measurement. Am J Med Genet 21(2):231–242
77. Kumar P, Radhakrishnan J, Chowdhary MA, Giampietro PF (2001) Prevalence and patterns of presentation of genetic disorders in a pediatric emergency department. Mayo Clin Proc 76 (8):777–783
78. Boone P, Wiszniewski W, Lupski J (2011) Genomic medicine and neurological disease. Hum Genet 130(1):103–121
79. Biesecker LG (2010) Exome sequencing makes medical genomics a reality. Nat Genet 42 (1):13–14
80. Lo YMD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CWG et al (1997) Presence of fetal DNA in maternal plasma and serum. Lancet 350(9076):485–487
81. Rossa WKC, Ranjit A, Yama WLZ, Tak YL, Hao S, Chan KCA et al (2011) Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ 342:c7401
82. Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R et al (2011) Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting. Am J Obstet Gynecol 204(3):205.e1–e11
83. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M et al (2011) DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med 13(11):913–920
84. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF et al (2010) Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2(61):61ra91
85. Ottesen EA, Hong JW, Quake SR, Leadbetter JR (2006) Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. Science 314(5804):1464–1467
86. Spurgeon SL, Jones RC, Ramakrishnan R (2008) High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. PLoS One 3(2):e1662
87. Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, David PH et al (2009) Microdroplet-based PCR enrichment for large-scale targeted sequencing. Nat Biotechnol 27(11):1025–1031
88. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME (2007) Microarray-based genomic selection for high-throughput resequencing. Nat Methods 4(11):907–909
89. Okou DT, Locke AE, Steinberg KM, Hagen K, Athri P, Shetty AC et al (2009) Combining microarray-based genomic selection (MGS) with the illumina genome analyzer platform to sequence diploid target regions. Ann Hum Genet 73(5):502–513
90. Clark MJ, Chen R, Lam HYK, Karczewski KJ, Chen R, Euskirchen G et al (2011) Performance comparison of exome DNA sequencing technologies. Nat Biotechnol 29(10):908–914

91. Sulonen A-M, Ellonen P, Almusa H, Lepisto M, Eldfors S, Hannula S et al (2011) Comparison of solution-based exome capture methods for next generation sequencing. Genome Biol 12(9):R94
92. Asan, Xu Y, Jiang H, Tyler-Smith C, Xue Y, Jiang T et al (2011) Comprehensive comparison of three commercial human whole-exome capture platforms. Genome Biol 12(9):R95
93. Parla J, Iossifov I, Grabill I, Spector M, Kramer M, McCombie WR (2011) A comparative analysis of exome capture. Genome Biol 12(9):R97
94. Robinson PN, Krawitz P, Mundlos S (2011) Strategies for exome and genome sequence data analysis in disease-gene discovery projects. Clin Genet 80(2):127–132
95. Tucker EJ, Mimaki M, Compton AG, McKenzie M, Ryan MT, Thorburn DR (2012) Next-generation sequencing in molecular diagnosis: NUBPL mutations highlight the challenges of variant detection and interpretation. Hum Mutat 33(2):411–418
96. De Leeneer K, De Schrijver J, Clement L, Baetens M, Lefever S, De Keulenaer S et al (2011) Practical tools to implement massive parallel pyrosequencing of PCR products in next generation molecular diagnostics. PLoS One 6(9):e25531
97. The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449(7164):851–861
98. Vissers LELM, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P et al (2010) A de novo paradigm for mental retardation. Nat Genet 42(12):1109–1112
99. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S et al (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43(6):585–589
100. Lindblom A, Robinson PN (2011) Bioinformatics for human genetics: promises and challenges. Hum Mutat 32(5):495–500
101. Medvedev P, Stanciu M, Brudno M (2009) Computational methods for discovering structural variation with next-generation sequencing. Nat Methods 6(11s):S13–S20
102. Wei Q, Wang L, Wang Q, Kruger WD, Dunbrack RL (2010) Testing computational prediction of missense mutation phenotypes: Functional characterization of 204 mutations of human cystathionine beta synthase. Proteins 78(9):2058–2074
103. Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT et al (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328 (5978):636–639
104. Robinson P (2010) Whole-exome sequencing for finding de novo mutations in sporadic mental retardation. Genome Biol 11(12):144
105. Ross JS (2011) Next-generation pathology. Am J Clin Pathol 135(5):663–665
106. Lucy Raymond F, Whittaker J, Jenkins L, Lench N, Chitty LS (2010) Molecular prenatal diagnosis: the impact of modern technologies. Prenat Diagn 30(7):674–681
107. Robin NH (2011) Dysmorphology in the era of whole exome sequencing. Curr Opin Pediatr 23(6):579–580
108. Yngvadottir B, MacArthur D, Jin H, Tyler-Smith C (2009) The promise and reality of personal genomics. Genome Biol 10(9):237
109. Lam HYK, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M et al (2011) Performance comparison of whole-genome sequencing platforms. Nat Biotechnol 30 (1):78–82
110. Greenberg CR, Dilling LA, Thompson GR, Seargeant LE, Haworth JC, Phillips S et al (2009) The paradox of the carnitine palmitoyltransferase type Ia P479L variant in Canadian Aboriginal populations. Mol Genet Metab 96(4):201–207
111. Collins SA, Sinclair G, McIntosh S, Bamforth F, Thompson R, Sobol I et al (2010) Carnitine palmitoyltransferase 1A (CPT1A) P479L prevalence in live newborns in Yukon, Northwest Territories, and Nunavut. Mol Genet Metab 101(2–3):200–204
112. Bick D, Dimmock D (2011) Whole exome and whole genome sequencing. Curr Opin Pediatr 23(6):594–600

113. Mayer AN, Dimmock DP, Arca MJ, Bick DP, Verbsky JW, Worthey EA et al (2011) A timely arrival for genomic medicine. Genet Med 13(3):195–196
114. Andrew BS (2011) Exome sequencing: a transformative technology. Lancet Neurol 10 (10):942–946
115. Teer JK, Mullikin JC (2010) Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet 19(R2):R145–R151
116. Jackson L, Pyeritz RE (2011) Molecular technologies open new clinical genetic vistas. Sci Transl Med 3(65):65ps2

http://www.springer.com/978-3-642-39941-1

Chemical Diagnostics
From Bench to Bedside
L.S. Tang, N.; Poon, T. (Eds.)
2014, VII, 200 p. 50 illus., 24 illus. in color., Hardcover
ISBN: 978-3-642-39941-1