medical sciences

Review

Application of Massively Parallel Sequencing in the Clinical Diagnostic Testing of Inherited Cardiac Conditions

Ivone U. S. Leong [1], Jonathan R. Skinner [2] and Donald R. Love [1],*

[1] Diagnostic Genetics, LabPlus, Auckland City Hospital, PO Box 110031, Auckland 1142, New Zealand; E-Mail: IvoneL@adhb.govt.nz

[2] Green Lane Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Private Bag 92024, Auckland 1142, New Zealand; E-Mail: jskinner@adhb.govt.nz

* Author to whom correspondence should be addressed; E-Mail: donaldl@adhb.govt.nz; Tel.: +64-9-3074949 (ext. 22013); Fax: +64-9-3074939.

Abstract: Sudden cardiac death in people between the ages of 1–40 years is a devastating event and is frequently caused by several heritable cardiac disorders. These disorders include cardiac ion channelopathies, such as long QT syndrome, catecholaminergic polymorphic ventricular tachycardia and Brugada syndrome and cardiomyopathies, such as hypertrophic cardiomyopathy and arrhythmogenic right ventricular cardiomyopathy. Through careful molecular genetic evaluation of DNA from sudden death victims, the causative gene mutation can be uncovered, and the rest of the family can be screened and preventative measures implemented in at-risk individuals. The current screening approach in most diagnostic laboratories uses Sanger-based sequencing; however, this method is time consuming and labour intensive. The development of massively parallel sequencing has made it possible to produce millions of sequence reads simultaneously and is potentially an ideal approach to screen for mutations in genes that are associated with sudden cardiac death. This approach offers mutation screening at reduced cost and turnaround time. Here, we will review the current commercially available enrichment kits, massively parallel sequencing (MPS) platforms, downstream data analysis and its application to sudden cardiac death in a diagnostic environment.

Keywords: sudden cardiac death; massively parallel sequencing; genetic screening; diagnostic

1. Introduction

Sudden cardiac death (SCD) in people between the ages of 1–40 years is common [1] and is a devastating event in any family. There are several heritable cardiac disorders that can cause sudden death in the young divided broadly into two groups: the cardiac ion channelopathies and heart muscle disorders or cardiomyopathies. The conditions most commonly implicated in those under 20 years of age are the ion channelopathies, particularly long QT syndrome (LQTS), catecholaminergic polymorphic ventricular tachycardia (CPVT), Brugada syndrome (BrS) and, more rarely, short QT syndrome (SQTS). Genetic testing is especially important in these conditions, because they can never be diagnosed otherwise at autopsy; the cardiac morphology and histology are normal. From the teenage years upwards the cardiomyopathies become a progressively more common cause of sudden death. The most common are hypertrophic cardiomyopathy (HCM), arrhythmogenic right ventricular cardiomyopathy (ARVC) and dilated cardiomyopathy (DCM). Although these have typical morphological and histological features, they may be subtle or even absent, particularly in the very young.

A diagnosis of a familial cardiac condition can be achieved in 22%–50% of sudden death victims either through genetic analysis of DNA (the "molecular autopsy") and/or cardiac evaluation of relatives of the deceased [2–4]. The current molecular genetic screening system in most diagnostic laboratories use a Sanger-based sequencing approach, which is the prevailing gold standard for sequence-based testing of Mendelian disease. However, this method involves a one-by-one approach, where only a single sequence is interrogated at a time, which makes screening disorders with more than a handful of genes inefficient and costly. In the case of SCD, there are many underlying disorders, each of which comprises at least three associated genes, and the autopsy and clinical history may give little or no clue where to start. One of the more common presentations is sudden death at night, for example, and there is frequently no previous medical history. If the disease-causing mutation is in a gene that is not part of a speculative cardiac disease screen, then the molecular diagnostic test will return with an uninformative result. Furthermore, even when the diagnosis is known, many are genotype negative; in long QT syndrome, for example, this is about 20% and in HCM, 40% [5].

The development of massively parallel sequencing (MPS) has made it possible to produce millions of sequence reads simultaneously and is potentially an ideal approach to screen for mutations in genes that are associated with sudden cardiac death. Several larger diagnostic laboratories already offer the screening of large gene panels for cardiac/SCD patients [6,7]; however, smaller laboratories are challenged in offering a similar service, due to limitations in accessing relevant technology and fewer referrals compared to larger laboratories.

2. Massively Parallel Sequencing

Sanger-based sequencing chemistry is part of the first generation of sequencing technology, which also includes the sequence by cleavage method developed by Maxam *et al*. [8]. The capillary-based, semi-automated implementation of Sanger-based chemistry [9,10] has become the gold standard sequencing technique involving chain-termination. With each subsequent reaction cycle, the template is terminated by the incorporation of fluorescently-labelled dideoxynucleotides (ddNTPs), which

corresponds to the identity of the nucleotide at the terminal position of that particular fragment (Figure 1). The sequence of the template is determined by high-resolution electrophoretic separation of end-labelled extension products in a capillary-based polymer [11]. The fluorescent labels are excited by lasers, and this is coupled to a four-colour detection system, which provides the readout that is represented in a Sanger sequencing trace, and the software translates these traces into a DNA sequence (Figure 1) [11].

Figure 1. Sanger-based sequencing chemistry. In each sequencing reaction cycle, the amplified product is terminated by the incorporation of fluorescently-labelled dideoxynucleotides (represented by yellow, blue, red and green circles), which generates a ladder of differently-sized products. These products are subjected to high-resolution electrophoretic separation, and the four-colour detection system translates it to sequencing traces. The image is reproduced with permission from [11].
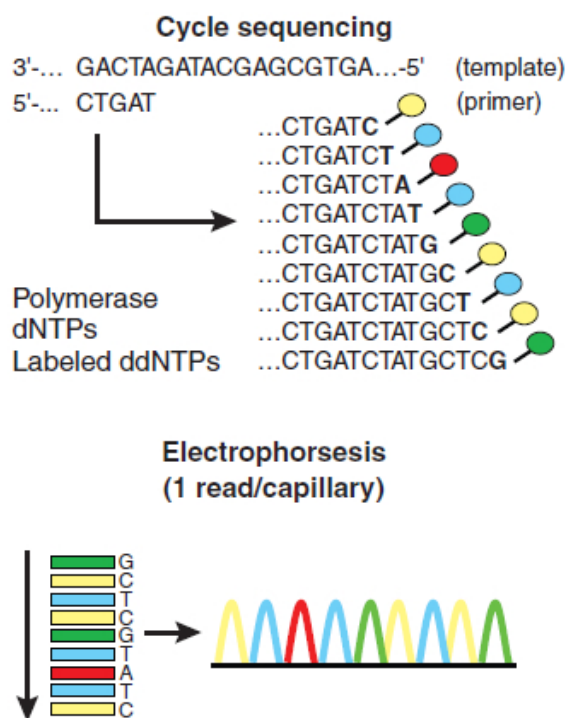

In contrast, the basic workflow of MPS consists of four stages: target DNA enrichment, library preparation, MPS run and downstream data analysis (Figure 2). The target DNA enrichment step can be divided into two categories: gene list-focused and whole exome. In both categories, the aim is to capture from the whole genome that which needs to be sequenced, and so, it is distinct from whole genome sequencing (WGS) in which target DNA enrichment is not needed [12]. Of relevance here is that mutations that cause disease generally lie in the coding regions of genes; therefore, the non-coding sequences gained from WGS may not be relevant for molecular diagnostic screening [13].

Enriching for only those genes that comprise a relevant list for diseases or syndromes is a preferred method for molecular diagnostic screening. The advantage of this approach is that it is tailored to the clinical referral, and also, it allows for low-cost sequencing per patient and avoids discovering mutations in genes for which consent has not been obtained. In contrast, whole exome sequencing (WES) involves enriching for the entire coding region of an individual's genome and is a good method

for discovering new genes that are associated with diseases [13], as well as offering diagnostic outcomes, although at a greater cost compared to a gene list-focused enrichment.

Figure 2. Basic workflow of massively parallel sequencing (MPS).
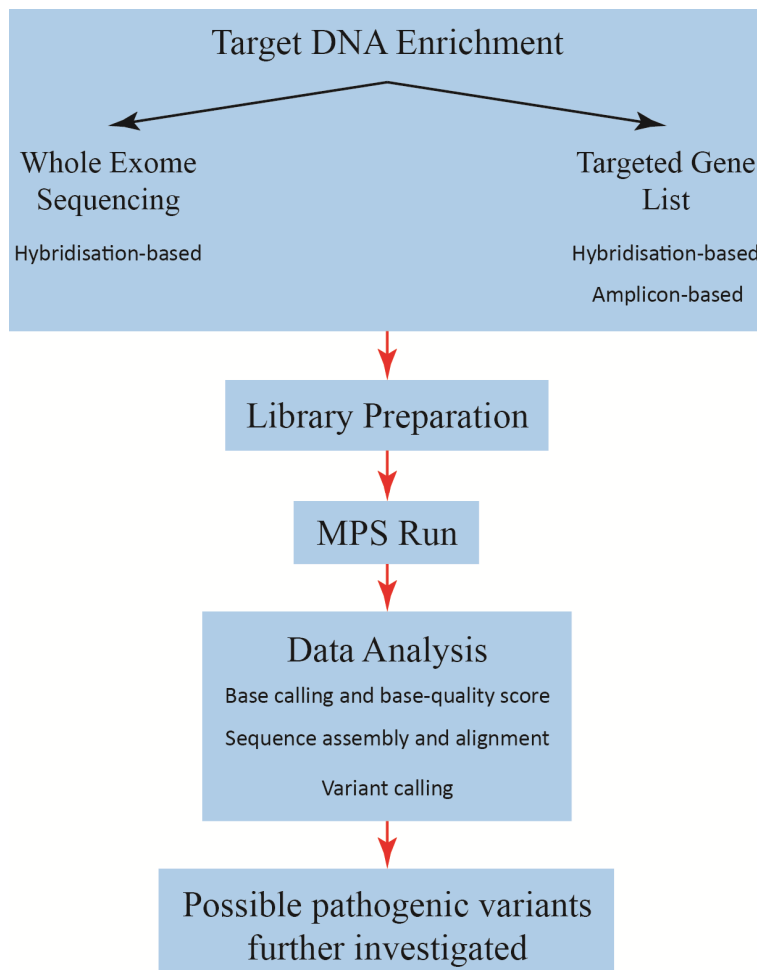
2.1. Target Enrichment Methods

Target enrichment methods can be divided into two main categories: amplicon/multiplex PCR-based and hybridisation capture-based. There are several different methods for these two different systems; however, the ultimate goal is the same: to capture relevant genomic DNA for subsequent MPS. The decision regarding which enrichment method to use depends on many factors: the MPS platform, the amount of input DNA, the clinically agreed turnaround time and the cost (largely labour) of the assay. The basic workflow of these methods is shown in Figure 3.

Figure 3. The basic workflow of amplicon-based and capture/hybridisation-based methods. The blue lines in the capture/hybridisation-based method section represent target-specific probes; the orange/yellow lines represent the adaptors. The differently-coloured lines in the amplicon-based method section represent the different amplicons being amplified.
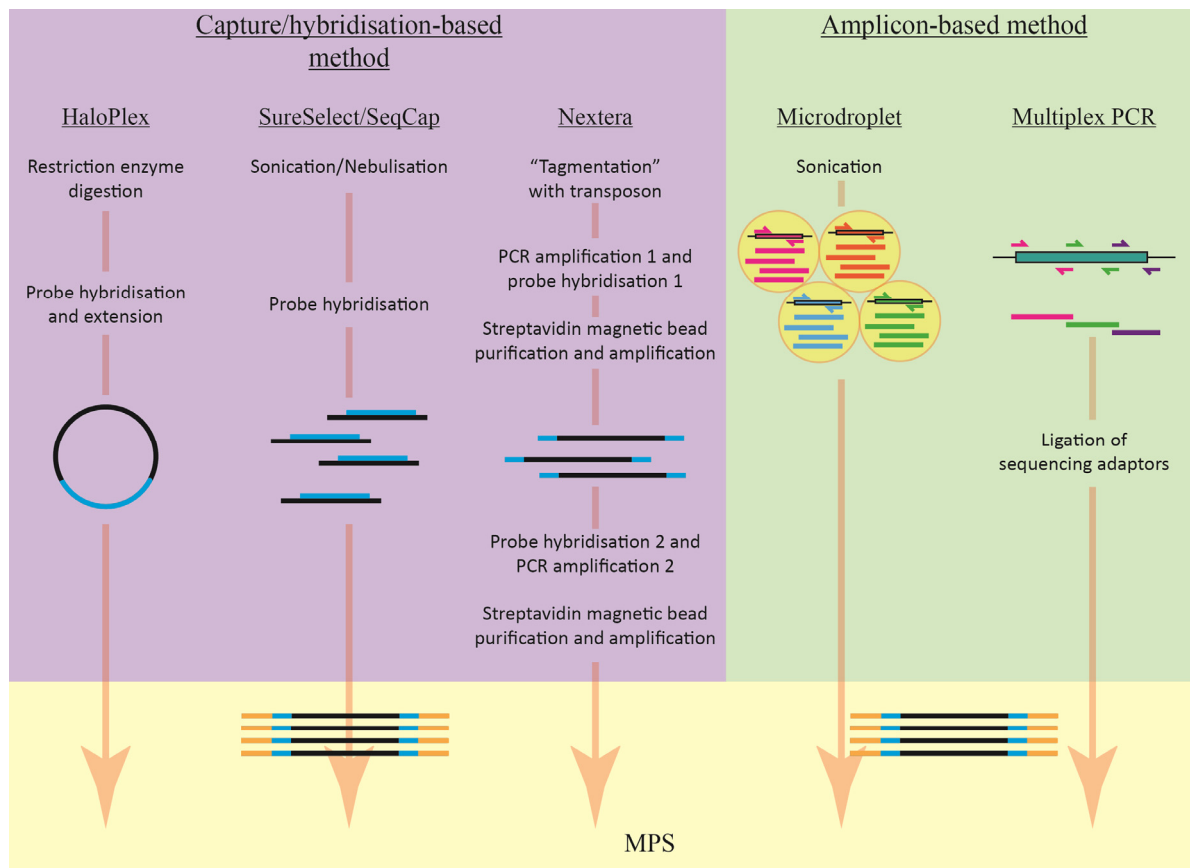
2.1.1. Amplicon-Based Enrichment

PCR is a standard technique in the diagnostic environment, and it can easily be used to enrich for desired genes for subsequent MPS. An advantage of the enrichment method over the capture/hybridization-based method, discussed later, is that it is more specific and avoids/minimizes the amplification of pseudogenes, as primers can be designed only to bind to regions of interest. The amplicon-based method begins with the amplification of the regions of interest using sequence-specific primers. Following amplification, the amplicons are pooled to form a library for MPS.

There are several commercially available amplicon-based kits, including the Ion AmpliSeq (Life Technology, Carlsbad, CA, USA), TruSeq (Illumina, San Diego, CA, USA), Microdroplet PCR (RainDance Technology, Billerica, MA, USA) and Access Array (Fluidigm, San Francisco, CA, USA). All kits offer custom design options and can be further optimized for processing formalin-fixed paraffin-embedded (FFPE) samples. Both the Ion AmpliSeq and TruSeq systems offer pre-designed, off-the-shelf products. A summary of these kits is shown in Table 1.

Table 1. Commercially available custom enrichment systems.

Enrichment system	Company	Amplicon or hybridisation	Target size	# Amplicons	Input DNA
Ion AmpliSeq DNA Custom Kit	Life Technologies	Amplicon	5 Mb	12–6,144	10 ng per pool
TruSeq Custom Amplicon	Illumina	Amplicon	4–650 Kb	16–1,536	50 ng
Microdroplet PCR Custom gene panel	RainDance	Amplicon		20,000	250 ng
Access Array 48.48	Fluidigm	Amplicon		48–480	50 ng
SeqCap EZ Choice Library	NimbleGen Roche	Hybridisation	7–50 Mb	N/A	500 ng
SureSelect Target Enrichment Kit	Agilent Technologies	Hybridisation	200 kb–24 Mb	N/A	500 ng–3 μg
HaloPlex Target Enrichment Kit	Agilent Technologies	Hybridisation	1 kb–5 Mb	N/A	200 ng–250 ng
Nextera Rapid Capture Custom Enrichment Kit	Illumina	Hybridisation	500 kb–15 Mb	N/A	50 ng

The Ion AmpliSeq System (Life Technology) uses a proprietary ultra-high multiplex PCR technology to amplify up to 5-Mb regions of interest that can be pooled for MPS in a single tube. For the custom designed kits, the primers are separated into two pools to minimize non-specific primer-primer interactions. Only 10 ng of DNA per pool of primers is required for enrichment, and up to 6.144 primers can be designed per pool [14]. The genomic DNA is subjected to two multiplex-PCR amplifications: the first round involves amplification with sequence-specific primers, and the second round involves attaching adapters needed for subsequent MPS. The Ion AmpliSeq off-the-shelf products include two cancer panels (Ion AmpliSeq Cancer Hotspot Panel v2 and Ion AmpliSeq Comprehensive Cancer Panel) and an Ion AmpliSeq Inherited Disease Panel that comprises approximately 300 genes associated with 700 inherited diseases. Only the Ion AmpliSeq Cancer Hotspot Panel v2 has been clinically tested for diagnostic use [15,16]. The Ion AmpliSeq Cancer Hotspot Panel is designed to screen for hot spots in 50 cancer-related genes.

The TruSeq Amplicon System (Illumina) uses oligonucleotide probes that flank the region of interest and a proprietary extension-ligation step to hybridise the probes to the region of interest. This is subsequently followed by PCR amplification. Fifty nanograms of input DNA are required for enrichment, and up to 1536 amplicons can be amplified in one multiplex-PCR reaction [17]. Illumina also offers some off-the-shelf products, which includes the TruSeq Amplicon-Cancer Panel, which screens 48 cancer-related genes; this approach is optimized for the Illumina MPS platforms. The TruSeq Custom Enrichment System has been used in proof-of-principle screens for diagnostic analysis of Fanconi anaemia [18].

Microdroplet PCR (RainDance Technologies Inc., Billerica, MA, USA) uses picolitre-sized droplets to partition genomic DNA samples into individual reaction vessels, which allows over one million unique PCRs to be performed per sample [19]. The primer pairs that target the regions of interest are individually encapsulated, and these droplets are mixed together in equal portions to ensure even representation for library construction [20]. Genomic DNA is fragmented, biotinylated, purified and then mixed with PCR components. This mixture is then made into droplets, and one template droplet is merged with one primer droplet before emulsion PCR occurs [20]. Once the amplification is finished, the oil-water emulsion is broken to release the amplicons for purification and then MPS. This enrichment technique requires some specialized equipment and may not be suited for all laboratories. RainDance Technologies Inc. offers custom gene panels and also off-the-shelf gene panels, including

the Cancer Hotspot Panel (targets 54 cancer gene hot spots and covers 13,000 mutations), the ASDSeq Panel (targets 62 autism-associated genes) and the XSeq Panel (targets 802 genes on the X-chromosome linked to autism and other intellectual disabilities). Custom-designed assays have been used to assess proof-of-principle screens for the diagnostic analysis of mitochondrial disorders [21], monogenic diabetes and obesity [22] and congenital muscular dystrophy [23].

The Access Array system developed by Fluidigm uses a microfluidic chip (48.48 Access Array Integrated Fluidic Circuit). This allows the amplification of up to 480 unique amplicons across 48 samples in a multiplex-PCR. As little as 50 ng of input DNA is required, and primers contain sample-specific barcodes and universal adapters. Before PCR is performed, the samples and primers are combined automatically in the pre-PCR IFC Controller AX, which is then placed into the FC1 Cycler for amplification. This technology requires specialized equipment. The 48.48 Access Array has been used in proof-of-principle screens for the diagnostic analysis of nephronophthisis-associated ciliopathy [24] and familial hypercholesterolemia [25].

Despite their ease of use and shorter overall library preparation time, the amplicon-based method is more suited for screening a small number of genes across a large number of patient samples, due to the difficulties associated with primer design for multiplex purposes. As the number of primers in the reaction increases, the level of non-specific amplification caused by the interaction between the primers increases [26]. Some problems are associated with amplicon-based methods. First, the large number of PCR cycles needed to amplify the regions of interest may give rise to sequence variants due to the lack of high fidelity proof-reading by some thermophilic DNA polymerases [27]. Secondly, the addition of more genes into an established gene list requires primer redesigns in order to accommodate the new genes. Finally, the large number of PCR cycles required cannot reliably detect copy number changes [27].

2.1.2. Capture/Hybridisation-Based Enrichment

There are several capture/hybridization-based enrichment systems that are commercially available, and the workflows for many of these methods are similar. Genomic DNA is fragmented; the DNA is hybridized or "captured" with biotinylated probes that target specific regions of interest, and these regions are isolated by streptavidin bead binding.

After further clean-up, the captured products are enriched via PCR, and usually at this stage, the necessary barcodes and tags are attached. The amplicons are then pooled and subjected to MPS. An advantage of the capture/hybridisation-based method is that it can be scaled up to capture more regions of interest. There are two main methods of capture: array-based or in-solution-based. Only the in-solution-based enrichment methods will be discussed here. A summary of these kits are shown in Table 1.

The SeqCap EZ Choice Library system (NimbleGen Roche, Madison, WI, USA) is an enrichment method that tiles the region of interest with many 80–105 mer DNA probes [28]. This ensures that there is enough redundancy and uniform capture, with a capture size between 7–50 Mb. After genomic DNA (500 ng) is fragmented by nebulisation, the biotinylated DNA probes are hybridized during a 72 h incubation before the desired regions can be isolated by magnetic pull-down. The isolated products are amplified by PCR. The NimbleGen capture system also offers a pre-designed exome kit that can enrich

for untranslated regions [28]. The custom-designed SeqCap EZ Choice Library system has been used in proof-of-principle screens for diagnostic analysis of heritable disorders. These include high-throughput screens for phenylketonuria and tetrahydrobiopterin-deficient hyperphenylalaninemia [29], cystic fibrosis [30] and retinitis pigmentosa-linked genes [31].

The Agilent SureSelect Target Enrichment kits (Agilent Technologies, Santa Clara, CA, USA) is another in solution hybridization-based system that has been widely used. Unlike the NimbleGen system, the SureSelect kits uses 120 nucleotide biotinylated RNA probes instead of DNA probes, as the bond between RNA-DNA hybrids are stronger than double-stranded DNA [30]. The capture size is between 200 kb to 24 Mb, and there is an option of post-capture indexing (SureSelect XT) or pre-capture indexing (SureSelect XT2) formats. The post-capture indexing format enriches individual samples prior to pooling, while pre-capture indexing pools different DNA samples before enrichment occurs. The latter format improves processing efficiency; however, the post-capture indexing format allows greater flexibility in the number of samples that can be processed. Genomic DNA (500 ng–3 μg) is fragmented by sonication, and the hybridization period for the SureSelect kit is only 24 h. After hybridization, the regions of interest are captured by magnetic pull-down, and the products are enriched via PCR and samples pooled [32]. Agilent SureSelect Target Enrichment offers several off-the-shelf kits, including the SureSelect All Exon Kit. The custom designed SureSelect Target Enrichment kit has been used in proof-of-principle screens for the diagnostic analysis of mitochondrial diseases [33], hereditary hearing loss [34,35], familial hypercholesterolemia [36] and aortopathies [37].

Agilent Technologies has also released the HaloPlex Target Enrichment System, which uses a different chemistry compared to the SureSelect kits. Genomic DNA (200–250 ng) is fragmented using restriction enzymes and denatured. The HaloPlex probes are then added and allowed to hybridise to their respective targets during 3–16 h of incubation [38]. The HaloPlex probes are oligonucleotides designed to bind to each end of targeted DNA fragments, thereby forming circular DNA molecules. The probes are biotinylated and contain sample-specific barcode sequences. Following circularization, the regions of interest are isolated from the pool with magnetic streptavidin beads, and the circular molecules are closed by ligation. These circular molecules are subsequently amplified by PCR to enrich for the regions of interest [38]. As well as the custom design kits, HaloPlex offers two off-the-shelf research panel kits: a cancer research panel and a cardiomyopathy research panel. Four other made-to-order pre-designed kits are also available: HaloPlex Arrhythmia, HaloPlex Noonan Syndrome, HaloPlex Connective Tissue Disorder and HaloPlex X Chromosome Disorder. All of these kits are for research use only.

Illumina has also released their own hybridization-based enrichment system: the Nextera Rapid Capture Custom Enrichment Kit. The Nextera kit simultaneously fragments and tags DNA with appropriate identifiers using transposomes (transposon/transposase complexes) [39]. This "tagmentation" technology does not require mechanical shearing of genome DNA. This is followed by the first round of PCR amplification and the hybridization of biotinylated probes specific to the regions of interest. These are subsequently purified using streptavidin magnetic beads, followed by a second round of hybridization, purification, PCR amplification and PCR clean-up. The double hybridization and PCR amplification ensures the specificity of the capture system [39]. Using Nextera technology, Illumina has manufactured several pre-designed research panels: TruSight Cancer, TruSight Tumor, TruSight Cardiomyopathy, TruSight Inherited Disease and TruSight Autism.

One of the disadvantages of using the capture/hybridization-based method compared to the amplicon-based method is the likelihood of non-specific binding by the capture probes. Another disadvantage is that it is more labour-intensive than the amplicon-based method. The decision of which capture/hybridisation method to choose depends on the length of the targeted region, the amount of input DNA and the genomic architecture of the region [26]. However, unlike the amplification-based capture method, the expansion of an established gene list can be easily achieved, as redesign is not required for the old genes.

Comparisons between the Agilent SureSelect and NimbleGen SeqCap methods [40,41] show that they are comparable. Both companies have released exome-enrichment kits, and the updated target designs are based on hg19 (GRCh37), RefSeq (67.0 Mb) and CCDS (Consensus Coding Sequence project, 31.1 Mb). Despite being updated, there are still some genomic regions that are still poorly covered or not captured [40]. Sulonen *et al.* [41] found a larger percentage of high quality reads aligned to the captured regions from the NimbleGen enrichment method. The libraries prepared by the Agilent kits contained fewer duplicated reads, and the alignment to the reference library was equal to the NimbleGen kit; however, the Nimblegen kit had more high-quality reads and deeper coverage in targeted regions [41]. A study that compared three commercially available exome-enrichment kits (Agilent SureSelect, NimbleGen SeqCap and Illumina TruSeq) found that the NimbleGen platform was able to cover the largest proportion of its target regions with the least amount of sequencing [42]. However, Agilent and Illumina were able to detect a greater number of variants when more sequencing was performed. Out of the three platforms, only Illumina captured untranslated regions, as the other two platforms did not target these regions [42].

Despite the range of enrichment methods (amplicon-based and capture/hybridization-based), there remain issues regarding the even capture of targeted regions and the subsequent uneven sequence coverage [43]. This is more common in GC-rich regions or regions where the DNA structure is susceptible to DNA fragmentation [44]. These regions still require Sanger-based sequencing to fill in the gaps.

2.2. Second Generation Sequencing

Several different second generation sequencing platforms are currently available, and each one employs different sequencing chemistry to achieve its goal. Unlike Sanger-based sequencing, second generation sequencing, termed MPS, allows millions of DNA templates to be sequenced and read at the same time. These sequencers rely on polymerase-based clonal replication of single DNA molecules that are separated on a solid support matrix and cyclic sequencing chemistries [12]. Currently, there are three different commercially available MPS technologies suitable for the diagnostic environment. These are pyrosequencing (Roche 454), reversible dye terminator (Illumina/Solexa) and sequencing by ligation (Life Technology). There are extensive reviews available that address the different MPS technologies [11,45–47]; therefore, only a brief description will be given here (Table 2).

Table 2. Second generation sequencing platforms.

Platform	Amplification method	Chemistry	Read length (bp)	Throughput	Run time	Sequencing homopolymer regions	# Sequence reads/run
Roche 454-GS Junior	Emulsion PCR	Pyrosequencing	200–400	35 Mb	10 h	Prone to errors	>70,000 (amplicon sequencing)
Illumina-MiSeq	Bridge PCR	Reversible dye terminator	35–150	>120 Mb (single-end sequencing, 1× 35 bp)	4 h	More accurate	>3.4 million single-end reads
				>680 Mb (paired-end sequencing, 2× 100 bp)	19 h		>6.8 million paired-end reads
				>1 Gb (paired-end sequencing, 2× 150 bp)	27 h		
Life Technologies –IonTorrent	Emulsion PCR	Sequence-by-ligation	100–200 bp	Chip314: >10 Mb	All three chips take <2 h	More accurate	Chip314 (>1 million wells) Chip316 (>6 million wells) Chip 318 (>11 million wells); The number of reads is approximately 30%–40% of the available wells for each chip
				Chip316: >100 Mb			
				Chip318: >1 Gb			

The Roche 454 sequencing uses emulsion PCR to amplify target sequences. This involves the denaturation of target sequences, which are captured by amplification beads, and these are compartmentalized into water-in-oil microvesicles. Clonal expansion of the target sequence takes place during the emulsion PCR (Figure 4a) [48]. Once amplified, the water-in-oil microvesicles are broken, and the beads are placed in a picotitre plate and the sequencing performed by the pyrosequencing method [49]. This method uses luciferase to generate light when individual nucleotides that complement the template strand are incorporated into the nascent DNA, and the intensity of the light production is measured and translated to sequence data [49]. A major limitation of the 454 technology is its high error rate when the sequence contains homopolymers (consecutive instances of the same base, e.g., AAA or GGG) longer than 6 bps. The Roche 454 GS Junior platform is capable of generating >70,000 sequence reads per run by amplicon sequencing with a read length of >400 bp and a throughput of 35 Mb per run [43].

Figure 4. Two clonal amplification methods used in MPS. (**a**) Emulsion PCR used by Roche 454 and IonTorrent MPS platforms. Enriched DNA products with attached adaptors (yellow and turquoise adaptors flanking the DNA sequences) are combined with beads that have one of the PCR primers tethered to its surface. The PCR amplification takes place in a water-in-oil emulsion vesicle with only one template present in each compartment. The amplicons are captured on the surface of the bead. Once amplification is complete, the emulsion compartments are broken, and the amplified products will be selectively enriched; (**b**) The Illumina platform uses the bridge PCR to clonally amplify its products. Enriched DNA products with attached adaptors (yellow and turquoise adaptors flanking the DNA sequences) are placed on a chip that is densely coated with both adaptor primers tethered to a solid surface. As PCR takes place, amplicons from a given template will remain tethered close to the point of origin. When the PCR is complete, each clonal cluster contains ~1,000 copies of a single template. The image is reproduced with permission from [11].
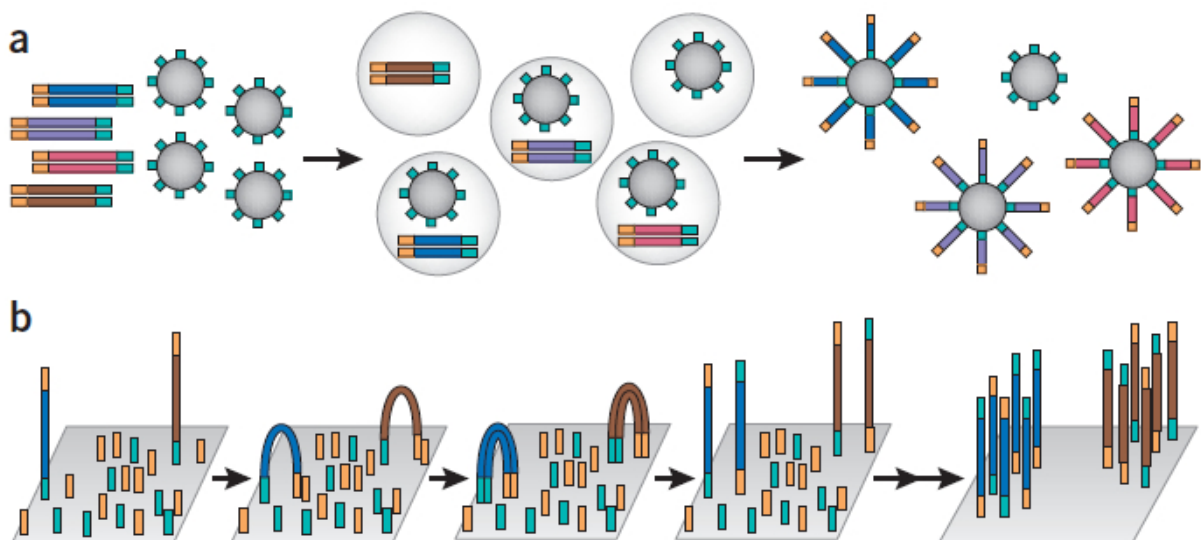
The Illumina/Solexa platform (HiSeq and MiSeq) uses cluster target sequence amplification (or bridge PCR) on a solid surface (Figure 4b). The forward and reverse PCR primers are attached to a solid surface, so that products amplified from any templates will remain immobilised and clustered to a single location on the array [50]. The Illumina platform uses reversible dye terminator sequencing-by-synthesis chemistry, which involves a single base extension with a modified DNA polymerase and a mixture of four modified nucleotides during each cycle. These nucleotides are reversible terminators, meaning that a cleavable component at the 3' hydroxyl position allows only a single-base to be incorporated in each cycle, and they are also fluorescently labelled. After a single-base extension, the fluorescent output is imaged, the cleavable component is removed and the next cycle occurs [51]. The MiSeq system can generate >3.4 million single-end reads or >6.8 million paired-end reads, which enables both ends of the DNA fragment to be sequenced [43]. There are three different throughputs, which vary from sequencing library and sequence read lengths: >120 Mb (single-end sequencing with a 35-bp read length), >680 Mb (paired-end sequencing with a 100-bp read length) and >1 Gb (paired-end sequencing with a 150-bp read length) [43].

The Life Technologies IonTorrent platform uses emulsion PCR to amplify the template and uses a sequence-by-synthesis method that is similar to the Roche 454 system. Instead of detecting light production (Roche 454), however, the IonTorrent system is an ultra-sensitive pH meter that detects the release of hydrogen ions when nucleotides are incorporated during DNA synthesis [19]. The IonTorrent platform comprises a PGM sequencer and semi-conductor sequencing chip, which is a high-density array of wells where sequencing is performed [43]. Within each well an ion-sensitive layer overlaying a proprietary ion sensor. The IonTorrent PGM sequencer provides the chip with one type of nucleotide after another [43]. The number of sequence reads depends on the number of wells per chip loaded with beads attached to DNA fragments. There are three different chips, chip314, chip316 and chip318, that contain approximately 1 million, 6 million and 11 million wells, respectively. Approximately 30%–40% of the available wells are filled in a sequence run [41]. Therefore, the minimum sequence throughput is estimated to be >10 Mb (chip314), >100 Mb (chip316) and >1 Gb (chip 318) [43].

3. Downstream Data Processing

Data analysis is a critical step of MPS, and the size of the data files makes it a challenging task. Data processing can be divided into three steps: base calling and generating the base quality score; assembly and alignment; and variant calling and annotation. Each individual platform comes with its own proprietary analysis software that calls the bases and generates the quality scores. Despite the base calling algorithms being different between platforms, the quality score system is based on a Phred score [52], which relates to the base-calling error probability [52]. The sequences may need the ends trimmed to improve the sequence quality scores. The platform-specific software is a convenient method for base-calling and generating quality scores; however, there are also other base-calling programs that use more advanced software and statistical techniques [47]. These alternative programs include the incorporation of ambiguous bases into the reads and the improved removal of poor-quality bases from read ends [53]. One such software is FastQC [54]. These features have reduced read error and improved alignment [47].

As well as the base quality scores, the overall coverage of the set target should also be considered. This will provide some information about how well and how consistent the enrichment system has performed across all targeted genes [55]. For diagnostic purposes, most regions with 30× coverage should give reliable variant calling. Regions that are poorly covered (lower than average coverage or zero coverage) should be identified so that Sanger-based sequencing of these regions could be implemented [55]. The average coverage can be calculated by multiplying the read length by the number of reads and dividing by the total length of the capture.

The second step of the data analysis is to align the sequence data to a reference library. For whole genome sequencing, there is no current reference library to align against; however, as this review is focused on the use of MPS in the diagnostic environment, the targeted genes will be present in a reference library. The alignment and assembly of MPS data is more difficult than for Sanger-based sequencing data, as the read lengths are shorter for MPS. Commercially available alignment software with proprietary algorithms can perform the alignment and assembly step in a data processing pipeline, and there are also others that are available, such as BWA, MAQ, Bowtie2 and Novoalign [56–59].

Some alignment programs are better suited for variant detection (e.g., BWA and Novoalign), while others are better at detecting indels (e.g., MAQ and Bowtie) [12]. The accuracy of MPS is accomplished by sequencing a given region multiple times, with each sequence contributing to the read depth [47].

In order for MPS data to be assembled and aligned, an adequate number of overlapping reads, or coverage, needs to be achieved [47]. Theoretically, reads are randomly distributed across the captured templates; however, in practice, the coverage across the sequenced regions is variable [56]. Therefore, it is important to ensure that adequate coverage has been achieved in all regions, as inadequate coverage can cause the failure to detect actual nucleotide variations, which would lead to false-negative results for heterozygotes [57,58]. The coverage of less than 20- to 30-fold reduces the accuracy of single nucleotide variant calls in data on the Roche 454 platform [59]. For the Illumina platforms, the coverage of less than 20- to 30-fold may be enough for certain MPS applications; however, coverage depths of 50- to 60-fold may be better to improve the alignment, assembly and accuracy [60]. Once the alignment and assembly has finished, alignment maps can be generated to visualize the sequence reads in a genome browser [12].

Variant calling is the next step, and the most commonly used programs are SAMtools, GATK Unified Genotyper and SOAPsnp [61–63]. These programs are used to detect variants in the sequenced data with respect to the reference genome. Custom-designed or commercially available programs can also be used to call variants. From this list of variants, the single nucleotide variants that are benign changes need to be identified so that disease-causing variants can be investigated. Sanger-based sequencing is usually performed to confirm the variant. There are publicly available databases that are useful for variant annotation, such as the 1,000 Genomes Project (http://www.1000genomes.org/) and the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/). Commonly used alignment and variant calling software packages are listed in Table 3.

Table 3. Computational tools for MPS data analysis. Reproduced and modified with permission from [19].

Program	Functions	URL	Reference
Bowtie2	Alignment	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	[64]
BWA	Alignment	http://bio-bwa.sourceforge.net	[65]
SOAP2	Alignment	http://soap.genomics.org/cn/soupaligner.html	[66]
MAQ	Alignment and assembly	http://maq.sourceforge.net	[67]
Novoalign	Alignment	http://www.novocraft.com	[68]
SAMtools	Variant calling	http://samtools.sourceforge.net	[62]
VARiD	Variant calling	http://compbio.cs.utoronto.ca/varid	[69]
VarScan2	Variant calling	http://varscan.sourceforge.net	[70]
GATK Unified Genotyper	Variant calling	http://www.broadinstitute.org/gatk/index.php	[61]
SOAPsnp	Variant calling	http://soap.genomics.org.cn/soapsnp.html	[63]

Possible disease-causing variants need to be classified as pathogenic, benign or of unknown clinical significance. There are many publicly or commercially available tools and databases that can be used for this, which are listed in Table 4. The software can be divided into two categories: missense tools and splice-site tools. The missense tools are used to predict whether the change in an amino acid caused by a missense mutation will affect protein function. The software can be further divided into those that use sequence and evolutionary conservation-based methods, protein sequence and structure-based methods or supervised learning methods (Table 4). The splice-site tools predict whether a variant will affect the splicing of the transcript if the variant is close to known splice sites or might activate cryptic splice sites, thereby causing incorrect splicing to occur.

As already stated, missense mutations cause a change in amino acids, and these comprise approximately 5% of the Human Gene Mutation Database (HGMD); about 55% of these are "disease-associated" [71], which shows the importance of missense mutations in affecting the normal function of proteins [72]. The sequence and evolutionary conservation-based methods of protein analysis are based on the evolutionary conservation of the amino acids within protein families, such that highly conserved amino acids are intolerant to substitution, and positions with a lower degree of conservation are more tolerant to change [72]. These methods use multiple sequence alignments to determine highly conserved regions; however, they are highly sensitive to the multiple sequence alignment that the user provides [73]. The protein sequence and structure-based methods consider the protein structure and whether a missense mutation will disrupt the overall integrity of the protein. However, some of the resources for these methods require a good understanding of structural information in order to interpret the results [73]. In contrast, supervised-learning methods use algorithms, such as neural networks, that can be "trained" to distinguish pathogenic variants from non-pathogenic variants [73]. This requires a large collection of both pathogenic and non-pathogenic variants to "train" the program. Once this has been achieved, then a query variant could be analysed using the parameters it has "learned" to determine whether the variant is pathogenic or not [73]. The requirement for large datasets of variants poses a problem.

As these are predictive tools, the results should be carefully considered. Each method has its advantages and disadvantages, so it is good practice to use a variety of resources from the different methods to assess a candidate variant. However, this last step is time-consuming, as the same variant needs to be assessed by different programs.

Downstream MPS data analysis is a challenge due to the large volume of data that is produced, which must be managed and stored, so this needs to be considered when using this technique in a diagnostic environment. Despite there being many commercially and publicly available software, all tools have their limitations, due to the different data provided by the different MPS platforms, the different reference sequences used for alignments and the different databases used for variant annotation and filtering [19]. Therefore, the use of existing tools for diagnostic screening should be evaluated for the specific needs. The minimum depth of coverage required for assays and the thresholds for data quality need to be determined when validating the assays.

Table 4. Resources for predicting mutant protein function and variant interpretation (as listed on http://www.ngrl.org.uk/Manchester/projects/bioinformatic-tools).

Type		Program	URL
Missense tools	Sequence and evolutionary conservation-based methods	SIFT [72]	http://sift.jcvi.org
		Align-GVGD [74]	http://agvgd.iarc.fr/index.php
		Mutation assessor [75]	http://mutationassessor.org/v1
		PANTHER [76]	http://www.pantherdb.org/tools/csnpScoreForm.jsp
		MAPP [77]	http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html
		PROVEAN [78]	http://provean.jcvi.org/index.php
	Protein sequence and structure-based methods	Polyphen-2 [79]	http://genetics.bwh.harvard.edu/pph2
		LS-SNP/PDB [80]	http://ls-snp.icm.jhu.edu/ls-snp-pdb/
		SNPeffect [81]	http://snpeffect.switchlab.org/
	Protein stability-based methods	MUpro [82] (http://www.ics.uci.edu/~baldig/mutation.html)	
		FoldX [83] (http://foldx.crg.es/)	
		PoPMuSiC [84] (http://babylone.ulb.ac.be/popmusic/)	
		SDM [85] (http://www-cryst.bioc.cam.ac.uk/~sdm/sdm.php)	
	Supervised-learning methods	PMUT [86]	http://mmb2.pcb.ub.es:8080/PMut/
		SNAP [87]	http://www.rostlab.org/services/SNAP/
		PhD-SNP [88]	http://gpcr2.biocomp.unibo.it/cgi/predictors/PhD-SNP/PhD-SNP.cgi
		SNPs&Go [89]	http://snps.uib.es/snps-and-go/
		Parepro [90]	http://www.mobioinfor.cn/parepro/contact.htm
		CanPredict [91]	http://www.cgl.ucsf.edu/Research/genentech/canpredict/index.html
		nsSNPAnalyzer [92]	http://snpanalyzer.uthsc.edu/
		MutPred [93]	http://mutpred.mutdb.org/
		Hansa [94]	http://hansa.cdfd.org.in:8080/
		MutationTaster [95]	http://www.mutationtaster.org/
		I-Mutant2.0 [96]	http://folding.biofold.org/i-mutant/i-mutant2.0.html
Splice-site tools		GeneSplicer [97]	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml
		Human Splice Finder [98]	http://www.umd.be/HSF/
		MaxEntScan [99]	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
		NetGene2 [100]	http://www.cbs.dtu.dk/services/NetGene2/
		NNSplice [101]	http://www.fruitfly.org/seq_tools/splice.html
		ESEFinder [102]	http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home
		ASSP [103]	http://wangcomputing.com/assp/index.html
Databases		HGVS	http://www.hgvs.org/mutnomen
		RefSeq	http://ncbi.nlm.nih.gov/RefSeq
		dbSNP [104]	http://www.ncbi.nlm.nih/gov/projects/SNP
		HGMD [105]	http://www.biobase-international.com/product/hgmd

Resources in blue font are used for publication purposes, and those that are bold and blue are used for publication purposes and routinely used in our diagnostic laboratory.

False Positive Variants

False positives can be a major problem when using MPS, and this can be due to shorter reads not aligning well or properly and sequencing errors [12]. Shorter reads can align to paralogous or repetitive regions. The human genome assembly is also not perfect, and there are gaps and misassemblies that can cause the MPS reads to misalign [106,107]. Longer read-lengths and using paired-end reads can address this problem [12]. Sequencing errors causing false positive single nucleotide variants can be caused by the wrong nucleotides being incorporated during PCR amplifications or sequence detection errors. The former error can be due to duplicate reads; however, duplicate reads are routinely removed at the analysis step [12].

4. Challenges and Limitations of MPS

As mentioned previously, a large volume of data is produced from a single MPS run. Many research facilities that use MPS have access to, or employ, their own bioinformatic specialists, who handle the data analysis and data storage. Establishing a semi-automated data analysis pipeline should make data analysis more efficient for the diagnostic environment. Many of the programs mentioned earlier that are used in downstream data analysis are online, and it is time consuming to go through each program individually to investigate sequence variants that may be novel (not lodged in any mutation database or described in the literature). A program that can access all of these programs simultaneously should reduce some of the data analysis time.

As MPS is still a relatively new technique, the use of this technology in a diagnostic environment is still a challenge and requires stringent validation. The problem of uneven coverage and uneven sequencing depth remains a concern for implementing MPS in the diagnostic field. While Sanger-based sequencing can provide "patches" if a small number of regions are poorly covered by MPS, this should be viewed as the less favourable option compared to developing improvements in sequence capture.

5. Validation of MPS Method for Diagnostic Use

Before the MPS method can be used as a test in the diagnostic environment, stringent validation tests must be satisfied. The evaluation of the analytical performance of an MPS run should include the depth of coverage, the uniformity of the distribution of read coverage, insufficiently covered regions/bases, the quality of base calls and the ability to detect large deletion events [44].

The validation process should include an evaluation of the platform that is used and the downstream data analysis [108]. Platform validation establishes the performance of the sequencing platform and the different variants that can be detected by the assay [109]. Downstream data analysis validation establishes the software parameters required to accurately read sequence data and to detect the variants [109]. These validation steps should establish the sensitivity, specificity and limitations of a particular assay.

6. MPS for Sudden Cardiac Death Screening

We have identified 81 genes implicated in inherited cardiac disorders linked to sudden cardiac death that could comprise an MPS panel for molecular autopsy in such cases. The analysis of the variants detected in each gene would require an appropriate automated filtering process to distil the number of variants down to reasonable candidates that would at least require confirmation by Sanger-based sequencing. This task is time consuming, so a more limited gene list may provide a way forward, where only those genes that are likely to carry a sequence-detectable mutation could be screened. Our own studies have suggested that a list of 23 genes could comprise a clinically relevant panel to screen for mutations implicated in six heritable cardiac disorders associated with sudden cardiac death (Table 5). The inclusion criterion for a gene in this panel (Figure 5) is if there is mutation incidence of at least 1%. While AVRC (arrhythmogenic right ventricular cardiomyopathy) was previously mentioned as one of the causes of SCD, there is not enough evidence showing that there is a high incidence of these genes being the cause of SCD; therefore, it has not been included in the 23 genes screen. If a patient sample still remains genotype-negative after screening this gene list, then screening the remaining cardiac-associated genes could be the next step (tier one testing; Figure 5). This strategy would allow a small diagnostic laboratory to offer MPS services to its local cardiology community, while making the MPS workload and data analysis manageable. Depending on what enrichment system is being implemented, copy number variants (CNV, deletions and duplication) may not be detected. However, as enrichment systems improve, CNVs should be detectable in patient samples without the use of a dedicated microarray assay to complement a sequence-based approach.

For a larger laboratory, patients who remain genotype-negative after screening with the full disease-specific gene list could proceed to WES (tier two testing; Figure 5). This would be a gene discovery task, and a larger laboratory would have more resources that could be dedicated to this. The necessary caveat to this approach is that improvements in capture technology, data analysis and sequencing costs may make the tier one type of screens redundant in that all samples could be processed following a generic WES approach with downstream bioinformatic filtering, allowing targeted genes to be analysed (Figure 6). Currently, there are two studies that have used WES to determine the gene mutation responsible for the cause of SCD [109,110]. After WES enrichment, both studies filtered the data to only show unique variants that are found in only cardiac-related genes. The process involved setting up three filters after an annotated list of all possible single nucleotide variants (SNVs) and indels is generated. The first filter excludes all non-coding and synonymous variants, which are DNA nucleotide changes that do not alter the amino acid sequence of the protein (Figure 6). The second filter is the gene-specific filter. In the case of SCD screening in our own laboratory, non-synonymous variants that are found in genes listed in Table 5 would be included. The third filter involves sorting through the gene-specific variants to find those that may be pathogenic. If a variant is not present in a large panel of ethnically-matched controls and three of the publicly available exome databases, then it can be considered possibly pathogenic [111]. These databases include the 1000 Genomes Project [112], the National Heart, Lung and Blood Institute Grand Opportunity Exome Sequencing project (NHLBI GO Exome Project) [113] and the 12,000-gene Exome Chip Design [114]. Using this approach, all patient samples can be processed through the same pipeline with a different set of filters used for different disorders.

Figure 5. Diagrammatic representation of the different "tiers" used for molecular diagnostic screening. Tier 1 testing can be conducted by a small diagnostic laboratory, where patients are first screened using the incidence gene list. If a patient still remains genotype-negative, then the patient can be screened using the disease gene list, where all known disease associated genes are screened. Tier 2 testing is better suited for a larger diagnostic laboratory, where patient samples are first screened using the full disease gene list. If a patient still remains genotype-negative, then the patient can have their whole exome sequenced. This will be a gene discovery task. The green background represents the number of genes that will be screened in each of the three tests with the incidence gene list screening the smallest proportion of genes and whole exome sequencing screening all of the genes.
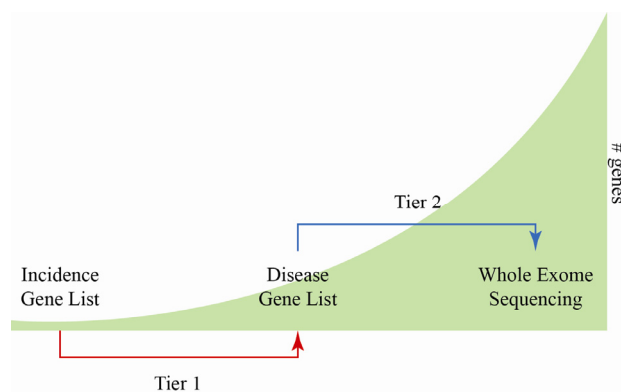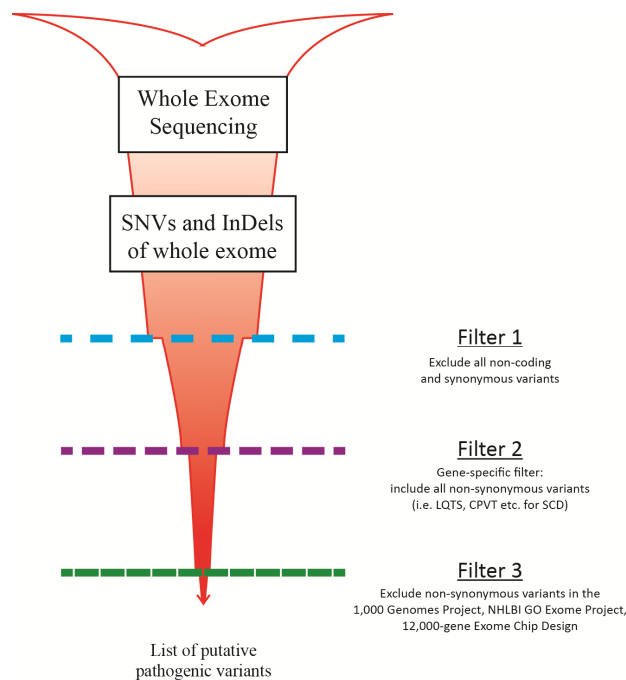

Figure 6. General overview of the three different filters used for the whole exome sequencing approach for sudden cardiac death variant screening.

MPS is a new technology that will make diagnostic screening increasingly more efficient and cost effective. Critically, we consider that improvements in data analysis will be essential in order to facilitate the full implementation of MPS in the diagnostic environment.

Table 5. The list of high incidence genes associated with five heritable cardiac disorders associated with sudden death.

Gene	Description	HCM	DCM	BrS	LQT	SQT	CPVT
BAG3	BAG family molecular chaperone regulator 3		2%–4% [115]				
CACNA1C	Voltage-dependent L-type calcium channel, α1c subunit			6%–7% [116]	Rare [117]	Limited data [118]	
CACNB2	Voltage-dependent L-type calcium channel, β2 subunit			4%–5% [116]			
CASQ2	Calsequestrin-2 precursor						1%–2% [119]
GLA	α-galactosidase A precursor	0.5%–3% [120,121]					
KCNA5	Potassium voltage-gated channel subfamily A, member 5						
KCNE1	Potassium voltage-gated channel subfamily E, member 1				Rare [117]		
KCNE2	Potassium voltage-gated channel subfamily E, member 2				Rare [117]		
KCNH2	Potassium voltage-gated channel subfamily H, member 2				25%–30% [117]	Limited data [118]	
KCNQ1	Potassium voltage-gated channel subfamily Q, member 1				30%–35% [117,122]	Limited data [118]	
LMNA	Lamin A/C		4%–8% [115]				
MYBPC3	Myosin-binding protein C, cardiac-type	15%–30% [123,124]	2%–4% [125,126]				
MYH6	Myosin heavy-chain 6	Rare [127]	4% [126]				
MYH7	Myosin heavy-chain 7	15%–30% [123,124]	4% [125,128]				
MYL2	Myosin regulatory light chain 2	<2% [123,124]					
RBM20	RNA-binding motif protein 20		3%–6% [115]				
RYR2	Ryanodine receptor 2						50%–55% [119]

Table 5. *Cont.*

Gene	Description	HCM	DCM	BrS	LQT	SQT	CPVT
SCN1B	Sodium channel protein type 1, β subunit			1%–2% [116]			
SCN5A	Sodium channel protein type 5, α subunit		1%–2% [115]	11%–18% [116]	5%–10% [117]		
TNNI3	Troponin I type 3, cardiac type	<2% [123,124]	Rare [125,126]				
TNNT2	Troponin T type 2, cardiac type	2%–5% [123,124]	3% [125,128]				
TPM1	Tropomyosin α1	2% [123,124]	1%–2% [115,125,126]				
TTN	Titin	Rare [127]	15%–25% [115]				
SCN5A	Sodium channel protein type 5, α subunit		1%–2% [115]	11%–18% [116]	5%–10% [117]		
TNNI3	Troponin I type 3, cardiac type	<2% [123,124]	Rare [125,126]				
TNNT2	Troponin T type 2, cardiac type	2%–5% [123,124]	3% [125,128]				
TPM1	Tropomyosin α1	2% [123,124]	1%–2% [115,125,126]				
TTN	Titin	Rare [127]	15%–25% [115]				

HCM, hypertrophic cardiomyopathy; DCM, dilated cardiomyopathy; BrS, Brugada syndrome; LQT, long QT syndrome; SQT, short QT syndrome; CPVT, catecholaminergic polymorphic ventricular tachycardia.

Rare mutations are found in <1% cases. As SQTS is rare, data on its prevalence and demographics are limited [118]. Dark blue represents genes with the highest incidence rates for each disorder (for LQT, the top three most prevalent genes have been highlighted); blue represents genes with medium incidence rates for each disorder; light blue represents genes that are rare.

7. Conclusions

The implementation of MPS will make molecular diagnostic testing more efficient and cost effective; however, there are still many issues that need to be addressed before this new technique will run smoothly in a diagnostic environment. As improvements on the chemistry, computational and bioinformatic analysis behind MPS improve, the issues mentioned in this review will be resolved and the adaptation of this new technique will be beneficial to patients.

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Doolan, A.; Semsarian, C.; Langlois, N. Causes of sudden cardiac death in young Australians. *Med. J. Aust.* **2004**, *180*, 110–112.
2. Behr, E.R.; Dalageorgou, C.; Christiansen, M.; Syrris, P.; Hughes, S.; Tome Esteban, M.T.; Rowland, E.; Jeffery, S.; McKenna, W.J. Sudden arrhythmic death syndrome: Familial evaluation identifies inheritable heart disease in the majority of families. *Eur. Heart J.* **2008**, *29*, 1670–1680.
3. Skinner, J.R.; Crawford, J.; Smith, W.; Aitken, A.; Heaven, D.; Evans, C.A.; Hayes, I.; Neas, K.R.; Stables, S.; Koelmeyer, T.; *et al.* Prospective, population-based long QT molecular autopsy study of postmortem negative sudden death in 1 to 40 year olds. *Heart Rhythm* **2011**, *8*, 412–419.
4. Tan, H.L.; Hofman, N.; van Langen, I.M.; van der Wal, A.C.; Wilde, A.A.M. Heritability and diagnostic yield of cardiological and genetic examination in surviving relatives. *Circulation* **2005**, *112*, 207–213.
5. Boczek, N.J.; Best, J.M.; Tester, D.J.; Giudicessi, J.R.; Middha, S.; Evans, J.M.; Kamp, T.J.; Ackerman, M.J. Exome sequencing and systems biology converge to identify novel mutations in the L-type calcium channel, *CACNA1C*, linked to autosomal dominant long QT syndrome. *Circ. Cardiovasc Genet.* **2013**, *6*, 279–289.
6. Next Generation Sequencing-Oxford University Hospitals. Available online: http://www.ouh.nhs.uk/services/referrals/genetics/genetics-laboratories/molecular-genetics-laboratory/next-generation-sequencing.aspx (accessed on 28 January 2014).

7. VCGS-Genetic testing for genetic heart conditions: Patient information sheet. Available online: http://www.vcgs.org.au/pathology/downloads/molecular/patient%20information%20sheet%2018_3_13.pdf (accessed on 28 January 2014).

8. Maxam, A.M.; Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 560–564.

9. Hunkapiller, T.; Kaiser, R.J.; Koop, B.F.; Hood, L. Large-scale and automated DNA sequence determination. *Science* **1991**, *254*, 59–67.

10. Swerdlow, H.; We, S.L.; Harke, H.; Dovichi, N.J. Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *J. Chromatogr.* **1990**, *516*, 61–67.

11. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145.

12. Meldrum, C.; Doyle, M.A.; Tothill, R.W. Next-generation sequencing for cancer diagnostics: A practical perspective. *Clin. Biochem. Rev.* **2011**, *32*, 177–195.

13. Robinson, P.N.; Krawitz, P.; Mundlos, S. Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin. Genet.* **2011**, *80*, 127–132.

14. Life Technologies-Ion AmpliSeq Panels. Available online: http://www.lifetechnologies.com/nz/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-select-targets/ampliseq-target-selection.html?icid=ampliseq-panels (accessed on 27 January 2014).

15. Singh, R.R.; Patel, K.P.; Routbort, M.J.; Reddy, N.G.; Barkoh, B.A.; Handal, B.; Kanagal-Shamanna, R.; Greaves, W.O.; Medeiros, L.J.; Aldape, K.D.; *et al.* Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J. Mol. Diagn.* **2013**, *15*, 604–622.

16. Tsongalis, G.J.; Peterson, J.D.; de Abreu, F.B.; Tunkey, C.D.; Gallagher, T.L.; Strausbaugh, L.D.; Wells, W.A.; Amos, C.L. Routine use of the Ion Torrent AmpliSeq™ Cancer Hotspot Panel for identification of clinically actionable somatic mutations. *Clin. Chem. Lab. Med.* **2013**, *13*, 1–8.

17. Illumina-TruSeq Custom Amplicon Guide. Available online: http://supportres.illumina.com/documents/myillumina/b718c350-b3b2–4234-b71a-0b832f14cda3/truseq_custom_amplicon_libraryprep_ug_15027983_b.pdf (accessed on 30 January 2014).

18. Chandrasekharappa, S.C.; Lach, F.P.; Kimble, D.C.; Kamat, A.; Teer, J.K.; Donovan, F.X.; Flynn, E.; Sen, S.K.; Thongthip, S.; Sanborn, E.; *et al.* Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. *Blood* **2013**, *121*, e138–e148.

19. Chang, F.; Li, M.M. Clinical application of amplicon-based next-generation sequencing in cancer. *Cancer Genet.* **2013**, pii:S2210-7762(13)00142-7.

20. Tewhey, R.; Warner, J.B.; Nakano, M.; Libby, B.; Medkova, M.; David, P.H.; Kotsopoulos, S.K.; Samuels, M.L.; Hutchison, J.B.; Larson, J.W.; *et al.* Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat. Biotechnol.* **2009**, *27*, 1025–1031.

21. Dames, S.; Chou, L.-S.; Xiao, Y.; Wayman, T.; Stocks, J.; Singleton, M.; Eilbeck, K.; Mao, R. The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associtted with mitochondrial disorders. *J. Mol. Diagn.* **2013**, *15*, 526–534.

22. Bonnefond, A.; Philippe, J.; Durand, E.; Muller, J.; Saeed, S.; Arslan, M.; Martínez, R.; de Graeve, F.; Dhennin, V.; Rabearivelo, I.; *et al.* Highly sensitive diagnosis of 43 monogenic forms of diabetes or obesity, through one step PCR-based enrichment in combination with next-generation sequencing. *Diabetes Care* **2014**, *37*, 460–467.

23. Valencia, C.A.; Ankala, A.; Rhodenizer, D.; Bhide, S.; Littlejohn, M.R.; Keong, L.M.; Rutkowski, A.; Sparks, S.; Bonnemann, C.; Hegde, M. Comprehensive mutation analysis for congenital muscular dystrophy: A clinical PCR-based enrichment and next-generation sequencing panel. *PLoS One* **2013**, *8*, e53083.

24. Halbritter, J.; Diaz, K.; Chaki, M.; Porath, J.D.; Tarrier, B.; Fu, C.; Innis, J.L.; Allen, S.J.; Lyons, R.H.; Stefanidis, C.J.; *et al.* High-throughput mutation analysis in patients with a nephronophthisis-associated ciliopathy applying multiplexed barcoded array-based PCR amplification and next-generation sequencing. *J. Med. Genet.* **2012**, *49*, 756–767.

25. Hollants, S.; Redeker, E.J.W.; Matthijs, G. Microfluidic amplification as a tool for massive parallel sequencing of the familial hypercholesterolemia genes. *Clin. Chem.* **2012**, *58*, 717–724.

26. Mamanova, L.; Coffey, A.J.; Scott, C.E.; Kozarewa, I.; Turner, E.H.; Kumar, A.; Howard, E.; Shendure, J.; Turner, D.J. Target-enrichment strategies for next-generation sequencing. *Nat. Methods* **2010**, *7*, 111–118.

27. Hagemann, I.S.; Cottrell, C.E.; Lockwood, C.M. Design of targeted, capture-based, next generation sequencing tests for precision cancer therapy. *Cancer Genet.* **2013**, *206*, 420–431.

28. NimbleGen-NimbleGen SeqCap EZ Library LR User's Guide. Available online: http://www.nimblegen.com/products/lit/06560881001_SeqCapEZLibraryLR_Guide_v2p0.pdf (accessed on 28 January 2014).

29. Trujillano, D.; Perez, B.; González, J.; Tornador, C.; Navarrete, R.; Escaramis, G.; Ossowski, S.; Armengol, L.; Cornejo, V.; Desviat, L.R.; *et al.* Accurate molecular diagnosis of phenylketonuria and tetrahydrobiopterin-deficient hyperphenylalaninemias using high-throughput targeted sequencing. *Eur. J. Hum. Genet.* **2014**, *22*, 528–534.

30. Trujillano, D.; Ramos, M.D.; González, J.; Tornador, C.; Sotillo, F.; Escaramis, G.; Ossowski, S.; Armengol, L.; Casals, T.; Estivill, X. Next generation diagnostics of cystic fibrosis and *CFTR*-related disorders by targeted multiplex high-coverage resequencing of *CFTR*. *J. Med. Genet.* **2013**, *50*, 455–462.

31. Schorderet, D.F.; Iouranova, A.; Favez, T.; Tiab, L.; Escher, P. IROme, a new high-throughput molecular tool for the diagnosis of inherited retinal dystrophies. *BioMed Res. Int.* **2013**, *2013*, 198089.

32. Agilent Technologies-SureSelect Target Enrichment System for Illumina Paired-End Sequencing Library. Available online: https://www.genomics.agilent.com/files/Manual/G3360-90020_SureSelect_Indexing_1.0.pdf (accessed on 28 January 2014).

33. Falk, M.J.; Pierce, E.A.; Consugar, M.; Xie, M.H.; Guadalupe, M.; Hardy, O.; Rappaport, E.F.; Wallace, D.C.; LeProust, E.; Gai, X. Mitochondrial disease genetic diagnostics: Optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. *Discov. Med.* **2012**, *14*, 389–399.

34. Mutai, H.; Suzuki, N.; Shimizu, A.; Torii, C.; Namba, K.; Morimoto, N.; Kudoh, J.; Kaga, K.; Kosaki, K.; Matsunaga, T. Diverse spectrum of rare deafness genes underlies early-childhood hearing loss in Japanese patients: A cross-sectional, multi-center next-generation sequencing study. *Orphanet J. Rare Dis.* **2013**, *8*, 172.

35. Shearer, A.E.; DeLuca, A.P.; Hildebrand, M.S.; Taylor, K.R.; Gurrola, J.N.; Scherer, S.; Scheetz, T.E.; Smith, R.J. Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21104–21109.

36. Vandrovcova, J.; Thomas, E.R.A.; Atanur, S.S.; Norsworthy, P.J.; Neuwirth, C.; Tan, Y.; Kasperaviciute, D.; Biggs, J.; Game, L.; Mueller, M.; *et al.* The use of next-generation sequencing in clinical diagnosis of familial hypercholesterolemia. *Genet. Med.* **2013**, *15*, 948–957.

37. Wooderchak-Donahue, W.; O'Fallon, B.; Furtado, L.; Durtschi, J.; Plant, P.; Ridge, P.; Rope, A.; Yetman, A.; Bayrak-Toydemir, P. A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel- clinical diagnostics perspective. *BMC Med. Genomics* **2012**, *5*, 50.

38. Agilent Technologies-HaloPlex Target Enrichment System. Available online: http://www. chem.agilent.com/library/usermanuals/Public/G9900-90001.pdf (accessed on 28 January 2014).

39. Nextera Rapid Capture Enrichment Guide. Available online: http://supportres.illumina.com/ documents/documentation/chemistry_documentation/samplepreps_nextera/nexterarapidcapture/ nextera-rapid-capture-enrichment-guide-15037436-f.pdf (accessed on 28 January 2014).

40. Parla, J.S.; Iossifov, I.; Grabill, I.; Spector, M.S.; Kramer, M.; McCombie, W.R. A comparative analysis of exome capture. *Genome Res.* **2011**, *12*, R91.

41. Sulonen, A.M.; Ellonen, P.; Almusa, H.; Lepistö, M.; Eldfors, S.; Hannula, S.; Miettinen, T.; Tyynismaa, H.; Salo, P.; Heckman, C.; *et al.* Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.* **2011**, *12*, R94.

42. Clark, M.J.; Chen, R.; Lam, H.Y.; Karczewski, K.J.; Chen, R.; Euskirchen, G.; Butte, A.J.; Snyder, M. Performance comparison of exome DNA sequencing technologies. *Nat. Biotechnol.* **2011**, *29*, 908–914.

43. Ku, C.-S.; Wu, M.; Cooper, D.N.; Naidoo, N.; Pawitan, Y.; Pang, B.; Iacopetta, B.; Soong, R. Technological advances in DNA sequence enrichment and sequencing for germline genetic diagnosis. *Expert Rev. Mol. Diagn.* **2012**, *12*, 159–173.

44. Zhang, W.; Cui, H.; Wong, L.-J. Application of next generation sequencing to molecular diagnosis of inherited diseases. *Top. Curr. Chem.* **2014**, *336*, 19–45.

45. Hui, P. Next generation sequencing: Chemistry, technology and applications. *Top. Curr. Chem.* **2014**, *336*, 1–18.

46. Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; Law, M. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, *2012*, 251364.

47. Voelkerding, K.V.; Dames, S.A.; Durtschi, J.D. Next-generation sequencing: From basic research to diagnostic. *Clin. Chem.* **2009**, *55*, 641–658.

48. Dressman, D.; Yan, H.; Traverso, G.; Kinzler, K.W.; Volgelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8817–8822.

49. Ronaghi, M.; Karamohamed, S.; Pettersson, B.; Uhlén, M.; Nyrén, P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **1996**, *242*, 84–89.

50. Adessi, C.; Matton, G.; Ayala, G.; Turcatti, G.; Mermod, J.J.; Mayer, P.; Kawashima, E. Solid phase DNA amplification: Characterisation of primer attachment and amplifcation mechanisms. *Nucleic Acids Res.* **2000**, *28*, E87.

51. Turcatti, G.; Romieu, A.; Fedurco, M.; Tairi, A.P. A new class of cleavable fluorescent nucleotides: Synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res.* **2008**, *36*, e25.

52. Ewing, B.; Green, P. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **1998**, *8*, 186–194.

53. Rougemont, J.; Amzallag, A.; Iseli, C.; Farinelli, L.; Xenarios, I.; Naef, F. Probabilistic base calling of Solexa sequencing data. *BMC Bioinforma* **2008**, *9*, 431.

54. Andrews, C. FastQC. Available online: http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc (accessed on 30 January 2014).

55. Allcock, R.J. Production and analytic bioinformatics for next-generation DNA sequencing. *Methods Mol. Biol.* **2014**, *1168*, 17–29.

56. Sims, D.; Sudbery, I.; Ilott, N.E.; Heger, A.; Ponting, C.P. Sequencing depth and coverage: Key considerations in genomic analysis. *Nat. Rev. Genet.* **2014**, *15*, 121–132.

57. Bentley, D.R.; Balasubramanian, S.; Swerdlow, H.P.; Smith, G.P.; Milton, J.; Brown, C.G.; Hall, K.P.; Evers, D.J.; Barnes, C.L.; Bignell, H.R.; *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **2008**, *456*, 53–59.

58. Wheeler, D.A.; Srinivasan, M.; Egholm, M.; Shen, Y.; Chen, L.; McGuire, A.; He, W.; Chen, Y.J.; Makhijani, V.; Roth, G.T.; *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nat. Biotechnol.* **2008**, *452*, 872–876.

59. Brockman, W.; Alvarez, P.; Young, S.; Garber, M.; Giannoukos, G.; Lee, W.L.; Russ, C.; Lander, E.S.; Nusbaum, C.; Jaffe, D.B. Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res.* **2008**, *18*, 763–770.

60. Dohm, J.C.; Lottaz, C.; Borodina, T.; Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **2008**, *36*, e105.

61. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **2011**, *43*, 491–498.

62. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, H.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079.

63. Li, R.; Li, Y.; Fang, X.; Yang, H.; Wang, J.; Kristiansen, K.; Wang, J. SNP detection for massively parallel whole-genome resequencing. *Genome Res.* **2009**, *19*, 1124–1132.

64. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25.

65. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760.

66. Li, R.; Yu, C.; Li, Y.; Lam, T.-W.; Yiu, S.-M.; Kristiansen, K.; Wang, W. SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* **2009**, *25*, 1966–1967.

67. Li, H.; Ruan, J.; Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **2008**, *18*, 1851–1858.

68. Novocraft.com Novoalign. Available online: http://www.novocraft.com (accessed on 30 January 2014).

69. Dalca, A.V.; Rumble, S.M.; Levy, S.; Brudno, M. VARiD: A variation detection framework for color-space and letter-space platforms. *Bioinformatics* **2010**, *26*, i343–i349.

70. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **3012**, *22*, 568–576.

71. Stenson, P.D.; Mort, M.; Ball, E.V.; Howells, K.; Phillips, A.D.; Thomas, N.S.; Cooper, D.N. The Human Gene Mutation Database: 2008 update. *Genome Med.* **2009**, *1*, 13.

72. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the effects of coding non-synonymouse variants on protein function using the SIFT algorithm. *Nat. Protocols* **2009**, *4*, 1073–1081.

73. National Genetics Reference Laboratory. Available online: http://www.ngrl.org.uk/Manchester/ projects/bioinformatic-tools (accessed on 30 January 2014).

74. Tavtigian, S.V.; Deffenbaugh, A.M.; Yin, L.; Judkins, T.; Scholl, T.; Samollow, P.B.; de Silva, D.; Zharkikh, A.; Thomas, A. Comprehensive statistical study of 452 *BRCA1* missense substitutions with classification of eight recurrent substitutions as neutral. *J. Med. Genet.* **2006**, *43*, 295–305.

75. Reva, B.; Antipin, Y.; Sander, C. Predicting the functional impact of protein mutations: Applications to cancer genomics. *Nucleic Acids Res.* **2011**, *39*, e118.

76. Brunham, L.R.; Singaraja, R.R.; Pape, T.D.; Kejariwal, A.; Thomas, P.D.; Hayden, M.R. Accurate prediction of the functional significance of nucleotide polymorphisms and mutations in the *ABCA1* gene. *PLoS Genet.* **2005**, *1*, e83.

77. Stone, E.A.; Sidow, A. Physicochemical constraint violation by missense substitution mediates impairment of protein function and disease severity. *Genome Res.* **2005**, *15*, 978–986.

78. Choi, Y.; Sims, G.; Murphy, S. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **2012**, *7*, e46688.

79. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A method and server for predicting damaging missense mutations. *Nat. Methods* **2010**, *7*, 248–249.

80. Ryan, M.; Diekhans, M.; Lien, S.; Liu, Y.; Karchin, R. LS-SNP/PDB: Annotated non-synonymous SNPs mapped to protein data bank structures. *Bioinformatics* **2009**, *25*, 1431–1432.

81. De Baets, G.; van Durme, J.; Reumers, J.; Maurer-Stroh, S.; Vanhee, P.; Schymkowitz, J.; Rousseau, F. SNPeffect4.0: Online prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Res.* **2012**, *40*, D935–D939.

82. Cheng, J.; Randall, A.; Baldi, P. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins* **2005**, *62*, 1125–1132.

83. Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: An online force field. *Nucleic Acids Res.* **2005**, *33*, W382–W388.

84. Dehouck, Y.; Kwasigroch, J.M.; Gilis, D.; Rooman, M. PoPMuSiC 2.1: A web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinforma.* **2011**, *12*, 151.

85. Worth, C.L.; Preissner, R.; Blundell, T.L. SDM—A server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* **2011**, *39*, W215–W222.

86. Ferrer-Costa, C.; Orozco, M.; de la Cruz, X. Sequence-based prediction of pathological mutations. *Proteins* **2004**, *57*, 811–819.

87. Bromberg, Y.; Tachdav, G.; Rost, B. SNAP predicts effect of mutations on protein function. *Bioinformatics* **2008**, *24*, 2397–2398.

88. Capriotti, E.; Calabrese, R.; Casadio, R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* **2006**, *22*, 2729–2734.

89. Calabrese, R.; Capriotti, E.; Fariselli, P.; Martelli, P.L.; Casadio, R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* **2009**, *30*, 1237–1244.

90. Tian, J.; Wu, N.; Guo, X.; Guo, J.; Zhang, J.; Fan, Y. Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinforma.* **2007**, *8*, 450–464.

91. Kaminker, J.S.; Zhang, Y.; Waugh, A.; Haverty, P.M.; Peters, B.; Sebisanovic, D.; Stinson, J.; Forrest, W.F.; Bazan, F.; Seshagiri, S.; *et al.* Distinguishing cancer-associated missense mutations from common polymorphisms. *Cancer Res.* **2007**, *67*, 465–473.

92. Bao, L.; Cui, Y. nsSNPAnalyzer: Identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* **2005**, *33*, W480–W482.

93. Li, B.; Krishnan, V.G.; Mort, M.E.; Xin, F.; Kamati, K.K.; Cooper, D.N.; Mooney, S.D.; Radivojac, P. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* **2009**, *25*, 2744–2750.

94. Acharya, V.; Nagarajaram, H.A. Hansa: An automated method for discriminating disease and neutral human nsSNPs. *Hum. Mutat.* **2012**, *2*, 332–337.

95. Schwarz, J.M.; Rödelsperger, C.; Schuelke, M.; Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **2010**, *7*, 575–576.

96. Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **2005**, *33*, W306–W310.

97. Pertea, M.; Lin, X.; Salzberg, S.L. GeneSplicer: A new computational method for splice site prediction. *Nucleic Acids Res.* **2001**, *29*, 1185–1190.

98. Desmet, F.O.; Hamroun, D.; Lalande, M.; Collod-Beroud, G.; Claustres, M.; Beroud, C. Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **2009**, *37*, e67.

99. Yeo, G.; Burge, C.B. Maximum entropy modelling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **2004**, *11*, 377–394.

100. Hebsgaard, S.M.; Korning, P.G.; Tolstrup, N.; Engelbrecht, J.; Rouzé, P.; Brunak, S. Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. *Nucleic Acids Res.* **1996**, *24*, 3439–3452.

101. Reese, M.G.; Eeckman, F.H.; Kulp, D.; Haussler, D. Improved Splice Site Detection in Genie'. *J. Comput. Biol.* **1997**, *4*, 311–323.

102. Cartegni, L.; Wang, J.; Zhu, Z.; Zhang, M.Q.; Krainer, A.R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **2003**, *31*, 3568–3571.

103. Wang, M.; Marín, A. Characterization and prediction of alternative splice sites. *Gene* **2006**, *366*, 219–227.

104. Coordinators, N.R. Database resources of the national Center for Biotechnology Information. *Nucleic Acids Res.* **2007**, *35*, D5–D12.

105. Cooper, D.N.; Stenson, P.D.; Chuzhanova, N.A. The Human Gene Mutation Database (HGMD) and its exploitation in the study of mutational mechanisms. *Curr. Protocols Bioinforma* **2006**, doi:10.1002/0471250953.bi0113s12.

106. Church, D.M.; Schneider, V.A.; Graves, T.; Auger, K.; Cunningham, F.; Bouk, N.; Chen, H.C.; Agarwala, R.; McLaren, W.M.; Ritchie, G.R.; *et al.* Modernizing reference genome assemblies. *PLoS Biol.* **2011**, *9*, e1001091.

107. Kidd, J.M.; Sampas, N.; Antonacci, F.; Graves, T.; Fulton, R.; Hayden, H.S.; Alkan, C.; Malig, M.; Ventura, M.; Giannuzzi, G.; *et al.* Characterization of missing human genome sequences and copy-number polymorphic insertions. *Nat. Methods* **2010**, *7*, 365–371.

108. Gargis, A.S.; Kalman, L.; Berry, M.W.; Bick, D.P.; Dimmock, D.P.; Hambuch, T.; Lu, F.; Lyon, E.; Voelkerding, K.V.; Zehnbauer, B.A.; *et al.* Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat. Biotechnol.* **2012**, *30*, 1033–1036.

109. Wong, L.-J. Challenges of bridging next generation sequencing technologies to clinical molecular diagnostics laboratories. *Neurotherapeutics* **2013**, *10*, 262–272.

110. Bagnall, R.D.; Das, K.J.; Duflou, J.; Semsarian, C. Exome analysis-based molecular autopsy in cases of sudden unexplained death in the young. *Heart Rhythm* **2014**, *11*, 655–662.

111. Loporcaro, C.G.; Tester, D.J.; Maleszewski, J.J.; Kruisselbrink, T.; Ackerman, M.J. Confirmation of cause and manner of death via a comprehensive cardiac autopsy including whole exome next-generation sequencing. *Arch. Pathol. Lab. Med.* **2013**, doi:10.5858/arpa.2013-0479-SA.

112. Clarke, L.; Zheng-Bradley, X.; Smith, R.J.; Kulesha, E.; Xiao, C.; Toneva, I.; Vaughan, B.; Preuss, D.; Leinonen, R.; Shumway, M.; *et al.* The 1000 Genomes Project: Data management and community access. *Nat. Methods* **2012**, *9*, 459–462.

113. Exome Variant Server, NHLBI Exome Sequencing Project (ESP). Available online: http://evs.gs.washington.edu/EVS/ (accessed on 30 January 2014).

114. Exome Chip Design. Available online: http://genome.sph.umich.edu/wiki/Exome_Chip_Design (accessed on 30 January 2014).

115. Lakdawala, N.K.; Winterfield, J.R.; Funke, B.H. Arrhythmogenic disorders of genetic origin. *Circ. Arrhythmia Electrophysiol.* **2013**, *6*, 228–237.

116. Berne, P. Brugada syndrome 2012. *Circulation* **2012**, *76*, 1563–1571.

117. Giudicessi, J.R.; Ackerman, M.J. Genotype- and phenotype-guided management of congenital long QT syndrome. *Curr. Probl. Cardiol.* **2013**, *38*, 417–455.

118. Perrin, M.J.; Gollob, M.H. Genetics of cardiac electrical disease. *Can. J. Cardiol.* **2013**, *29*, 89–99.

119. Napolitano, C.; Priori, S.G.; Bloise, R. Catecholaminergic polymorphic ventricular tachycardia. Available online: http://www.ncbi.nlm.nih.gov/books/NBK1289/ (accessed on 28 January 2014).

120. Elliott, P.; Baker, R.; Pasquale, F.; Quarta, G.; Ebrahim, H.; Mehta, A.B.; Hughes, D.A.; ACES study group. Prevalence of Anderson-Fabry disease in patients with hypertrophic cardiomyopathy: The European Anderson-Fabry Disease survey. *Heart* **2011**, *97*, 1957–1960.

121. Havndrup, O.; Christiansen, M.; Stoevring, B.; Jensen, M.; Hoffman-Bang, J.; Andersen, P.S.; Hasholt, L.; Nørremølle, A.; Feldt-Rasmussen, U.; Køber, L.; *et al.* Fabry disease mimicking hypertrophic cardiomyopathy: Genetic screening needed for establishing the diagnosis in women. *Eur. Heart J.* **2010**, *12*, 535–540.

122. Giudicessi, J.R.; Ackerman, M.J. Prevalence and potential genetic determinants of sensorineural deafness in *KCNQ1* homozygosity and compound heterozygosity. *Circ. Cardiovasc. Genet.* **2013**, *6*, 193–200.

123. Keren, A.; Syrris, P.; McKenna, W.J. Hypertrophic cardiomyopathy: The genetic determinants of clinical disease expression. *Nat. Clin. Pract. Cardiovasc. Med.* **2008**, *5*, 158–168.

124. Van Driest, S.L.; Ommen, S.R.; Tajik, A.J.; Gersh, B.J.; Ackerman, M.J. Sarcomeric genotyping in hypertrophic cardiomyopathy. *Mayo Clin. Proc.* **2005**, *80*, 463–469.

125. Hershberger, R.E.; Morales, A.; Siegfried, J.D. Clinical and genetic issues in dilated cardiomyopathy: A review for genetics professionals. *Genet. Med.* **2010**, *12*, 655–667.

126. Hershberger, R.E.; Norton, N.; Morales, A.; Li, D.; Siegfried, J.D.; Gonzalez-Quintana, J. Coding sequence rare variants identified in *MYBPC3*, *MYH6*, *TPM1*, *TNNC1*, and *TNNI3* from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ. Cardiovasc. Gene.* **2010**, *3*, 155–161.

127. Tian, T.; Liu, Y.; Zhou, X.; Song, L. Progress in the molecular genetics of hypertrophic cardiomyopathy: A mini-review. *Gerontology* **2012**, *59*, 199–205.

128. Møller, D.V.; Andersen, P.S.; Hedley, P.; Ersbøll, M.K.; Bundgaard, H.; Moolman-Smook, J.; Christiansen, M.; Køber, L. The role of sarcomere gene mutations in patients with idiopathic dilated cardiomyopathy. *Eur. J. Hum. Gene.* **2009**, *17*, 1241–1249.