The impact of next generation sequencing technologies on haematological research - A review

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Review article

The impact of next generation sequencing technologies on haematological research — A review

Jessica S. Black [a], Manuel Salto-Tellez [a, b], Ken I. Mills [a], Mark A. Catherwood [a, c, *]

[a] Centre for Cancer Research and Cell Biology (CCRCB), Queen's University Belfast, UK
[b] Tissue Pathology Department, Belfast Health and Social Care Trust, UK
[c] Haematology Department, Belfast Health and Social Care Trust, UK

ARTICLE INFO

ABSTRACT

Next-generation sequencing (NGS) technologies have begun to revolutionize the field of haematological malignancies through the assessment of a patient's genetic makeup with a minimal cost. Significant discoveries have already provided a unique insight into disease initiation, risk stratification and therapeutic intervention. Sequencing analysis will likely form part of the routine diagnostic testing in the future. However, a number of important issues need to be addressed for that to become a reality with regard to result interpretation, laboratory workflow, data storage and ethical issues. In this review we summarize the contribution that NGS has already made to the field of haematological malignances. Finally, we discuss the challenges that NGS technologies will bring in relation to data storage, ethical and legal issues and laboratory validation. Despite these challenges, we predict that high-throughput DNA sequencing will redefine haematological malignancies based on individualized genomic analysis.

1. Introduction

The completion of the human genome sequence project heralded a new era of medical and genetic research. With the publication of the draft genome in 2001 and the completion of the euchromatic sequence in 2004, the true potential of genetic research was realised [1—3]. Clinicians began to see personalised medicine as the future of hospitals and researchers seen the potential for discovery and better understanding of many diseases [4]. The technique which made this new era a reality was Sanger Sequencing, developed in the late 70s by Frederick Sanger and his team [5]. This technique was the crux of the Human Genome Project, securing Sanger sequencing's place in laboratories across the globe as the "Gold Standard" method for validation, research, and diagnostics for the past three decades.

Research began in earnest into the Human Genome for discovery of new biomarkers, pathways of disease and inherited genes. However, discoveries did not come as fast as many presumed. The main limiting step for the advancement in research was that of Sanger sequencing itself due to its expense, low throughput, large fragment requirements and time intensive protocols [6]. Despite being the driving force of many major discoveries and projects, it could not keep up with the ever increasing demands of researchers and clinicians, with its lack of multiplexing, limit of fragment sizes, and painstaking repetitiveness, many researchers began asking for a new, improved sequencing paradigms [7].

Several techniques began to appear such as sequencing by hybridization, parallel signature sequencing and pyrosequencing to name a few [8—10]. These three methods addressed the need for shorter fragments, multiplexing, larger throughput and higher sensitivity, but it was pyrosequencing which was to become the first of the Next Generation of sequencers with the release of the 454 system by 454/Roche in 2005 [11]. This opened the floodgates for more next generation sequencers with several platforms entering the market in a relatively short period of time. Each of these machines shared similar performance on throughput, accuracy and cost in comparison to Sanger and so after the success of these initial platforms, new machines and techniques were rapidly developed to keep up with the demands of research laboratories [12—15]. With rapid improvements in sequencing methods, several of the initial techniques used have already been superseded with simpler, more accurate techniques [16], but in an ever developing field, no technique remains the same for long.

* Corresponding author. Molecular Haematology, Haematology Department, Belfast Health and Social Care Trust, 51 Lisburn Road, Belfast, BT9 7AB, UK.
E-mail address: mark.catherwood@belfasttrust.hscni.net (M.A. Catherwood).

Over the last 10 years, the field of sequencing has developed at a break neck speed with the ever expanding interest and knowledge of the importance of the underlying genetics of many diseases. One such field that has benefitted greatly from this is that of haematological malignancies. Next generation sequencing has made large advancements in the understanding of the underlying mechanisms of several haematological cancers over the last 10 years which would not have been possible with previous investigative techniques [17,18] (Fig. 1).

1.1. The next generation

Multiple platforms have been released over the past 8 years which have opened the doors to high throughput, high sensitivity, quick whole genome sequencing. Each platform offers various sequencing approaches, such as whole genome sequencing (WGS), whole exome sequencing (WES), amplicon sequencing to name a few, alongside their own advantages and disadvantages and so a decision on which platform to ultimately go with is down to the needs of the researcher and the data produced, namely transcriptomics, genomics, and epigenomics [17,19]. Due to the large number of platforms currently available, below is a table constructed to compare the most popular models available based on their accuracy, method of sequencing and other criteria (Table 1). For those wishing for a more in depth knowledge of the platforms, please see reviews [11,14,15,19,20].

Two of the most common sequencing approaches taken are that of Whole Genome Sequencing (WGS) and Whole Exome Sequencing (WES). Both are very powerful methods but both carry their own benefits and pitfalls. One such study which benefited from WGS was that of sequencing the Acute Myeloid Leukaemia (AML) genome which allowed for the discovery of new candidate genes which were associated in the pathogenesis of this disease [25] however, the massive amounts of data produced from this level of sequencing proved to be almost impossible to analyse in a reasonable amount of time leading to only 2% of the genome being used for validation [26]. WES would prove to be a good answer to this colossal data problem however in doing this, ~98% of the genome will be excluded from sequencing which could in turn eliminate important genetic discoveries after the findings of the Encyclopedia of DNA Elements group who have reported that >80% of the genome is biochemically active [27].

Another increasingly popular sequencing method is that of amplicon sequencing, a technique which selectively enriches chosen genomic areas of interest for sequencing [28]. This is a good alternative to the massive data outputs of whole genomes which can run into the 90 gigabyte territory with deep sequencing making analysis a major problem and also introduces data storage issues. Amplicon sequencing in comparison to WES and WGS requires much less DNA, involves only one PCR reaction saving time, it is also more cost effective that WGS and WES if you wish to look at only a particular region of interest. Another major advantage is that this has allowed sequencing to occur in smaller laboratories due to the minimal requirements for this method over WGS/WES but as with every technique, there are disadvantages such as time required for library preparation, and with the ever decreasing costs of WGS there will come a point where this method is no longer financially feasible.

1.2. Impact on haematological malignancy research

Research in this field developed with the introduction of cytogenetics in the 1960s allowing for the discovery of important chromosomal translocations in Chronic Myeloid Leukaemia (CML) and Acute Myeloid Leukaemia (AML) such as that of the infamous Philadelphia
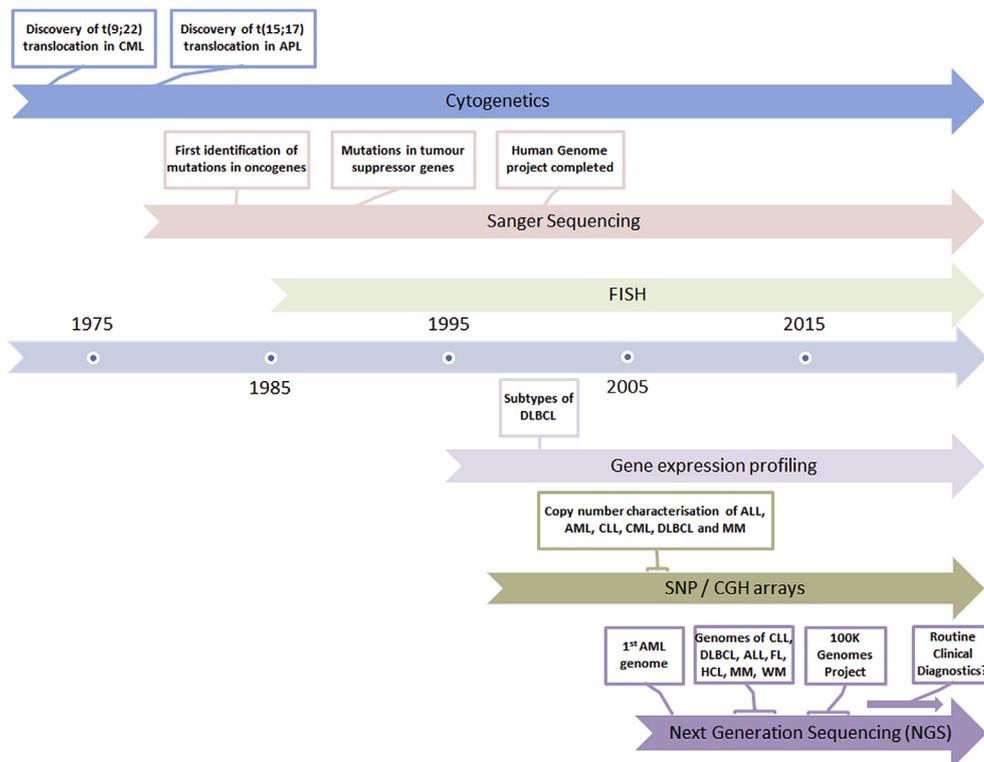

Fig. 1. Timeline of genetic discoveries.

Table 1
Summary of commonly used sequencing platforms.

Platform	Sequencing	Amplification	Read length	Data output	Run time	Accuracy
Roche 454 GS Junior	Sequencing by synthesis – pyrosequencing	emPCR	400 bp	35 Megabases	10 h	99%
Illumina HiSeq 2000	Synthesis by sequencing – reverse dye terminator	Bridge PCR	36–101 bp	≤300 Gigabytes	2.5–11 days	90–99.9%
Illumina MiSeq	Synthesis by sequencing – reverse dye terminator	Bridge PCR	36–151 bp	>1 Gigabyte	4–27 h	75–90%
Life Technologies Ion Torrent	Synthesis by sequencing – hydrogen ion detection	emPCR	Dependent on chip size – >100bp	>1 Gigabyte	2 h>	99%

NB: All information derived from platform websites [21–24].

Chromosome [29–31]. With the addition of new techniques such as Sanger sequencing, FISH, GEP and SNP Arrays, further discoveries helped evolve our understanding of the underlying mechanisms of haematological cancer. As NGS technologies became commercially available during the mid-2000s, there was a sudden inflation in the rate of major discoveries due to the increased sensitivity of NGS as well as a better understanding of the importance of the genetics of these conditions which led to as many breakthroughs in the last 10 years as there has been in the previous 30 years [17]. (Fig. 1)

As NGS importance has become more profound, and with its associated costs dropping exponentially each year, larger and more encompassing studies have begun into haematological cancers to unearth new pathways and biomarkers, to enhance our understanding of its development as well as provide new diagnostic and treatable markers for translation into the clinic to improve patient care and survival.

1.3. Acute Myeloid Leukaemia (AML)

One of the most compelling studies showcasing the need for NGS is that of the WGS of AML, the most common myeloid malignancy in adults [25,26]. The tumour genome of a patient presenting with the most common subtype, French-American British (FAB) M1, of AML with normal cytogenetics was sequenced, alongside the patient's normal skin cells to provide an unbiased portrait of the mutations present within the tumour in 2008. Both genomes were sequenced with a high degree of coverage (32.7 fold tumour genome, 13.9 skin genome) [25]. The results obtained revealed 10 mutated genes of which 2 are previously described and 8 newly discovered; *CDH24, PCLKC/PCDH24, GPR123, EBI2/GPR183, PTPRT, KNDC1, SLC15A1*, and *GRINL1B* were novel.

Due to the massive data handling associated with WGS studies, the data selected for further analysis was defined by comparing the single nucleotide variants (SNVs) within the tumour and skin genomes which showed 97.6% of the SNVs were present in both genomes, leaving 2.4% for further investigation where the 10 described mutations were found. This study highlighted NGS as a powerful tool for the discovery of novel mutations for future study as biomarkers and therapeutic targets as within the previous two decades little had changed in the therapy of AML due to the lack of understanding of its genetic pathogenesis.

In 2009, a study using similar data elimination methods on a different patient with the same AML subtype uncovered recurrent *IDH1* and *IDH2* mutations, a gene encoding isocitrate dehydrogenase 1, which were previously only reported in gliomas [32,33]. *IDH1* is a suggested tumour suppressor gene which is inactivated by dominant mutations of Arginine 132 (*R132*) [34] . *IDH1* mutations were found in up to 50% of AML patients which further strengthened the support for NGS as a mutation discovery tool. Another finding of this study was that the mutations found in the AML tumours were in different frequencies than that found in glioblastoma samples; *R132C* mutations were present in 50% of AML samples in comparison to 4% of gliomas. Further studies showed these mutations to be present in high occurrence in intermediate-risk AML cytogenetic profiles and associated with a poor prognosis [35].

Ley and his team re-sequenced the genome of their cytogenetically normal AML M1 patient in their seminal 2008 paper using advanced NGS technology providing deeper sequencing coverage [25]. Mutations within the *DNMT3A* gene, encoding DNA methyltransferases, were subsequently discovered resulting in the further study of 188 normal matched AML tumours to find that was a frequently occurring event associated with poor prognosis. Relationships between *DNMT3A* mutations as other commonly mutated genes within AML were also resolved showing *FLT3, NPM1* and *IDH1* mutations were significantly enriched in the presence of this mutation.

A recent study carried out in 2012 identified 9 mutations which were already present in haematopoietic cells which required an initiating event to drive oncogenesis; an unusual discovery as in many cancers, it is thought a key event creates these mutations [36]. This study was carried out on M1 and M3 AML subtypes and provided interesting results showing *DNMT3A, NPM1, IDH1*, and *TET2* were mutations unique to M1 genomes and occur rarely in M3 indicating different initiating mutation events for each subtype. There was also evidence of shared initiation mutations through *FLT3 ITD* which is present in both subtypes.

Another major discovery which deserves mentioning involving NGS and AML is that of clonal evolution pathways in relapsed AML patients and associated novel mutations [37], which further shows NGSs potential in the understanding of the underlying mutations and pathways in this highly complex disease.

1.4. Myelodysplasia

Splicing machinery defects were a prevalent discovery using WES in this condition with evidence of six novel RNA splicing machinery genes containing mutations affecting the 3'-splice site recognition in pre-mRNA processing [38]. Major effects of this are impaired haematopoiesis and dysfunctional mRNA splicing which were found to be caused by the following 6 genes; *PRPF40B, U2AF35, SRSF2, ZRSR2, SF3A1*, and *SF3B1*. These were also found to be mutually exclusive of one another and, depending on disease subtype, occurring in different frequencies. *U2AF2* and *SF1* mutations hold some significance as being exclusively found in myelodysplastic syndromes [39]. Mutations were associated in only 44% of cases with increased sideroblasts and up to 88% of cases with minimal sideroblasts. This was the first indication that mutations resulting in changes to the splicing machinery could play a role in human pathogenesis.

The results of this study were further substantiated several months later revealing the importance of a spliceosomal subunit *SF3B1*, part of the small nuclear ribonucleoprotein (snRNP), as a recurrently mutated gene of a frequency of ~20% within patients of myeleodysplastic

syndromes and association with better prognosis [40]. A major finding of this particular paper is the ability to detect this mutation within peripheral blood samples providing a potential biomarker for future use within the clinic.

1.5. Chronic lymphocytic leukaemia (CLL)

As is the case for many haematological malignancies, the molecular pathogenesis of CLL, a common clinically heterogeneous cancer, remained largely unknown until a novel study approach combining NGS WES and Copy Number Analysis was able to shed some light on the mutations driving this disease [41]. This study found 32 genes to be associated with the initiation of CLL however, a novel gene, *NOTCH1*, was the focus for further study as it was found to be present in 8.3% of the CLL cases at diagnosis. Although previously reported in various studies, the frequency of this mutation of the transmembrane protein gene family was considered to be a rare event [42]. As the disease progressed to a more aggressive form and Richter's Transformation, this frequency increased to 31.0% providing a diagnostic predictor of prognosis for clinicians at presentation of CLL.

Further studies on the *NOTCH1* gene provided evidence of functional relevance of this mutation in CLL, alongside another mutation as oncogenic activating events [43]. Initial WGS on 4 patients with CLL provided 46 somatic mutations which may affect gene function; this led to a cohort of 363 CLL tumours being sequenced to discover 4 frequently mutated genes: *NOTCH1, XPO1, MYD88,* and *KLHL6. MYD88*, a myeloid differentiation primary response gene, plays a role in the signalling pathways of the immune response as well as being implicated in lymphomas, promoting the view of the relevance of this mutation in lymphoid neoplasms [44]. An association with lower age of diagnosis is linked with this mutation. This study successfully highlights the functional relevance of these two mutations for the formation of a favourable environment for the survival of CLL cells as well as adverse clinical outcome. Not only were potential treatable targets revealed, but also new factors to take into consideration in the clinic for diagnosis and prognosis.

With the importance of splicing machinery defects becoming clear in studies for myeloid neoplasms, the discovery of these mutations in CLL came in 2012. WES of a group of 105 CLL tumours revealed 78 genes with potential function altering mutations [45]. Of these genes, *SF3B1* was mutated in 9.7% of cases which was coupled with poor clinical outcome, faster disease progression and low survival rates in affected individuals. This paper also substantiated the discovery of NOTCH1 as a recurrently mutated gene in CLL [43,44] as well as reinforcing the notion of targeting the mRNA splicing pathway for treatment. The importance of *SF3B1* as a clinically relevant gene became apparent with its relationship with aggressive disease progression, the opposite of findings in myelodysplastic syndromes, as well as being an independent prognostic factor [40,46].

Other novel mutations have been found including *ZMYM3, FBXW7, MAPK1* and *DDX3X* in work carried out in 2011 showing their interaction with the commonly mutated CLL genes *NOTCH1, SF3B1* and *MYD88* [47]. Nine genes in total, including previously identified *TP53, ATM* and *MYD88* genes, were identified as being driver mutations in the oncogenesis of CLL, supported by their frequency in conserved sites as well as showing functional importance. This study was unusual in its design as rather than several genomes being sequenced for detection of potential somatic mutations for further analysis in expanded cohorts, the approach of WGS and WES of a large set of patients (91 in total) was taken to increase the probability of discovering a much broader range of mutations to give a better picture of the pathogenesis of this disease.

The importance of *SF3B1* and *NOTCH1* mutations were not overlooked with a clinical trial of 494 patients analysed for their mutational status of these genes and then monitored in relation to these for treatment response, survival and biological variables [48]. The findings supported previous suggestion of clinical relevance with *NOTCH1* patients (10% overall) having reduced, progression free survival and *SF3B1* patients (17% overall) having shorter overall survival and no associated *TP53* mutations. *TP53* was still the most informative mutation as indicator of poor survival [49] however *NOTCH1* and *SF3B1* were still added sources of information as independent prognostic markers for shorter survival.

1.6. Other haematological malignancies

One leukaemia with poor survival is a relapsed paediatric patient with an Acute Lymphocytic Leukaemia which has a 30% survival rate. Clinical and biological differences between diagnosed and relapsed cells were well established however the molecular trigger for relapse remained unknown until a study of Copy Number Analysis (CNA) on 61 diagnosis-relapse matched paediatric cases was carried out [50]. B-ALL and T-ALL cases were included in this study showing a majority of relapsed cases (88.5%) contained CNAs present at diagnosis leading to the conclusion that the CNAs responsible for relapse are present at diagnosis in small clone quantities and are also therapeutically immune. Potential pathways were identified for these CNAs including the previously defined *ETV6-RUNX1* pathway present at both diagnosis and relapse [51] as a potential treatment candidate. Other findings in this paper showed the complexity of CNA changes between diagnosed and relapsed samples which will require further research.

Good understanding of several of the translocations and mutations causing structural aberrations in the subgroups of Diffuse Large B-Cell Lymphoma (DLBCL) were well known but the genetic lesions responsible for malignant transformation remained unclear. Combined WES and CNA techniques revealed a number of previously unrecognised pathways and mutations in this malignancy [52]. The main aim of this study was to characterise the complexity of the DLBCL genome but other interesting findings were that of unknown alterations to the *TNF AIP3* tumour suppressing gene and a translocation between *CDKN2A-CDKN2B*. The main finding of this study is the mutations of the *MLL2* gene encoding for a trimethyltranferase. This gene with a well-known influence over other genes has a suggested role as a haploinsufficient tumour suppressor and so that abnormalities in *MLL2* could trigger an oncogenic event within DLCBL, this claim is further substantiated by the role of *MLL2* in the pathogenesis of acute leukaemia and other cancers [53–55].

Whole exome sequencing of Hairy Cell Leukaemia (HCL) tumour cells with matched normal cells from a patient showed previously undiscovered BRAF mutations, further strengthening the argument for the increased sensitivity of NGS [56]. As with many other haematological malignancies, the clinco-pathological features of this disease were well characterized however its underlying development was poorly understood until the discovery of recurrent *BRAF V600E* mutations in all patients. Further studies showed this mutation to be targeted by *PLX-4720*, a small inhibitor molecule, which provided strong evidence for further research of the pathogenesis of HCL, diagnosis and development of targeted therapeutics. Further research conducted by Dietrich and his team [57], confirmed these findings as well as

showing *BRAF* to be a good diagnostic marker as this mutation is not present in other B-cell malignancies with similar features to HCL, such as splenic marginal zone lymphoma.

Finally, a WES study into the genetics of T-Cell Large Granular Lymphocytic Leukaemia provided evidence of *STAT3* mutations in 40% (31 of 77) of cases, indicating this mutation is the underlying cause of the pathogenesis of this cancer [58]. Aside from aberrant *STATs* signalling, further hotspots were found in exon 21 of the 31 *STAT3* mutant positive cases, an area where dimerization and activation of the *STAT* protein occurs, indicates this exon could provide information on the development of this disease.

Taking all of these diverse studies together, the impact of NGS technologies on the deciphering of haematological malignancies has been profound, with a greater understanding of the development of these disease found within the last six years than there has been in the last two decades. The discoveries discussed here are just a small snapshot of the power this technology has in cancer research.

1.7. Clinical relevance of NGS

As outlined, the developments within haematological cancer research have been significant from the introduction of NGS with discoveries coming much faster than before and the understanding of this group of diseases becoming clearer with each mutation detected. Clinicians are now seeing the benefits of genetic based detection for both diagnosis and treatment but the question remains as to whether this is a feasible technique or do the current tests already do what is required? Can NGS transit from a discovery tool in haematological malignancies to a cost-effective molecular diagnostic tool?

Most diagnostic tests for haematological malignancies were immunologically, cytogenetically, and PCR based. Cytogenetics was a particularly powerful tool for gaining prognostic information and flow cytometry alongside immunocytochemistry for detection of aberrantly expressed proteins [59]. However, with the introduction of PCR for detection of translocations, microsatellites, clonality and mutational analysis; it became an increasingly useful tool in the diagnostic laboratory [60]. Sanger sequencing is currently considered the "Gold Standard" of mutational analysis in diagnostics and has been for the past three decades, routinely sequencing for the status of *KRAS, NRAS, EGFR* and *BRAF* within multiple cancers however, its low throughput, high cost, low sensitivity and slow result turnaround are no longer suitable for diagnostics with our increased understanding of genetics in cancer and hence the need for more sensitive, faster, cheaper techniques [61,62].

As outlined in a study of NGS in a clinical diagnostic setting, massively parallel sequencing is a feasible option for future diagnostics with its low cost, high throughput, high sensitivity, minimal DNA input and range of information to be gained from results such as mutational status, translocation information and analysis of multiple genes or hotspots in one experiment [63]. The ability to barcode samples with Multiplex Identifiers (MIDs) also allows for multiple patients to be sequenced at once which is a key need for a diagnostic laboratory. Another benefit of NGS is the ability to detect multiple abnormalities such as point mutations, insertions, deletions and chromosomal rearrangements using a single sequencing run [64], an ability not currently available through any other test. In a novel approach Bouamar and colleagues used a capture based analysis for the identification of IGH rearrangements in a series of DLBCL (REF). In this series they confirmed the presence of know fusions and discovered novel IGH partners in DLBCL. Whilst this was performed only on fresh frozen material it is envisaged this will be applicable to FFPE material [65].

A major advantage of this technology is its sensitivity as Sanger sequencing has difficulties in detecting mutations which occur lower than around 10% in tumours [66]. One study was able to detect low levels of TP53 mutations, with a clone size of 11% in 20% of patients with early low risk MDS allowing for better follow up of these patients who now present a risk of increased leukaemic evolution [67].

With its high throughput, sensitivity, speed, cost and information capabilities, the future of NGS as a clinical tool is strong and is likely to replace current diagnostic standards such as Sanger sequencing however, there are still several hurdles for this technology to cross before this goal can be realised.

1.8. Challenges of NGS technologies

With all the advantages and potential NGS brings to research and diagnostics, it also has several pitfalls which need to be addressed.

The first problem encountered in developing NGS for diagnostics was that of the massive amounts of data produced from WGS and WES, most of which would not be relevant in a diagnostic setting where information from a particular gene of interest is sufficient. One such study which is addressing this problem is that of the Interlaboratory Robustness of next generation sequencing consortium (IRON). This group is designing diagnostically suitable deep sequencing amplicon assays which will provide information on disease classification, patient stratification, and allow for monitoring of a patient's minimal residual disease from a sequencing run for various haematological cancers [68]. These assays are composed of multiple clinically and diagnostically relevant genes and hotspots of interest within each cancer to give this range of information. As mentioned before, MIDs are used in these assays to allow for multiple patients to be screened at once. Amplicon assays such as this allow for quick gain of relevant information without the need for in depth bioinformatics analysis.

Other challenges faced are that of the costs associated with gaining the equipment, software and consumables required for NGS. Current estimates of the start-up capital required are around $500,000 (~£300,000) [69] but this can be offset by the cost per base of sequencing where Sanger is more expensive and a much larger data output is also achieved.

In terms of data output, another problem is encountered through analysis and storage of this data. The amount of data produced per sequencing run on NGS platforms runs into the gigabytes, which will require specialist high power computers to allow for quick, effective processing and analysis of this level of data. One such suggestion to remedy this is the use of cloud computing to create a galaxy-based analysis pipeline to detect mutations [70]. The computing infrastructure required to carry out NGS is also a large demand requiring a minimum of an 8 quad core, 32 gig RAM, 10 terabyte hard disk computer. In order to set up and maintain such an infrastructure will also require qualified ICT and bioinformaticians which can immediately out price some research and diagnostic facilities [71]. However, with computers becoming more prevalent in the everyday lives of researchers, newly graduated scientists are equipped with the knowledge required to carry out maintenance and repair of most computer problems.

A major area of concern within NGS is that of ethical and legal issues. With WGS being able to give an unbiased overview of a person's full genome, known cancer/disease driving genes can be seen as well as their potential to develop and this information could then be shared to

insurance companies and employers. The Genetic Information Non-discrimination Act (GINA) 2008 has begun to address such concerns through making it illegal to pass genetic information onto insurance companies or potential employers to ensure no genetic discrimination can take place [71]. Alongside this, the conundrum of deciding to tell a patient what potential disease causing genes they have is still a highly debated issue. There have been detailed discussion amongst researchers about whether or not to tell patients about potential disease causing genes in their genomes, as not only would affect them, but also relatives who could potentially carry the same genes themselves [72,73]. Overall, the current consensus is that it is up to the individual researcher to decide how much information to pass on.

Validation is another important aspect to address in terms of using this technology for diagnosis and disease monitoring in a hospital setting. As with any new diagnostic test, several questions need to be answered about its practicality and relevance within a clinical setting as well as ensuring it can cope with that particular laboratory's workload, samples and turnaround times while providing data which is at the same level, if not better, than the Gold Standard. In regards to NGS technologies, there are many questions to be answered such as should every target be validated, should accuracy and specificity for every target for every sample type be validated, should the bioinformatics side of this technology also be validated. The need to establish clear international validation guidelines has been clearly identified [74]. These suggested guidelines would address the multiple validation questions raised by such an encompassing technology as well as provide advice on how to make NGS as cost effective as possible in a clinical setting. This need for harmonization also emphasizes the need for a more in depth discussion amongst clinical bodies to ensure NGS can be brought into the clinical setting sooner rather than later due its ability to lower the complexity of the level tests required for particular haematological conditions currently required.

With the technology rapidly evolving to eliminate these challenges, and our better understanding of it in the research and diagnostic settings, the massively parallel sequencing technology has already become the true gold-standards in many centres.

2. Conclusion

Since the completion of the Human Genome Project, the true potential in genetic research was truly realised beginning the development and advancement of next generation sequencing technologies. This field has grown at an exponential rate with multiple new companies emerging dedicating their services to the field, as well as many well established scientific companies developing their own technologies [19,75]. The growth and dominance of this set of technologies has caused a shift in the framework of scientific research, opening the doors to a new era of oncology research and diagnostics, as well as requiring scientists with stronger molecular knowledge alongside computational skills to be able to carry out this work.

As the importance of this technology has been realised, research using these machines boomed with many major breakthroughs within a short few years, as outlined in this review, but also in other areas such as that of sequencing ancient DNA which could potentially open the doors to full high quality genome sequencing from a single cell, a development which could have major implications for diagnostic testing [76], an area where this technology has been heavily validated for future use due to its vast array of uses, robustness, sensitivity, accuracy, minimal DNA input, speed and information output. One such development in NGS with clinical implications in mind is that of amplicon sequencing allowing for targeted sequencing of genes of interest to minimise data output and provide fast, expansive results on multiple genes for particular disease to allow for a more comprehensive understanding of that patient and their clinical progress [68]. With the quick advances in the speed, accuracy, and cost of these systems, the potential for using NGS as a clinical tool has become more feasible over the last three years, supported by the multitude of validation research papers available today.

Overall, we have been challenged in our research by NGS with its ever evolving protocols, chemistry, and machinery with scientists playing a game of catch up to keep up to date with the technology but despite this, our understanding of the underlying genetic mechanisms of many diseases both within and outside of haematology has proved NGS to be an essential tool in modern research and diagnostics.

Conflict of interest
None.

Abbreviations

PGM	personal genome machine
FISH	fluorescence in situ hybridization
NGS	next generation sequencing
GEP	gene expression profiling
WGS	whole genome sequencing
SNP	single nucleotide polymorphisms
WES	whole exome sequencing
FAB	French/American/British Subtype
PCR	polymerase chain reaction
SNVs	single nucleotide variants
DNA	deoxyribonucleic acid
mRNA	messenger ribonucleic acid
RNA	ribonucleic acid
CLL	chronic lymphocytic leukaemia
CML	chronic myeloid leukaemia
ALL	acute lymphocytic leukaemia
AML	acute myeloid leukaemia
CNA	copy number analysis
DLBCL	diffuse large B-cell leukaemia

References

[1] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931—45.

[2] Stein DS. Human Genome: end of the beginning. Nature 2004;431:915—6.

[3] Venter TC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001;291(5507):1304—51.

[4] Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. Nature 2003;422:835—47.

[5] Sanger F, Nicklen S, Coulson AR. DNA Sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74(12):5463—7.

[6] Chaisson M, Pevzner P, Tang H. Fragment assembly with short reads. Bioinformatics 2004;20(13):2067—74.

[7] Martinez DA, Nelson MA. The next generation becomes the now generations. PLoS Genet 2010;6(4):e10000906.

[8] Bains W, Smith GC. A novel method for nucleic acid sequence determination. J Theor Biol 1988;135(3):303—7.

[9] Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, et al. In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differently expressed cDNAs. Proc Natl Acad Sci 2000;97:1665—70.

[10] Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 1996;242:84—9.

[11] Liu L, Yinhu L, Siliang L, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomedicine Biotechnol 2012 [Article ID 251364].

[12] http://www.illumina.com/systems/miseq/performance_specifications.ilmn.

[13] http://www.lifetechnologies.com/uk/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrentnext-generation-sequencing-run-sequence/ion-pgm-system-for-next-generation-sequencing.html.

[14] Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosystems and Illumina MiSeq Sequencers. BMC Genomics 2012;13:341.

[15] Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 2012;30:434—9.

[16] http://allseq.com/knowledgebank/sequencing-platforms/roche-454.

[17] Braggio E, Egan JB, Fonseca R, Stewart AK. Lessons from next-generation sequencing analysis in haematological malignancies. Blood Cancer J 2013;3:e127.

[18] Kohlmann A, Grossmann V, Nadarahjah N, Haferlach T. Next generation sequencing — feasibility and practicality in hematology. BJH 2013;160(6):736—53.

[19] Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet 2010;11:31—46.

[20] Cullum R, Alder O, Hoodless PA. The next generation: using new sequencing technologies to analyse gene regulation. Respirology 2010;16:210—22.

[21] http://454.com/downloads/GSJuniorSystem_Brochure.pdf.

[22] http://res.illumina.com/documents/systems/hiseq/datasheet_hiseq_systems.pdf.

[23] http://res.illumina.com/documents/products/datasheets/datasheet_miseq.pdf.

[24] https://tools.lifetechnologies.com/content/sfs/brochures/PGM-Specification-Sheet.pdf.

[25] Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A mutations in acute myeloid leukemia. N Eng J Med 2010;363:2424—33.

[26] Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukemia genome. Nature 2008;456:66—72.

[27] The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57—74.

[28] Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods 2010;7:111—8.

[29] Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinaerine fluorescence and Giemsa staining. Nature 1973;243(5405):290—3.

[30] Rowley JD, Golomb HM, Dougherty C. 15/17 translocation: a consistent chromosomal change in acute promyelocytic leukaemia. Lancet 1977;309(8010):549—50.

[31] Rudkin GT, Hungerford DA, Nowell PC. DNA contents of chromosome Ph1 and chromosome 21 in human chronic granulocytic leukemia. Science 1964;144(3623):1229—32.

[32] Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med 2009;361:1058—66.

[33] Kang MR, Kin MS, Oh JE, Kim YR, Song SY, Seo SI, et al. Mutational analysis of IDH1 codon 132 in glioblastomas and other common cancers. Int J Cancer 2009;125:353—5.

[34] Zhao S, Lin Y, Xu W, Jiang W, Zha Z, Wang P, et al. Glioma-derived mutation in IDH1 domain inhibit IDH1 catalytic activity and in HIF-1alpha. Science 2009;324:261—5.

[35] Paschka P, Schlenk RF, Gaidzik VI, Habdank M, Kronke J, Bullinger L, et al. IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplications. J Clin Oncol 2010;28:3636—43.

[36] Welch JS, Ley TJ, Link DC, Millar CA, Larson DE, Koboldt DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell 2012;150(2):264—78.

[37] Ding L, Ley TJ, Larson DE, Millar CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukemia revealed by whole genome sequencing. Nature 2012;481:506—10.

[38] Yoshida K, Sonada M, Shiraishi K, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011;478:64—9.

[39] Visconte V, Makishima H, Jankowska A, Szpurka H, Traina F, Jerez A, et al. SF3B1; a splicing factor is frequently mutated in refractory anaemia with ring sideroblasts. Leukemia 2012;26:542—5.

[40] Papaemmanuil E, Cazzola M, Boultwood J, Malcovatil L, Vyas P, Bowen D, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med 2011;305:1384—95.

[41] Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabanian H, Ma J, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. JEM 2011;208(7):1389—401.

[42] Sportoletti P, Baldoni S, Cavalli L, Del Papa B, Bonifacio E, Ciurnelli R, et al. NOTCH1 PEST domain mutation is adverse prognostic factor in B-CLL. Br J Haematol 2010;151:404—6.

[43] Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature 2011;475:101—5.

[44] Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao S, Lim K, et al. Oncogenically active MYD88 mutations in human lymphoma. Nature 2011;470:115—9.

[45] Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Gen 2012;44:47—52.

[46] Quesada V, Ramsay AJ. Chronic lymphocytic leukemia with SF3B1 mutation. N Engl J Med 2012;366:2530.

[47] Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med 2011;365:2497—506.

[48] Oscier DG, Rose-Zerilli MJJ, Winkelmann N, Je Castro GD, Gomez B, Forster J, et al. The clinical significance of NOTCH1 and SF3B1 mutations in UK LRF CLL4 trial. Blood 2013;121:468—75.

[49] Gonzalez D, Martinez P, Wade R, Hockley S, Oscier D, Matutes E, et al. Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. J Clin Oncol 2011;29(16):2223—9.

[50] Mullighan GC, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. Science 2008;322(5906):1377—80.

[51] Zuna J, Ford AM, Peham M, Patel N, Saha V, Eckert C, et al. TEL deletion analysis supports a novel view of relapse in childhood acute lymphoblastic leukemia. Clin Cancer Res 2004;10:5355—60.

[52] Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet 2011;43:830—7.

[53] Krivtsov AV, Armstran SA. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer 2007;7:823—33.

[54] Parsons DW, Meng L, Zhang X, Jones S, Leary RJ, Lin JC, et al. The genetic landscape of the childhood cancer medulloblastoma. Science 2011;331:435—9.

[55] Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 2010;463:360—3.

[56] Tiacci E, Trifonav V, Schiavoni G, Holmes A, Kern W, Martelli MP, et al. BRAF mutations in hairy-cell leukemia. N Engl J Med 2011;463:2305—15.

[57] Dietrich S, Glimm H, Andrulis M. BRAF inhibition in refractory hairy-cell leukemia. N Engl J Med 2012;366:2038—40.

[58] Koskella HL, Eldfors S, Ellonen P, van Adrichem AJ, Kuusanmaki H, Andersson EI, et al. Somatic STAT3 mutations in large granular lymphocytic leukemia. N Engl J Med 2012;366:1905—13.

[59] Morgan GJ, Pratt G. Modern molecular diagnostics and the management of haematological malignancies. Clin Lab Haem 1998;20:135—41.

[60] Godley LA. Profiles in leukemia. N Engl J Med 2012;366(12):1152—3.
[61] McCourt CM, McArt DG, Maxwell P, Catherwood MA, Mills KI, Waugh DJ, et al. Validation of next-generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. PLoS One 2013;8:e69604.
[62] Voelkerding K, Dames SA, Durtschi JD. Next generation sequencing: from basic research to diagnostics. Clin Chem 2009;55(4):641—58.
[63] Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med 2012;366:1079—89.
[64] Grossman V, Kohlmann A, Klein HU, Schindela S, Schnittger S, Dicker F, et al. Targeted next-generation sequencing detects point mutations, insertions, deletions and balanced chromosomal rearrangements as well as identifies novel leukemia-specific fusion genes in a single procedure. Leukemia 2011;25:671—80.
[65] Bouamar H, Abbas S, Lin AP, Wang L, Jiang D, Holder KN. A capture-sequencing strategy identifies IRF8, EBF1, APRIL as Nov IGH fusion partners in B-cell lymphoma 2013;122:726—33.
[66] Querings S, Altmuller J, Ansen S, Zander T, Seidel D, Gabler F, et al. Benchmarking of mutation diagnostics in clinical lung cancer specimens. PLoS One 2011;6(5):e19601.
[67] Jadersten M, Saft L, Smith A, Kulasekararaj A, Pomplun S, Gohring G, et al. TP53 mutations in low risk myleodysplastic syndromes with del(5g) predict disease progression. J Clin Oncol 2011;29(15):1971—9.
[68] Kohlmann A, Grossman V, Haferlach T. Integration of next-generation sequencing into clinical practice: are we there yet? Seminars Oncol 2012;39(1):26—36.
[69] Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. J Mol Diagnostics 2008;10(6):484—92.
[70] Onsong G, Erdmann J, Spears MD, Chilton J, Beckman KB, Hauge A, et al. Implementation of cloud based next generation sequencing data analysis in a clinical laboratory. BMC Res Notes 2012;7:314.
[71] Desai AN, Jere A. Next-generation sequencing: ready for the clinics? Clin Genet 2012;81(6):503—10.
[72] Lyon GJ, Jiang T, Van Wijk R, Wang W, Bodily PM, Xing J, et al. Exome sequencing and unrelated findings in the context of complex disease research: ethical and clinical implications. Discov Med 2011;12(62):41—55.
[73] McGuire AL, Lupski JR. Personal genome research: what should the participants be told? Trends Genet 2010;26(5):199—201.
[74] Salto-Tellez M, Gonzalez de Casto D. Next-generation sequencing: a change of paradigm in molecular diagnostic validation. J Pathology 2014;234(1):5—10.
[75] Schloss JA. Hot to get genomes at one ten-thousandth the cost. Nat Biotech 2008;26:1113—5.
[76] Briggs AW, Good JM, Green RE, Krause J, Marieic T, Stenzel U, et al. Targeted Q4 retrieval and analysis of five Neanderthal mtDNA genomes. Science 2009;325(5938):318—21.