INTRODUCTION
Over the past few years, a remarkable effort has been underway to identify the genetic basis of hematological malignancies catalyzed by increasing availability and more refined sequencing technologies (Figure 1). The growth in high-throughput sequencing, which has facilitated this effort, has been exponential with a dramatic increase in efficiency and correlating drop in price per base pair. Modern platforms can now perform whole-genome sequencing (WGS) of an individual for less than $5000 and a few days of work; notable progress compared with the resources and time that were used just a few years ago by an international consortium when completing the first human genome. Strikingly, this technological revolution occurred in <10 years.

A representative example of how increasingly more powerful technologies continuously improve our knowledge of the genetic basis of hematological malignancies comes from the study of the genome of an acute myeloid leukemia (AML) patient with normal cytogenetics, which was studied twice in a period of 2 years. The initial WGS analysis study only revealed small insertions and deletions affecting two genes, and nonsynonymous somatic mutations, in another eight genes. Two years later, the same genome was resequenced utilizing more advanced sequencing technology and analytical methods resulting in the detection of a previously unidentified frameshift deletion in DNMT3A. After the initial discovery, DNMT3A mutations were screened in large cohorts and it is now known that 22–30% of AML patients have mutations in this gene, being currently one of the most relevant mutations found in AML.

Next-generation sequencing (NGS) encompasses several different methodologies that allow the investigation of genomics, transcriptomics, and epigenomics. A summary of the different sequencing approaches is briefly described below and summarized in Table 1. For more in-depth information, we direct the reader to a number of excellent reviews.4–10

WHOLE-GENOME SEQUENCING
Two major approaches are utilized in the preparation of DNA libraries for WGS. The first is called paired-end sequencing, where ~100 bp are sequenced from each end of ~400-bp DNA fragments. By this method, single nucleotide variants (SNV), insertions and deletions and copy-number changes can be identified. Paired-end WGS needs low-input quantities of DNA (<1 μg) for generating the libraries, which is a critical advantage in the study of hematological malignancies, where the amount of tumor tissue is usually scarce.

The second approach used in the DNA library preparation is named mate-pair sequencing. Mate-pair is based on the generation of much larger DNA fragments than paired-end sequencing with fragments ranging in length from 1 to 10 kb. Longer distance between the read pairs enables improved detection of structural rearrangements because the read pairs can span repeat and duplicate regions, thereby capturing regions not adequately captured with smaller insert sizes utilized with paired-end sequencing. Very low coverage of the genome is enough for studies focused on the detection of structural abnormalities, thus reducing costs and complexity of the analysis. On the other hand, if the genome is covered in enough depth (>30X mean coverage), mate-pair sequencing can be used for the simultaneous detection of mutations, copy-number changes and structural abnormalities. A disadvantage of the mate-pair approach is that quite a large quantity of DNA is required for the library preparation, thus limiting its use in a significant number of tumors.

WHOLE-EXOME SEQUENCING
Whole-exome sequencing (WES) is useful for those interested in studying only what lies within the exome (coding genome) and untranslated regions. This method is based on an initial enrichment step of exonic regions followed by targeted
sequencing. As the exome represents only 1.4% of the genome, multiple samples can be pooled and sequenced together in a single instrument run. The major weakness of WES is the inability of the available enrichment kits to capture the totality of the exome. Ideally a non-tumoral, reference DNA sample from the individual patient is simultaneously analyzed with each tumor. The amount of normal variation between individuals is in the order of thousands of variants, and performing a paired analysis enables subtraction of the nontumor-specific from the tumor-specific

Table 1. Summary of high-throughput sequencing methods

| Method          | Minimum input quantity | Strengths                                                                 | Weaknesses                                                                 |
|-----------------|------------------------|---------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Whole genome    | 10 ng–1 μg genomic DNA | Small input DNA requirement                                                | Lower DNA input may reduce library complexity and representation          |
|                 |                        | Variant detection in all regions of the genome                             | PCR duplicates can impact accuracy of variant detection software          |
| Mate pair       | 5–10 μg genomic DNA    | Identification of large structural rearrangements                        | Large input DNA requirement                                               |
| Whole exome     | 1 μg genomic DNA       | Deep coverage of exome enabling precise interrogation of coding regions   | High false discovery rate                                                 |
|                 |                        | Multiple samples can be pooled and run together reducing time and cost per sample | Non-coding regions excluded                                               |
| mRNASeq        | 100–400 ng total RNA<sup>b</sup> | Dynamic range of expression detection can be much broader than using microarrays | RNA fragmentation methods can bias the resulting library                   |
|                 |                        | Detection of rare and hybrid transcripts                                   | Artifacts from amplified cDNA libraries<sup>b</sup>                       |
|                 |                        | Precise quantitation of highly expressed transcripts and multiple isoforms | Appropriate normal controls may be difficult to obtain for tumor/normal comparison |
| ChIPSeq         | 10 ng ChIP enriched DNA | Detection of DNA–protein interactions                                     | Quality of sequencing results dependent on the quality of ChIP assay       |
|                 |                        | Discovery of new interactions in regions not represented on microarray chips | Library preparation can introduce GC-rich region bias                      |
|                 |                        | Avoids hybridization problems associated with array-based ChIP assays     |                                                                           |
| Single molecule | 1 μg genomic DNA       | No amplification step resulting in no PCR duplicates                      | High error rate                                                            |
|                 |                        | Long-read length (> 1 kb)                                                 | Throughput not comparable to current platforms                           |

Abbreviations: ChIPSeq, chromatin immunoprecipitation sequencing; UTR, untranslated region. Cost per sample is highly variable depending on the platform and on the amount of multiplexing utilized. *May vary by platform and approach. **May require polyA RNA- or rRNA-depleted total RNA.
variants. In the event that normal tissue is not available for comparison, an increased number of publicly available databases, such as dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), HapMap (http://hapmap.ncbi.nlm.nih.gov/) and 1000 genomes (http://www.1000genomes.org/), can be utilized to identify and clean previously reported variants in the general population that are normal genetic variation rather than somatically acquired mutations.

**MESSENGER RNA SEQUENCING**

Besides the detection of mutations, messenger RNA sequencing is also a powerful tool for gene expression analysis. The dynamic range of expression obtained by messenger RNA sequencing can be much broader than that obtained by gene expression microarrays, allowing the detection of rare transcripts and more precise quantitation of expressed transcripts.\(^\text{5,11}\) In addition, messenger RNA sequencing can be utilized for the identification of hybrid transcripts and quantitation of multiple isoforms resulting from alternative splicing. Major weaknesses include the biases in library preparation caused by the RNA fragmentation methods utilized, introducing artifacts into the resulting reads.

**NEW APPROACHES AND FRONTIERS**

It is still the subject of debate which approach is the most appropriate for studying the cancer genome. WGS is the most inclusive approach, but major limitations remain related to the high cost and the difficulty associated with managing the data storage and intensive computational analysis. The study of the exome reduces these limitations. WES is a well-established strategy for analyzing coding regions at low cost, making this approach the most popular in the analysis of the tumor genome nowadays. However, eliminating 98% of the genome from the analysis brings the obvious risks associated with omitting crucial information. This concern is supported by recent findings performed by the Encyclopedia of DNA Elements Project, where integrated analysis demonstrated that >80% of the genome is biochemically active.\(^\text{12}\) This new paradigm will require reconsideration of the best strategic approach to optimize the cost/benefit ratio in the analysis of the cancer genome.

Single-molecule, long-read, sequencing approaches are now available and allow the simultaneous search for single-allele mutations and methylation profiles. Furthermore, several new platforms are currently under development that promise to sequence the whole genome in few hours for less than $1000. These technological advances open a new world of opportunities and soon will put the use of WGS within reach of the clinical labs.

**CHALLENGES IN DATA ANALYSIS**

Data generation is just a small facet of the much bigger challenge associated with data analysis. The goal of data analysis is to utilize bioinformatics tools in a data analysis ‘pipeline’ (Figure 2) to transform the raw data into results that can ultimately be seen in a user-friendly visualization tool. Detailed information about the most commonly utilized alignment and functional prediction tools are beyond the scope of this manuscript, and thus we direct the readers’ attention to several informative manuscripts.\(^\text{13–16}\)

The ability to quickly generate large quantities of data at relatively low cost is limited by these constantly evolving data analysis pipelines, failure to report analytical methods with the level of detail expected from traditional experimental data and the lack of consensus regarding which tools to use when transforming the data into a useable form. Nekrutenko and Taylor\(^\text{17}\) reviewed 50 papers that used the Burrows–Wheeler Aligner for mapping sequencing reads, and they found that most of the papers neither provide access to the raw data nor specify the parameters utilized and soon will put the use of WGS within reach of the clinical labs.

**SEQUENCING IN HEMATOLOGICAL MALIGNANCIES**

At this point, where medium-sized cohorts of most of the hematological malignancies have been sequenced, it is time to wonder what have these data revealed thus far in this group of malignancies and what are the opportunities ahead? As we expected, genes such as TP53, ATM and RAS among others were confirmed as mutated in a wide variety of malignancies. However, promising and exciting findings came from the discovery of a completely novel group of genes and pathways impaired in hematological malignancies. A few malignancies seem to be driven by mutations in only one or a few genes, suggesting a unique pathway, pathognomonic to the disease. However, most of the malignancies show considerable genetic heterogeneity, with multiple genes and pathways affected. In this review, we aim to summarize the current knowledge of the genetic background
on different hematological malignancies and how this knowledge could facilitate targeting of dysregulated signaling pathways by therapeutic targets. The most recurrent novel somatic genetic mutations per malignancy are summarized in Figure 3 and Table 2.

SINGLE CAUSATIVE MUTATIONS: HAIRY CELL LEUKEMIA AND WALDENSTROM’S MACROGLOBULINEMIA AS PARADIGMS

Probably, the most representative example of a single hit identified by sequencing are the hairy cell leukemias (HCL). Initially, WES was performed on a single HCL tumor/normal pair with somatic mutations identified in five genes: *BRAF*, *CSMD3*, *SLCSA1*, *CNTN6* and *OR8J1*. Another 47 HCL cases were screened for *BRAF* mutations and strikingly, the *BRAF* V600E substitution was found in all 47 patients evaluated. Conversely, *BRAF* mutations were absent in related peripheral B-cell lymphomas and chronic lymphocytic leukemia (CLL), and were only found in a small subset of multiple myeloma (MM) patients (4%). The same activating mutation and its damaging effect has been previously reported in solid tumors such as in melanoma and papillary thyroid cancer.

The presence of a common mutation across HCL provides a central novel therapeutic avenue in HCL based on V600E BRAF inhibitors alone or in combination with MEK or ERK inhibitors. The success of vemurafenib, a BRAF inhibitor, in the treatment of V600E BRAF-mutated melanoma patients led two groups to investigate the effectiveness of this small molecule inhibitor in one HCL case study each. In both cases, including one with a biallelic V600E BRAF mutation, treatment with vemurafenib resulted in successful disease treatment, thus providing evidence for clinical trials to evaluate the use of BRAF inhibitors in HCL.

A similar situation was found in Waldenström’s macroglobulinemia (WM). Remarkably, a *MYD88* L265P-activating mutation was recently found in 90% of WM cases. *MYD88* encodes for an adapter protein that affects the interleukin-1 and toll-like receptor pathway; with the L265P mutation leading to the dysregulation of the nuclear factor-κB and the JAK-signaling pathways. The same mutation has been found, but to a lesser extent, in additional B-cell lymphomas, such as diffuse large B-cell lymphomas (DLBCL) of the ABC type (~40%), MALT lymphomas and CLL (<10%), supporting the key role of *MYD88* in the pathogenesis of these neoplasias. Interestingly, a recent study evaluating the association between *MYD88* L265P and clinical characteristics of WM patients reported more involvement of the bone marrow disease, higher serum IgM, and lower IgA and IgG levels. Another group conducting a case-control study evaluating the association between *MYD88* L265P and IgM MGUS patients progressing to WM or other lymphoproliferative disorders reported a trend toward progression in patients with the presence of the mutation when compared with patients with wild-type *MYD88*. These findings highlight the potential value of *MYD88* as a potential biomarker of disease progression in WM.

In CLL, Velusamy et al. recently identified the presence of a YPELS-PP1TCB RNA fusion in 95% of CLL patients screened. Interestingly, WGS in the two index cases possessing the chimera did not reveal the presence of a gene fusion at the DNA level. These findings emphasize the importance of concurrently utilizing multiple methodologies such as WGS and RNASeq when studying tumors to better screen for genetic abnormalities.

One of the recurrent findings of sequencing research efforts has been the epistatic nature of discoveries. This notion reinforces the thought of classifying disease more along the line of functional aberrant pathways, rather than on specific genetic changes. In fact, excluding HCL and WM, the majority of the malignancies show a considerable genetic heterogeneity, affecting multiple genes and pathways. Presented here are some of the most remarkable recent discoveries.

MUTATIONS AFFECTING THE SPlicing MACHINERY

Recent sequencing studies identified recurrent mutations affecting genes of the splicing machinery in myelodysplastic syndrome (MDS). Interestingly, six of these genes (*SF3A1*, *SF3B1*, *SRSF2*, *U2AF35*, *ZRSR2* and *PRPF40B*) affect the initial steps of RNA splicing; thus, mutations leading to the impaired recognition of the 39 splice site result in the production of abnormal mRNA splicing. Mutations of the spliceosome are highly prevalent in MDS and other myeloproliferative disorders, ranging from 44% of cases...
### Table 2. Summary of high-throughput sequencing studies performed so far in hematological malignancies

| Disease | Discovery cohort (N) | Validation cohort (N) | Method | Platform | Mean coverage depth (X) | Highlights | Year | Reference |
|---------|----------------------|-----------------------|--------|----------|-------------------------|------------|------|-----------|
| ALL     | 1                    | 24                    | mRNASeq | GAII     | NR                      | DPEP1 (4%), longitudinal detection of PLXNB2 and CXorf21 | 2012  | 80 |
| ALL(Phlike) | 15                  | 231                   | WGS/mRNASeq | GAIIx/HiSeq | NR                      | NUP214-ABL1 fusion (2%), IK2F1 (67%) | 2012  | 81 |
| ALL (ETP) | 12                   | 94                    | WGS     | GAII     | 33                      | RAS pathway (67%), hematopoiesis and lymphoid development (58%), histone modification (42%) | 2012  | 40 |
| ALL (T) | 11                   | —                     | WGS     | HiSeq    | 55/15                   | DNMT3A (17%), JARID2 (8%), IDH2 (8%), EZH2 (17%) | 2012  | 82 |
| AML     | 1                    | —                     | WES     | HiSeq2000 | NR                      | Leukemic transformation from SCN to AML | 2012  | 83 |
| AML     | 8                    | —                     | WGS/deep sequencing | GAIIx  | 25/590                  | Clonal evolution | 2012  | 73 |
| AML     | 5                    | 160                   | WES     | GAIIx    | NR                      | GATA2 (39%) with biallelic CEBA mutation | 2012  | 84 |
| AML     | 2                    | 3                     | ChIPSeq | GAII/HiSeq | NR                      | Differential H3K4me3/H3K27me3 gene enrichment of stem and progenitor cells | 2012  | 85 |
| AML-CN  | 1                    | 95                    | mRNASeq | GAIIx    | NR                      | TLE4 (2%), SHKBP1 (2%) | 2011  | 86 |
| AML-CN  | 1                    | 262                   | WES     | GAIIx    | 69                      | BCR (4%), DNMT3A (13%) | 2011  | 87 |
| AML-CN  | 7                    | 230                   | mRNASeq | HiScanSQ  | 36                      | CBF2AT3-GLI2 fusion (6%) | 2013  | 88 |
| AML-M1  | 1                    | —                     | WGS     | GAII     | 33                      | First genome sequenced | 2008  | 2  |
| AML-M1  | 1                    | 187                   | WGS     | GAII     | 23                      | IDH1 (8%) | 2009  | 50 |
| AML-M1  | 1                    | 281                   | WGS     | GAII     | 39                      | DNMT3A (22%) | 2010  | 3  |
| AML-M5  | 14                   | 98                    | WES     | GAIIx    | 97                      | DNMT3A (21%) | 2011  | 89 |
| sAML    | 7                    | 200                   | WGS/WES | GAIIx/HiSeq | 34                      | UMODL1 (29%), SMC3 (14%), CDH23, ZSWIM4 (14%) | 2012  | 90 |
| BL      | 28                   | 78                    | mRNASeq | HiSeq2000 | NR                      | ID3 (59%), TCF3 (29%), CCND3 (15%) | 2012  | 91 |
| BL      | 4                    | 97                    | WGS/WES | GAIIx/HiSeq | 32/121                  | ID3 (42%) | 2012  | 92 |
| BL      | 14                   | 45                    | WES     | GAIIx/HiSeq | 47                      | ID3 (34%) | 2012  | 62 |
| CLL     | 4                    | 363                   | WGS/WES | Genome Sequencer FLX | 40/119                 | NOTCH1 (12%), MYD88 (3%) | 2011  | 59 |
| CLL     | 5                    | 226                   | WES     | Genome Sequencer FLX | 10                       | NOTCH1 (17%) | 2011  | 63 |
| CLL     | 3/88                 | 101                   | WGS/WES | GAII     | 38/132                  | SF3B1 (15%), MYD88 (10%) | 2011  | 36 |
| CLL     | 105                 | 279                   | WES     | GAIIx    | 62                      | SF3B1 (10%), NOTCH1 (10%) | 2012  | 37 |
| CLL     | 7                    | 103                   | RNASeq/WGS | GAII  | NR/12                  | YPELS-PP1/ICB fusion (95%) | 2013  | 32 |
| CLL     | 160                 | —                     | WES     | GAIIx/HiSeq | 112                      | Patterns of clonal evolution | 2013  | 76 |
| DLBCL   | 6                    | 105                   | WES     | Genome Sequencer FLX | 10                       | ML2 (24%), regulation of immune response (63% ABC, 31% CCB) | 2011  | 29 |
| DLBCL   | 13/83               | 37                    | WGS/mRNASeq | GAIIx/HiSeq | 32/41                  | ML2 (32%), MEF2B (11%), histone modification (13%), lymphocyte activation, differentiation, and apoptosis | 2011  | 28 |
| DLBCL   | 49                  | —                     | WES     | HiSeq    | 150                     | Histone H1 proteins (69%), ACTB (10%), P2RY8 (12%), PCLO (35%) | 2012  | 38 |
| DLBCL   | 34                  | 39                    | WGS/WES | GAIIx/HiSeq | 29/47                  | Signal transduction, chromatin modification | 2013  | 39 |
| FL      | 1/12                | 35                    | mRNASeq | GAIIx/HiSeq | 9/28                    | ML2 (13%), histone modification (15%), lymphocyte activation, differentiation and apoptosis | 2011  | 28 |
| HCL     | 1                    | 47                    | WES     | GAIIx    | 71                      | BRAF V600E (100%) | 2011  | 18 |
| MCL     | 18                  | 108                   | mRNASeq | GAII     | NR                      | NOTCH1 (12%), CCND1 (19%) | 2012  | 60 |
| MM      | 23/16               | 161                   | WGS/WES | GAII     | 33/104                  | Protein translation (42%), HOX9 pathway (29%) | 2011  | 19 |
| MM      | 22                  | 127                   | WES     | GAIIx    | 61                      | Distinct mutation patterns between t(4;14) and t(11;14) | 2012  | 75 |
| MM      | 1                   | —                     | WGS     | SOLiD/HiSeq | 30                      | Genomic evolution and clonal tides over course of disease | 2012  | 74 |
| MDS     | 9                    | 354                   | WES     | GAIIx    | NR                      | SF3B1 (67%) | 2011  | 34 |
| MDS     | 29                  | 582                   | WES     | GAIIx/HiSeq | 134                     | RNA splicing (55%) | 2011  | 33 |
| MDS     | 1                   | 150                   | WGS     | GAIIx/HiSeq | 39                      | U2AF1 (9%) | 2012  | 43 |
| MDS/MPN | 15                  | 310                   | WESmRNASeq | HiSeq  | NR                      | RNA splicing, SRSF2 (24%), clinical outcomes associated with mutations | 2012  | 35 |
almost 90% of cases. The vast majority of mutations had an association with deletion 11q (P < 0.001). Clinically, SF3B1 mutations were associated with fewer cytopenias and longer event-free survival. The high prevalence of SF3B1 mutations in diseases with ring sideroblasts and the confirmation that the mutation can be identified in peripheral blood suggest that SF3B1 could potentially be used as a biomarker.

SF3B1 mutation was also one of the most significant discoveries in MDS, found in 10–15% of cases. Mutations in SF3B1 were associated with deletion 11q (P = 0.004). Moreover,SF3B1 mutations and/or deletion 11q were predictive markers of an earlier need for treatment (P < 0.0001). Altogether, these results indicate that mutations of the spliceosome are involved in hematological malignancies and offer a novel therapeutic avenue for MDS and CLL.

### MUTATIONS MODULATING TRANSCRIPTION AND TRANSLATION

One of the most interesting themes arising from the study of hematological malignancies is that alterations of genes modulating transcription and expression are a recurrent finding. Sequencing studies in DLBCL and follicular lymphomas reported mutations in the histone modification process, particularly DNMT3A, a methyltransferase, is the most commonly mutated gene in follicular lymphoma, affecting more than 40% of cases. The vast majority of mutations had an inactivating effect and included missense and frameshift mutations affecting or truncating the C-terminal domains, including the SET domain. These findings place MLL2 collectively with the t(14;18)(q32;q21), as the two most common abnormalities in follicular lymphoma.

In addition, genes involved in histone modification were collectively identified in ~20–40% of DLBCL and early T-cell precursor acute lymphoblastic leukemia (ETP ALL). Furthermore, EZH2, which is involved in histone methylation, is mutated in DLBCL and follicular lymphoma, with the mutations occurring in a critical SET domain. Preclinical studies in DLBCL have found the inhibition of EZH2 an effective therapeutic approach for tumors containing activating mutations, thus presenting a novel therapeutic target for the treatment of DLBCL.

Chromatin modifiers are also recurrently affected in MM. MMSET, a histone methyltransferase transcriptional repressor, is over-expressed in ~15% of MM as a consequence of the t(4;14)(p16;q32). Sequencing studies show that other chromatin modifiers are mutated in a significant subset of MM, including KDM6A and HOXA9. In addition, in the analysis of MM there was an enrichment of mutations within genes involved in protein translation. Thus, 42% of MM cases had mutations in this pathway, mainly affecting FAM46C (13%), DIS3 (11%) and LRRK2 (8%).

A major finding in MDS and AML was the identification of mutations in a set of genes associated with DNA methylation. DNMT3A, a methyltransferase, is the most commonly mutated gene in AML found in around 20–30% of AML cases. Interestingly, no mutations were found in the related genes DNMT1, DNMT3B or DNMT3L. DNMT3A mutations were associated with poor survival (P < 0.001). In addition, mutations have been identified in U2AF1 in MDS patients, and those harboring U2AF1 mutations were more likely to progress to secondary AML.

### OTHER BREAKTHROUGH DISCOVERIES BY NGS: IDH1 AND IDH2 MUTATIONS IN AML

Another major discovery in AML was the identification of mutations in IDH1, which encodes isocitrate dehydrogenase 1, and the related IDH2 gene. IDH1 mutations have been observed in DLBCL and cartilaginous tumors, whereas IDH2 mutations have been reported in astrocytoma. Interestingly, patients with grade II astrocytoma who have IDH1 mutations show significantly shorter progression-free survival than tumors with wild-type IDH1. Studies in AML have reported mutations in 10–15% of cases, preferentially found in the intermediate-risk cytogenetic group, and their association with worse prognosis in a subset of AML patients that have been confirmed. Mutations in IDH1 and IDH2 are mutually exclusive and primarily affect IDH1 at codons R132 and IDH2 at codons R140 or R172. Mutations in IDH1 were enriched in cases possessing DNMT3A mutations. Conversely, IDH2 mutations are rarely found together with other known recurrent mutations. In addition, mutations in IDH1 and IDH2 seem to be mutually exclusive with TET2 mutations.
or IDH2 disrupted TET2 function and led to a hypermethylation phenotype with impaired hematopoietic differentiation.33

It becomes clear then that the morphological and clinicopathological classification of AML is now challenged by these new genetic findings. How many subcategories of AML exist? How does this heterogeneity exist, or not, in the better defined entities at the chromosome level (for example, M3)? In short, the various new perspectives to classify AML may ultimately lead more toward a molecular and pathway approach, but in some cases they might still have very significant resemblance to older cytogenetic classification.

OTHER BREAKTHROUGH DISCOVERIES BY NGS: NOTCH MUTATIONS

Aberrant NOTCH1 signaling has been identified in both solid and hematological tumors, and is a therapeutic target of interest currently in preclinical and clinical trials.54,55 NOTCH1 encodes a transcription factor that transduces extracellular signals into expression changes in targets genes, including MYC and P3K–AKT signaling pathways.57 NOTCH receptors are involved in cell fate determination, having a critical role in T-cell development. In fact, impaired NOTCH1 results in a block at the earliest stages of cell fate determination, having a critical role in T-cell development.58,59 Mutations in NOTCH1 lead to an active protein isoform lacking the C-terminal domain, and have been identified in over 50% of T-cell ALL and, to a lesser extent, in CLL, MCL and Burkitt’s lymphoma.56–62 These mutations mainly target the PEST domain, which is required for NOTCH1 interaction with FBW7, and subsequent NOTCH1 targeting for proteosomal degradation.

Data suggest that NOTCH1 mutations are a progressive event in CLL, increasing in prevalence from newly diagnosed CLL to chemorefractory CLL to CLL patients with Ritcher syndrome that underwent transformation to DLBCL.63 NOTCH1 mutations were associated with trisomy 12 (P = 0.009) and with IGHV-unmutated status.30,59,64 In addition to the association with more advanced stages of the disease and with transformation to DLBCL, NOTCH1 mutations were associated with adverse biological course and worse overall survival in CLL (P = 0.03)59,63 and MCL (P = 0.003).60

In T-cell ALL, NOTCH1 mutations were associated with improved response to glucocorticoid therapy; however, the association of NOTCH activation and clinical outcome seems to be therapy dependent.65–67

On the other hand, recurrent NOTCH2-activating mutations were identified in 21–25% splenic marginal zone lymphomas, but only rarely in nonsplenic MZLs and other low-grade B-cell lymphomas and leukemias.58,60 Although these studies evaluated the association of NOTCH2 mutations and clinical outcomes, the findings are conflicting and additional work is necessary to clarify the potential clinical impact of mutations in NOTCH2.58,61 Small molecule pan-NOTCH inhibitors have not shown significant effects as single agents targeting T-cell ALL, but there is an improved antileukemia effect when used in combination with inhibitors of P3K–AKT–mTOR pathway or CDK inhibitors.60,61

NGS AS A TOOL FOR DISCRIMINATION OF RELATED DISEASES

Besides the importance of identifying pathogenic mutations and pathways, sequencing is also a powerful tool to differentiate related entities. Overall, 67% of ETP ALL had mutations in the RAS signaling pathway (BRAF, JAK1, JAK3, KRAS, NRAS) or cytokine receptors (IL7R), which was significantly higher than that in non-ETP ALL (19%; P = 0.0001).40 Furthermore, genes involved in hematopoiesis and lymphoid development (RUNX1, IKZF1, ETV6, GATA3 and EP300) were also more frequently mutated in ETP ALL (58%) than in non-ETP ALL (17%; P < 0.0001). Altogether, 81% of ETP ALL cases have mutations in either of these pathways compared with 31% of non-ETP ALL cases (P < 0.0001). A similar enrichment was identified in genes involved in histone modification (EED, EZH2 and SUZ12), which were more commonly mutated in ETP ALL (42%) compared with non-ETP ALL (12%; P = 0.0001).

Mutations in genes affecting the RAS pathway, cytokine receptor and epigenetic modification are common in AML, but are rare in B- and T-cell neoplasias.40,72 These findings together with previous data demonstrate that ETP ALL has a gene expression signature closer to leukemic stem cells and granulocyte precursors, suggesting that ETP ALL is a distinct entity from non-ETP ALL with a less mature phenotype that retains the potential to become a myeloid cell.

GENOMIC SEQUENCING IN THE ANALYSIS OF CLONAL ARCHITECTURE AND CLONAL EVOLUTION

Genomic sequencing performed in high-coverage depth is a useful tool for characterizing the clonal architecture and analyzing the clonal evolution in disease progression and in response to therapy. Ding et al.73 have provided a good example of the power of genomic sequencing in sequential analysis. WGS was performed in eight AML cases utilizing normal skin biopsies paired with tumor samples collected at diagnosis and after relapse. Candidate somatic events were analyzed by deep sequencing with a median of 590X coverage. In five out of eight cases, the primary sample was characterized by up to four mutation clusters, thus indicating the existence of multiple (sub)clones. Two major patterns of clonal evolution were identified when comparing primary versus relapse samples. Either the original clone in the primary tumor sample acquired additional mutations and evolved into the relapse clone, or most of the (sub)clones were eradicated by therapy leaving one clone. This clone is usually observed at a low frequency in the primary sample, it then survives the initial therapy, gains additional mutations and expands, becoming the predominant clone at relapse. Another interesting finding was obtained by comparing the transition with transversion mutation rate between primary and relapse samples. The data obtained strongly suggest that the chemotherapy regimen used (cytarabine and anthracycline for induction and additional cytotoxic chemotherapy for consolidation) had a significant effect in the origin of novel mutations in the AML relapse sample.

In our sequencing analysis performed in MM, we have observed the presence of single nucleotide variants waxing and waning over the course of several longitudinally collected samples from a single patient.74 This shift in the presence of single nucleotide variants suggests the presence of multiple clones rising and falling in dominance over time. In the work by Walker et al.,75 a single MM patient, from whom they had WES data, identified three populations containing mutations in four genes: ATM, FSP2, CLTC and GLMN. When they then evaluated which mutations were shared in a single cell, they identified one population with only an ATM mutation, a second with ATM and FSP2 and a third with ATM, CLTC and GLMN mutations.72 The observed presence of these different clones suggests that if these patients were to be followed longitudinally, clonal dominance would likely shift as the tumor evolves with time and treatment.

The study of clonal complexity and clonal evolution is an old field that has been reinvigorated since the introduction of NGS. We believe NGS will help us to elucidate several unanswered questions such as what are the driver-initiating mutations in the different hematological malignancies? What are the specific mutations associated with disease progression in the different hematological malignancies? What mutations are the primary contributors to chemoresistance? Does clonal heterogeneity need to be considered in the context of determining therapeutic options? Do all the clones need to be targeted?

Some of these questions have been at least partially answered in a recent study.76 The authors analyzed 149 CLL cases, including 18 that were analyzed at two time points, using WES and SNP arrays. Data obtained from this study confirm previous findings.

© 2013 Macmillan Publishers Limited

Blood Cancer Journal
showing the existence of linear and multibranching clonal evolution in CLL. Furthermore, the authors were able to infer the order of genetic changes occurring in CLL pathogenesis. Thus, it was suggested that the clonal driver mutations, which are proposed to be initiating events, mainly affect genes that selectively affect B cells, such as MYD88 and del13q, whereas subclonal driver mutations associated with disease progression affect genes more ubiquitously involved in carcinogenesis such as TP53 and ATM. The number of subclonal mutations increases in treated compared with untreated cases; thus, the therapy would be a trigger for natural selection leading to the emergence of more aggressive subclones. Furthermore, the study shows the importance of subclonal driver mutations as an independent risk factor for rapid disease progression and poor outcome. Thus, this study suggests that dissecting the clonal architecture of CLL is crucial not only for developing novel risk-stratification algorithms but also for designing novel therapeutic approaches, considering the presence of driver mutations as well as the genomic landscape.

We expect to see similar efforts in several other hematological malignancies, which will ultimately help to elucidate the clonal complexity and its importance in each particular disease.

HOW ARE NEW DISCOVERIES TRANSLATING INTO NOVEL THERAPEUTIC APPROACHES?

Several novel somatic mutations, such as SF3B1, IDH1, IDH2, DNMT3A, MYD88 and MLH2 have been identified as a consequence of NGS efforts, leading to the discovery of previously unrecognized genes and molecular processes/pathways with pathogenic effects. The genomic profiling of each individual cancer will potentially have a key role clinically assisting in early disease diagnosis, risk stratification, longitudinal analyses of tumor evolution and selection of the most favorable and personalized therapeutic intervention.

One of the most emblematic examples is AML. The unprecedented characterization of the AML cancer genome may substantially affect the clinical management and the therapeutic decisions. The prior characterization of mutations in FLT3, NPM1, RUNX1 and CEBPA together with the recent identification of mutations in IDH1, IDH2, DNMT3A and TET2 encourage the incorporation of genomic studies as part of routine clinical tests and may enable optimization of therapeutic plans based on this patient-specific genomic background.

However, the genetic characterization of AML will not improve patient survival per se, unless it is synchronized with the development of alternative therapeutic approaches. One of the major limitations in the treatment of AML is the intrinsic drug resistance of the tumor cells. Standard induction chemotherapy regimens, consisting of cytarabine and anthracycline combinations, have remained largely unchanged in the treatment of AML over decades. Thus, the major challenge is to provide the AML patients with alternative drug combinations targeting novel genes/pathways discovered in chemoresistant cases.

The discovery of novel genes/pathways not only increases our understanding of the pathogenesis of the disease but also opens new therapeutic avenues. The existence of potential ‘Achilles’ heels to be exploited for generating a unifying targeted therapy for all patients is very provocative and opens an exciting era for translational research. Exploiting this knowledge is critical in hematological malignancies when considering that most of them are still incurable and more effective therapies are urgently needed.

So far, we have discovered a different range of genetic heterogeneity across tumor types. We have learnt that some malignancies have a mutated gene or pathway that affect most or all cases. An excellent example is provided by the BRAF V600E, common to all HCL patients, or MYD88 L265P, found in most WM. For example, V600E BRAF can be targeted with BRAF inhibitors alone or in combination with MEK or ERK inhibitors.

Conversely, the majority of hematological malignancies are characterized by considerable tumor heterogeneity, making the search for therapeutic targets more difficult. One of the biggest challenges is to reduce the complexity of the generated data by first, distinguishing the driver over the passenger mutations and clones, and second, generating systematic and more sophisticated approaches for data analysis integration, thus unifying the vast genomic heterogeneity of these cancers into more homogeneous groupings based on cellular pathways rather than on single genes. As in the case of single gene mutations, the disruption of specific pathways may be exploited therapeutically.

With the dawn of the $1000 genome drawing close, we anticipate an ever-increasing role of genomic sequencing in the diagnosis and treatment of patients. As this technology moves ever closer to widespread clinical application, there are several challenges that must be addressed. First, the management of the data obtained from sequencing must be addressed. Not only does the physical storage of the data present a challenge but how the information obtained is reported to the patient, retained over time and/or destroyed, are important issues that also must be discussed. Second, incidental findings of mutations in genes unrelated to the medical reason, a patient is seeking genome sequencing can result in legal and ethical dilemmas for the care providers. Furthermore, knowledge about genomics and disease is rapidly expanding, thus communication of what the patient’s genome should be re-evaluated at a later time must be considered.

As our focus shifts from large population-based studies with large cohorts to the ‘N of one’ with individualized genomic medicine, it is imperative that the recommendations made to patients be based on evidence from well-designed functional studies. Translating this genetic data into the clinic is challenging and a significant amount of functional work is still required to better understand the biological significance of these hits using both in vitro and in vivo models. In the near future, we anticipate a standard of care for personalized medicine that involves sending samples for sequencing at the time of biopsy. A variant report will be generated for the physician who will then base treatment decisions on the findings from sequencing in addition to pathology and clinical diagnostics.

The ultimate goal in the post-genomic era will be to extend to other hematological malignancies the successful transition from gene discovery to therapeutic intervention observed in the paradigmatic BCR-ABL CML cases treated with imatinib.
55 Wei P, Wallis M, Oiu M, Ding R, Denlinger RH, Wong A et al. Evaluation of selective gamma-secretase inhibitor PF-03084014 for its antitumor efficacy and gastrointestinal safety to guide optimal clinical trial design. Mol Cancer Ther 2010; 9: 1618–1628.

56 Palomero T, Lim WK, Odom DT, Sulis ML, Real PJ, Margolin A et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. Proc Natl Acad Sci USA 2006; 103: 18261–18266.

57 Ciofani M, Zuniga-Pflucker JC. Notch promotes survival of pre-T cells at the beta-selection checkpoint by regulating cellular metabolism. Nat Immunol 2005; 6: 881–888.

58 Love C, Sun Z, Jima D, Li G, Zhang J, Miles R et al. Whole-transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. Blood 2012; 119: 1963–1971.

59 Weng AP, Ferrando AA, Lee W, JPT Morris, Silverman LB, Sanchez-Irizarry C et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science 2004; 306: 269–271.

60 Kridel R, Meissner B, Rogic S, Boyle M, Telenius A, Woolcock B et al. NOTCH1 and NOTCH2 mutations of t(4;14) and t(11;14) myeloma. Blood 2012; 119: 1321–1325.

61 Tanigaki K, Honjo T. Regulation of lymphocyte development by Notch signaling. Blood 2004; 103: 3469–3472.

62 Egan JB, Shi CX, Tembe W, Margolin A et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. Proc Natl Acad Sci USA 2006; 103: 18261–18266.

63 Beekman R, Valkhof MG, Sanders MA, van Strien PM, Haanstra JR, Broeders L et al. Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. Blood 2012; 119: 5071–5077.

64 Kadowaki T, Arai Y, Kidokoro Y, Soda M, Yamaguchi Y, Shima H et al. The genetic landscape of leukemia reveals complex evolution of clonal architecture at disease progression and at the time of relapse. Leukemia 2012; 26: 1698–1701.

65 Knight SJ, Yau C, Clifford R, Timbs AT, Sadghii Akha E, Dreau HM et al. Quantification of subclonal distributions of recurrent genomic aberrations in paired treatment and relapse samples from patients with B-chronic lymphocytic leukemia. Leukemia 2012; 26: 1564–1575.

66 Xia L, Sohani AR, Arcaini L, Hunter Z, Yang G, Zhou Y et al. A somatic variant in MYD88 (L265P) revealed by whole genome sequencing differentiates lymphoplasmacytic lymphoma from marginal zone lymphomas. Blood (ASH Annual Meeting Abstr) 2011; 118: Abstract 261.

67 Iacobucci I, Ferrarini A, Sazzini M, Giacomelli E, Lonetti A, Xumerle L et al. Application of the whole-transcriptome shotgun sequencing approach to the study of Philadelphia-positive acute lymphoblastic leukemia. Blood Cancer J 2012; 2: e61.

68 Roberts KG, Morin RD, Jiang J, Hirst M, Zhao Y, Su X et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. Cancer Cell 2012; 22: 153–166.

69 Simon C, Chagraoui J, Krol J, Gendron P, Wilhelm B, Lemieux S et al. A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. Genes Dev 2012; 26: 651–656.

70 Michalik L, Veldman T, Nowell PC, Schlette WE, Schiller GN et al. The coding network promoting leukemic cell growth. Nat Immunol 2007; 8: 881–888.

71 Bartram CR, Ziegler-Heitbrock HW, Kindermann HW, Siebert R. Clonal evolution and at the time of relapse. Leukemia 2012; 26: 1698–1701.

72 Braggio E, Kay NE, Vanwier S, Tschumper RC, Smoley S, Eckel-Passow JE et al. Longitudinal genome-wide analysis of patients with chronic lymphocytic leukemia reveals complex evolution of clonal architecture at disease progression and at the time of relapse. Leukemia 2012; 26: 1698–1701.

73 Knight SJ, Yau C, Clifford R, Timbs AT, Sadghii Akha E, Dreau HM et al. Quantification of subclonal distributions of recurrent genomic aberrations in paired treatment and relapse samples from patients with B-chronic lymphocytic leukemia. Leukemia 2012; 26: 1564–1575.

74 Xia L, Sohani AR, Arcaini L, Hunter Z, Yang G, Zhou Y et al. A somatic variant in MYD88 (L265P) revealed by whole genome sequencing differentiates lymphoplasmacytic lymphoma from marginal zone lymphomas. Blood (ASH Annual Meeting Abstr) 2011; 118: Abstract 261.

75 Iacobucci I, Ferrarini A, Sazzini M, Giacomelli E, Lonetti A, Xumerle L et al. Application of the whole-transcriptome shotgun sequencing approach to the study of Philadelphia-positive acute lymphoblastic leukemia. Blood Cancer J 2012; 2: e61.

76 Love C, Sun Z, Jima D, Li G, Zhang J, Miles R et al. The genetic landscape of mutations in Burkitt lymphoma. Nat Genet 2012; 44: 1321–1325.

77 Fabbri G, Rasi S, Rossi D, Trifonov V, Khlabianin H, Ma J et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med 2011; 208: 1389–1401.

78 Xiao X, Wei P, Wallis M, Oiu M, Ding R, Denlinger RH et al. Whole-genome sequencing identifies recurrent NOTCH1 mutations in mantle cell lymphoma. Blood 2012; 119: 1321–1325.

79 Weng AP, Ferrando AA, Lee W, JPT Morris, Silverman LB, Sanchez-Irizarry C et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science 2004; 306: 269–271.

80 Love C, Sun Z, Jima D, Li G, Zhang J, Miles R et al. The genetic landscape of mutations in Burkitt lymphoma. Nat Genet 2012; 44: 1321–1325.

81 Fabbri G, Rasi S, Rossi D, Trifonov V, Khlabianin H et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med 2011; 208: 1389–1401.

82 Xiao X, Wei P, Wallis M, Oiu M, Ding R, Denlinger RH et al. Whole-genome sequencing identifies recurrent NOTCH1 mutations in mantle cell lymphoma. Blood 2012; 119: 1321–1325.

83 Beekman R, Valkhof MG, Sanders MA, van Strien PM, Haanstra JR, Broeders L et al. Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. Blood 2012; 119: 5071–5077.

84 Greaf PA, Dufour A, Konstandin NP, Sizemery B, Zeilmeier E, Tizazu B et al. GATA2 zinc finger 1 mutations associated with biallelic CEGBPA mutations define a unique genetic entity of acute myeloid leukemia. Blood 2012; 120: 395–403.

85 Yamazaki J, Estecio MR, Lu Y, Long H, Malouf GG, Graber D et al. The epigenome of AML stem and progenitor cells. Epigenetics 2013; 8: 92–104.

86 Greaf PA, Eck SH, Konstandin NP, Benet-Pages A, Sizemery B, Dufour A et al. Identification of recurrent tumor-specific somatic mutations in acute myeloid leukemia by transcriptome sequencing. Leukemia 2011; 25: 821–827.

87 Groszmann V, Tiacci E, Holmes AB, Kohlmann A, Martelli MP, Kern W et al. Whole-exome sequencing identifies somatic mutations of BCOR in acute monocytic leukemia but not restricted to FAB M7 subtype. Blood 2013; 121: 3469–3472.

88 Yan JX, Xu J, Gu ZH, Pan CM, Lu G, Shen Y et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet 2011; 43: 309–315.

89 Walter MJ, Shao J, Keboldt DC, Chen K et al. Clonal architecture of secondary acute myeloid leukemia. N Engl J Med 2012; 366: 1090–1098.

90 Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature 2012; 490: 116–120.

91 Richter J, Schlesner M, Hoffmann S, Kreuz M, Leich E, Burkhardt B et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nat Genet 2012; 44: 1316–1320.

92 Puda A, Milosevic JD, Berg T, Klamptf T, Harutyunyan AS, Gisslinger B et al. Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. Am J Hematol 2011; 87: 245–250.

93 Steidl C, Shah SP, Woolcock BW, Siebert R, Schmitz R et al. Clonal evolution and at the time of relapse. Leukemia 2012; 26: 1698–1701.

94 Steidl C, Shah SP, Woolcock BW, Habbiba N, Rossetto M, Lennert K et al. Recurrent mutations of MYD88 and TBL1XR1 in primary central nervous system lymphomas. Clin Cancer Res 2012; 18: 5203–5211.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/