

A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects

Bas J. Wouters,¹ Bob Löwenberg,¹ and Ruud Delwel¹

¹Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

The past decade has shown a marked increase in the use of high-throughput assays in clinical research into human cancer, including acute myeloid leukemia (AML). In particular, genome-wide gene expression profiling (GEP) using DNA microarrays has been extensively used for improved understanding of the diagnosis, prognosis, and pathobiology of this heterogeneous disease. This review discusses the progress that has been made, places the technologic limitations in perspective, and highlights promising future avenues. (Blood. 2009;113:291-298)

Introduction

Acute myeloid leukemia (AML) is characterized by a maturation block and accumulation of myeloid progenitor cells.^{1,2} Clinically, it has been recognized as a heterogeneous disorder.^{1,2} Laboratory support for that notion has come from various directions. Chromosomal abnormalities and gene mutations are common in AML, many of which are apparent in particular subtypes.^{3,4}

Classification of AML subtypes is clinically relevant, as particular abnormalities are associated with distinct clinical behavior.¹ For instance, recurring reciprocal translocations t(15;17)(q22;q21), t(8;21)(q22;q22), or inv(16)(p13q22)/t(16)16(p13;q22), further abbreviated as t(15;17), t(8;21), and inv(16), respectively, predict favorable prognosis, whereas other chromosomal aberrations are associated with inferior outcome.¹ Likewise, sequence mutations in certain genes are associated with either favorable or unfavorable response to treatment.⁴

Insight into cytogenetic and genetic aberrations is invaluable for diagnosis, and it may also allow for better understanding of the pathobiology. Furthermore, it may enable the development and application of specific treatment modalities targeted to underlying oncogenic abnormalities. The efficacy of such drugs as all-*trans* retinoic acid for the treatment of t(15;17) AML and imatinib for BCR-ABL-positive chronic myeloid leukemia offer well established examples.⁵

Despite great progress, much of the heterogeneity of AML remains to be resolved. A significant proportion of human AML appears cytogenetically and genetically normal, which implies that the underlying molecular abnormalities are still unknown. Furthermore, it is likely that in AMLs carrying recognizable aberrations additional hits await to be uncovered, as one lesion usually does not appear sufficient for full leukemic transformation.⁶

In recent years, DNA microarrays, together with the availability of the complete nucleotide sequence of the human genome, have spurred the search for abnormalities in cancer, including AML.⁷ The accessibility of these tools allows assessment of abnormalities and variations on a genome wide basis, covering various molecular levels.⁸ Gene expression profiling (GEP) is one of these technologies, in which DNA microarrays containing cDNAs or oligonucle-

otide probes are used to simultaneously measure levels of many different mRNA transcripts.⁷⁻¹⁰ In an early landmark study in 1999, Golub and colleagues were able to use GEP to discriminate a collection of AML from acute lymphoblastic leukemia (ALL) specimens.¹¹ Their study suggested 3 important potential applications of GEP: class discovery, class prediction and class comparison. Class discovery refers to the identification of new subgroups while for class prediction one uses gene expression data to predict already defined subgroups. These 2 applications therefore have diagnostic implications. The third proposed application, class comparison, refers to the identification of genes that are deregulated in certain subgroups, and may address biologic questions. Given these proposed possibilities, investigators have used DNA microarrays to investigate gene expression in clinical AML samples in the past years.¹²⁻¹⁴

Now, some 10 years later, we shall discuss the results of a decade of experience with GEP as regards diagnosis, prognosis and biology of AML in the context of clinical studies. Is it possible to identify new clinically relevant subgroups of AML using GEP? Does GEP deserve a place in clinical diagnosis? Does GEP allow prediction of prognosis? And is it possible to extract new insights into the pathobiology of AML from clinical GEP data?

Can GEP identify new subgroups of AML?

A straightforward way of using GEP is to compare expression profiles between cases of AML and to search for similarities and differences. In an unsupervised approach this is done in an unbiased way, ie without the use of external information such as mutations or karyotypic subtypes. This procedure is therefore representative of class discovery. The grouping of cases according to similar gene expression signatures is often referred to as clustering.^{15,16} The underlying assumption is that cases with the same gene expression profiles may carry the same genetic abnormality. Support for this hypothesis has come from observations, now confirmed by several research groups, that particular cytogenetic

Submitted April 29, 2008; accepted August 13, 2008. Prepublished online as *Blood* First Edition paper, August 14, 2008; DOI 10.1182/blood-2008-04-153239.

The publication costs of this article were defrayed in part by page charge

payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2009 by The American Society of Hematology

AML subtypes (eg, AMLs with t(8;21), t(15;17), and inv(16)) each share distinctive GEP profiles (Figure 1A).¹⁷⁻²³ Likewise, mutations in CCAAT/enhancer binding protein alpha (*CEBPA*), and to a lesser extent nucleophosmin (*NPM1*) (Figure 1A), correlate with gene expression signatures that appear after unsupervised clustering.^{22,24}

The potential of GEP to uncover new subgroups has been illustrated for several types of cancer, such as cutaneous malignant melanoma, diffuse large B-cell lymphoma, breast cancer, and acute lymphoblastic leukemia.²⁵⁻²⁸ In AML, several GEP cohort studies have been performed in the last few years. In many of those, new subgroups were discerned. In a cohort of 166 cases of AML, 2 subgroups of normal karyotypes with distinguishable expression profiles were identified.²¹ The investigators postulated that this subdivision of the normal karyotype group could be diagnostically relevant, but noted an association of the 2 subgroups with known factors: internal tandem duplications (ITDs) in fms-like tyrosine kinase 3 (*FLT3*) and FAB-M4 and M5 monocytic leukemias, respectively. In another dataset, of 285 AMLs, 16 subgroups were recognized, several of which lacked previously known denominators.²² A different study divided 170 cases of AML, mostly patients

of older age lacking favorable cytogenetic features, into 6 subgroups, some of which appeared novel.²⁹

Unsupervised analyses have also been performed to uncover heterogeneity within established AML subtypes. In a set of 166 AMLs, the core binding factor (CBF) AMLs, ie cases with t(8;21) or inv(16), each could be split into subgroups merely based on the GEP data, which was subsequently reproduced in another study.^{21,30} These observations suggested that CBF leukemias in terms of gene expression patterns may represent heterogeneous entities. In 130 cases of pediatric AML, heterogeneity within CBF leukemias was seen as well, particularly within AML inv(16).²³

In each of the above studies, GEP revealed previously unrecognized heterogeneity of AML. But is this diversity relevant? As cluster algorithms, by definition, focus on similarities between cases, a newly identified group of AMLs does not necessarily have biologic or clinical importance. The biologic significance of a newly identified subgroup would be convincingly substantiated by the subsequent discovery of a related underlying defect or correlation with a characteristic clinical phenotype or treatment response. Alternatively, validation of a signature in one or more independent datasets would provide support that the novel subgroup is stable.

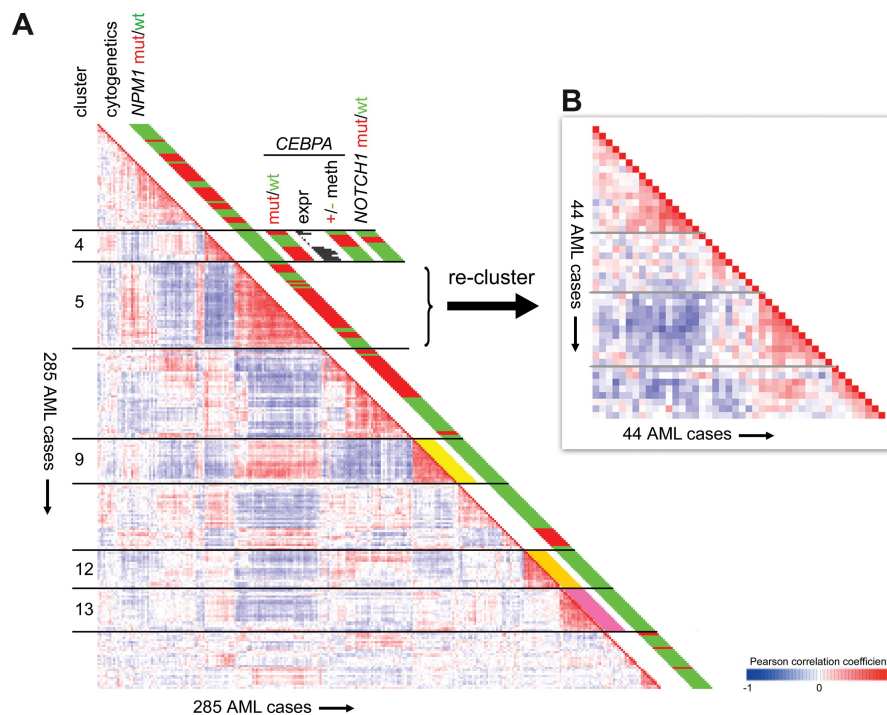


Figure 1. Summary of GEP findings in a cohort of 285 cases of AML. (A) A previous study of 285 cases of AML revealed 16 subgroups (clusters) of cases based on similarities in gene expression profiles.²² Pairwise correlations between these AML cases are shown on the left. The cells in the visualization are colored by Pearson correlation coefficient values, with deeper colors depicting higher positive (red) or negative (blue) correlations, as indicated by the scale bar. Five of the 16 clusters have been labeled as clusters 4, 5, 9, 12, and 13. One finding of the original study was the tight aggregation into distinct clusters of AML cases with cytogenetic abnormalities that predict good risk. For those cases, cytogenetic status is color-coded in the cytogenetics column: inv(16) is yellow, next to cluster 9; t(15;17) is orange, next to cluster 12; and t(8;21) is pink, next to cluster 13. A subsequent study in the same patient cohort identified *NPM1* mutations in 95 of 285 cases. *NPM1* mutational status is depicted next to each case (red indicates *NPM1* mutant; green, *NPM1* wild-type).²⁴ The figure illustrates that *NPM1* mutations were not randomly distributed over the 16 previously defined clusters, but enriched in several of them. Cluster 4 was found to associate with *CEBPA* mutations (red indicates *CEBPA* mutant). However, a subset of 6 patients in this cluster did not show any *CEBPA* mutation (green indicates *CEBPA* wild-type). It was found that these cases differed in their *CEBPA* mRNA expression as compared with the *CEBPA* mutant AMLs, as indicated by the histograms depicting signal intensity values for the *CEBPA* probe set on the microarray. In fact, whereas *CEBPA* mutant AMLs highly expressed *CEBPA* mRNA, expression was silenced in the cases lacking mutations. This silencing was associated with *CEBPA* DNA promoter hypermethylation (red indicates methylation; green, no methylation). In addition, *NOTCH1* mutations were found as common characteristics of this subgroup (red indicates *NOTCH1* mutation; green, *NOTCH1* wild-type).³¹ (B) In the original analysis of 285 AML cases (panel A left), the 44 cases in cluster 5 aggregated very tightly, as indicated by the deep red colors representing positive Pearson correlation coefficients. Most of these 44 cases showed a monocytoid morphology (FAB-M4 or -M5).²² This raises the possibility that a significant part of the clustering effect was caused by specific up- or down-regulation of genes that are important in monocytic differentiation, resulting in a different signature than the remaining, mostly nonmonocytoid, cases of AML in the study. To answer whether gene expression profiling would enable identification of potential heterogeneity within this apparently homogeneous subgroup, in panel B the 44 cases were reclustered as an isolated cohort. For this analysis, only probe sets that showed a variable expression within these 44 AML cases were taken into account, as defined by a fold change of 3.5 to the mean in log₂ scale in at least 1 case. The resulting cluster image shows that several potentially interesting subgroups can indeed be identified within these 44 AML cases, which have been indicated by gray lines.

Evidence indicating that the detection of a distinctive gene expression subtype indeed can lead to the discovery of a biologically and clinically significant subgroup has recently been demonstrated by the identification of a distinct form of leukemia characterized by epigenetic *CEBPA* silencing and an immature myeloid/T-lymphoid phenotype (Figure 1A).³¹ Importantly, these findings could be confirmed in an independent cohort of human AML. Studies like these show how complementing technologies can be used to make full use of the wealth of GEP data for subgroup discovery. Evidently, for successful subgroup discovery it is important to have access to sufficiently large series of cases that represent the variable subtypes of AML. Compatibility of platforms and data sharing through online repositories can facilitate the latter.^{32,33}

Interpretation of the results of GEP studies for class discovery

While GEP has been demonstrated to offer robust and reproducible technology in the analysis of cohorts of AML,²¹⁻²³ one should remain well aware of the factors that may affect the results. We discuss here some of the most important of those factors in the context of class discovery, although many will also affect class prediction and class comparison.

Differences in study design may determine the probability of the discovery of disease heterogeneity. Variations in the selection of study populations, in terms of size as well as demographic diversity, will have a direct impact on experimental results. Furthermore, various technical differences between studies may influence results, ranging from sample processing and mRNA isolation to microarray hybridization and analysis. Interstudy variations with regard to the bioinformatic and biostatistical approaches, which involve choices regarding data normalization, gene filtering, and clustering procedures, may exert marked effects on outcome of the analysis.^{10,34} Most cluster algorithms present results in a 2-dimensional way, in which slight changes in calculation may move samples from one cluster to another. Such analytical differences will most likely not have a major influence on subgroups with very distinctive signatures, but may have an effect on the identification of more subtle differences.

It is important to keep in mind that GEP based clustering is driven by similarities and differences in gene expression profiles between samples. The similarities *could* be caused by shared underlying genomic defects—which is what most researchers are primarily interested in—but could also be caused by factors that are not directly related to pathogenetic mechanisms, for example, similarities in maturation phenotype of the leukemias. This has indeed been observed in AML. In 2 relatively large studies, AMLs with monocytoid morphologies (FAB-M4-M5 leukemias) showed a tendency to aggregate according similar gene expression patterns.^{21,22} The differences in predominant maturation stage of the leukemias may also have influenced the aggregation of *CEBPA* mutant cases into 2 gene expression clusters, one cluster including cytologically more immature leukemias.^{22,31} Such pitfalls in the analysis of GEP for subgroup discovery may explain why the number of novel subtypes of AML reliably identified by GEP have remained relatively limited.

How can the potential of GEP for discovery of novel subgroups of AML be enhanced? In the search for pathogenetically relevant differences between leukemias, one would wish to avoid the

interference of effects of phenotypic differences. One way of achieving this would be to restrict the analysis to relatively homogeneous populations. The feasibility of avoiding unwanted background differences that might obscure interesting pathogenetic differences in a selected predefined AML group was demonstrated by investigators who established 2 gene expression profiles of non-Down syndrome-associated acute megakaryoblastic leukemia.³⁵ Similarly, it has been noted that distinct subgroups can be found within the preselected t(15;17) AML subtype.³⁶ Likewise, the 44 AMLs of a previously established cluster that strongly associated with FAB-M4 and M5 leukemias exhibited notable internal heterogeneity when studied as an isolated population (Figure 1B). Studies dealing with purified progenitor cells will also be instrumental for exclusion of interfering transcriptional background. Accumulating evidence suggests the existence of leukemia stem cells.^{37,38} Profiling of those cells, instead of total blast populations, may enhance the possibilities of GEP for subgroup discovery. Focusing on the leukemic stem cell may also reveal stronger transcriptional profiles that may be buried only in subclones. However, this approach directly depends on the accepted definition of immunophenotypic markers of leukemic stem cells and suffers from the technical drawback of rare stem cell numbers. Another opportunity for class discovery lies in the application of improved analytical procedures. A notable example is the use of pathway oriented analyses, which allow the discovery of distorted functional networks of genes as opposed to gene-based approaches.^{39,40}

Do genome-wide gene expression analyses deserve a place in clinical diagnosis?

Several techniques are currently used in the initial diagnosis of AML, including cytology, immunophenotyping, karyotyping, polymerase chain reaction (PCR), and fluorescence in situ hybridization (FISH). Because a GEP-based approach allows detection of many transcripts at the same time, it provides a transcriptional snap shot of the leukemia. Several research groups have investigated the possibility to define specific gene expression classifiers (also referred to as class predictors) for disease subtypes, eg a discriminative set of genes for AML with the t(8;21) translocation. This procedure of class prediction through the generation of classifiers differs from class discovery, as discussed in the first section, in one important aspect: it makes use of external information, such as absence or presence of t(8;21), to derive a signature that can subsequently be used for prediction of samples of leukemia of which the status is not known yet. This type of approach is, therefore, often referred to as supervised.

AMLs defined by t(8;21), AML inv(16) and acute promyelocytic leukemia with t(15;17) have consistently been found to be predictable using gene expression classifiers with almost 100% accuracy.^{17,22,23,41} For AMLs with 11q23 rearrangements involving the mixed lineage leukemia (*MLL*) gene, reported prediction accuracy in a study within several types of human leukemia was approximately 90%, while this was 95% in a study on pediatric AML.^{23,41} Efforts to derive gene expression classifiers for AMLs with other chromosomal abnormalities have as yet met little success. Thus, for instance AML with chromosomal abnormalities involving trisomy 8, complex karyotype and 3q appear not to be accurately predictable by GEP in representative cohorts of AML.^{20,21,41}

Prediction of mutations in *CEBPA* and *NPM1* has also been pursued.^{22,24} In one study, correct prediction of most *NPM1* mutation positive cases was possible, albeit at the expense of a significant number of false positives, resulting in a positive predictive value of 70% to 75%.²⁴ This would create a hurdle when GEP would be used for diagnostic purposes. For *CEBPA* mutations, prediction accuracy may be dependent on the type of mutation. While biallelic mutations, which are relatively frequent, appear to be predictable with a gene expression classifier with high positive and negative predictive values, heterozygous monoallelic mutations may not be accurately predicted.⁴²

Likewise, abnormalities in signaling molecules such as *FLT3* and *RAS* appear not to be readily predictable within diverse AML cohorts.²² This may not be too surprising taking into account their less direct role in transcriptional modulation. In some studies, reasonably successful prediction of *FLT3*-ITD status was achieved when cytogenetically normal AMLs were selectively analyzed.^{43,44} A recently reported classifier for *FLT3*-ITD mutation status in normal karyotype AML showed only a modest performance in predicting *FLT3*-ITD status in a validation cohort of 72 normal karyotype cases, with both a relatively high number of false positives and false negatives (sensitivity 73%, specificity 85%).⁴⁵

Taken together, these observations suggest that, using current methodology, definition of GEP-based classifiers is only feasible for selected AML subtypes in which the underlying molecular abnormality is not too distantly involved in transcriptional modulation. Those abnormalities, involving t(8;21), inv(16), and t(15;17), can be identified by rapid and widely used methods such as PCR as well. This may raise the important question what the ultimate clinical utility of GEP classifiers will be. The particular value of GEP-based classification lies in its comprehensiveness (ie, the opportunity to perform many tests simultaneously), for instance using specifically designed diagnostic DNA microarrays. Evaluation of larger representative AML patient series will be needed to reveal whether it is possible to define additional classifiers for less frequent cytogenetic or molecular subtypes. Definition of such additional classifiers will enhance the attractiveness of GEP as a diagnostic assay.

in a set of adult AML cases. The same investigators were not able to confirm the value of a prognostic signature for pediatric AML that had been proposed by a different group.⁴⁹

While the above prognostic predictors were constructed without any a priori biologic assumption, a hypothesis-driven approach was used to construct a predictor consisting of 11 genes associated with a stem cell–like expression pattern. Those genes were chosen because of their relationship to *BMI1* activation in a murine prostate cancer model and in human cancer samples.⁵⁰ The predictor recognized adverse outcome in several types of human cancer including AML, but its value awaits independent evaluation.

Will these studies lead to the introduction of GEP-based prognostic tests for AML in clinical practice? Experience in the field of breast cancer and diffuse large B-cell lymphoma may exemplify the use of GEP-based prognostication.^{51–56} For breast cancer, 2 prognostic predictors were reported recently by independent groups.^{53,54} Initially these GEP-based predictors met some skepticism because of their minimal overlap in genes. Independent validation studies, however, subsequently confirmed prognostic value for both signatures, independent of other available prognostic markers.^{52,57} The very limited overlap in genes between the 2 predictors is most likely explained by the fact that in a typical GEP experiment, many transcripts are more or less similarly correlated to outcome. Consequently, several combinations of genes are equally informative for prognosis.⁵⁸ The success of the breast cancer predictors has led to the recent approval by the Food and Drug Administration of a commercial test that is based on one of them.^{56,59}

From currently available evidence it appears that GEP predictors for prognosis can probably be established for AML. This will have the advantage of a comprehensive prognostic test that can substitute several currently used prognostic markers. However, it is still necessary to better assess whether such GEP based predictors can add information over currently available cytogenetic and molecular markers for prognosis of AML.^{46,60} Validation of any prognostic signature in sufficiently large series will be necessary before accepting this as a solid and reliable indicator of prognosis.^{46,58,61–65}

Is genome-wide gene expression analysis useful for predicting prognosis?

Several attempts have been made to derive prognostic signatures for AML. This approach is similar to class prediction, as discussed in the previous section, but instead of subgroup status (eg, t(8;21)), outcome (eg, overall survival) is used as the end point to define a prognostic predictor.^{46,47}

In one study, a set of 133 genes was demonstrated to predict survival among adults with normal karyotype AML.²¹ A second group of investigators converted this signature into a prognostic predictor and confirmed its prognostic ability as regards overall and disease-free survival in another series of normal karyotype AML.⁴⁸ Because of differences in DNA microarray platforms, the investigators could only verify 81 of the original 133 genes, so that a complete validation of the original signature was not possible.³⁴ A significant part of the prognostic effect was associated with *FLT3*-ITD mutations.⁴⁸

For relapse of pediatric AML, a 2-gene predictor has been proposed.²³ The predictive value of this indicator appeared to be modest in a small validation subset of the pediatric cohort as well as

What can genome-wide gene expression analyses tell about the biology of AML?

An additional quality of genome-wide assessment of mRNA levels of tens of thousands of genes is that it may allow for the discovery of pathobiologic pathways. Unbiased genome-wide GEP in AML may identify critical downstream targets of known oncogenes or tumor suppressors or identify novel causative genetic abnormalities. This concept has been illustrated to be particularly promising in animal or cell line cancer models in which an oncogene of interest had been introduced. Recent examples include models for acute promyelocytic leukemia,⁶⁶ mutant *Cebpa*⁶⁷ and *Mll-Af9*.⁶⁸ Indeed, these studies demonstrate that biologic information is captured within the thousands of transcripts measured.

Although the search for critical abnormalities in clinical AML sounds straightforward, in practice there may be nontrivial hurdles on the way to the discovery of genomic abnormalities playing a key role in the pathophysiology. There may be numerous differences in gene expression between AML cases, but only few of them may be truly disease pathogenesis-related. How to extract the biologically important transcriptional differences from the large amount of

data? One possible approach is the supervised class comparison strategy, which typically involves 3 steps. First, a subgroup of AML of interest is defined. Next, the gene expression profiles of cases in this subgroup are compared with those of control cases, yielding a list of differentially expressed genes. And finally, a selection of several genes from that list is made for further study, based on their presumed biologic impact. For example, in one recent study, gene expression profiles from clinical acute promyelocytic leukemia samples with the t(11;17) translocation expressing both PLZF-RARA and the reciprocal fusion protein RARA-PLZF were compared with those from samples expressing the PLZF-RARA fusion only. This led to the identification of *CRABP1* as a specific target of the reciprocal fusion product. *CRABP1* was subsequently shown to play a functionally important role in retinoid resistance.⁶⁹ Comparisons of *FLT3* ITD and *FLT3* TKD signatures to *FLT3* wild-type signatures have led to the discovery of potentially relevant downstream targets as well, although these remain to be functionally validated.⁴⁴

While these examples show the potential of this approach, the disadvantage of supervised comparisons with subsequent selection of candidate genes is restriction due to inherent selection bias. One way around the selection bias problem may be employment of pathway-oriented bioinformatic analyses. Using specialized software, it has become possible to investigate the differential expression of sets of genes known to function in the same biologic pathways or to identify genes that have frequently been linked in the literature.^{39,40,70} These analyses facilitate the identification of the most promising candidate genes and their selection for functional validation in either in vitro or in vivo model systems. At the same time, it should be kept in mind that relative levels of mRNA expression do not necessarily reflect biologic activity, as the latter may be highly dependent on other factors, such as posttranslational modifications.

Nevertheless, a recent report on a novel subgroup of immature leukemias with myeloid and T-lymphoid characteristics demonstrated the value of the application of biologic pathway analysis to clinical GEP data.³¹ In a clinical AML GEP study, these leukemias were found to carry expression profiles similar to AML cases with *CEBPA* mutations, but such mutations were not present (Figure 1A). Subsequent experiments elucidated the likely explanation for this phenomenon, as in the novel subgroup *CEBPA* expression was silenced, frequently through promoter hypermethylation. Pathway analysis then revealed that the leukemias carried both myeloid and T-lymphoid features. Mouse modeling subsequently demonstrated that lack of *CEBPA* expression induced the expression of certain T-cell genes in immature hematopoietic cells. Thus, the analysis of human AML GEP using pathway analysis in combination with experiments in a representative mouse model uncovered part of the pathobiology of a subgroup of leukemia characterized by a myeloid/T-lymphoid phenotype, *CEBPA* silencing, and, in fact, frequent *NOTCH1* mutations.

Another strategy for distilling biologically significant genes from human AML gene expression data makes use of possible associations with large sets of genes involved in animal retroviral insertion leukemogenesis. A study that compared integration sites in retrovirally induced leukemias in mice with human AML datasets demonstrated that mouse cancer genes were frequently deregulated in the human AMLs.⁷¹ Moreover, pathway analysis defined several biologic networks that associated with particular AML subsets. Thus, comparisons of human AML GEP data with high-throughput results from dedicated experimental models pro-

vide valuable opportunities to identify candidate disease genes. As pointed out before, these experimental approaches could also utilize GEP of cell lines or animal models,²⁵⁻²⁷ or, alternatively, use techniques such as chromatin immunoprecipitation on DNA microarray chips (ChIP-chip),⁷² ChIP-sequencing⁷³ and RNA interference libraries to uncover target genes. A successful example of the latter approach comes from diffuse large B-cell lymphoma, for which 2 previously GEP-defined disease subgroups were functionally investigated using an RNA interference library to search for specific targets inhibiting tumor growth.^{26,74}

Another viable strategy is based on the correlation of putative disease genes from in vivo or in vitro models to specific human leukemia subtypes. The strength of this approach is that the gene choice is based on experimental data. This allows for a rapid correlation of basic research findings to human AML datasets: is transforming gene X, that induces a leukemic phenotype in an experimental mouse model, differentially expressed in particular forms of human AML? Using this approach, researchers have provided evidence that genes such as *TrkA* and *Trib2*, which are involved in murine leukemia, may also be engaged in t(8;21) AML and *CEBPA*-silenced AML, respectively.^{75,76}

These studies demonstrate the abilities of GEP to resolve questions related to the biology of AML when combined with appropriate other experimental and analytical tools. At the same time, there are various other experimental options to pinpoint key genomic abnormalities through GEP in human AML samples.

Concluding remarks and future perspectives

Recent years have shown an increase in high-throughput applications apart from GEP.⁸ In this respect profiling of microRNA (miRNA) levels, chromosomal copy number changes, epigenetic modifications, and DNA sequencing offer interesting opportunities.

MiRNAs are small noncoding RNAs that play a role in transcriptional or posttranscriptional regulation of genes involved in numerous biologic processes, including differentiation and proliferation.⁷⁷ Profiling of miRNA expression levels in AML cohorts has indicated that, similar to mRNA profiling, distinct miRNA signatures are associated with specific subgroups of AML.⁷⁸⁻⁸⁰ As a single miRNA can play a role in concomitant regulation of multiple genes, closer comparisons between miRNA and mRNA profiles may provide clues for significant defects.

Changes in expression levels of critical genes may be due to small DNA amplifications or deletions undetected by conventional cytogenetics. Platforms to study those small chromosomal aberrations include single nucleotide polymorphism (SNP) arrays and array-based comparative genomic hybridization (CGH).^{81,82} The power of the use of SNP arrays for this type of analysis was recently illustrated in a study of 242 cases of pediatric ALL.⁸³ The investigators identified focal abnormalities in lymphocyte differentiation related genes in 40% of cases, including 30% in the *PAX5* gene. As yet, only studies of limited size have been performed in AML.⁸⁴⁻⁸⁹ A study in which array CGH (bacterial artificial and P1-derived artificial chromosome clones) was used in a series of 60 cases of AML with complex karyotypes identified several recurrent lesions.⁹⁰ It is not clear yet what the overall frequency and distribution of these genomic alterations in AML are. A notable observation from several SNP array-based studies is the relatively common occurrence of copy number-neutral loss of heterozygosity through segmental uniparental disomy.^{84,87,91} This phenomenon can

lead to homozygous mutations or deletions of leukemia-related genes, including *FLT3*, *WT1*, *RUNX1*, and *CEBPA*.^{91,92}

Variations in gene expression may also be caused by yet unknown gene sequence mutations. Whole genome sequencing is emerging as a means to address this issue on a global scale.^{93,94} A challenge in this regard will be to distinguish functionally relevant mutations from the abundant so-called passenger mutations—unimportant genetic changes caused by genomic instability of cancer cells—that will be picked up at the same time. Studies on tyrosine kinase abnormalities have underscored the need for validation of the biologic effects of novel mutations identified by high-throughput nucleotide sequencing.^{95,96}

In addition to genetic alterations, epigenetic modifications, including DNA and histone methylation, play a pivotal role in gene regulation. Several platforms have been developed to study alterations in these mechanisms on a genome-wide scale.^{97,98} The first results of genome-wide CpG methylation profiling of AML cell lines and primary cells have now been reported, and it is likely that more extensive investigations will follow.^{99,100} Given the direct association between epigenetics and expression, much is to be expected from this field.

As these fields have just started to emerge, there is still a need for bioinformatics to keep up with the technical developments, and to develop software platforms that allow the integration of the massive amounts of data. It is likely that the combination of GEP

with complementing technologies, such as those outlined above, will provide challenging opportunities to address questions that cannot be resolved by GEP alone.

Acknowledgments

We thank Dr Peter Valk for valuable comments on the manuscript.

This work was supported by the Dutch Cancer Society Konigin Wilhelmina Fonds (Amsterdam, The Netherlands) and by the National Institutes of Health (Bethesda, MD; CA118316).

Authorship

Contribution: B.J.W., B.L., and R.D. wrote the paper.

Conflict-of-interest disclosure: R.D. and B.L. have declared ownership interests in Skyline, a spinoff company of Erasmus University Medical Center (Erasmus MC), held in a Special Purpose Foundation of Erasmus MC. B.J.W. declares no competing financial interests.

Correspondence: Bas Wouters or Ruud Delwel, Erasmus University Medical Center, Department of Hematology, rooms Ee1330B / Ee1342, PO Box 2040, 3000 CA Rotterdam, The Netherlands; e-mail: b.wouters@erasmusmc.nl or h.delwel@erasmusmc.nl.

References

- Estey E, Dohner H. Acute myeloid leukaemia. *Lancet*. 2006;368:1894-1907.
- Lowenberg B, Downing JR, Burnett A. Acute myeloid leukemia. *N Engl J Med*. 1999;341:1051-1062.
- Mrozek K, Heinonen K, Bloomfield CD. Clinical importance of cytogenetics in acute myeloid leukaemia. *Best Pract Res Clin Haematol*. 2001;14:19-47.
- Mrozek K, Marcucci G, Paschka P, Whitman SP, Bloomfield CD. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? *Blood*. 2007;109:431-448.
- Tallman MS, Gilliland DG, Rowe JM. Drug therapy for acute myeloid leukemia. *Blood*. 2005;106:1154-1163.
- Dash A, Gilliland DG. Molecular genetics of acute myeloid leukaemia. *Best Pract Res Clin Haematol*. 2001;14:49-64.
- Pollack JR. A perspective on DNA microarrays in pathology research and practice. *Am J Pathol*. 2007;171:375-385.
- Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet*. 2006;7:200-210.
- Elvidge G. Microarray expression technology: from start to finish. *Pharmacogenomics*. 2006;7:123-134.
- Quackenbush J. Microarray analysis and tumor classification. *N Engl J Med*. 2006;354:2463-2472.
- Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999;286:531-537.
- Bullinger L, Valk PJ. Gene expression profiling in acute myeloid leukemia. *J Clin Oncol*. 2005;23:6296-6305.
- Ebert BL, Golub TR. Genomic approaches to hematologic malignancies. *Blood*. 2004;104:923-932.
- Haferlach T, Kohlmann A, Kern W, Hiddemann W, Schnittger S, Schoch C. Gene expression profiling as a tool for the diagnosis of acute leukemias. *Semin Hematol*. 2003;40:281-295.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. 1998;95:14863-14868.
- D'Haeseleer P. How does gene expression clustering work? *Nat Biotechnol*. 2005;23:1499-1501.
- Schoch C, Kohlmann A, Schnittger S, et al. Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc Natl Acad Sci U S A*. 2002;99:10008-10013.
- Vey N, Mozziconacci MJ, Groulet-Martinec A, et al. Identification of new classes among acute myelogenous leukaemias with normal karyotype using gene expression profiling. *Oncogene*. 2004;23:9381-9391.
- Debernardi S, Lillington DM, Chaplin T, et al. Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events. *Genes Chromosomes Cancer*. 2003;37:149-158.
- Virtaneva K, Wright FA, Tanner SM, et al. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc Natl Acad Sci U S A*. 2001;98:1124-1129.
- Bullinger L, Dohner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med*. 2004;350:1605-1616.
- Valk PJ, Verhaak RG, Beijin MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004;350:1617-1628.
- Ross ME, Mahfouz R, Onciu M, et al. Gene expression profiling of pediatric acute myelogenous leukemia. *Blood*. 2004;104:3679-3687.
- Verhaak RG, Goudswaard CS, van Putten W, et al. Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood*. 2005;106:3747-3754.
- Bittner M, Meltzer P, Chen Y, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*. 2000;406:536-540.
- Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403:503-511.
- Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747-752.
- Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. 2002;1:133-143.
- Wilson CS, Davidson GS, Martin SB, et al. Gene expression profiling of adult acute myeloid leukemia identifies novel biologic clusters for risk classification and outcome prediction. *Blood*. 2006;108:685-696.
- Bullinger L, Rucker FG, Kurz S, et al. Gene-expression profiling identifies distinct subclasses of core binding factor acute myeloid leukemia. *Blood*. 2007;110:1291-1300.
- Wouters BJ, Jorda MA, Keeshan K, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced *CEBPA* and mutations in *NOTCH1*. *Blood*. 2007;110:3706-3714.
- Brazma A, Parkinson H, Sarkans U, et al. ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res*. 2003;31:68-71.
- Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30:207-210.
- Michiels S, Koscielny S, Hill C. Interpretation of microarray data in cancer. *Br J Cancer*. 2007;96:1155-1158.
- Bourquin JP, Subramanian A, Langebrake C, et al. Identification of distinct molecular phenotypes

- in acute megakaryoblastic leukemia by gene expression profiling. *Proc Natl Acad Sci U S A*. 2006;103:3339-3344.
36. Marasca R, Maffei R, Zucchini P, et al. Gene expression profiling of acute promyelocytic leukemia identifies two subtypes mainly associated with FLT3 mutational status. *Leukemia*. 2006;20:103-114.
 37. Huntly BJ, Gilliland DG. Leukaemia stem cells and the evolution of cancer-stem-cell research. *Nat Rev Cancer*. 2005;5:311-321.
 38. Wang JC, Dick JE. Cancer stem cells: lessons from leukemia. *Trends Cell Biol*. 2005;15:494-501.
 39. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545-15550.
 40. Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. *Genes Dev*. 2007;21:1010-1024.
 41. Haferlach T, Kohlmann A, Schnittger S, et al. Global approach to the diagnosis of leukemia using gene expression profiling. *Blood*. 2005;106:1189-1198.
 42. Verhaak RGW, Wouters BJ, Erpelinck CAJ, et al. Prediction of molecular subtypes in acute myeloid leukemia (AML) based on gene expression profiling. *Haematologica*; in press.
 43. Lacayo NJ, Meshinchi S, Kinnunen P, et al. Gene expression profiles at diagnosis in de novo childhood AML patients identify FLT3 mutations with good clinical outcomes. *Blood*. 2004;104:2646-2654.
 44. Neben K, Schnittger S, Brors B, et al. Distinct gene expression patterns associated with FLT3- and NRAS-activating mutations in acute myeloid leukemia with normal karyotype. *Oncogene*. 2005;24:1580-1588.
 45. Bullinger L, Dohner K, Kranz R, et al. An FLT3 gene-expression signature predicts clinical outcome in normal karyotype AML. *Blood*. 2008;111:4490-4495.
 46. Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol*. 2005;23:7332-7341.
 47. Boelstad HM, Nygard S, Storvold HL, et al. Predicting survival from microarray data—a comparative study. *Bioinformatics*. 2007;23:2080-2087.
 48. Radmacher MD, Marcucci G, Ruppert AS, et al. Independent confirmation of a prognostic gene-expression signature in adult acute myeloid leukemia with a normal karyotype: a Cancer and Leukemia Group B study. *Blood*. 2006;108:1677-1683.
 49. Yagi T, Morimoto A, Eguchi M, et al. Identification of a gene expression signature associated with pediatric AML prognosis. *Blood*. 2003;102:1849-1856.
 50. Glinsky GV, Berezovska O, Glinskii AB. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *J Clin Invest*. 2005;115:1503-1521.
 51. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med*. 2002;8:68-74.
 52. Foekens JA, Atkins D, Zhang Y, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. *J Clin Oncol*. 2006;24:1665-1671.
 53. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005;365:671-679.
 54. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530-536.
 55. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347:1999-2009.
 56. van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature*. 2008;452:564-570.
 57. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*. 2006;98:1183-1192.
 58. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst*. 2007;99:147-157.
 59. Glas AM, Floore A, Delahaye LJ, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics*. 2006;7:278.
 60. Kattan MW. Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst*. 2003;95:634-635.
 61. Clarke R, Ransom HW, Wang A, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat Rev Cancer*. 2008;8:37-49.
 62. Tinker AV, Boussioutas A, Bowtell DD. The challenges of gene expression microarrays for the study of human cancer. *Cancer Cell*. 2006;9:333-339.
 63. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet*. 2003;362:1439-1444.
 64. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet*. 2005;365:488-492.
 65. Dobbin KK, Zhao Y, Simon RM. How large a training set is needed to develop a classifier for microarray data? *Clin Cancer Res*. 2008;14:108-114.
 66. Yuan W, Payton JE, Holt MS, et al. Commonly dysregulated genes in murine APL cells. *Blood*. 2007;109:961-970.
 67. Kirstetter P, Schuster MB, Bereshchenko O, et al. Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. *Cancer Cell*. 2008;13:299-310.
 68. Chen W, Kumar AR, Hudson WA, et al. Malignant transformation initiated by MLL-AF9: gene dosage and critical target cells. *Cancer Cell*. 2008;13:432-440.
 69. Guidez F, Parks S, Wong H, et al. RARalpha-PLZF overcomes PLZF-mediated repression of CRABP1, contributing to retinoid resistance in t(11;17) acute promyelocytic leukemia. *Proc Natl Acad Sci U S A*. 2007;104:18694-18699.
 70. Jelier R, Jenster G, Dorssers LC, et al. Text-derived concept profiles support assessment of DNA microarray data for acute myeloid leukemia and for androgen receptor stimulation. *BMC Bioinformatics*. 2007;8:14.
 71. Erkeland SJ, Verhaak RG, Valk PJ, Delwel R, Lowenberg B, Touw IP. Significance of murine retroviral mutagenesis for identification of disease genes in human acute myeloid leukemia. *Cancer Res*. 2006;66:622-626.
 72. Wu J, Smith LT, Plass C, Huang TH. ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res*. 2006;66:6899-6902.
 73. Robertson G, Hirst M, Bainbridge M, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*. 2007;4:651-657.
 74. Ngo VN, Davis RE, Lamy L, et al. A loss-of-function RNA interference screen for molecular targets in cancer. *Nature*. 2006;441:106-110.
 75. Keeshan K, He Y, Wouters BJ, et al. Tribbles homolog 2 inactivates C/EBPalpha and causes acute myelogenous leukemia. *Cancer Cell*. 2006;10:401-411.
 76. Mulloy JC, Jankovic V, Wunderlich M, et al. AML1-ETO fusion protein up-regulates TRKA mRNA expression in human CD34+ cells, allowing nerve growth factor-induced expansion. *Proc Natl Acad Sci U S A*. 2005;102:4016-4021.
 77. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;6:857-866.
 78. Jongen-Lavrencic M, Sun SM, Dijkstra MK, Valk PJ, Lowenberg B. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. *Blood*. 2008;111:5078-5085.
 79. Mi S, Lu J, Sun M, et al. MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia. *Proc Natl Acad Sci U S A*. 2007;104:19971-19976.
 80. Garzon R, Volinia S, Liu CG, et al. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood*. 2008;111:3183-3189.
 81. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet*. 2007;39:S16-21.
 82. Dutt A, Beroukhi R. Single nucleotide polymorphism array analysis of cancer. *Curr Opin Oncol*. 2007;19:43-49.
 83. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007;446:758-764.
 84. Raghavan M, Lillington DM, Skoulakis S, et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res*. 2005;65:375-378.
 85. Suela J, Alvarez S, Cifuentes F, et al. DNA profiling analysis of 100 consecutive de novo acute myeloid leukemia cases reveals patterns of genomic instability that affect all cytogenetic risk groups. *Leukemia*. 2007;21:1224-1231.
 86. Karnan S, Tsuzuki S, Kiyoi H, et al. Genomewide array-based comparative genomic hybridization analysis of acute promyelocytic leukemia. *Genes Chromosomes Cancer*. 2006;45:420-425.
 87. Gorletta TA, Gasparini P, D'Elios MM, Trubia M, Pelicci PG, Di Fiore PP. Frequent loss of heterozygosity without loss of genetic material in acute myeloid leukemia with a normal karyotype. *Genes Chromosomes Cancer*. 2005;44:334-337.
 88. Tyybakinoja A, Elonen E, Piippo K, Porkka K, Knuutila S. Oligonucleotide array-CGH reveals cryptic gene copy number alterations in karyotypically normal acute myeloid leukemia. *Leukemia*. 2007;21:571-574.
 89. Paulsson K, Heidenblad M, Strombeck B, et al. High-resolution genome-wide array-based comparative genome hybridization reveals cryptic chromosome changes in AML and MDS cases with trisomy 8 as the sole cytogenetic aberration. *Leukemia*. 2006;20:840-846.
 90. Rucker FG, Bullinger L, Schwaenen C, et al. Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. *J Clin Oncol*. 2006;24:3887-3894.
 91. Fitzgibbon J, Smith LL, Raghavan M, et al. Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res*. 2005;65:9152-9154.

92. Wouters BJ, Sanders MA, Lugthart S, et al. Segmental uniparental disomy as a recurrent mechanism for homozygous CEBPA mutations in acute myeloid leukemia. *Leukemia*. 2007;21:2382-2384.
93. Ley TJ, Minx PJ, Walter MJ, et al. A pilot study of high-throughput, sequence-based mutational profiling of primary human acute myeloid leukemia cell genomes. *Proc Natl Acad Sci U S A*. 2003;100:14275-14280.
94. Thomas RK, Baker AC, Debiasi RM, et al. High-throughput oncogene mutation profiling in human cancer. *Nat Genet*. 2007;39:347-351.
95. Loriaux MM, Levine RL, Tyner JW, et al. High-throughput sequence analysis of the tyrosine kinase in acute myeloid leukemia. *Blood*. 2008;111:4788-4796.
96. Frohling S, Scholl C, Levine RL, et al. Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell*. 2007;12:501-513.
97. Khulan B, Thompson RF, Ye K, et al. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res*. 2006;16:1046-1055.
98. Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet*. 2005;37:853-862.
99. Gebhard C, Schwarzfischer L, Pham TH, et al. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res*. 2006;66:6118-6128.
100. Yuan E, Haghighi F, White S, et al. A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. *Cancer Res*. 2006;66:3443-3451.