A four-gene LincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients

D Beck[1,2], J A I Thoms[1,16], C Palu[1,16], T Herold[3], A Shah[1], J Olivier[4], L Boelen[5], Y Huang[1], D Chacon[1], A Brown[6], M Babic[6], C Hahn[6], M Perugini[6], X Zhou[7], B J Huntly[8], A Schwarzer[9], J-H Klusmann[9], W E Berdel[10], B Wörmann[11], T Büchner[10], W Hiddemann[3,12], S K Bohlander[13], L B To[14], H S Scott[6], I D Lewis[6,14], R J D'Andrea[6,1,4], J W H Wong[1] and J E Pimanda[1,2,15]
[1]Adult Cancer Program, Lowy Cancer Research Centre, Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales, Australia
[2]Centre for Health Technologies and the School of Software, University of Technology, Sydney, NSW, Australia
[3]Department of Internal Medicine III, University Hospital Grosshadern, Ludwig Maximilians Universität, Munich, Germany
[4]School of Mathematics, UNSW Australia, Sydney, NSW, Australia
[5]Department of Medicine, Faculty of Medicine, Imperial College London, London, UK
[6]Centre for Cancer Biology, SA Pathology, University of South Australia, Adelaide, SA, Australia
[7]School of Medicine, Wake Forest University, Winston-Salem, NC, USA
[8]Department of Haematology, Cambridge Institute for Medical Research, Cambridge, UK
[9]Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany
[10]Department of Medicine, Hematology and Oncology, University of Münster, Münster, Germany
[11]Department of Medicine, Hematology, Oncology, Tumor Immunology, Charité—Universitätsmedizin Berlin, Berlin, Germany
[12]Laboratory for Leukemia Diagnostics, Ludwig Maximilians Universität, Munich, Germany
[13]Department of Molecular Medicine and Pathology, University of Auckland, New Zealand
[14]Department of Haematology, SA Pathology, Royal Adelaide Hospital, Adelaide, SA, Australia
[15]Department of Haematology, Prince of Wales Hospital, Randwick, NSW, Australia
Correspondence: Dr D Beck or Dr JWH Wong or Professor JE Pimanda, Adult Cancer Program, Lowy Cancer Research Centre, Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales 2052, Australia. E-mail: d.beck@unsw.edu.au or jason.wong@unsw.edu.au or jpimanda@unsw.edu.au
[16]These authors contributed equally to this work.

Abstract
Prognostic gene expression signatures have been proposed as clinical tools to clarify therapeutic options in acute myeloid leukemia (AML). However, these signatures rely on measuring large numbers of genes and often perform poorly when applied to independent cohorts or those with older patients. Long intergenic non-coding RNAs (lincRNAs) are emerging as important regulators of cell identity and oncogenesis, but knowledge of their utility as prognostic markers in AML is limited. Here we analyze transcriptomic data from multiple cohorts of clinically annotated AML patients and report that (i) microarrays designed for coding gene expression can be repurposed to yield robust lincRNA expression data, (ii) some lincRNA genes are located in close proximity to hematopoietic coding genes and show strong expression correlations in AML, (iii) lincRNA gene expression patterns distinguish cytogenetic and molecular subtypes of AML, (iv) lincRNA signatures composed of three or four genes are independent predictors of clinical outcome and further dichotomize survival in European Leukemia Net (ELN) risk groups and (v) an analytical tool based on logistic regression analysis of quantitative PCR measurement of four lincRNA genes (LINC4) can be used to determine risk in AML.

Introduction
Acute myeloid leukemia (AML) is a heterogeneous myeloid neoplasm that develops in hematopoietic stem and progenitor cells (HSPCs) with altered ability to self-renew, proliferate and differentiate.[1] Leukemic clones often carry cytogenetic abnormalities including translocations, inversions, deletions, monosomies and trisomies.[1] Recurrent mutations and aberrant expression profiles of coding genes and non-coding microRNAs are also commonly found in AML patients with normal cytogenetics (CN-AML). The karyotype,[2] recurrent lesions in the nucleophosmin gene (*NPM1*),[3] tandem internal duplications of the fms-related tyrosine kinase 3 (*FLT3-ITD*)[4] and mutations in the transcription factor CCAAT/enhancer-binding protein alpha (*CEBPA*)[5] are variably associated with patient remission, relapse and survival, and have been integrated into AML prognostic indices such as those developed by the United Kingdom Medical Research Council[6] or the European Leukemia Net (ELN).[7] However, the utility of these guidelines for clinical decision-making has only been validated in ~40% of patients

younger than 60 years (NCI Surveillance, Epidemiology, and End Results Program), and, even within this subpopulation, there is substantial prognostic heterogeneity, particularly within the intermediate-risk group that constitutes the majority of patients.[8, 9, 10]

The human genome consists of more than three billion nucleotides that encode a relatively small percentage of coding (~1.5%) and a larger percentage of non-coding (~75%) transcripts.[11] The associations between coding genes and the biology, pathogenesis and clinical outcomes of AML have been extensively studied. However, the contributions from the large pool of non-coding RNAs to these events is largely unknown. In particular, the group of long non-coding RNAs (lncRNAs; >200 nucleotides) are emerging as key regulators of an increasing number of molecular processes and their aberrant expression is correlated with the development of solid tumors and clinical outcome (reviewed in Shen *et al.*[12] and Ulitsky and Bartel[13]). In the hematopoietic system, lncRNAs regulate stem cell activity and function in both myeloid and lymphoid differentiation, but their role in leukemia is poorly understood. However, lncRNAs associated with the *HOX* cluster[14, 15, 16] and the *RUNX1* locus have attracted attention in the pathogenesis of acute leukemias.[17, 18, 19] To accelerate research in this area, data from large multicentre cohort studies are needed to help guide targeted experimental investigations and the adoption of lncRNAs as useful clinical biomarkers.

Here we present data on 1664 long intervening/intergenic non-coding RNAs (lincRNAs), a subgroup of genes not overlapping coding exons, in 736 patients across three independent patient cohorts. We show for the first time that distinct lincRNA expression profiles were associated with recognized cytogenetic and mutational subgroups of AML. We also show that an expression signature composed of just four lincRNAs (*ENSG00000153363*, *XLOC_013473*, *ENSG00000255571* and *ENSG00000260257*) had independent prognostic value across all three patient cohorts in a multivariable analysis that included age, gender and risk recomendations made by the ELN, which incorporates karyotype and the analysis of specific gene mutations. Furthermore, using this signature, patients in each ELN risk group could be dichotomized into cohorts with significantly better or worse prognosis. This is particularly relevant to patients in ELN intermediate-risk groups, where optimal clinical management is often uncertain. To facilitate clinical application, we developed a regression-based tool to measure and normalize the expression of these four lincRNA and further show its prognostic utility in an independent test cohort of 115 clinically annotated AMLs.

Materials and methods

Patient samples and publicly available microarray data sets

The patient samples from the Netherlands (NL; *N*=419), Germany (GER; *N*=139), the United States of America (USA; *N*=178) and the United Kingdom sets have previously been published, and informed consent was given at the corresponding hospitals at the time of diagnosis.[20, 21, 22, 23] The study protocols were approved by the institutional review boards of corresponding institutes and hospitals. The patient samples from the Australia (AUS) set (*N*=115) were assembled using bone marrow mononuclear cell preparations taken at diagnosis with consent via the South Australian Cancer Research Biobank with approval from relevant hospital ethics committees in accordance with the Declaration of Helsinki. Normal bone marrow samples (*N*=6) were obtained using an approved collection process through the Royal Adelaide Hospital. Research ethics approval was obtained through the Royal Adelaide Hospital human ethics committee. Patients were treated using standard induction and consolidation chemotherapy protocols. Briefly, patients less than 60 years of age received chemotherapy regimens with idarubicin and high-dose cytarabine, and patients over 60 years received idarubicin and standard dose cytarabine. Consolidation therapy consisted of high-dose cytarabine alone or standard dose cytarabine with idarubicin. An overview of the clinical and molecular characteristics of patient cohorts is provided in Supplementary Table S1.

Affymetrix U133 Plus 2.0 (HG133P2) array data have previously been published for patients in the NL, USA and GER cohorts, and samples from the USA set have also been profiled using the Illumina HiSeq2000 (HS2000) platform making HG133P2 and HiSeq2000 data available for the same 159 samples from the USA set.

Data set description

A total of nine expression data sets $D=\{d_1,...,d_8\}$ were analyzed. The data sets d_1, d_2 and d_3 were profiled using the Affymetrix HG-133 Plus 2 (HG133P2) microarray technology, d_4, d_5, d_6, d_7 and d_8 were profiled using RNA sequencing (RNA-seq) and the data set d_8 was profiled using high-throughput reverse transcriptase polymerase chain reaction (RT-PCR). (1) AML cohort- NL

set/HG133P2, (2) AML cohort—USA set/HG133P2, (3) AML cohort—GER set/HG133P2, (4) AML cohort—USA set/ RNA-seq, (5) Normal hematopoietic cell fractions—BLUEPRINT/RNA-seq, (6) AML Cell Lines—Cancer Cell line Encyclopedia/RNA-seq, (7) HSPC—GSE48846/RNA-seq, (8) HSPC—GSE63569 and (9) AML cohort—AUS set/RT-PCR. We denote the *i*th sample in a data set *d* as $S^d = \{S_i^d \mid 1 \leq i \leq N_d\}$, where N_d is the size of the data set *d*.

Estimation of lincRNA expression from expression array

A flowchart of the lincRNA estimation routine (LER) is depicted in Supplementary Figure S1. We analyzed the three expression data sets d_1,d_2,d_3 using the LER. In each data set a finite number of probes $p^d=\{p_1|1 \leq 1 \leq O_d\}$ (O_d is the total number of probes tested in data set *d*) is measured for each sample S^d. The raw probe-expression levels were pre-processed using the Partek Genomic Suite including background correction, quantile normalization and log2 transformation. The ComBat algorithm[24] was used to remove experimental variation associated with different array batches, as well as biological variation from a gender bias observed in the data. The core of the LER algorithm is the association between these expression probes *p* and known lincRNA transcripts. To generate this catalog of probe-to-lincRNA associations, we retrieved the genomic coordinates of currently annotated lincRNA transcripts and compared these with the location of the strand-specific HG133P2 expression probes.[25] We considered only lincRNA transcripts *T* that were probed by at least four unique and non-repetitive expression probes in each of the data sets d_1, d_2, d_3; for example, probes matching to multiple genes were excluded for the analysis. Given this reduced set of lincRNA transcripts, we determined for each lincRNA/a set $T_{r(l)}=\{T/T$ is a transcript of l$\}$. There were 1664 lincRNAs for which this set was non-empty. To define the LER expression value of a lincRNA *l*, we apply a two-tiered approach. Firstly, for a transcript *T*, we define its expression *E* (*T*) as the median polished expression value of the probes *p* overlapping *T*; for example, if a transcript *T* overlaps *n* probes $p_1,...,p_n(n \geq 4)$, with normalized expression levels $e_1,...,e_n$, the expression level for the transcript *E(T)*=median polish$\{e_1,...,e_n\}$. Secondly, we define the expression value of the lincRNA/as $\max\{E(T)T \in T_{r(l)}\}$. We adopt the notation *E*(l,d,j) to denote the LER-defined expression value of lincRNA/in patient *j* of data set *d*.

LincRNA expression profiling using Fluidigm RT-PCR

To rationally select one target lincRNA signatures (lincRNA-sig) for experimental validation, we first calculate the frequency of each lincRNA within the 201 lincRNA-sig (for example, the highest ranked lincRNA *XLOC_013473* was part of 76/201 lincRNA-sigs) and then scored each lincRNA-sig by the mean frequency of all lincRNAs within the lincRNA-sig (Supplementary Table S3). The lincRNA-sig consisting of *ENSG00000153363*, XLOC_013473, ENSG00000255571 and ENSG00000260257 (LINC4) scored highest with a cumulative frequency of 233 and mean frequency of 58.3. We also reanalyzed the NL, GER and the USA sets to identify lincRNA genes with low expression variance spanning the expression range of the target lincRNA genes (Supplementary Figure S2). Primers were designed for the following reference lincRNA genes: *ENSG00000245614*, *XLOC_004584*, *ENSG00000232388*, *ENSG00000234608* as well as the reference gene *EMC7*(ref. 26)(Supplementary Table S5). See Supplementary Methods for details.

Results

Repurposing microarrays to profile lincRNA expression in AML

Large clinically annotated AML patient cohorts from the NL (*N*=419), USA (*N*=178), and GER (*N*=139) have been profiled using standard expression arrays and are currently the largest existing resource to study leukemia transcriptomics.[20, 21, 22] LincRNAs are typically capped, polyadenylated and often spliced and can be profiled using standard expression assays.[13] Indeed, typical microarray platforms contain probe sets that fully or partially overlap the genomic coordinates of annotated lincRNA genes (Figure 1a), and some of these can be repurposed to estimate lincRNA expression.[25, 27, 28, 29] To measure lincRNA expression levels in large existing leukemia data sets, we implemented the LER (Materials and Methods, Supplementary Figure S1).

Repurposing microarrays to profile LincRNA expression in AML. (**a**) UCSC genome browser visualization of lincRNA LINC00925 and flanking coding genes. *LINC00925* is robustly expressed in some patients (for example, AB-2803 and 2861) but not in others (for example, AB-2915 and 2808). *LINC00925* overlaps with probes of two probe sets on Affymetrics HG133P2 arrays. (**b**) Correlations between lincRNA expression values estimated from Affymetrics HG133P2 array and measured by

RNA-seq in the USA set. (i) Pearson correlation coefficients. (ii) −log10 (*P*-values). (**c**; i) Correlation between expression levels of 1664 lincRNAs in patient AB-2921 in the USA set estimated from an Affymetrics HG133P2 array and measured by RNA-seq. (ii) C_T values of six selected lincRNAs measured by RT-PCR in 27 AML patients (UK set). Note, a high C_T corresponds to low expression while a low C_T corresponds to high expression. These lincRNAs were chosen for their low (1, 2), medium (3, 4) and high (5, 6) expression in a. (**d**) Hierarchical clustering of lincRNA expression levels (top) and visualization of correlated coding gene expression levels (bottom) in the USA set. (**e**) Network graph showing lincRNA-to-coding gene relationships generated by the ingenuity pathway analysis (IPA). In red/green coding genes positively/negatively correlated with lincRNA genes. The network is associated with the ingenuity pathway analysis terms 'Cancer', 'Hematological Disease' and 'Immunological Disease'.

Interrogation of five expression array platforms identified the Affymetrix HG-133 Plus 2 (HG133P2) platform as the most comprehensive resource to study large numbers of lincRNA genes across multiple large patient cohorts (Supplementary Table S6). Application of the LER to three existing HG133P2 AML data sets revealed that expression levels of 1664 known lincRNA genes could be estimated for 736 patients (Supplementary Table S7). This is approximately a fivefold increase compared with the 338 lncRNA genes annotated by Affymetrix on the standard HG133P2 array. To evaluate the accuracy of these estimates, we first reanalyzed genome-wide expression data of 159 primary AML samples that have been profiled using both the HG133P2 array and RNA-seq in the USA set from The Cancer Genome Atlas project.[21] We found that expression levels of lincRNA genes measured by these two methods were concordant (*P*<0.001, Student's *t*-test; mean *r*=0.61, s.d.=0.04; Figure 1b). Given the relatively low expression of lincRNAs compared with coding genes,[13] these correlations were somewhat lower but in a similar range when compared with those from protein-coding genes (Supplementary Figure S3 and S4). To ensure that this association was not isolated to the USA set, we selected two low, two medium and two high expressed lincRNAs from this cohort for RT-PCR validation in an independent cohort from the UK[30] (Figure 1c). These data confirmed that the six lincRNAs, selected by their LER-estimated expression levels (that is, low, medium or high), were expressed in the same relative levels when assayed by RT-PCR (that is, qPCR cycle threshold). There was a greater concordance of the coefficient of variation when expression was measured by strand-specific assays (for example, HG133P2) in different patient cohorts compared with the measurements using non-strand-specific assays (for example, RNA-seq) in the same cohort (Supplementary Table S8). Taken together, these data suggest that the lincRNA level estimates from the HG133P2 array accurately reflect expression levels and that these data could be interrogated for new insights in AML.

Expression levels of lincRNAs and their neighboring coding genes are correlated

The function of most lincRNA genes is not known, but multiple pathways have been suggested in the literature.[13, 31] One such pathway active in the hematopoietic and other cell systems is the recruitment of epigenetic modifiers to enhance or repress the expression of coding genes in close proximity (reviewed in Ulitsky and Bartel[13]). For example, these include the regulation of *HOXA* genes by *HOTAIRM1* (antisense to *HOXA*; epigenetically activates *HOXA1*/*A4*/*A5*[14]). We investigated data from the USA set (HG133P2 and RNA-seq) to examine whether these lincRNA–target relationships were supported in primary AMLs and found consistent positive correlations between the expression levels of *HOTAIRM1* and its previously reported targets *HOXA1* (*R*~0.68, *P*<0.001), *HOXA4* (*R*~0.53, *P*<0.001) and *HOXA5* (*R*~0.54, *P*<0.001; Supplementary Table S9). We confirmed these correlations in the NL and GER sets, suggesting that the *HOTAIRM1*–*HOXA1*/*A4*/*A5* relationship is also present in primary AMLs (Supplementary Table S9).

To extend this analysis to currently unknown lincRNA–mRNA correlations, we performed an equivalent analysis for each of the 1664 lincRNAs with their two closest neighboring coding genes (up/downstream) in three independent patient sets, that is, NL, USA and GER (see Materials and Methods). We found that the expression levels of a subset of 263 lincRNAs were significantly correlated with the expression of 286 neighboring genes (*P*<0.05; a total of 299 significant correlations; Supplementary Table S2) in all three data sets, which was significantly higher than expected by chance (*P*<0.001, confidence interval (132.866, 134.556); Materials and methods). In some cases, the directionality of lincRNA–mRNA correlations was not concordant across all three patient cohorts, and we excluded such cases (~23%) from further analysis (that is, to 230 genes; *P*<0.001, confidence interval (132.866, 134.556); Materials and Methods; Figure 1d and Supplementary Table S2). The reduced set included epigenetic modifiers (for example, *HMGN4* and

SETD1B), transcriptional regulators (for example, *MED13L*) and hematopoietic transcription factors (for example, *CEBPB*). Indeed, network analysis (ingenuity pathway analysis) of these lincRNA targets identified 11 networks associated with 'Hematological System Development', 'Hematological Disease', Cancer and 'Cell Death and Survival' (Figure 1e, Supplementary Figure S5 and Supplementary Table S10), suggesting that some of the lincRNAs quantified by the standard array platform have an important role during hematopoiesis and in the regulation of gene expression.

Cytogenetic and molecular subtypes of AML have distinct patterns of lincRNA expression

Clinical outcomes in AML vary depending on cytogenetic and mutation status as well as patterns of mRNA and microRNA expression.[20, 21] Extensive cytogenetic and mutational profiles were available for 419 patients in the NL and 178 patients in the USA sets, and we evaluated whether the expression of lincRNAs was also associated with these AML subtypes. Interestingly, an initial analysis revealed a surprising association of lincRNA expression patterns and gender across many of the established subgroups. Therefore, we computationally removed gender effects and identified 235 lincRNA genes with significant variation across the available cytogenetic and mutational profiles (analysis of variance; fold change >1.5; *P*-value<0.05). Hierarchical clustering of the 235 lincRNA genes in the NL (Figure 2a) and USA sets (Figure 2b) identified distinct grouping effects in both cohorts, which were comparable to an equivalent analysis of coding genes (Supplementary Figures S6 and S7). In particular, we found that patients with t(8:21) and t(15;17)/FAB M3 leukemia and those with mutations in *CEBPA*, *NPM1* and/or *FLT3-ITD* had distinct lincRNA profiles in both cohorts. We also found associations that were unique to one or the other data set, that is, FAB M2 and M5 in the NL and inv(16) in the USA set. Notably, patients with *TP53* mutations also showed a distinct lincRNA expression profile (available only in the USA set, Figure 2b).

Cytogenetic and molecular subtypes of AML have distinct patterns of lincRNA expression. (**a**) Hierarchical clustering of 235 lincRNAs in the NL set. (**b**) Hierarchical clustering of 235 lincRNAs in the US set. (**c**) Unsupervised NMF clustering of 1664 lincRNAs in the NL set (left) and patient subtypes enriched in each cluster (Fisher exact test; right). (**d**) Unsupervised NMF clustering of 1664 lincRNAs in the USA set (left) and patient subtypes enriched in each cluster (Fisher exact test; right). *LR, low cytogenetic risk; *IR, intermediate cytogenetic risk; *HR, high cytogenetic risk; *Complex karyotype, greater or equal to three distinct karyotypic abnormalities.

The full set of 1664 lincRNAs was further analyzed using unsupervised non-negative matrix factorization clustering in the NL and USA sets. The NL cohort optimally separated into four novel groups that were enriched for specific AML subtypes (Figure 2c; *P*<0.01; Fisher exact test; Supplementary Table S11). Similarly, the USA cohort was optimally separated into five groups, including cluster three (C3), which was associated with older patients, complex karyotype, del(7q), those in FAB M0 and mutations in *EZH2*, *RUNX1* and *TP53*, while cluster five (C5) was associated with younger patients, chromosomal translocation t(8,21) and t(15;17)/FAB M3 and mutations in *CEBPA* (Figure 2d; *P*<0.01; Fisher's exact test; Supplementary Table S12). These data showed that lincRNA profiles segregated with clinical subgroups, and some clusters were enriched for specific patient groups with either good or poor prognosis; together this suggested that lincRNAs might be of prognostic significance in AML.

LincRNA expression and clinical outcome

Although the predictive value of coding gene expression signatures has been reported across large independent multicentre cohorts,[22, 32] evaluation of lincRNA expression signatures has thus far been limited to a single report in CN-AML in older (>60 years) patients.[19] To redress this, we evaluated whether the expression levels of the 1664 lincRNAs were predictive of overall survival (OS) in patients from the NL (*N*=392), USA (*N*=167) and GER (*N*=135) cohorts for whom survival data were available. Univariate Cox-regression analysis identified 78 lincRNAs in the NL, 109 in the USA and 87 in the GER sets individually associated with OS (*P*<0.05; Figure 3a and Supplementary Table S13; Materials and Methods). *ENSG00000260257* was the sole lincRNA associated with survival across all three patient cohorts. To examine these correlations with an alternate method, we reanalyzed these data using an unsupervised clustering approach (Materials and Methods). Assessment of OS by the Kaplan–Meier model identified 115 lincRNAs in the NL, 111 in the USA and 92 in the GER sets (*P*<0.05; Figure 3b and Supplementary Table S13; Materials and methods). While *ENSG00000260257* was again identified, *ENSG00000236537* emerged as a second lincRNA individually associated with OS across all three patient cohorts (Figure 3c). The expression of *ENSG00000260257* and *ENSG00000236537* correlated with adjacent coding genes *MAPRE1* and

TULP4. However, neither of these coding genes significantly associated with patient survival across all three patient cohorts (that is, mean *P*-values from Kaplan–Meier analysis: 0.28 and 0.11). The limited overlaps between these three patient cohorts were comparable (for example, there was no significant differences to these overlaps when coding genes instead of lincRNAs were considered *P*=0.5456, Fisher's exact test) using an equivalent analysis performed with all 54 677 HG133P2 probes (that is, including probes for coding genes) instead of the 1664 lincRNAs. Importantly, these data are consistent with earlier reports correlating individual coding gene expression levels with clinical data across multiple patient cohorts.[33, 34]

LincRNA expression is associated with clinical outcome. (**a**) Venn diagram of univariate Cox-regression analysis of patient OS in the NL, US and GER sets. *ENSG00000260257* is shared between all three data sets. (**b**) Clustering/Kaplan–Meier analysis of patient OS. *ENSG00000260257* and ENSG00000236537 are shared between all three data sets. (**c**) Kaplan–Meier plots stratified by the two shared lincRNA genes *ENSG00000260257* (top) and *ENSG00000236537* (bottom).
Full figure and legend (174K)

By contrast, the expression levels of several genes in combination are a more powerful predictor of clinical outcome, and this has been harnessed to generate prognostic coding gene signatures in AML.[22, 32, 35, 36, 37] However, these coding gene signatures require measuring between 17 and 133 coding genes. Therefore, we set about identifying and evaluating combinations of lincRNAs, that is, lincRNA-sigs, as predictors of clinical outcome that apply across the US, NL and GER cohorts. Particularly, we aimed to evaluate whether measuring a smaller number of two to four lincRNAs could serve as an alternative to measuring large numbers of coding genes.

LincRNA gene signatures are independent predictors of clinical outcome across multiple patient cohorts

The expression levels of 25 and 28 individual lincRNAs were associated with OS in at least two of the three patient cohorts, assessed by Cox-regression and Kaplan–Meier statistics, respectively (*P*<0.05; Figures 3a and b). These included 43 unique lincRNAs (Supplementary Table S13) and, given their performance as individual prognosticators, we reasoned that combinations of these genes were likely to reveal lincRNA-sigs composed of only a small number of lincRNAs, which are predictive for OS across the NL, USA and GER cohorts.

With the objective to minimize the number of lincRNAs in each lincRNA-sig, we incrementally screened combinations of two, three or four lincRNAs selected from the set of 43 lincRNAs. This limited the search space to 136 697 combinations (Figure 4a and Supplementary Figure S8; Materials and methods). Univariate Cox-regression and Kaplan–Meier analysis revealed that ~12% of the tested lincRNA-sigs (that is, 16491) were significantly associated with OS across all three patient cohorts, including 74, 1328 and 15 089 signatures composed of two, three and four lincRNAs, respectively (*P*<0.05; Figure 4a and Supplementary Table S14). In addition to outcome data, extensive clinical, cytogenetic and molecular data were available for 390 patients in the NL, 159 patients in the USA and 116 patients in the GER cohorts. Therefore, we further interrogated the identified lincRNA-sigs in the context of these parameters (for example, age, gender and white blood cell count) and clinically useful cytogenetic and mutational data (for example, the four risk groups defined by the ELN[7]). Multivariate Cox-regression analysis showed that 201 or ~1.2% (of the 16 491) candidates remained independent predictors of OS across all three patient cohorts, including 15 and 186 signatures composed three and four lincRNAs, respectively (Figure 4a and Supplementary Table S15). These lincRNA-sigs conferred increased risk of death that ranged from 29 to 62% (hazards ratio (HR): 1.29–1.62), 45 to 192% (HR: 1.45–2.92) and 55 to 142% (HR: 1.55–2.42) in the NL, USA and GER sets, respectively (Supplementary Tables S15).

LincRNA gene signatures are independent predictors of clinical outcome. (**a**) Flowchart of the lincRNA-sig discovery pipeline. (**b**) LincRNAs were scored according to their frequency in the 201 lincRNA-sigs significantly associated with patient overall survival. LincRNA-sigs were scored according to the mean frequency of their participating lincRNAs. The frequency distribution of the lincRNA-sig scores is shown. A single signature (LINC4) consisting of the four genes *ENSG00000153363*, *XLOC_013473*, *ENSG00000255571* and *ENSG00000260257* had the highest score (marked in red). (**c**) Application of LINC4 in the NL, USA and GER sets. (i) Principle component analysis of patients grouped (G1 and G2) by expression levels of these lincRNAs. (ii) Kaplan–Meier overall survival plots of patient groups G1 and G2. (iii) Multivariate Cox-regression OS tables for this

lincRNA-sig (the prefix a was used to indicate the adjusted model parameters). (**d**) Application of LINC4 to patient subsets with ELN-FR, ELN INT-1, ELN INT-2 and ELN-AR in the NL cohort subtypes further dichotomizes patients with favorable and adverse risk profiles. (**e**) Overall survival of the TCGA cohort grouped using (i) the LSC17 signature. The low (ii) and high (iii) risk patients were further subgrouped using LINC4. (**f**) Overall survival of the German cohort grouped using (i) the LSC17 signature. The low (ii) and high (iii) risk patients were further subgrouped using LINC4.
Full figure and legend (470K)

Interestingly, we found that *XLOC_013473* was part of 38% (that is, 76), while *ENSG00000227946* and *ENSG00000246985* only featured in one of the 201 lincRNA-sigs, indicating that some lincRNAs might contribute more to the observed prognostication than others (Supplementary Table S3). Therefore, we scored each of the 201 lincRNA-sig by the frequency of its contributing lincRNAs and found that the signature consisting of *ENSG00000153363*, *XLOC_013473*, *ENSG00000255571* and *ENSG00000260257* achieved the highest score (cummulative frequency of 233, ~29%; mean frequency 58.25; Figure 4b and Supplementary Table S4). Cluster analysis using expression levels of these four lincRNAs in healthy CD34+ cells and AML cell lines from CCLE showed distinct patterns of expression between healthy and leukemic cells (Supplementary Figures S9 and S10). In addition, we used data from BLUEPRINT (*N*=82; ref. 38) and the Cancer Cell Line Encyclopedia (*N*=31; ref. 39) to map the expression patterns of these lincRNAs in normal human stem and progenitor subsets and in AML cell lines. We then performed differential expression analysis using EdgeR for well-defined early hematopoietic transitions including HSC→ MPP, MPP → CMP, CMP → MEP, CMP → GMP and two MEP → MK/ERY, and compared the pool of CD34+ with CD34− populations and the pool of the CD34+ population with AML cell lines (Supplementary Table S16). We found that none of these six lincRNAs was associated with defined early hematopoietic transitions. However, ENSG00000255571 was higher expressed in CD34− and AML cells when comparted with CD34+ cells. Similarly, ENSG00000153363 was more expressed in AML cells when compared with CD34+ cells. The lincRNA-sig composed of these four genes (hereafter referred to as LINC4) was then applied to the NL, USA and GER AML patient cohorts to evaluate its prognostic potential. Patients in all three cohorts clustered into distinct groups (Figure 4c(i)) and had significantly different OS (Figure 4c(ii)) with HRs that were often close to those achieved when the ELN IR-1, IR-2 and AR were assessed in reference to ELN-FR (for example, patients with favorable prognosis including t(8;21) or inv(16); Figure 4c(iii)). Importantly, when LINC4 was applied independently to patients classified as ELN-FR, IR-1, IR-2 or AR, it was able to further dichotomize patients in each ELN group into cohorts with significantly better or worse OS (Figure 4d and Supplementary Figure S11).

Most recently, a study reported that a set of 17 coding genes was higher expressed in leukemic compared with normal blood stem cells and that a score comprising these 17 coding genes (that is, LSC17) was prognostic in multiple patient cohorts.[32] We re-calculated the LSC17 score for the validation cohorts from the US and GER, and found that it dichotomized survival (Figure 4e(i) and f(i)). Strikingly, we found that application of LINC4 to the patient subsets with either high or low LSC17 score further dichotomized patient survival in both the US (Figure 4e(ii, iii)) and GER cohorts (Figure 4f(ii, iii)). This further suggests that LINC4 might capture pathways associated with patient survival, which are independent of the LSC17 score and hence LSC biology.

RT-PCR-based validation of the four-gene signature
Motivated by these results and to validate and facilitate application of LINC4, we sourced an additional patient cohort (AUS) to profile lincRNA expression using an orthogonal profiling method. In brief, we developed a Fluidigm RT-PCR-based assay that included a regression-based approach for data normalization (Materials and methods). We designed primer pairs for each of the LINC4 lincRNAs *ENSG00000153363*, *XLOC_013473*, *ENSG00000255571* and *ENSG00000260257* and five reference genes that showed stable expression across the dynamic range of the 201 lincRNA-sigs (Supplementary Table S5 and Supplementary Figure S2, Materials and Methods).

We applied this pipeline to a fourth large independent set of 113 primary AMLs with extensive clinical, cytogenetic and molecular data (AUS set, Supplementary Table S1). Expression profiles were in the range expected from the array analysis of the NL, USA and GER cohorts (Figure 5a), and patients were well separated into two distinct groups (Figure 5b). Logistic regression with leave-one-out and 10-fold cross-validation confirmed that patients could be dichotomize into those with favorable (LINC4 Fav; longer OS) and adverse (LINC4 Adv; shorter OS) prognosis with perfect sensitivity and specificity (Figure 5c) following:

$\frac{\pi_{LincFav}}{\pi_{LincAdv}} = 10.88 - 0.76G1 - 1.75G2 - 11.47G3 + 0.42G4,$, where G1 is the normalized expression of *ENSG00000153363*, G2 of *XLOC_013473*, G3 of *ENSG00000255571* and G4 of *ENSG00000260257.* Further interrogation of the two identified patient groups using multivariable Cox-regression analysis (including gender, age, wbc and ELN) confirmed that LINC4 was also an independent predictor of OS in the AUS cohort conferring an increased risk of death of 197% (HR: 2.96; *P*-value=0.00548; Figure 5d).

RT-PCR-based assay confirms the association of lincRNA-sigs in AUS cohort. (**a**) Expression levels of *ENSG00000153363*, *XLOC_013473*, *ENSG00000255571* and *ENSG00000260257* (LINC4) and five control primers in the AUS set measured by RT-PCR. (**b**) Application of LINC4 to the AUS cohort identifies two patient groups. (**c**) Logistic regression was used to delineate the two patient groups. The regression model was validated using cross-validation and receiver operating characteristic analysis. (**d**) Multivariate Cox-regression table of patient OS (the prefix a was used to indicate the adjusted model parameters).
Full figure and legend (150K)

Discussion

LincRNAs are emerging as key regulators of many biological processes and are widely studied in solid tumors.[12, 13] However, very little is known about their expression, function or potential as prognosticators in leukemias. To gain insights into their role and to evaluate whether existing data warrant further investigation of these genes in AML, we developed and validated a novel approach to quantify the expression of 1664 lincRNAs in 736 AML patients from three well-characterized patient cohorts. Interrogation of this database with matched expression profiles of coding genes reconfirmed previously known regulatory relationships, and uncovered 263 novel lincRNA–coding gene relationships. This unbiased list of *in silico* predictions was valid across three patient cohorts and warrants further experimental/functional validation.

In a univariate survival analysis, we found that 43 lincRNAs were associated with OS in at least two patient cohorts. Screening our database for biomarker panels of up to four lincRNAs, we identified 201 distinct lincRNA-sigs associated with OS across all three patient cohorts. A recent study reported the association of 48 lncRNAs (including 10 lincRNAs), with OS in a cohort of 148 CN-AMLs aged over 60 years.[19] Four of these 10 lincRNAs were among the 1664 lincRNAs detectable by HG133P2 arrays, but none was significantly associated with OS in the three AML cohorts analyzed in our study (Supplementary Table S17). Overall, these data suggest that prognostic markers derived merely from *in silico* analysis of single trial cohorts risk identifying trial-specific rather than broadly applicable and biologically significant risk factors. In contrast, we report that a prognostic marker composed of four lincRNAs—*ENSG00000153363*, *XLOC_013473*, *ENSG00000255571* and *ENSG00000260257*—was an independent predictor of OS across four patient cohorts, including an independent test cohort of 113 AML patients profiled by high-throughput RT-PCR.

However, given the poor functional annotation of lincRNAs in general, it is still unclear whether these particular lincRNAs have a direct role in the pathogenesis and/or maintenance of leukemia or are merely a surrogate readout for a transcriptional milieu that makes leukemic cells sensitive to current treatment. We want to note that, although the expression levels of all four components of the lincRNA-sig constitute its prognostic power, understanding their individual functions will further provide insights into their value as therapeutic targets. For example, we have profiled the expression of these genes at six transitions during normal hematopoiesis, and could show that neither of these genes associated with well-established transition points. However, *ENSG00000153363* and *ENSG00000255571* are specifically upregulated in leukemic cells, and this warrants further functional experiments to test whether reducing the expression of *ENSG00000153363* and/or *ENSG00000255571* affects leukemic cells.

The ELN 2017 iteration incorporates recent updates to the World Health Organization classification,[40] but unlike ELN 2010, which was used in this study, is yet to be applied to AML cohorts with longitudinal data. It now employs a three-group classification (favorable, intermediate and adverse), by merging of intermediate-I and -II groups, where the distinctions were mostly genetic rather than prognostic. AML with biallelic mutated *CEBPA* is now recognized as favorable, mutations in *RUNX1*, *ASXL1* and *TP53* as well as monosomal karyotype are recognized as adverse and consideration is given to the *FLT3-ITD* mutant/allele ratio.[41] Treatment recommendations for AML patients with

favorable (ELN-FR) risk would usually be standard chemotherapy unlike those with an unfavorable (ELN-AR) risk who would be considered for allogenic bone marrow transplantation.[42] On the other hand, patients with intermediate-risk profiles (ELN IR-1 and ELN IR-2) pose a clinical conundrum. The NL cohort was unique among the four AML cohorts in having sufficiently large patient numbers in each ELN risk group for us to apply the lincRNA-sig to all groups (Figure 4d). On the evidence shown here, the lincRNA-sig could be a useful adjunct to clinical decision-making. The adverse arms in the intermediate-risk groups as dichotomized by the lincRNA-sig mirrors ELN adverse risk, and these patients may need to be considered for allogeneic bone marrow transplantation. Conversely, those dichotomized into the favorable arm could possibly be spared an unnecessary transplant.

The purpose of this study was to explore the expression of lincRNAs in the best clinically annotated AML data sets that have formed the foundation of a large body of existing research, and thus limited our analysis to evaluate 1664 poly (A) + lincRNAs. Like others, we show that lincRNAs are often relatively low expressed transcripts and consequently might not be not be well represented in non-targeted and low-coverage transcriptomic studies using RNA-seq (that is, including poly (A) + RNA-seq by the TCGA). In addition, lincRNAs might also be transcribed from antisense loci of coding genes, and thus be difficult to resolve by non-strand-specific transcriptomes measured by RNA-seq (that is, RNA-seq by the TCGA). Interestingly, the ENCODE project reported the widespread expression of non-coding poly (A)–transcripts, suggesting that poly (A)–lincRNAs might also exist in AML.[11] This suggests that an unknown number of novel lincRNAs is expressed in AML, and our investigation of a small subset of lincRNAs further implies that some of these are relevant during leukemogenesis and/or have prognostic value. We believe that deep sequencing followed by targeted sequencing of strand-specific libraries from poly (A) + and poly (A)–transcripts, which fulfill the criteria for lincRNAs, in prospective clinical trials will further refine and translate the signatures that we have described in this manuscript.[43]

This work was supported by research grants, fellowships and scholarships from the National Health and Medical Research Council of Australia (JP, JWHW, DB, RJD'A, HSS, LBT and IDL), Australian Research Council (JWHW), Cancer Institute NSW (DB), Cure Cancer Australia Foundation (DB), Anthony Rothe Memorial Trust (JAIT and DB), Ian Potter Foundation (DB), UNSW Australia (YH and DC), the Translational Cancer Research Network of the Cancer Institute of NSW (DB, YH and DC) and the Wilhelm-Sander-Stiftung (TH). Biospecimens and/ or clinical data were provided by the South Australian Cancer Research Biobank (SACRB), which is supported by the Cancer Council SA Beat Cancer Project, Medvet Laboratories Pty Ltd and the Government of South Australia. We thank Diana Iarossi, Silke Danner and Ing Soo Tiong for maintaining the Acute Leukaemia Laboratory AML database and Ruud Delwel and Peter Valk for helpful discussion and the provision of expression and clinical data from the Dutch-Belgian Hemato-Oncology Cooperative Group.

Author contributions

DB, JAIT, CP, TH, AS, JO, LB, YH, DC, AB, MB, MP, performed research, contributed to study design, analysed and interpreted data; CH, XZ, BJH, AS, J-H K, WEB, BW, TB, WH, SKB, LBT, HSS, IDL, RJD'A. provided vital reagents and data, DB, JWHW and JEP designed the study and wrote the paper.

References

1 Dohner K, Dohner H. Molecular characterization of acute myeloid leukemia. Haematologica 2008; 93: 976–982.
2 Grimwade D, Walker H, Oliver F, Wheatley K, Harrison C, Harrison G et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. Blood 1998; 92: 2322–2333.
3 Thiede C, Koch S, Creutzig E, Steudel C, Illmer T, Schaich M et al. Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). Blood 2006; 107: 4011–4020.
4 Frohling S, Schlenk RF, Breitruck J, Benner A, Kreitmeier S, Tobis K et al. Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. Blood 2002; 100: 4372–4380.
5 Preudhomme C, Sagot C, Boissel N, Cayuela JM, Tigaud I, de Botton S et al. Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA). Blood 2002; 100: 2717–2723.

6 Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH et al. Refinement of cytogenetic classification in acute myeloid leukemia: determina- tion of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood 2010; 116: 354–365.
7 Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK et al. Diagnosis and management of acute myeloid leukemia in adults: recommenda- tions from an international expert panel, on behalf of the European LeukemiaNet. Blood 2010; 115: 453–474.
8 Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrozek K et al. ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. Blood 2011; 118: 6920–6929.
9 Metzeler KH, Maharry K, Radmacher MD, Mrozek K, Margeson D, Becker H et al. TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol 2011; 29: 1373–1381.
10 Kihara R, Nagata Y, Kiyoi H, Kato T, Yamamoto E, Suzuki K et al. Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. Leukemia 2014; 28: 1586–1595.
11 Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK et al. Defining functional DNA elements in the human genome. Proc Natl Acad Sci USA 2014; 111: 6131–6138.
12 Shen XH, Qi P, Du X. Long non-coding RNAs in cancer invasion and metastasis. Mod Pathol 2015; 28: 4–13.
13 Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell 2013; 154: 26–46.
14 Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M et al. A myelo- poiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. Blood 2009; 113: 2526–2534.
15 Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007; 129: 1311–1323.
16 Bertani S, Sauer S, Bolotin E, Sauer F. The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chro- matin. Mol Cell 2011; 43: 1040–1046.
17 Benetatos L, Hatzimichael E, Dasoula A, Dranitsaris G, Tsiara S, Syrrou M et al. CpG methylation analysis of the MEG3 and SNRPN imprinted genes in acute myeloid leukemia and myelodysplastic syndromes. Leuk Res 2010; 34: 148–153.
18 Emmrich S, Streltsov A, Schmidt F, Thangapandi VR, Reinhardt D, Klusmann JH. LincRNAs MON and MIR100HG act as oncogenes in acute megakaryoblastic leukemia. Mol Cancer 2014; 13: 171.
19 Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci USA 2014; 111: 18679–18684.
20 Verhaak RG, Wouters BJ, Erpelinck CA, Abbas S, Beverloo HB, Lugthart S et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. Haematologica 2009; 94: 131–134.
21 Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013; 368: 2059–2074.
22 Li Z, Herold T, He C, Valk PJ, Chen P, Jurinovic V et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. J Clin Oncol 2013; 31: 1172–1181.
23 Friedberg EC. The Writing Life of James D. Watson. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2005; xvii, 193.
24 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007; 8: 118–127.
25 Michelhaugh SK, Lipovich L, Blythe J, Jia H, Kapatos G, Bannon MJ. Mining Affy- metrix microarray data for long noncoding RNAs: altered expression in the nucleus accumbens of heroin abusers. J Neurochem 2011; 116: 459–466.
26 Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends Genet 2013; 29: 569–574.
27 Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci USA 2008; 105: 716–721.
28 Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011; 25: 1915–1927.

29 Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res 2011; 39: 3864–3878.
30 Beck D, Diffner E, Gudgin E, Thoms J, Knezevic K, Pridans C et al. Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia. Blood 2013; 121: 2289–2300.
31 Alvarez-Dominguez JR, Hu W, Gromatzky AA, Lodish HF. Long noncoding RNAs during normal and malignant hematopoiesis. Int J Hematol 2014; 99: 531–541.
32 Ng SW, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature 2016; 540: 433–437.
33 Bullinger L, Valk PJ. Gene expression profiling in acute myeloid leukemia. J Clin Oncol 2005; 23: 6296–6305.
34 Miller BG, Stamatoyannopoulos JA. Integrative meta-analysis of differential gene expression in acute myeloid leukemia. PLoS ONE 2010; 5: e9466.
35 Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R et al. Use of gene- expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004; 350: 1605–1616.
36 Radmacher MD, Marcucci G, Ruppert AS, Mrozek K, Whitman SP, Vardiman JW 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.
37 Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland MC et al. An 86-probe-set gene-expression signature predicts survival in cytogeneti- cally normal acute myeloid leukemia. Blood 2008; 112: 4193–4201.
38 Chen L, Kostadima M, Martens JHA, Canu G, Garcia SP, Turro E et al. Transcrip- tional diversity during lineage commitment of human blood progenitors. Science 2014; 345: 1251033.
39 Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012; 483: 603–607.
40 Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood 2016; 127: 2375–2390.
41 Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2016; 129: 424–447.
42 Dohner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. N Engl J Med 2015; 373: 1136–1152.
43 Clark MB, Mercer TR, Bussotti G, Leonardi T, Haynes KR, Crawford J et al. Quan- titative gene profiling of long noncoding RNAs with targeted RNA sequencing. Nat Methods 2015; 12: 339–342.