Predicting interactome network perturbations in human cancer: application to gene fusions in acute lymphoblastic leukemia

Leon Juvenal Hajingabo, Sarah Daakour, Maud Martin, Reinhard Grausenburger, Renate Panzer-Grümayer, Franck Dequiedt, Nicolas Simonis, and Jean-Claude Twizere

ABSTRACT Genomic variations such as point mutations and gene fusions are directly or indirectly associated with human diseases. They are recognized as diagnostic, prognostic markers and therapeutic targets. However, predicting the functional effect of these genetic alterations beyond affected genes and their products is challenging because diseased phenotypes are likely dependent on complex molecular interaction networks. Using as models three different chromosomal translocations—ETV6-RUNX1 (TEL-AML1), BCR-ABL1, and TCF3-PBX1 (E2A-PBX1)—frequently found in precursor-B-cell acute lymphoblastic leukemia (pre-B-ALL), we develop an approach to extract perturbed molecular interactions from gene expression changes. We show that the MYC and JunD transcriptional circuits are specifically deregulated after ETV6-RUNX1 and TCF3-PBX1 gene fusions, respectively. We also identified the bulk mRNA NXF1-dependent machinery as a direct target for the TCF3-PBX1 fusion protein. Through a novel approach combining gene expression and interactome data analysis, we provide new insight into TCF3-PBX1 and ETV6-RUNX1 acute lymphoblastic leukemia.

INTRODUCTION

The development of every cancer is characterized by frequent genomic aberrations. Investigations focused on specific human neoplasms have identified numerous sequence variants in which mutations are implicated in oncogenesis. These human cancer genes are listed in the Cancer Genome Project database, with genes encoding protein kinase and transcriptional regulation domains highly represented (Futreal et al., 2004). Characterization of the biological properties of some mutated genes, such as the breakpoint cluster region-v-Abelson murine leukemia viral oncogene homologue 1 (BCR-ABL1), has led to the development of successful targeted therapies (Lynch et al., 2004; Gazdar, 2009; Quintas-Cardama and Cortes, 2009; Agrawal et al., 2010; Kaufuss et al., 2013). The most prevalent category among the known cancer genes are chromosomal translocations, often involving immunoglobulin, T-cell receptor, and transcription factor genes (Futreal et al., 2004). Although these rearrangements represent important diagnostic markers that are used to define cancer subtypes (Mitelman et al., 2004; Maher et al., 2009), their molecular interactions and the pathways affected by the result of gene fusions are poorly characterized.

Genes and their products do not act in isolation but as part of complex molecular networks in which most genes play their roles through several molecular functions or interactions. The changes induced by gene fusions and other genetic alterations, as well as modifications of expression levels, do not lead to a complete loss of the gene products and are thus very likely to alter the different interactions of the same gene or protein in distinct fashions (Zhong et al., 2009). Classically, genome-wide transcriptomic studies have been used to identify genes or gene expression signatures in order to...
characterize and classify cancer types or subtypes (Golub et al., 1999; Andersson et al., 2005; Gandemer et al., 2007; Den Boer et al., 2009; Li et al., 2009; Fuka et al., 2011). Although very useful to identify oncogenes and for diagnostic purposes, these methods are limited in their ability to understand the underlying molecular biology, as they are focused on genes, transcripts, and proteins, neglecting the interactions between them.

In this study, we propose a strategy that uses gene expression profiles to identify genes, molecular interactions, and pathways that are important in a specific genetic alteration. We use as models two chromosomal translocations found in precursor-B-cell acute lymphoblastic leukemia (preB-ALL) and involving key specific transcription factors regulating hematopoietic development: 1) the Ets transcription factor variant 6 (ETV6)–runt-related transcription factor 1 (RUNX1) fusion (also known as TEL-AML1) and 2) the transcription factor 3 (TCF3)–pre-B-cell leukemia homeobox 1 (PBX1) fusion (also known as E2A-PBX1; Okuda et al., 1996; Zhou et al., 2012; Tijchon et al., 2013). These chromosomal rearrangements alone are insufficient for leukemogenesis but may support leukemia when additional molecular perturbations are present (Andreasson et al., 2001; Seto, 2010). We thus extracted perturbed molecular interactions and showed that MYC and JunD interactomes are specifically deregulated after ETV6-RUNX1 and TCF3-PBX1 gene fusions, respectively. Furthermore, we demonstrated that the TCF3-PBX1 fusion could impair the normal mRNA export machinery.

RESULTS
Predicting perturbed interactions linked to gene fusions
To predict perturbed molecular interactions specifically linked to ETV6-RUNX1, TCF3-PBX1, and BCR-ABL1 gene fusions, we used the human B-cell interactome (HBCI; Lefebvre, 2007, 2010) and expression data sets from two microarray series (Den Boer et al., 2009; Mullighan et al., 2009), including 24 samples with BCR-ABL1 fusion, 77 with ETV6-RUNX1 fusion, 16 with TCF3-PBX1 fusion, and 248 samples with other different genetic subtypes. Expression data were first normalized by frozen robust multiarray analysis (fRMA; McCall and Irizarry, 2011). For each interaction in HBCI, we computed the difference between the correlation of expression profiles in a group of samples exhibiting a genotype of interest and in the control samples (groups of samples with other genotypes). Because interacting genes/proteins are likely to be involved in similar biological processes and are likely coexpressed (Ge et al., 2001), we selected interactions with significant differences of correlation as deregulated (corrected p < 0.05; Figure 1A).

We detected 2550 perturbed interactions (~4.5% of interactions in the HBCI, involving 664 human genes) and 3334 (~5.8% of the HBCI, involving 1022 human genes) in the ETV6-RUNX1 and TCF3-PBX1 ALL samples, respectively (Supplemental Tables S1 and S2). We found only 74 (0.13%) overlapping interactions between ETV6-RUNX1 and TCF3-PBX1 ALL samples, showing the specificity of the method (Figure 1B). For BCR-ABL1 genotype, which does not involve direct translocation of a transcription factor–coding gene, we detected only 10 (~0.018%) potentially perturbed interactions (Supplemental Table S3). Our next analyses thus will compare perturbed networks for ETV6-RUNX1 and TCF3-PBX1 fusions.

We ranked proteins/genes according to the number of perturbed interactions, and identified MYC (~46% of HBCI) as the most perturbed in the ETV6-RUNX1 subtype of preB-ALL. To confirm the direct link between MYC network alteration and the presence of ETV6-RUNX1 fusion protein, we used HEK293 cells stably expressing ETV6-RUNX1 and control cells expressing similar amounts of MYC (Figure 2A). We performed chromatin immunoprecipitation followed by high-throughput sequencing (ChiP-seq) in cells expressing the ETV6-RUNX1 fusion protein to detect the MYC-binding sites at a genome scale. We identified 557 MYC target genes in both cell lines (Figure 2B and Supplemental Table S4, HEK293 +ETV6-RUNX1 anti MYC and HEK293 anti MYC), representing 19% of MYC target genes reported in the human B-cell interactome (Lefebvre, 2007, 2010). As predicted, this experiment showed a high modification of MYC targets in the presence of ETV6-RUNX1 fusion, with ~88% (489 of 557) of the targets being different between the two cell lines. Among these, 52% were also identified as MYC–perturbed interactions by our method (Figure 2C and Supplemental Table S1), further supporting the use of differences of correlation between expression profiles to predict perturbed interactions.

Topological analysis of the perturbed networks
To determine whether the structure of the network is modified after ETV6-RUNX1 or TCF3-PBX1 fusions, we analyzed network topology perturbations using three metrics: characteristic path length (cpl), edge betweenness centrality (ecb), and edge-clustering coefficient (ecc).

We sequentially removed edges corresponding to perturbed interactions by decreasing order of significance, calculated the cpl, average ecb, and average ecc of the resulting network at each step, and compared these metrics to those obtained by removing random edges (Figure 3, red lines). For ETV6-RUNX1 fusion, we observed a significant increase of cpl and ecb, whereas ecc decreased, indicating that edge perturbations in ETV6-RUNX1 fusion leads to a less compact network but with a globally higher, more evenly distributed communication potential and a lower local connectivity on high-degree nodes than expected at random (Figure 3, compare green to red lines). In the case of TCF3-PBX1 fusion, on the contrary, the perturbed network becomes more compact, with a slightly lower communication potential and local connectivity than expected at random (Figure 3, compare blue and red lines). We also compared ebc, ecc, and the edge shortest path length (espl) of the network composed of perturbed edges with the rest of the network (the network of not-perturbed edges). The espl was computed as the mean of all shortest path lengths between the vertices of an edge and all other vertices in the network. It appeared that perturbed edges after ETV6-RUNX1 fusion are characterized by significantly higher ebc, higher ecc, and lower espl than other edges in the network (Table 1A). Similar local metrics for the TCF3-PBX1 fusion showed minor changes (Table 1B).

Taken together, our network topology analysis suggests that ETV6-RUNX1 chromosomal translocation may lead to disruption of molecular interactions important for B-cell communication circuits, whereas TCF3-PBX1 fusion only slightly modifies the structure of the network.

Specific deregulation of transcription factor networks
We ranked nodes based on the proportion of their perturbed interactions in HBCI, and highlighted the 10 most deregulated nodes in ETV6-RUNX1 and TCF3-PBX1 fusion subtypes of ALL. It appeared that, for both ALL subtypes, top deregulated nodes correspond to diverse transcription regulators (Figure 4, A and B). This result suggests that ETV6-RUNX1 and TCF3-PBX1 fusions support oncogenesis mostly by specifically deregulating other transcriptional regulators. We therefore analyzed the interaction networks of transcription factors (TFs) that have at least one interaction predicted as deregulated and categorized the TFs according to the number of perturbed interactions and the published classification of human TFs.
and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (...). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos, (Wingender et al., 2013). We found that two classes of transcription factors, basic helix-loop-helix (bHLH) and leucine zipper (bZIP), account for the majority of perturbed interactions for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Figure 5, A and B, red arrows). In particular, the majority of bZIP members of the activating protein-1 (AP-1) complexes, including JunD, JunB, c-Jun, c-Fos,
result indicates a p53-independent mode of regulation in ETV6-RUNX1 ALL samples. We thus hypothesized that the ETV6-RUNX1 fusion could compete with the formation of MYC/Miz1 complex or the binding to its down-regulated targets, including the CDKN1A promoter. To test the latter hypothesis, we analyzed the CDKN1A promoter sequence (~499 to +100 base pairs around the start site) using the eukaryotic promoter database (EPD; epd.vital-it.ch) and the TFsearch program (Heinemeyer et al., 1998) and identified two putative RUNX1-binding sites. The ETV6-RUNX1 chromosomal translocation conserves the RUNX1 DNA-binding domain (Supplemental Figure 1A), suggesting that any putative RUNX1-binding site would also be a potential target sequence for ETV6-RUNX1 fusion protein, as previously demonstrated (Wotton et al., 2008; Krapf et al., 2010; Kaindl et al., 2014). We therefore performed ChIP assays using leukemic REH cells (or REH cells silenced in ETV6-RUNX1 expression by short hairpin RNA [shRNA; REH-G]; Fuka et al., 2012; Supplemental Figure 1B) and a specific antibody to ETV6 transcription factor. We specifically amplified a genomic CDKN1A promoter fragment encompassing both putative RUNX1-binding sites, indicating that ETV6-RUNX1 fusion protein directly binds the CDKN1A promoter (Supplemental Figure 5C). In a transcriptional reporter assay, we also demonstrated that the activation of the CDKN1A promoter by a phorbol ester (PMA; Zeng and el-Deiry, 1996) could be inhibited by overexpression of ETV6-RUNX1 fusion protein (Supplemental Figure 1D), similar to the previously reported repressor effect of MYC on CDKN1A promoter (Seoane et al., 2002; Wu et al., 2003). Finally, using HEK293 cells stably expressing ETV6-RUNX1, we compared MYC and ETV6-RUNX1 ChIP-seq data sets and showed that both transcription regulators could target similar genes (Supplemental Table S9). Taken together, these results support the idea that the ETV6-RUNX1 fusion protein might interplay with the MYC transcriptional network.

**Perturbations in signaling pathways**

To explore signaling pathways potentially disturbed after ETV6-RUNX1 and TCF3-PBX1 gene fusions, we adopted two strategies. First, we performed a gene enrichment pathway analysis using the 664 or 1024 human genes involved in ETV6-RUNX1 or TCF3-PBX1 perturbed interactions, respectively. As stated earlier, this data set is restricted to genes involved in B-cell interactions reported in the HBCI database. We performed a functional enrichment analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID; Huang da et al., 2009). We linked perturbed interactions in ETV6-RUNX1 or TCF3-PBX1 ALL samples to similar signaling pathways, including pathways in cancer, T- and B-cell receptor signaling, Toll-like and growth factor signaling, mitogen-activated protein kinase signaling, and cell cycle and cell adhesion (Figure 7, A and D). However, the proportions of perturbed interactions (Figure 7, compare B and E) and the pathway subnetwork profiles (Figure 7, compare C and F) are very different. Of interest, pathways involved in B-cell migration (chemokine receptor signaling) and cell adhesion were specific to the ETV6-RUNX1 fusion, consistent with previous studies showing that ALL cells, including REH, are highly motile and capable of rapid migration within lymphoid tissues (Makrynikola et al., 2012).
et al., 1994; Gandemer et al., 2007). Alternatively, we predicted NF-κB pathway deregulation for TCF3-PBX1, in accordance with the significant proportion of perturbed interactions involving NF-κB RELA (p65) and REL subunits (16 and 14% of HBCI reported interactions, respectively).

![Graphs showing characteristic path length, edge betweenness centrality, and edge clustering coefficient](image)

**FIGURE 3:** Topological analysis of the perturbed BCI network. We show the evolution of three network metrics while removing predicted perturbed interactions in their order of significance (highest difference of correlation first) for ETV6-RUNX1 and TCF3-PBX1 fusions. Abscissa, number of removed edges (perturbed interactions); ordinate, values of the metric. Green, blue, and red curves represents the distributions of the values of the metric removing edges sequentially in ETV6-RUNX1, TCF3-PBX1, and randomly (100 random iterations), respectively. Vertical bars, SEs of random iterations. (A) Characteristic path length (cpl). (B) Edge-betweenness centrality (ebc). (C) Edge-clustering coefficient (ecc).

Next we used Pathway commons (www.pathwaycommons.org), which is a collection of pathways from multiple sources and organisms. Compared to the cell-context HBCI database, we reasoned that Pathway Commons might allow us to uncover unexpected novel functions for TCF3-PBX1 or ETV6-RUNX1 fusion proteins. We thus considered Pathway Commons as a single network and predicted disrupted interactions in the same way as for the HBCI. This analysis revealed 61 and 45 perturbed pathways for ETV6-RUNX1 and TCF3-PBX1 fusions, respectively (Supplemental Tables S6 and S7). Confirming our foregoing results, several pathways linked to MYC transcription factor were predicted as perturbed by the ETV6-RUNX1 fusion. Of interest, we highlighted a potential deregulation of pathways linked to RNA transport machinery after TCF3-PBX1 (Table 2). Perturbed interactions involve several proteins important for RNA processing, including mRNA export proteins such as the eukaryotic translation factor 4A3 (eIF4A3), the nuclear pore complex (NCP/NUP), and the nuclear export receptor NXF1/TAP (Siddiqui and Borden, 2012; Supplemental Table S8). Eukaryotic mRNA is exported from the nucleus either by the bulk export NXF1-dependent pathway or via more specialized factors such as the chromosome region maintenance 1 (CRM1, also called exportin-1 [XPO1]; Hutten and Kehlenbach, 2007; Siddiqui and Borden, 2012). Because most TCF3-PBX1 perturbed interactions involved the NXF1 rather than the CRM1 pathway, we analyzed potential interaction between TCF3-PBX1 fusion protein and NXF1 by examining the subcellular localization of both proteins. We showed that both NXF1 and TCF3-PBX1 colocalize in the nucleoplasm (Supplemental Figure S2A), indicating functional interplay. To test whether TCF3-PBX1 fusion protein could interfere with RNA localization, we visualized RNA molecules in cells transfected with the TCF3-PBX1 fusion and observed colocalization between T3-PBX1 and RNA molecules, and, most important, RNA was delocalized from the nucleoli to the nucleoplasm (Supplemental Figure S2B). Together these results suggest a potential deregulating role of TCF3-PBX1 in the mRNA export machinery.

**DISCUSSION**

As genome-wide expression profiling and interactomic data accumulate and are stored in public databases, the integration to drive interpretation of genotype–phenotype relationships and identify genes and pathways associated with specific diseases remains challenging. Several approaches have been conducted on cancer samples to identify tumor markers and gene expression signatures and to classify cancer types or subtypes. However, functional perturbations arising from expression changes are rarely interpreted in the context of molecular network perturbations, which may be sensitive to subtle transcriptional changes.

In this study, we integrated data from gene expression in B-cell ALL subtypes, molecular interaction networks from the human B-cell interactome, and Pathway Commons databases to provide novel

---

**TABLE 1:** Comparison of local metrics between perturbed edges and the rest of the network for (A) ETV6-RUNX1 and (B) TCF3-PBX1 fusion.

| Metrics            | Perturbed edges | Other edges | p         |
|--------------------|-----------------|-------------|-----------|
| A. ETV6-RUNX1      |                 |             |           |
| Edge shortest path length (espl) | 2.084 | 2.288 | <2.22E-16 |
| Edge betweenness (ebc) | 1838 | 572 | <2.22E-16 |
| Edge clustering coefficient | 0.683 | 0.451 | <2.22E-16 |
| B. TCF3-PBX1       |                 |             |           |
| Edge shortest path length (espl) | 2.259 | 2.280 | 3.5219E-13 |
| Edge betweenness (ebc) | 614 | 628 | 1.859E-12 |
| Edge clustering coefficient (ecc) | 0.457 | 0.462 | 0.4265 |

Difference of the means (p) is assessed through a Mann–Whitney U test.
outcome on the downstream signaling transduction molecules (Pendergast et al., 1993; Skorski et al., 1995; Ren, 2005); and 2) the common changes are not observable in BCR-ABL1 ALL samples and may have been missed. Those unidentified changes found in all subtypes could define the major networks implicated in BCR-ABL1, ETV6-RUNX1, and TCF3-PBX1 leukemia subtypes.

Although we showed that our method is useful in the identification of previously unknown mutant-specific deregulated biological processes, this strategy, like any other system biology model that predicts perturbations, presents some limitations: 1) To predict interactome network perturbations, we calculated the difference of correlation between expression profiles of two genes coding for proteins involved in a protein–protein or protein–DNA interaction. High-throughput data concerning other variables that may influence an interaction, such as mutations in coding sequences, proteins localization and translocation, protein modifications (phosphorylation, acetylation, glycosylation, etc.), and mRNA processing (transport, degradation, stability, etc.), were not included. 2) We applied our methodology to the analysis of ET6-6RUNX1 or TCF3-PBX1 chromosomal rearrangements restricted to precursor-B-cell leukemia, whereas we interrogated a mature-B-cell interactome data set (HBCI), which is an interaction network assembled from a collection of 254 B-cell gene expression profiles representing 24 distinct phenotypes of normal and diseased B cells (Lefebvre et al., 2010). To the best of our knowledge, similar cell-context interactomes for all stages of B-cell development, including precursor B cells, are not yet available. Some important interactions specific for precursor B cells may be missed and some irrelevant interactions may be included in our analysis. 3) The accuracy of our predictions depends on the technical quality of transcriptome and interactome data sets. In our study, transcriptome data sets were from published microarray hybridization data (Den Boer et al., 2009; Mullighan et al., 2009). High-throughput RNA sequencing should provide more precise measurement of gene expression levels and increase the accuracy of our predictions (Wang et al., 2009). In addition, the human B-cell interaction data were obtained by either reverse engineering of transcriptome data or literature curation of interactions. For technical reasons, both methods capture a number of false-positive and false-negative interactions and do not give a complete view of interactomes (Cusick et al., 2009; Dreze et al., 2010; Lefebvre et al., 2010; Yu et al., 2011; Tsang et al., 2014). To summarize, our strategy, like other systems biology predicting models, will improve over time as more accurate cell-specific interactome and transcriptome data are available. In all cases, biological validations are necessary to confirm perturbations of interactome networks in cancer subtypes of interest.

The identification of the MYC network as specifically deregulated after ET6-6RUNX1 fusion could not be anticipated. Of interest, we did not observe dramatic changes in MYC transcript expression levels in ET6-6RUNX1 compared with other subtype of B-ALL, and at the protein level, we show that ectopic expression of ET6-6RUNX1 does not affect MYC expression. We thus speculate that the deregulation of MYC network may be attributed to functional interplay between MYC and ET6-6RUNX1 transcriptional activities. MYC forms highly stable heterodimers with MYC interacting factor X (Max) through their respective basic helix-loop-helix leucine zipper (bHLH2) domains, which specifically bind the E-box (5′-CACGTG-3′) DNA sequences (Nair and Burley, 2003) and recruit different cofactors for transcriptional activation or repression (Conacci-Sorrell et al., 2014; Diolaiti et al., 2014). Our analysis demonstrated that, in ET6-6RUNX1 samples, MYC/Max interaction was not affected, whereas some MYC interactions with cofactors, such as Miz-1, were

hypotheses about deregulated molecular interactions and pathways. We detected 0.018, 4.5, and 5.8% of perturbed interactions in the human B-cell interactome after chromosomal translocations of BCR-ABL1, ET6-6RUNX1, and TCF3-PBX1 fusions in ALL, respectively. Potential perturbed interactions were ranked according to the magnitude of change in gene expression for a pair of interacting partners (Supplemental Tables S1–S3).

The relatively low number of specific perturbed interactions for BCR-ABL subtype (Supplemental Table S3) is unexpected and implies that, for most pairs of interactions, 1) gene expression profiles are too different between BCR-ABL1 ALL samples, consistent with the fact that several breakpoints on chromosomes 9 (for the ABL gene) and 22 (for the BCR gene) may generate kinases with different
FIGURE 5: Network of TFs in (A) ETV6-RUNX1 and (B) TCF3-PBX1 fusion. Left, circles represent TFs that have at least one protein–DNA interaction (PDI) predicted as perturbed; colors correspond to the class of transcription factor. Red/gray edges represent the perturbed/not-perturbed interactions, respectively. Right, chart showing the number of perturbed interactions for each transcription factor class according to TF class (Wingender et al., 2013).
perturbed (Figure 6, red vs. blue lines). Perturbation of MYC/Miz-1 interaction may suggest that the ETV6-RUNX1 fusion could preferentially target the repression function of MYC. It is possible that ETV6-RUNX1 fusion interferes with formation of MYC/Miz-1 complex and recruitment of MYC to target gene transcriptional initiators, as previously reported for the interplay between transforming growth factor-β/Smad signaling pathway and MYC/Miz-1 complex to control p15INK4b and p21WAF1 CDK inhibitors (Seoane et al., 2001, 2002, 2004). In addition, ETV6-RUNX1 may also exert its effect on the MYC network by binding to its target promoters (Supplemental Figure 1, C and D).

Another important result from this study is the possible involvement of the TCF3-PBX1 fusion protein in mRNA transport. RNA export is a central process in gene expression regulation and is an exciting new field in cancer biology. Although overexpression of some components of the mRNA export machinery, such as nucleoporins Nup88 and Nup214 (von Lindern et al., 1992; Xu and Powers, 2009; Kohler and Hurt, 2010), CRM1 (Noske et al., 2008), elf4E (Borden and Culjkovic-Kraljacic, 2010), and GANP, the nuclear adapter for NXF1 (Fujimura et al., 2005), have been associated with other types of cancer, including B lymphomas, our data constitute the first report implicating TCF3-PBX1 in RNA localization and interaction with an export factor, NXF1 (Supplemental Figure S2). Similar to overexpression of elf4E being efficiently inhibited by ribavirin in acute myelogenous leukemia (Kentsis et al., 2004), targeting TCF3-PBX1/mRNA export pathway interactions could lead to effective ALL therapies.

In conclusion, our study establishes the feasibility of predicting specific perturbations of molecular interactions based on gene expression profiles from multiple experiments and different biological conditions. Of importance, we show that integration of interaction data with differences of correlation between expression profiles could classify subtypes within the same lineage and provide specific potential targets.

**MATERIALS AND METHODS**

**Experimental data**

We downloaded from Gene Expression Omnibus (GEO) the Affymetrix HG-U133A expression data sets (GSE13425, GSE12995), comprising 190 and 175 ALL samples, respectively. These data sets contain 24 samples with BCR-ABL1 fusion, 77 with ETV6-RUNX1 fusion, 16 with TCF3-PBX1 fusion, and 248 with various other genetic subtypes (Den Boer et al., 2009; Mullighan et al., 2009).

The list of protein interactions was retrieved from the HBCI (Lefebvre, 2007, 2010), among which were 21,156 protein–protein interactions (PPIs), 41,568 protein–DNA interactions (PDI), and 1925 transcription factor–modulator interactions (TFMIs). We also analyzed the network composed of known cellular pathways in Pathway Commons (Cerami et al., 2011) to predict affected cellular pathways. The Pathway Commons version of 27 October 2011 used in our study contains 2308 pathways collected from multiple sources (HumanCyc, Reactome, NCI-Pathways Interactions Database, Biocarta, and KEGG; Romero et al., 2005; Romero et al., 2008; Matthews et al., 2009; Schaefer et al., 2009; Kanehisa et al., 2012).

**Prediction of disrupted interactions**

Inspired by an oncogene prediction method (Mani et al., 2008), we detected changes in correlation of expression between gene pairs.
First, microarray expression profiles are normalized using FCMRMA (McCall and Irizarry, 2011). Second, for each genotype (fusion) and each interaction in the HBCI, we computed the difference of correlations of expression profiles between a genotype of interest (exhibiting in different groups of patients: gene pairs whose expression correlation values show significantly different values between a test group (e.g., the ALL associated to a particular gene fusion) and a control group of samples (e.g., all other ALL samples).
where $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$ are vectors representing expression profiles of two genes/proteins in interaction; $n$ is the number of samples; $i = \{1, \ldots, n\}$; $d_i = x_i - y_i$ represents the difference between ranks; and $\bar{x}$ and $\bar{y}$ are the sample means of the $X$ and $Y$ vectors, respectively.

The difference of correlations ($\Delta_{cor}$) of two genes between a genotype of interest ($gi$) and other genotypes ($og$) was computed as

$$\Delta_{cor} = \text{cor}(X, Y)_{gi} - \text{cor}(X, Y)_{og}$$

where cor(.,.) represents the correlation function (Spearman’s rank or Pearson’s correlation).

### Table 2: The 10 top perturbed pathways after (A) ETV6-RUNX1 and (B) TCF3-PBX1 gene fusion.

| Pathway name                                      | Number of perturbed interactions | Number of interactions in the current pathway | Corrected p     |
|--------------------------------------------------|----------------------------------|-----------------------------------------------|-----------------|
| A. ETV6-RUNX1                                    |                                  |                                               |                 |
| Validated targets of c-Myc transcriptional activation | 48                              | 489                                           | 2.92E-25        |
| Validated targets of c-Myc transcriptional repression | 27                              | 272                                           | 7.26E-14        |
| Regulation of nuclear SMAD2 3 signaling           | 41                              | 716                                           | 2.56E-13        |
| RNA polymerase II transcription termination       | 31                              | 576                                           | 3.43E-9         |
| Cleavage of growing transcript in the termination region | 31                              | 576                                           | 3.43E-9         |
| Postelongation processing of the transcript       | 31                              | 576                                           | 3.43E-9         |
| Postelongation processing of intron-containing pre-mRNA | 24                              | 406                                           | 6.79E-8         |
| mRNA 3′-end processing                            | 24                              | 406                                           | 6.79E-8         |
| c-Myc pathway                                     | 13                              | 106                                           | 1.15E-7         |
| Regulation of nuclear β-catenin signaling and target gene transcription | 22                              | 362                                           | 1.56E-7         |
| B. TCF3-PBX1                                      |                                  |                                               |                 |
| mRNA transport                                    | 209                             | 1742                                          | 5.55E-17        |
| Transport of mature transcript to cytoplasm       | 138                             | 1004                                          | 9.85E-16        |
| Transport of mature mRNA derived from an intron-containing transcript | 126                             | 934                                           | 8.65E-14        |
| Noncoding RNA metabolism                          | 101                             | 703                                           | 1.09E-12        |
| Small nuclear ribonucleoprotein assembly          | 101                             | 703                                           | 1.09E-12        |
| mRNA surveillance pathway                         | 78                              | 595                                           | 8.85E-8         |
| Transport of the stem-loop binding protein–dependent mature mRNA | 56                              | 378                                           | 2.87E-7         |
| Regulation of nuclear SMAD2 3 signaling           | 86                              | 716                                           | 6.49E-7         |
| Transport of mature mRNAs derived from intronless transcripts | 57                              | 404                                           | 9.90E-7         |
| TNFα                                              | 244                             | 2784                                          | 3.27E-6         |

Corrected $p$ value is computed using Benjamini–Hochberg multtesting correction.
Topological analysis

The characteristic path length (cpl) of a graph G is the average length of the shortest paths between all distinct pairs of vertices in the graph (Watts and Strogatz, 1998). In a nondirected graph, the characteristic path length L(G) is computed as follows:

\[
L(G) = \frac{1}{|V|^2} \sum_{v \in V} \sum_{v' \in V} d(v', v)
\]

where V is the set of vertices and d(v, v') is distance between vertices v and v', that is, the length of the shortest path joining them.

The edge-betweenness centrality (ebc) or B(e) is defined as

\[
B(e) = \sum_{s \neq v \neq t \neq e} \frac{\sigma_{st}(e)}{\sigma_{st}}
\]

where \( \sigma_{st}(e) \) is the number of shortest paths between vertex s and t that pass through the given edge (Newman, 2010).

Chromatin immunoprecipitation

We collected 10^7 REH-G1, REH-C, HEK293-E/R-V5, or HEK295-V5 cells, performed DNA–protein cross-linking using 1% formaldehyde for 8 min at room temperature, and stopped the fixation by adding 125 mM glycine for 5 min at room temperature. Cells were collected and lysed using lysis buffers iL1 and iL2 according to the manufacturer’s instructions (Diagenode, Liege, Belgium), and chromatin DNA was sheared by sonication for two or three runs of 10–30 cycles (depending on the cell line: two runs of 10 cycles for HEK293 and three runs of 30 cycles for REH cells), using the Bioruptor (Diagenode). DNA–protein complexes were immunoprecipitated overnight using validated specific ChIP antibodies for MYC or ETV6 proteins (Seitz et al., 2011; Torrano et al., 2011; N-262 and N-19, respectively; Santa Cruz Biotechnology, Santa Cruz, CA) and positive and negative control antibodies (histone H3 rabbit and normal rabbit immunoglobulin G). An aliquot (10%) was used for regular PCR amplification using specific primers, human RPMI 30 Exon 3 (2014; Cell Signaling, Danvers, MA) as positive control for histone H3 immunoprecipitation and CDKN1A-specific primers (forward, 5’-ACTGCCTTTATTTGGGAC-3’; and reverse, 5’-GATCACAATACCCGTGTCA-3’). The remaining samples (10–20 ng of immunoprecipitated DNA) were used for ChIP-seq library sample preparation and subjected to HiSeq Illumina sequencing according to the manufacturer’s instructions (Illumina, San Diego, CA).

The resulting reads were mapped to the human genome (GRCh37/hg19) using BWA, version 0.6. (Li and Durbin, 2009). We used SWEMBL, version 3.3.1 (ebi.ac.uk/-swilder/SWEMBL/), to identify regions of the genome where multiple reads align (peaks). We adjusted parameters for ChIP-seq and reference (Input) sequence relative to the number of reads in the samples, with a relative gradient of 0.002 (R parameter). The resulting peaks were submitted to GREAT, version 2.0.2 (McLean et al., 2010), to identify gene targets. We assigned each gene to a “regulatory domain” (Dostie et al., 2006; Lieberman-Aiden et al., 2009; Schoenfelder et al., 2010) of a minimum distance of 5.0 kb upstream and 1.0 kb downstream from its transcription start site. We set the extension of the regulatory domain up to 1000.0 kb in both directions. Then each DNA-binding region was associated with all genes whose regulatory domain it overlaps. Comparison between identified target gene lists and statistical analysis were performed using the R statistical package.

Immunofluorescence and confocal microscopy

HeLa cells were seeded onto coverslips in a 24-well plate and transfected with 1 μg of reporter plasmid (p4XAP1-luc, pkB-luc, CMV-luc, pCDKN1A-luc), different amounts of effector plasmids (pFlag-TCF3, pFlag-PBX1, pFlag-TCF3-PBX1 or pFlag-ETV6-RUNX1, PMX-MYC), and 100 ng of a control Renilla luciferase construct using polyethyleneimine (Polysciences Europe, Eppelheim, Germany) at 3 μg/μg of DNA. For CDKN1A promoter activation, cells were treated with 100 μg/ml of PMA. At 24 h post-transfection, cells were lysed and luciferase activities determined.
ACKNOWLEDGMENTS

We thank Jacques van Helden (Technological Advances for Genomics and Clinics, Université de la Méditerranée, Marseille, France) for careful and critical reading of the manuscript. We thank J. L. Hartley, D. Esposito, D.P. LeBrun, G. Leclerc, and C. G. Mullighan and J. R. Downing (St. Jude Children’s Research Hospital, Memphis, TN) for plasmid DNA constructs. We also thank the GI-GA-Interactomics, GI-GA-Imaging and Flow Cytometry, and GI-GA-Genotranscriptomics platforms for technical support. L.J.H., S.M.M., F.D., N.S., and T.J.C. are members of the Fonds National de la Recherche Scientifique. This project was supported by Fonds National de la Recherche Scientifique Televie Grant 7.4624.09 to T.J.C. and N.S.

and a Fondation contre le Cancer Technical Research Platform and Recherche Scientifique Televie Grant 7.4624.09 to T.J.C. and N.S.

We also thank the GIGA-Interactomics, D. Esposito, D.P. LeBrun, G. Leclerc, and C. G. Mullighan and J. R. Hartley, for their careful and critical reading of the manuscript. We thank J. L. Hartley, for her comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293.

REFERENCES

Agrawal M, Garg RJ, Cortes J, Quintas-Cardama A (2010). Tyrosine kinase inhibitors: the first decade. Curr Hematol Malig Rep 5, 70–80.

Andersson A, Olofsson T, Lindgren D,Nilsson B, Ritz C, Eden P, Lassen C, Rade J, Fontes M, Morse H, et al. (2005). Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations. Proc Natl Acad Sci USA 102, 13909–13914.

Andreasson P, Roberts DE (1975). Algorism AS 89: the upper tail probabilities of Spearman’s rho. J R Stat Soc C Appl Stat 24, 377–379.

Borden KL, Culjkovic-Kraljacic B (2010). Ribavirin as an anti-cancer therapy: acute myeloid leukemia and beyond? Leuk Lymphoma 51, 1805–1815.

Cerami EG, Grosse B, Demir E, Rodchenko V, Babur O, Anwar N, Schultz N, Bader GD, Sander C (2011). Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res 39 (Database issue), D685–D690.

Conacci-Sorrell M, McFerrin L, Eisenman RN (2014). An overview of MYC and its interactome. Cold Spring Harb Perspect Biol 6, a019437.

Culjak A, Escherich G, Gladdines JG, Peters ST, Van Zutven LJ, Beverloo HB, Van der Spek PJ, Lefebvre C, Lim WK, Basso K, Dalla Favera R, Califano A (2007). A context-dependent map of the 7-methyl guanosine mRNA cap. Proc Natl Acad Sci USA 101, 10305–10310.

Kohler A, Hurt E (2010). Gene regulation by nucleoporins and links to cancer. Mol Cell 38, 6–15.

Krafl G, Kaindl U, Kilbey A, Fuka G, Inthal A, Joas R, Mann G, Neil JC, Haas OA, Panzer-Grumayer R (2014). Blocking ETV6/RUNX1-induced MDM2 overexpression by Nutlin-3 reactivates p53 signaling in childhood leukemia. Leukemia 28, 600–608.

Kaneshisa M, Goto S, Yato Y, Furumichi M, Tanabe M (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40 (Database issue), D109–D114.

Kaulfuss S, Seemann H, Rampey R, Meyer J, Dressel R, Konig B, Scharf J, Virshup DM (2013). Blockade of the PDK4/CAK family together with SGC leads to diminished proliferation of colorectal cancer cells. Oncotarget 4, 1037–1049.

Kentsis A, Topisirovic I, Culjovic B, Shao L, Borden KL (2004). Ribavirin suppresses eIF4E-mediated oncogenic transformation by physical mimicry of the 7-methyl guanosine mRNA cap. Proc Natl Acad Sci USA 101, 10305–10310.

Coller H, Loh ML, Downing JR, Caligiuri MA, et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531–537.

Heinemeyer T, Wingender E, Reuter H, Herrmann H, Kel AE, Kel OV, Ignatieve EV, Ananko EA, Podkolodnaya OA, Kolpakov FA, et al. (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COM-PeLC. Nucleic Acids Res 26, 362–367.

Huang da W, Sherman BT, Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37, 1–13.

Hutton S, Kehlenbach RH (2007). CRM1-mediated nuclear export: to the core and beyond. Trends Cell Biol 17, 193–201.

Hyndman BD, Thompson P, Bayly R, Cote GP, LeBrun DP (2012). EZA proteins enhance the histone acetyltransferase activity of the transcriptional co-activators CBP and p300. Biochim Biophys Acta 1819, 446–453.

Kaindl U, Morak M, Porteous CM, Mecklenbrauker A, Kauer M, Zeininger M, Attarbaschis A, Haas OA, Panzer-Grumayer R (2014). Blocking ETV6/RUNX1-induced MDM2 overexpression by Nutlin-3 reactivate p53 signaling in childhood leukemia. Leukemia 28, 600–608.

Kaneshisa M, Goto S, Yato Y, Furumichi M, Tanabe M (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40 (Database issue), D109–D114.

Kaulfuss S, Seemann H, Rampey R, Meyer J, Dressel R, Konig B, Scharf J, Virshup DM (2013). Blockade of the PDK4/CAK family together with SGC leads to diminished proliferation of colorectal cancer cells. Oncotarget 4, 1037–1049.

Kentsis A, Topisirovic I, Culjovic B, Shao L, Borden KL (2004). Ribavirin suppresses eIF4E-mediated oncogenic transformation by physical mimicry of the 7-methyl guanosine mRNA cap. Proc Natl Acad Sci USA 101, 10305–10310.
responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350, 2129–2132.

Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palansamy N, Chinnaiyan AM (2009). Transcriptome sequencing to detect gene fusions in cancer. Nature 458, 97–101.

Makrynikola V, Bianchi A, Bradstock K, Gottlieb D, Hewson J (1994). Migration of acute lymphoblastic leukemia cells into human bone marrow stroma. Leukemia 8, 1734–1743.

Mangolini M, de Boer J, Waller-Vorderwülbecke V, Pieters R, den Boer ML, Williams O (2013). STAT3 mediates oncogenic addiction to TEL-AML1 in t(12;21) acute lymphoblastic leukaemia. Blood 122, 542–549.

Mani KM, Lefebvre C, Wang K, Lim WK, Basso K, Dalla-Favera R, Calilano A (2008). A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. Mol Syst Biol 4, 169.

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemnish J, Hermjakob H, Jassal B, et al. (2009). Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res 37 (Database issue), D619–D622.

McCall MN, Irizary RA (2011). Thawing frozen robust multi-array analysis (RMA). BMC Bioinformatics 12, 369.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28, 495–501.

Mitelman F, Johansson B, Mertens F (2004). Fusion genes and rearranged oncogenes as targets of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell 84, 321–330.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Flores N, Gomi K, Yamanishi Y, Aoshima M,ALLERY L, et al. (2000). KEGG: a database resource for the genome-scale analysis of cellular pathways. Nucleic Acids Res 28, 29–30.

Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR (1996). AML1, the target of multiple chromosomal translocations in human leukaemia, is essential for normal fetal liver hematopoiesis. Cell 84, 321–330.

Ojcius DM, Cao J, Schreiber DN, Lane WS, Hopfner KP, Williams J, Kassler M, Minn AJ, Kastan MB, et al. (2012). BCL-2 family proteins induce p21CIP1 expression via Miz-1-dependent interaction with the p21 core promoter. Oncogene 22, 351–360.

Okuda T, de Decker P, Tuckermann J, Hsieh W-L, Hiebert SW, de la Chapelle A, Grosveld G, Downing JR (1996). AML1, the target of multiple chromosomal translocations in human leukaemia, is essential for normal fetal liver hematopoiesis. Cell 84, 321–330.

Ojcius DM, Cao J, Schreiber DN, Lane WS, Hopfner KP, Williams J, Kassler M, Minn AJ, Kastan MB, et al. (2012). BCL-2 family proteins induce p21CIP1 expression via Miz-1-dependent interaction with the p21 core promoter. Oncogene 22, 351–360.