
Leon Juvenal Hajingabo¹, Sarah Daakour², Maud Martin², Reinhard Grausenburger³, Renate Panzer-Grümayer⁴, Franck Dequiedt², Nicolas Simonis¹, * and Jean-Claude Twizere²*

¹ Laboratoire de Bioinformatique des Génomes et des Réseaux (BiGRe), Université Libre de Bruxelles (ULB), B-1050 Bruxelles, Belgium
² Laboratory of Protein Signaling and Interactions, GIGA-Research, University of Liège, B-4000 Liège, Belgium
³ Department of Biotechnology, University of Natural Resources and Life Sciences Vienna, Austria
⁴ Children's Cancer Research Institute, St Anna Kinderkrebsforschung, Vienna, Austria

* These authors contributed equally.

Correspondence should be addressed to: Jean-Claude Twizere (jean-claude.twizere@ulg.ac.be) or Nicolas Simonis (nicolassimonis@gmail.com).

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 impact of these genetic alterations beyond affected genes and their products is challenging because diseased phenotypes are likely dependent of 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 (preB-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 following 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,

Supplemental Material can be found at: http://www.molbiolcell.org/content/suppl/2014/09/29/mbc.E14-06-1038v1.DC1
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 homolog 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, Kaufluss 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 a 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 used, 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 following 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 C 2007, Lefebvre et al. 2010) and expression datasets 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 others different genetic subtypes. Expression data were first normalized by 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). As interacting genes/proteins are likely to be involved in similar biological processes and are likely co-expressed (Ge et al. 2001), we selected interactions with significant differences of correlation as deregulated (corrected P-value <0.05 - Figure 1A).

We detected 2,550 perturbed interactions (~4.5% of interactions in the HBCI, involving 664 human genes), and 3,334 (~5.8% of the HBCI, involving 1,022 human genes), in the ETV6-RUNX1 and TCF3-PBX1 ALL samples, respectively (supplemental table 1 and 2). We only found 74 (0.13%) overlapping interactions between both 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 only detected 10 (~0.018%)
potentially perturbed interactions (supplemental table 3). 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 amount of MYC (Figure 2A). We performed a ChIP-seq experiment in cells expressing the ETV6-RUNX1 fusion protein to detect the MYC-binding sites at a genome scale. We identified a total of 557 MYC targets genes in both cell lines (Figure 2B and supplemental table 4, HEK293 +ETV6-RUNX1 anti MYC and HEK293 anti MYC), representing 19% of MYC target genes reported in the human B-cell interactome (Lefebvre C 2007, Lefebvre et al. 2010).

As predicted, this experiment showed a high modification of MYC targets in the presence of ETV6-RUNX1 fusion, with ~88% (489/557) 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 1), 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 following ETV6-RUNX1 or TCF3-PBX1 fusions, we analyzed network topology perturbations using 3 metrics: characteristic path length (cpl), edge betweenness centrality (ebc) and edge-clustering coefficient (ecc).

We sequentially removed edges corresponding to perturbed interactions by decreasing order of significance, calculated the cpl, average ebc 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 the cpl and ebc while the 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 the ebc, the 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 paths lengths between the vertices of an edge and all other vertices in the network. It appeared that perturbed edges following ETV6-RUNX1 fusion are characterized by significantly higher ebc, higher ecc and lower espl than other edges in the network (Table 1 A). Similar local metrics for the TCF3-PBX1 fusion showed minor changes (Table 1 B).

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 top 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 to the published classification of human TFs (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 5A 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, Fra-1 (FOSL1), FosB, ATF2, ATF3, CREB1 and CREBL2, seems affected following TCF3-PBX1 fusion (Figure 4B and 5B).

To experimentally validate our analysis, we investigated TCF3-PBX1-induced transcriptional regulation of the Jun/AP-1 pathway. HEK293 cells were cotransfected with a luciferase reporter construct harboring AP-1 binding sites (Samuel et al. 2008) and increasing amounts of expression vectors for TCF3-PBX1, TCF3 or PBX1 transcription factors. As shown on Figure 4C, AP-1 transactivation is significantly increased in the presence of the oncogenic fusion TCF3-PBX1, in
accordance with previous reports that demonstrated functional interplay between TCF3-PBX1 and the CREB binding protein (CBP) and p300 transcriptional co-activators (LeBrun 2003, Bayly et al. 2004, Denis et al. 2012, Hyndman et al. 2012) and, that CBP/p300 and AP-1 factors have many partners in common (Denis et al. 2012).

Among the bHLH TFs, the proto-oncogene MYC was the top ranked gene with ~ 46% of HBCI perturbed interactions (Figure 5A). By comparison, only a small minority (1%) of MYC interactions is perturbed in TCF3-PBX1 associated ALL samples, suggesting a specific interplay between MYC and ETV6-RUNX1 transcriptional regulation. This is also illustrated by the network of MYC transcriptional regulators showing that the majority of protein-DNA interactions (such as STAT5 and the MYC promoter) (Denis et al. 2012, Ott et al. 2012, Mangolini et al. 2013) and protein-protein interactions (such as SP1 and MYC co-regulation of some target genes) (Zhang et al. 2012, Wang et al. 2013, Wang et al. 2013) are specifically deregulated in ETV6-RUNX1 compared to TCF3-PBX1 ALL subtypes (Figure 6). To further validate MYC deregulation in ETV6-RUNX1 compared to TCF3-PBX1 subtypes of ALL, we analyzed the relative expression of MYC and its targets involved in cell cycle regulation, including cyclin-dependent kinase inhibitors (CKIs), which are known targets repressed by the MYC/Miz1 complex (Figure 6C and supplemental table 5). Among these key cell cycle regulators, TP53, MYC and CDKN3 were relatively under-expressed, while CDKN1A (p21WAF1/CIP1) expression was unchanged in ETV6-RUNX1 samples compared to other groups of samples. As recently reported for the MDM2 promoter (Kaindl et al. 2014), this result indicates a p53-independent mode of CDKN1A promoter 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 bp 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 chromatin
immunoprecipitation assays (ChIP) using leukemic REH cells (or REH cells silenced in ETV6-RUNX1 expression by shRNA (REH-G1) (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 1C). 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), similarly 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 datasets and showed that both transcription regulators could target similar genes (supplemental table 9). Taken together, these results support the idea that the ETV6-RUNX1 fusion protein might interplay with MYC transcriptional network.

**Perturbations in signaling pathways**

To explore signaling pathways potentially disturbed following ETV6-RUNX1 and TCF3-PBX1 gene fusions, we adopted two strategies:

First, we performed a gene enrichment pathway analysis using the 664 or 1,024 human genes involved in ETV6-RUNX1 or TCF3-PBX1 perturbed interactions, respectively. As stated above, this dataset 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 receptors signaling, Toll-like and growth factors signaling, MAPK signaling, 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 sub-network profiles (Figure 7, compare C and F) are very different. Interestingly, pathways involved in B cell migration (chemokine receptors 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. 1994, Gandemer et al. 2007). Alternatively, we predicted NF-kB pathway deregulation for TCF3-PBX1, in accordance
with the significant proportion of perturbed interactions involving NF-kB RELA (p65) and REL subunits (16% and 14% of HBCI reported interactions, respectively).

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 table 6 and 7). Confirming our above results, several pathways linked to MYC transcription factor were predicted as perturbed by the ETV6-RUNX1 fusion. Interestingly, we highlighted a potential deregulation of pathways linked to RNA transport machinery following 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 8). 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). Since 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 co-localize in the nucleoplasm (Supplemental figure 2A), indicating a functional interplay. To test whether TCF3-PBX1 fusion protein could interfere with RNA localization, we visualized RNA molecules in cell transfected with the TCF3-PBX1 fusion and observed a co-localization between TC3-PBX1 and RNA molecules, and, most importantly, RNA was delocalized from the nucleoli to the nucleoplasm (Supplemental figure 2B). Together, these results suggest a potential deregulating role of TCF3-PBX1 on 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 to identify genes and pathways associated with specific diseases remain challenging. Several approaches have been conducted on cancer samples to identify tumor markers, 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 networks perturbations, which may be sensitive to subtle transcriptional changes.

In this study, we have 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 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 following chromosomal translocations BCR-ABL1, ETV6-RUNX1 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 Table 1-3).

The relatively low number of specific perturbed interactions for BCR-ABL subtype (supplemental table 3) is unexpected, and implies that, for most pairs of interactions, (i) gene expression profiles are too different between BCR-ABL1 ALL samples, consistent with the fact that several breakpoints on chromosome 9 (for the ABL gene) and 22 (for the BCR gene) may generate kinases with different outcome on the downstream signaling transduction molecules (Pendergast et al. 1993, Skorski et al. 1995, Ren 2005) (ii) 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 predict perturbations, presents some limitations: (i) To predict interactome networks 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, proteins modifications (phosphorylation, acetylation, glycosylation,...) and mRNAs processing (transport, degradation, stability,...), were not included. (ii) We applied our methodology to the analysis of ETV6-RUNX1 or TCF3-PBX1 chromosomal rearrangements restricted to precursor-B cell leukemia whereas we interrogated a mature-B cell interactome dataset (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 cell 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. (iii) The accuracy of our predictions depends on the technical quality of transcriptome and interactome datasets. In our study, transcriptome datasets 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 interactions data were obtained either by reverse-engineering of transcriptome data, or by literature curation of interactions. For technical reasons, both methods capture a number of false positives and false negatives 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 network in cancer subtypes of interest.

The identification of the MYC network as specifically deregulated following ETV6-RUNX1 fusion could not be anticipated. Interestingly, we did not observe dramatic changes in MYC transcript expression levels in ETV6-RUNX1 compared to other subtype of B-ALL, and at the protein level, we show that ectopic expression of ETV6-RUNX1 does not affect MYC expression. We thus speculate that the deregulation of MYC network may be attributed to functional interplay between MYC and ETV6-RUNX1 transcriptional activities. MYC forms
highly stable heterodimers with MYC interacting factor X (Max) through their respective basic helix-loop-helix leucine zipper (bHLHZ) domains, which specifically bind the E-box (5’-CACGTG-3’) DNA sequences (Nair and Burley 2003) and recruits different co-factors for transcriptional activation or repression (Conacci-Sorrell et al. 2014, Diolaiti et al. 2014). Our analysis demonstrated that, in ETV6-RUNX1 samples, MYC/Max interaction was not affected while some MYC interactions with co-factors, such as Miz-1, were perturbed (Figure 6, red versus 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 the formation of MYC/Miz-1 complex and the recruitment of MYC to target genes transcriptional initiators, as previously reported for the interplay between TGF-β/Smad signaling pathway and MYC/Miz-1 complex to control p15INK4b and p21WAF1 CDK inhibitors (Seoane et al. 2001, Seoane et al. 2002, Seoane 2004). In addition, ETV6-RUNX1 may also exert its effect on the MYC network by binding to its target promoters (supplemental table 10), as shown here for the CDKN1A promoter (Supplemental Figure 1C 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), eIF4E (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 2). Similarly to overexpression of eIF4E efficiently inhibited by ribavirin in acute melogenous 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. Importantly, we show that integration of interactome 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 datasets (GSE13425, GSE12995), comprising 190 and 175 ALL samples, respectively. These datasets contain 24 samples with \textit{BCR-ABL1} fusion, 77 with \textit{ETV6-RUNXI} fusion, 16 with \textit{TCF3-PBX1} fusion and 248 samples with various other genetic subtypes (Den Boer et al. 2009, Mullighan et al. 2009)

The list of protein interactions was retrieved from the Human B-cell interactome (HBCI) (Lefebvre C 2007, Lefebvre et al. 2010) among which 21,156 protein-protein interactions (PPI), 41,568 protein-DNA interactions (PDI) and 1,925 transcription factor-modulator interactions (TFMI). 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 Oct 27, 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, 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 are detecting changes in correlation of expression between gene pairs in different groups of patients: gene pairs whose expression correlation values show significantly different values between a test group (for example the ALL associated to a particular gene fusion) and a control group of samples (for example, all other ALL samples).

First, microarray expression profiles are normalized using fRMA (McCall and Irizarry 2011). Second, for each genotype (fusion), each interaction in the HBCI, we computed the difference of correlations of expression profiles between a genotype of interest (exhibiting a genotype of interest) and other genotypes. We selected interactions showing significant differences of
correlation for both Pearson’s and Spearman’s rank correlation coefficients. In order to detect interactions with significant differences of correlation, we generated 10,000 data sets permuting expression values across the whole table of the original data set. For each random set, we computed the difference of correlation in the same way as for the original data, and computed the P-value and the corrected P-value (using Benjamini-Hochberg multiple testing correction) from the distribution of all difference of correlation values. Interactions with corrected P-value < 0.05 for both Pearson’s and Spearman’s correlation measures are predicted as perturbed.

**Difference of correlation between expression profiles**

We computed the Spearman’s rank correlation \( (\rho) \) (D. J. Best 1975) and the Pearson’s correlation \( (r) \) (Pearson 1895) using the following formulas:

\[
\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}
\]

\[
r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

Where \( X = (x_1, ..., x_n) \) and \( Y = (y_1, ..., y_n) \) vectors represent expression profiles of two genes/proteins in interaction; \( n \) the number of samples; \( i = \{1, ..., n\} \); \( d_i = x_i - y_i \) represents the difference between ranks; \( \bar{x}, \bar{y} \) are the sample means of \( X \) and \( Y \) vectors.

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

\[
\Delta_{cor} = cor(X,Y)_{gi} - cor(X,Y)_{og}
\]

Where \( cor(\ldots) \) represents the correlation function (Spearman’s rank or Pearson’s correlation).
Topological analysis

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

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

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

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

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

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

The edge clustering coefficient (ecc) is computed as the number of triangles to which a given edge belongs, divided by the number of triangles in which the edge may possibly participate at most, given the degrees of the adjacent nodes (Wang et al. 2012):

\[
EC(u,w) = \frac{\lambda_G(u,w)}{\min(d(u)-1,d(w)-1)}
\]

where \( \lambda_G \) denote the number of triangles that include the edge \((u, w)\), \( d(u) \) and \( d(w) \) are degree of \( u \) and \( w \) respectively. The \( \min(d(u)-1,d(w)-1) \) is the number of triangles in which the edge \((u,w)\) may possibly participate at most.
Cell culture and transfection

HeLa and HEK293 cells were cultured in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 2 mM glutamine and penicillin/streptomycin. The same medium was used to culture ETV6-RUNX1 (E/R)-expressing HEK 293T clones generated as previously described (Fuka et al. 2012). The ETV6-RUNX1- (E/R) positive leukemia cell line REH were cultured in RPMI 1640 supplemented with 10-20% FBS and antibiotics as recommended by the distributor (DSMZ, Germany). Knockdown of E/R was performed as previously described (Fuka et al. 2012) using validated lentiviral vector encoding a U6 promoter driven shRNA targeting the E/R fusion sequence. Knockdown of E/R was monitored by RT-qPCR using published primers and probe sets (Fuka et al. 2012) and cells with at least 50% reduction were used in ChIP and western blot experiments.

Plasmids

Plasmids pFlag-TCF3, pFlag-PBX1, pYFP-NXF1 were generated from the corresponding entry clones (human ORFeome 7.1) by LR recombination into pDEST1899 (Flag N-ter vector) or pDEST491 (YFP N-ter vector) (gift of James L. Hartley and Dominic Esposito, NCI-Frederick). TCF3-PBX1 (gift of Dr. David P. LeBrun, Queen's University) and ETV6-RUNX1 (gift of Dr. Guy Leclerc, University of Miami) cDNA clones were also subcloned by Gateway technology into pDEST1899 Flag N-ter vector.

Luciferase reporter assays

HEK293 cells were transfected with one microgram 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 GmbH, Germany) at 3µg / µg of DNA. For CDKN1A promoter activation, cells were treated with 100 µg/ml of PMA. Twenty-four hours post-transfection, cells were lysed and luciferase activities determined.
Chromatin immunoprecipitation (ChIP)

$10^7$ REH-G1, REH-C, HEK293-E/R-V5 or HEK295-V5 cells were collected, DNA-protein cross-linked using 1% formaldehyde for 8 min at room temperature, and the fixation was stopped by adding 125mM of glycine for 5 min at room temperature. Cells were collected and lysed using the lysis buffer iL1 and iL2 according to the manufacture instructions (Diagenode, Liege, Belgium) and chromatin DNA sheared by sonication for 2-3 runs of 10-30 cycles (depending on the cell line: 2 runs of 10 cycles for HEK293 and 3 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) (Santa Cruz, N-262 and N-19, respectively) positive and negative control antibodies (histone H3 rabbit and normal rabbit IgG). An aliquot (10%) was used for regular PCR amplification using specific primers, human RPL30 Exon 3 (#7014, Cell Signaling, Danvers, MA, USA) as positive control for histone H3 immunoprecipitation and CDKN1A specific primers (Forward: 5’-ACTGCCCTATTTGAGAC-3’ and Reverse: 5’-GATCACATACCTGTTCA-3’). The remaining samples (10 – 20 ng of immunoprecipitated DNA) were used for ChIP-seq library samples preparation and subjected to HiSeq Illumina sequencing according to the manufacturer instructions (Illumina, San Diego, CA, USA).

The resulting reads were mapped to the human genome (GRCh 37/hg19) using BWA version 0.6 (Li and Durbin 2009). We used SWEMBL version 3.3.1 (http://www.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 is associated with all genes whose regulatory domain it overlaps. The comparison between identified target genes lists and statistical analysis were performed using R statistical package.
Immunofluorescence and confocal microscopy

HeLa cells were seeded onto coverslips in a 24-well plate and transfected with 1µg of pFlag-TCF3-PBX1 and/or pYFP-NXF1 plasmids using lipofectamine2000 (Invitrogen). Twenty-four hours post-transfection, cells were washed with PBS, fixed in 3.7% formaldehyde for 20 min at room temperature, permeabilized with 0.5% triton X-100 for 15 minutes at room temperature and incubated with anti-flag M2 antibody (Sigma) diluted in immunofluorescence (IF) buffer (5% FBS, 0.1% tween-20 in PBS) for 1H at room temperature. After extensive washes, cells were incubated with Alexa 568 secondary antibody diluted in the IF buffer and for1H at room temperature. Where indicated, cells were also incubated with the SYTO RNASelect marker (Invitrogen) and mounted with the Prolong gold Antifade reagent with DAPI (Invitrogen). Slides were analyzed by confocal microscopy using the Nikon A1R system and images processed with the IMARIS software.

Acknowledgments

We thank Dr. Jacques van Helden (Technological Advances for Genomics and Clinics, Université de la Méditerranée, Campus de Luminy, 13288 Marseille Cedex 9, France) for his careful and critical reading of the manuscript.

We thank Dr. JL. Hartley and Dr. D Esposito (Protein Expression Laboratory, SAIC-Frederick Inc., NCI-Frederick, Frederick, MD 21702, USA), Dr. DP. LeBrun, (Queen's University, Kingston, Ontario K7L 3N6, Canada), Dr. G. Leclerc, University of Miami, Florida 33101, USA), Dr. CG. Mullighan and Dr. JR. Downing (St. Jude Children’s Research Hospital, Memphis, TN 38105-3678) for plasmid DNA constructs. We also thank the ‘GIGA-interactomics’, ‘GIGA-imaging and flow cytometry’, and ‘GIGA-Genotranscriptomics’ platforms for their technical support. LJH, SD, MM, FD, NS and TJC are members of the Fonds National de la Recherche Scientifique (FRS-FNRS). This project was supported by the FRS-FNRS Televie grant# 7.4624.09 to TJC and NS and the Fondation contre le Cancer Technical research platform & Heavy Equipment grant to FD and TJC.
References


Figure 1

A

Microarray data → Human B-cell Interactome (HBCI) → Pathway Commons interactions

Microarray data normalization using fRMA

Differential in correlation between genotypes

All samples

Other genotypes

Genotype of interest

Prediction of perturbed interactions

Corrected P-value < 0.05

Prediction of perturbed pathways

list of perturbed pathways

Randomize data

10,000 x

10,000 x

B

ETV6-RUNX1

BCR-ABL1

2411
77 samples

0
10
23 samples

3169
16 samples

TCF3-PBX1

74
0
0

0
Figure 1. Prediction of perturbed interactions. A. Flowchart of the method. Arrows show the flow of data analysis: black for microarrays, green and red for HBCI and Pathway commons interactome, respectively. For each interaction in B-cell or pathways interactome, we computed the differential in correlation between genotypes. Significance of the difference in correlation is estimated from randomized data. Interactions with corrected P-value < 0.05 are predicted as perturbed. Dashed lines represent perturbed interactions. B. Venn diagram representing the number of detected perturbed interactions (DPI) in the B-cell interactome for ETV6-RUNX1, BCR-ABL1 and TCF3-PBX1.
Figure 2. ETV6-RUNX1 expression perturbs MYC binding to its targets. A. HEK 293T expressing V5-ETV6-RUNX1 and control cells were subjected to western blot analysis using anti-MYC and anti-V5 antibodies. B. Chromatin immunoprecipitation was performed using an anti-MYC antibody followed by massively parallel sequencing of isolated DNA fragments. Venn diagrams indicate the comparison of MYC target genes identified in HEK versus HEK+ETV6-RUNX1 cell lines. C. Venn diagrams showing a comparison between the numbers of perturbed MYC targets identified by Chip-seq and those predicted by computing the differences of correlation between expression profiles.
Figure 3. **Topological analysis of the perturbed BCI network.** We show the evolution of 3 network metrics while removing predicted perturbed interactions in their order of significance (highest difference of correlation first) for ETV6-RUNXI and TCF3-PBXI fusions. In abscissa, we show the number of removed edges (perturbed interactions), and in ordinate the values of the metric. Green, blue and red curves represents the distributions of the values of the metric removing edges sequentially in ETV6-RUNXI, TCF3-PBXI and randomly (100 random iterations), respectively. Vertical bars represent the standard errors of random iterations. A. Characteristic path length (cpl) B. Edge betweenness centrality (ebc) and C. Edge-clustering coefficient (ecc).
**Figure 4.** Deregulation of transcription factors interactions. **A.** *ETV6-RUNX1* B. *TCF3-PBX1*. Top 10 most perturbed nodes, i.e. genes/proteins showing the highest number of perturbed interactions. Colors: red bars (pred_inter) represent perturbed interactions, blue bars (other HBCI) represent other interactions in the HBCI database. C. Transactivation of an AP-1 reporter by TCF3-PBX1. HEK293 cells were transfected with the reporter plasmid p4XAP1-luc and indicated amounts of effector plasmids (pFlag-TCF3-PBX1, pFlag-TCF3, pFlag-PBX1). Firefly luciferase data were normalized to renilla luciferase counts, and data are reported as mean and standard error of 3 independent experiments in triplicate.
Figure 5. Network of transcription factors (TF) in: A. ETV6-RUNX1 and B. TCF3-PBX1 fusion. Left panel: circles represent TFs that have at least one protein-DNA interaction (PDI) predicted as perturbed, colors correspond
to the class of transcription factor. Red/grey edges represent the perturbed/not-perturbed interactions respectively. Right panel: chart showing the number of perturbed interaction for each transcription factor class according to TFClass (Wingender et al. 2013).
Figure 6. **MYC regulators and the cell cycle.** Network showing HBCI interactions for transcription factors regulating Myc. Red edges represent interactions detected as perturbed, grey edges interactions not detected as perturbed. 

A. for *ETV6-RUNX1*  
B. for *TCF3-PBX1*.  
C. Network showing HBCI interactions for cyclins (CCN’s, light magenta), cyclin-dependent kinases (CDK, light blue) and cyclin-dependent kinase inhibitors (CDKN’s, light green), relative to MYC, RUNX1, ETV6 and MIZ-1.
Figure 7

ETV6-RUNX1

A

Chemokine receptors
Tol-like & growth factors
MAPK signaling
Pathways in cancer
Cell adhesion
Cell cycle
T and B cell receptors

B

D

NF-κB Signaling
MAPK signaling
Pathways in cancer
Cell adhesion
Cell cycle
T and B cell receptors

C

Pathways in cancer
T and B cell receptors
Tol-like & growth factors
MAPK signaling
Cell Cycle
Cell adhesion
Chemokine receptors

D

Tol-like & growth factors

E

Pathways in cancer
NF-κB Signaling
T and B cell receptors
MAPK signaling
Cell cycle
Cell adhesion

TCF3-PBX1

F

Pathways in cancer
T and B cell receptors
Tol-like & growth factors
MAPK signaling
Cell cycle
NF-κB Signaling
Figure 7. Pathways enrichment of genes having at least one interaction predicted as perturbed in: (1) *ETV6-RUNX1* and (2) *TCF3-PBX1* fusions. A, D. The size of the circle represents the number of genes involved in the pathway and the edges size represents the number of shared genes. The size of green/red circle represents the total number of interactions in HBCI/perturbed interactions between genes involved in the pathway, respectively. The edges size represents the shared interactions (grey for interactions in BCI and red for the perturbed ones) between pathways. B and E represents the proportion of perturbed interactions in each pathway. In C, F, we represent the networks of predicted perturbed interactions in each pathway. The circles represent the genes, the edges the interactions.
Table 1. Comparison of local metrics between perturbed edges and the rest of the network for: A ETV6-RUNX1 B TCF3-PBX1 fusion. Difference of the means (p-value) is assessed through a Mann-Whitney U-test.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Perturbed edges</th>
<th>Other edges</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge shortest path length (espl)</td>
<td>2.084</td>
<td>2.288</td>
<td>&lt; 2.22e-16</td>
</tr>
<tr>
<td>Edge betweenness (ebc)</td>
<td>1838</td>
<td>572</td>
<td>&lt; 2.22e-16</td>
</tr>
<tr>
<td>Edge clustering coefficient</td>
<td>0.683</td>
<td>0.451</td>
<td>&lt; 2.22e-16</td>
</tr>
</tbody>
</table>

Table 2. The 10 top perturbed pathways following: A. ETV6-RUNX1 and B. TCF3-PBX1 gene fusion. Corr P-value is computed using Benjamini and Hochberg multitesting correction.

<table>
<thead>
<tr>
<th>Pathway name</th>
<th>Number of perturbed interactions</th>
<th>Number of interactions in the current pathway</th>
<th>Corr P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validated targets of c-Myc transcriptional activation</td>
<td>48</td>
<td>489</td>
<td>2.92E-25</td>
</tr>
<tr>
<td>Validated targets of c-Myc transcriptional repression</td>
<td>27</td>
<td>272</td>
<td>7.26E-14</td>
</tr>
<tr>
<td>Regulation of nuclear SMAD2 3 signaling</td>
<td>41</td>
<td>716</td>
<td>2.56E-13</td>
</tr>
<tr>
<td>RNA Polymerase II Transcription Termination</td>
<td>31</td>
<td>576</td>
<td>3.43E-9</td>
</tr>
<tr>
<td>Cleavage of Growing Transcript in the Termination Region</td>
<td>31</td>
<td>576</td>
<td>3.43E-9</td>
</tr>
<tr>
<td>Post-Elongation Processing of the Transcript</td>
<td>31</td>
<td>576</td>
<td>3.43E-9</td>
</tr>
<tr>
<td>Post-Elongation Processing of Intron-Containing pre-mRNA</td>
<td>24</td>
<td>406</td>
<td>6.79E-8</td>
</tr>
<tr>
<td>mRNA 3-end processing</td>
<td>24</td>
<td>406</td>
<td>6.79E-8</td>
</tr>
<tr>
<td>c-Myc pathway</td>
<td>13</td>
<td>106</td>
<td>1.15E-7</td>
</tr>
<tr>
<td>Regulation of nuclear beta catenin signaling and target gene transcription</td>
<td>22</td>
<td>362</td>
<td>1.56E-7</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>Pathway name</th>
<th>Number of perturbed interactions</th>
<th>Number of interactions in the current pathway</th>
<th>Corr P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathway</td>
<td>Nodes</td>
<td>Edges</td>
<td>p-value</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>-------</td>
<td>---------</td>
<td>---------------</td>
</tr>
<tr>
<td>RNA transport</td>
<td>209</td>
<td>1742</td>
<td>5.55E-17</td>
</tr>
<tr>
<td>Transport of Mature Transcript to Cytoplasm</td>
<td>138</td>
<td>1004</td>
<td>9.85E-16</td>
</tr>
<tr>
<td>Transport of Mature mRNA derived from Intron-Containing Transcript</td>
<td>126</td>
<td>934</td>
<td>8.65E-14</td>
</tr>
<tr>
<td>Non-coding RNA Metabolism</td>
<td>101</td>
<td>703</td>
<td>1.09E-12</td>
</tr>
<tr>
<td>snRNP Assembly</td>
<td>101</td>
<td>703</td>
<td>1.09E-12</td>
</tr>
<tr>
<td>mRNA surveillance pathway</td>
<td>78</td>
<td>595</td>
<td>8.85E-8</td>
</tr>
<tr>
<td>Transport of the SLBP Dependant Mature mRNA</td>
<td>56</td>
<td>378</td>
<td>2.87E-7</td>
</tr>
<tr>
<td>Regulation of nuclear SMAD2 3 signaling</td>
<td>86</td>
<td>716</td>
<td>6.49E-7</td>
</tr>
<tr>
<td>Transport of Mature mRNAs Derived from Intronless Transcripts</td>
<td>57</td>
<td>404</td>
<td>9.90E-7</td>
</tr>
<tr>
<td>TNFalpha</td>
<td>244</td>
<td>2784</td>
<td>3.27E-6</td>
</tr>
</tbody>
</table>