

# **Hematopoietic cancers and Nup98 fusions: determining common mechanisms of malignancy**

**Ph.D. Candidacy examination**

**Juliana S. Capitanio**  
Department of Cell Biology  
Faculty of Medicine and Dentistry  
University of Alberta

Submitted: Tuesday, January 10<sup>th</sup>, 2012  
Oral exam: Tuesday, January 17<sup>th</sup>, 2012  
2:30 p.m. Room 5-10 Medical Sciences Building

Examining committee:  
Dr. Rick Wozniak  
Dr. Paul Melançon  
Dr. Mark Glover  
Dr. Tom Hobman  
Dr. Andrew MacMillan

### **Summary of proposal:**

Chromosomal aberrations are very frequent in leukemias and several recurring mutations capable of malignant transformation have been described. These mutations usually occur in hematopoietic stem cells (HSC), transforming them into leukemia stem cells. NUP98 gene translocations are an example of such chromosomal aberrations; these translocations produce a fusion protein containing the N-terminal portion of Nup98 and the C-terminal of a fusion partner. Over 75% of Nup98 fusions can interact with chromatin, and lead to changes in gene expression. Therefore, I hypothesize that nup98 fusions act as rogue transcriptional regulators in the cell.

Collecting previously published gene expression data (microarray) from HSCs expressing Nup98 fusions, we can generate some preliminary data to corroborate this hypothesis. Several different fusions affect the expression of similar genes; these are involved in a few biological processes in the cell: embryonic development, immune system formation and chromatin organization. Deregulated genes also present similar transcription factor binding sites in their regulatory regions. These putative regulatory transcription factors are highly interconnected through protein-protein interactions and transcriptional regulation among themselves, and they have important roles in cell cycle regulation, embryonic development, hematopoiesis, apoptosis and chromatin modification.

In this proposal I will extend this knowledge to all Nup98 fusions and identify the mechanisms behind the effects they have in gene expression. Using HSCs expressing a GFP tagged version of all known NUP98 gene fusions as a model system, I propose two Aims in order to achieve a network level understanding of the role NUP98 translocations play in leukemogenesis:

**- Aim 1: Constructing the interactome of Nup98 fusion oncoproteins.** Very few protein-protein interactions have been described for nup98 fusion proteins, most of which involve chromatin modifiers that can affect gene expression; in addition, several fusion proteins can interact with chromatin, but only one locus where this interaction occurs has been described. This indicates it is imperative that we expand our knowledge on the subject. This is addressed in two ways in Aim 1: 1) determining protein interaction networks for leukemogenic nup98 fusions – using affinity purification of protein complexes followed by high resolution liquid chromatography and tandem mass spectrometry for protein identification; 2) mapping genome-wide nup98 fusions interaction with chromatin – using a chromatin immunoprecipitation protocol followed by deep sequencing of all DNA regions with enriched binding to the fusions. The results of these two experiments will generate interaction networks of nup98 fusions in the cell. Careful analysis of these networks will identify altered interactions and several novel connections of these fusions to cellular pathways, hinting at their functional role. Given our initial hypothesis that nup98 fusion may act as rogue transcriptional regulator, Aim 2 attempts to further our understanding of the functional role these fusions have within this cellular network.

**- Aim 2: Effects of NUP98 translocations in the transcriptome of HSCs.** The effect of some nup98 fusions in gene expression has been explored through microarray experiments. However, several still remain to be evaluated, and broader comparisons between different fusions is still lacking. This will be achieved by purifying the complete RNA pool present in HSCs expressing different nup98 fusions and submitting all samples to deep sequencing. Results will provide information on the expression level of several mRNAs, splice isoform abundance, RNA editing events, preferential allele expression, non-coding RNA abundances, and presence of mutant transcripts. Merging these data with those obtained in Aim 1 will generate a regulatory network of gene expression, and help clarify the mechanisms by which nup98 fusions affect gene expression as well as the consequences this deregulation has in the cell.

Achieving the Aims described here I will broaden the understanding of how nup98 fusions fit in the cellular landscape and how they lead to the disease phenotype. This integrated differential disease network will be the scaffold from which to tease out affected pathways and identify important therapeutic targets. It will generate several novel testable hypotheses on functional principles and mechanisms underlying leukemia development, and it can be expanded as these new questions are answered making it a more robust and detailed model of this disease.

## **Background:**

Leukemia is a type of hematologic malignancy that originates in the bone marrow and leads to accumulation of immature hematopoietic cells with clonal origin. These leukemic cells can out-compete normal blood cells, replacing them in the bone marrow and spreading to extramedullary sites, and thus interfering with the normal function of the hematological tissue. The Canadian Cancer Society estimates 5000 new cases of leukemia for 2011, with approximately 2500 deaths.<sup>1</sup>

Leukemia is a heterogeneous disease that can be subdivided according to the cell lineage affected (myeloid or lymphoid) and the degree of differentiation of leukemic cells. More recent classifications by the World Health Organization also incorporate pathologic and genetic markers, achieving more biological significance.<sup>2</sup>

### **- Chromosomal aberrations and leukemogenesis: delineating causes and effects.**

Our knowledge of leukemogenesis is greatly influenced by the discovery of recurring chromosomal aberrations and/or gene mutations capable of malignant transformation of cells.<sup>3</sup> The target cell for these mutations is not always known, but increasing evidence indicates that leukemias originate in hematopoietic stem cells (HSCs) that are transformed into leukemia stem cells (LSC) by these chromosomal aberrations. LSCs and HSCs share two important characteristics, self-renewal and differentiation of new hematopoietic tissue. The clonogenic nature of leukemias is similar to that observed in normal hematopoiesis, and only a small specific subset of leukemic cells, LSCs, are capable of indefinite proliferation, as seen only for HSCs.<sup>4</sup>

If we consider HSCs as the cells of origin for most leukemias, we can better comprehend the high incidence of chromosomal aberrations present in this type of cancer. In a normal organism HSCs are usually quiescent and cycle very slowly; this is a protective mechanism to minimize DNA replication errors and the generation of toxic metabolic subproducts. However, quiescence of HSCs also presents a disadvantage. Once DNA damage does occur, specifically DNA double strand breaks (DSB), these cells must repair the lesions through the non-homologous end joining (NHEJ) pathway. NHEJ is an error-prone, mutagenic pathway that often causes chromosomal aberrations.<sup>5</sup> Given the self-renewal property of HSCs, mutations can be transmitted and accumulated, giving rise to LSCs with a high frequency of chromosomal aberrations; in fact, most recurring translocations in leukemia display a NHEJ repair signature.<sup>6</sup>

Most chromosomal translocations in leukemias occur in chromosomal regions where the DNA is more susceptible to double strand breaks. These susceptibility regions can be characterized by several chromatin structural elements, including topo II DNA cleavage sites, DNase I hypersensitive sites, scaffold/matrix attachment regions (S/MAR) and retrotransposon regions (LINE and SINE).<sup>6</sup> Chromosomal position within the nucleus also influences the frequency of translocations. It is likely that loci localized in close proximity inside the nucleus will more frequently translocate in cases of DSBs.<sup>7</sup> Finally, for these translocations to lead to transformation, they must generate fusion proteins that promote an advantage to the cell.

Thousands of chromosomal translocations have been detected in leukemia cells.<sup>8</sup> Most of these fall within two categories: type I mutations promote increased proliferation or survival, and type II mutations impair differentiation or enhance self-renewal. At least one mutation of each type seems to be necessary for leukemogenesis.<sup>9</sup> Leukemic chromosomal translocations generate chimeric fusion proteins, and many of these fusion proteins have similar characteristics: they localize to the nucleus, affect transcriptional regulation, contain a DNA binding domain and cause epigenetic modifications.<sup>10</sup>

In fact, epigenetic changes are a common occurrence in most acute leukemias.<sup>3</sup> If we consider that during normal hematopoiesis a complex program of epigenetic modifications takes place, it becomes clearer how altering epigenetic modifications can affect cell differentiation and self-renewal leading to leukemogenesis.<sup>11</sup> HOX genes are a prime example of epigenetic regulation. The pattern of HOX expression in cells is epigenetically regulated and inherited; each cell in the hematopoietic differentiation continuum displays a specific pattern of HOX genes expressed. More primitive hematopoietic cells have higher levels of HOX expression, as cells differentiate and lose their proliferative capabilities, HOX expression decreases until it becomes absent in completely differentiated cells. Overexpression of numerous HOX genes can induce leukemogenesis, in several cases changes in their expression levels are a result of histone modifications in the 5' HOXA gene.<sup>12</sup>

#### - Nucleoporin genes and cancer.

To date, the literature has revealed five nucleoporins involved in carcinogenesis, Tpr, Nup88, Nup98, Nup214<sup>13</sup> and Nup358.<sup>14</sup> Nup88 expression is up-regulated in several cancers, specially carcinomas. Increase in its protein levels is thought to deregulate NF- $\kappa$ B nuclear transport maintaining it constantly activated.<sup>15</sup> The remaining 4 nucleoporins are involved in carcinogenic gene fusions.

Tpr gene fusions with Met and NTrk1 have been described in gastric cancers and papillary thyroid carcinomas, respectively. In both cases the N-terminal coiled-coil domain of Tpr is juxtaposed with the tyrosine kinase domain of the fusion partner. This leads to dimerization independent of ligand, and to constitutive activation of kinase activity, causing deregulated signaling that leads to carcinogenesis.<sup>13</sup> A similar carcinogenic mechanism is seen in Nup358 fusions. In inflammatory myofibroblastic tumors, the N-terminal leucine zipper of Nup358 is fused to the tyrosine kinase domain of ALK, also leading to deregulated kinase activation.<sup>16</sup>

Gene fusion of NUP214 and NUP98 play a role in leukemogenesis. NUP214 gene fusions with ABL also promote constitutive kinase activation in T-ALL. The kinase domain of ABL is fused to the N-terminal coiled-coil motif of Nup214. Interaction between this motif and nup88 allows localization of the fusion to the nuclear pore complex (NPC), bringing the kinase domains to sufficient proximity for constitutive activation.<sup>17</sup> In very rare cases Nup214 also translocates with SET and DEK in T-ALL and AML, respectively. In these cases almost the full length SET or DEK protein is fused to the C-terminal FG repeat of Nup214.<sup>13</sup> SET-NUP214 can interact with HOXA gene promoters leading to their expression. HOXA expression only occurs in the earliest T-cell precursor, so the fusion blocks T cell differentiation.<sup>18</sup> DEK-NUP214 fusions seem to increase overall protein translation specifically in myeloid cells, possibly facilitating carcinogenesis.<sup>19</sup> SET and DEK are two histone interacting proteins that perform opposing roles in the regulation of access to chromatin. SET promotes and DEK restricts accessibility to chromatin by the transcriptional machinery.<sup>20</sup> It is possible that NUP214 translocation with these genes affects the balance between the two chromatin modifiers.

The NUP98 gene is fused to a wide range of partner genes, resulting in several hematopoietic disorders, especially acute myeloid leukemia (AML).

#### - NUP98 translocations in leukemias.

All NUP98 translocations described thus far generate a chimeric fusion protein that retains the N-terminal portion of Nup98 and the C-terminal of the fusion partner. Most chromosomal breaks take place between exons 11 and 13 of the NUP98 gene.<sup>21</sup> Interestingly, an enrichment of DNase I hypersensitivity sites and a strong prediction of S/MAR are present in this region (Fig. 1), corroborating its increased susceptibility to translocations.

The N-terminal portion of Nup98 (conserved in fusions) contains FG/GLFG repeats flanking a coiled-coil Rae1 interaction site (Fig. 2A). Even though a third of all nucleoporins contain FG repeats, Nup98 is the only GLFG repeat containing nucleoporin in humans.<sup>22</sup> At nuclear pore complexes (NPC), Nup98 interacts with transport molecules, mediating traffic through the NPC. The N-terminal of Nup98 interacts with XPO1, facilitating export of specific proteins from the nucleus,<sup>23</sup> and with TAP and Rae1, promoting mRNA export to the cytoplasm.<sup>24</sup>

Nup98 can also be found away from the NPC, dispersed through the nucleoplasm and in intranuclear structures called GLFG bodies (Fig. 2B).<sup>25</sup> In embryonic *Drosophila* cells, the intranuclear pool of Nup98 interacts with transcriptionally active genes and changes to the level of Nup98 present can modulate their expression, especially in developmental genes.<sup>26</sup> In human cells, the N-terminal GLFG repeats of nup98 have been shown to interact with histone acetyl transferases and histone deacetylases.<sup>27,28</sup>

At least 27 different genes have been found translocated with NUP98 in leukemic patients. Most of these gene fusions lead to myeloid malignancies (AML, CML, MDS); however, six fusions have so far been identified in T-ALL patients. Nup98 fusions are rare (approximately 2% AML cases) however, they usually indicate a poor prognosis. Over half of Nup98 fusions are detected in patients under 20 years of age and only 25% occur in patients with therapy related malignancies.<sup>21</sup> The karyotype of patients with Nup98 fusions is usually simple, with no more than 3 chromosomal aberrations,<sup>29</sup> indicating a strong transformation potential for the fusions and arguing against an increase in genetic instability.

Nup98 fusion partners can be divided into homeodomain and non-HD containing proteins. Fusions with HD containing proteins always maintain the N-terminal GLFG domain of Nup98 fused in frame to the C-terminal HD of the partner gene.<sup>30</sup> NUP98-HOXA9 was the first fusion detected in an AML patient<sup>31,32</sup> and it is currently the best-characterized Nup98 translocation. All non-HD containing partner genes encode putative coiled-coil motifs,<sup>33</sup> a domain usually involved in mediating protein-protein interactions. Chromatin recognition domains, such as plant homeodomain zinc fingers (PHD), are also recurrent in non-HD Nup98 fusion partners.<sup>34</sup>

So far, few common denominators have been identified when it come to the mechanism by which Nup98 fusions may lead to leukemogenesis. Characterizing their effects in altered gene expression indicates a few common targets: fusions of Nup98 with NSD1, KDM5A, PHF23, HOXA9, HOXD13, PRRX1, HHEX and DDX10 seem to increase the expression of HOXA cluster genes; NSD1, HOXA9 and DDX10 fusions also up-regulate the Hox co-factor Meis1; and HOXA9 or HOXD13 translocations increase the expression of interferon responsive genes.<sup>21</sup> A putative mechanism for how these fusions may alter gene expression has only been described for PHD domain containing translocations (NSD1, KDM5A and PHF23). These fusions seem to bind HOXA gene promoters (through the PHD finger) and recruit histone acetylases CBP/p300 (via GLFG domain) that modify chromatin into a transcriptionally active state.<sup>34,35</sup> Nup98 fusions with HD containing proteins are assumed to act directly as transcription factors; they can in some cases collaborate with Meis1, a Hox co-factor that increases specificity and binding to target DNAs.<sup>36</sup>

#### NUP98 translocation partner genes analysis:

Performing bioinformatic analysis of Nup98 fusion partner genes, I uncovered a few recurrent themes (Tab. I). Investigating the interaction profile of these partner proteins I noticed that chromatin and/or DNA binding was a characteristic of 75% of them. The partner genes usually displayed direct DNA interaction or recognition of histone post-translational modifications, with

50% of all fusion partners working as transcriptional regulators. Separating Nup98 fusions leading to myeloid malignancies from those causing T-ALL, an even clearer picture appears: over 95% of myeloid related fusions can interact with DNA/chromatin, with over 2/3 of them acting as transcriptional regulators. T-ALL related fusions show no transcriptional regulators among partner genes and only one of them can interact with DNA/chromatin, likely pointing towards distinct molecular mechanisms for myeloid and lymphoid causing Nup98 fusions.

Evaluating biological processes affected by Nup98 fusion partner genes, I determined that almost 60% of them participate in embryonic regionalization and development, with over half being involved in transcription. Other biological processes over-represented among the fusion partners were regulation of cell proliferation, cell differentiation and chromatin modification (Tab. I). Interestingly, in *Drosophila* cells, Nup98 itself seems to regulate the transcription of developmental and cell cycle genes.<sup>26</sup>

The data above indicate that Nup98 fusions might function as rogue transcriptional regulators, especially in myeloid malignancies. It is possible that these fusions can affect gene expression acting directly as transcription factors (TF), altering histone modifications or deregulating other TFs.

### **Hypothesis formulation:**

The goal of this project is to define common mechanisms by which these different Nup98 fusions lead to malignancy. Based on the background and on the preliminary analysis provided above *I hypothesize that Nup98 fusions, especially those leading to AML, might function as rogue transcriptional regulators, and that their deregulated target genes might impair cell differentiation and increase self-renewal, setting the stage for malignant transformation and acute myeloid leukemia.* Previous studies have attempted to clarify the molecular mechanism of a few Nup98 fusions, mainly by evaluating their effects in gene expression and BM cell transformation potential.<sup>21</sup> However, so far, no one has attempted a more ample comparison of the cellular networks affected by all the described Nup98 fusions.

*In this project, I propose an integrative functional study of the effects of nup98 translocations in bone marrow cells.* Using genome-wide high-throughput experiments and unsupervised data integration, I will construct global-discovery driven molecular networks and genome maps to identify cellular processes altered in Nup98 translocation containing cells. This will generate new predictions that can be tested through experimental perturbations and supervised data integration, leading to the discovery of specific pathways responsible for the disease phenotype. These relevant pathways can indicate key drug targets for this malignancy, and expected drug responses can be modeled in the existing networks, aiding in the development of new therapies for this disease.

### **Preliminary results: Nup98 fusions effects in gene expression.**

In order to further explore the possible role of Nup98 fusions as rogue transcriptional regulators, I collected microarray experiments of bone marrow cells transformed with different fusions for analysis. In order for this multi-experiment analysis to present biological significance, I only used results from experiments performed in similar conditions, ending up with 4 directly comparable sets: NUP98-HHEX, NUP98-HOXA9, NUP98-HOXA10, NUP98-HOXD13<sup>37,38</sup>.

Analyzing the genes whose expression was affected by all Nup98 fusions similarly, we see enrichment of a few biological processes<sup>39</sup>: embryonic development, immune system formation and chromatin organization (Fig. 3). More interestingly, mapping the regulatory regions and transcription factor binding sites (TFBS)<sup>40</sup> present in the deregulated genes, we can see that up and down-regulated genes present several regulatory regions in common, and 38 transcription factor binding site are enriched in both sets of genes (Tab. II). These transcription factors play relevant roles in cell cycle regulation, embryonic development, hematopoiesis, apoptosis and chromatin modifications (Fig. 4). They form a highly interconnected network,<sup>41</sup> indicating protein-protein interactions and transcriptional regulation among themselves (Fig. 5). More in-depth analysis of the role of a few of the TFs putatively regulating genes with altered expression in cells bearing NUP98 translocations can uncover possible novel mechanisms by which these fusions may lead to leukemogenesis.

Wilms Tumor 1 (WT1) is one of the most enriched TFBS in up and down regulated genes. WT1 is a TF with expression restricted to hematopoietic progenitor cells in the bone marrow with a role in their self-renewal. Mutations in WT1 indicate a worse prognosis in acute leukemias and can be found in approximately 10% of AML cases.<sup>42</sup> This transcription factor can work as both a tumor suppressor and an oncogene, and it can enhance or repress transcription of its target genes (such as MYC and BCL-2) depending on cellular conditions.<sup>43</sup> WT1 and N-terminal Nup98 (present in fusions) both interact with CBP, providing an interesting putative mechanism on how Nup98 fusions might be affecting this TF and its targets without affecting its expression level. As another example, CEBPA appears as an important regulatory TF in genes deregulated by Nup98 fusions. Interestingly its own expression is reduced nearly 2 fold in Nup98 fusion expressing cells. Decreased CEBPA expression in BM cells decreases differentiation and increases proliferation of myeloid progenitors leading to leukemia. Another leukemic fusion, AML1-ETO, has also been shown to down-regulate CEBPA expression, and CEBPA mutations that abrogate its function or generate dominant negatives have also been described as leukemogenic.<sup>44</sup> TFBS for MYC are also overrepresented in this gene set, albeit to a lesser extent. MYC expression increases 40% in the presence of Nup98 fusions, similarly to what is observed with several other leukemic chromosomal aberrations (AML1-ETO, PML-RARA, PLZF-RARA, FLT3-ITD) shown to induce c-myc activation. Overexpression of c-myc alone in BM cells can quickly induce fatal AML.<sup>45</sup>

Evaluating only those genes whose expression was increased in the presence of all NUP98 translocations, we see enrichment for regulation of transcription, cell proliferation and immune system development (Fig. S1B). On the other hand, genes with decreased expression in the presence of Nup98 fusions are overrepresented for embryonic development, RNA processing and chromatin modification (Fig. S1A). Transcription factors with enriched binding sites in these genes affect similar biological processes: embryonic and immune system development, cell differentiation and proliferation (Fig. S2). Networks of TFs regulating genes with increased expression or genes with decreased expression are less interconnected than that of TFs that are present in both; however, each network contains at least one module of a few highly interconnected TFs (Fig. S3 and S4).

Given the above results, we can hypothesize that these NUP98 translocations are deregulating transcription factors that control differentiation and self-renewal in primitive hematopoietic cells. The network of transcription factors regulating growth arrest and differentiation in a human myeloid cell line has already been described (Fig. 6A).<sup>46</sup> Superposing the microarray results obtained above into this previously published network (Fig. 6B), we notice that over 80% of the

TFs thought to regulate growth arrest and differentiation have decreased expression in cells containing Nup98 fusions. Additionally, 70% of the TFs represented in this network have enriched TFBS in genes with deregulated expression upon NUP98 translocations. This reinforces the idea that Nup98 fusions can deregulate key TFs in myeloid cells, leading to a cascade of changes in their gene expression profile that ultimately disrupts differentiation and proliferation, promoting leukemogenesis.

A final interesting observation is the enrichment of CpG islands in the promoters of genes with deregulated expression in Nup98 fusions, indicating an important role for epigenetic changes in their altered expression. Epigenetic regulation of gene expression is a hallmark of hematopoiesis, and several leukemic translocations have been shown to alter transcription by altering epigenetic markers in their target genes.<sup>3</sup> HOX genes are an example of this epigenetic regulation and they are highly enriched for CpG islands in humans.<sup>47</sup> Several Nup98 fusions promote HOX genes up-regulation,<sup>21</sup> some of them have been shown to alter histone post translational modifications in the HOX locus,<sup>34,35</sup> and it is possible that Nup98 fusions may also lead to epigenetic deregulation affecting DNA methylation in CpG islands.

### **Aims overview:**

*In this project I propose two Aims in order to achieve a network level understanding of the role NUP98 translocations play in leukemogenesis.* In Aim 1 I intend to produce an interactome for known Nup98 fusion oncoproteins. The experiments described in this section will identify protein interaction networks for Nup98 fusions and map genome-wide regions of its interaction with chromatin. This should provide a picture of the position occupied by these fusions in the cellular landscape. Aim 2 will focus on understanding how this cellular landscape altered by Nup98 fusions originates the disease phenotype seen in leukemia. These experiments will map transcriptome changes induced by NUP98 translocations, further explore their causes and identify their consequences in the cellular context.

Unsupervised integration of data collected from Aims 1 and 2 will uncover unknown mechanisms and pathways affected by Nup98 fusions generating novel hypothesis and predictions. These can be tested through supervised data integration incorporating new experiments and comparable external data sets. Repeating this cycle of hypothesis generation, predictions and tests will fine-tune our model of NUP98 translocation based leukemogenesis separating the key pathways altered from secondary effects.

### **Experimental model: human hematopoietic stem cells**

NUP98 translocations leading to leukemia likely originate in hematopoietic stem cells; therefore it seems reasonable to use this cell line as the experimental model for this proposal. Human HSCs are usually derived from two sources: umbilical cord blood cells (purified from umbilical cord of newborns) or bone marrows. This second pool of cells can be collected directly from the bone marrow, or they can be mobilized to the peripheral blood by cytokines such as granulocyte colony-stimulating factor, and purified from blood samples. Separation of HSCs from other cell types is usually achieved by sorting according to specific known surface markers, such as CD133+, CD34+, CD59+, Thy1/CD90+, CD38lo/-, C-kit/CD117+, lin- (lineage specific markers).<sup>48</sup> These cells can self-renew and originate all other hematopoietic cell lineages. Several protocols have been established for culturing these cells in vitro for long and short-term experiments.<sup>49</sup> They can also be transduced with lentiviral or retroviral vectors, allowing the expression of engineered transgenes of interest.

In this project I propose to use purified human HSCs. These cells will be maintained in vitro and transduced<sup>50</sup> with HIV based lentiviral vectors<sup>51</sup> containing a GFP tag fused to the constructs of interest (Tab. III). After 72 hours of transgene expression, cells will be FACS sorted, and fractions expressing GFP will be collected for the experiments described below.

### **AIM1: Constructing the interactome of Nup98 fusion oncoproteins.**

Very few proteins interacting with Nup98 fusions have been identified thus far, among them are CBP/p300, HDAC1 and AES. All of these binding partners interact with the GLFG domain of Nup98 and could theoretically interact with any of the Nup98 fusions known. Several of these fusions also have a DNA or chromatin interacting domain (C-terminal of partner gene), yet the interaction of only a couple fusions with the HOXA cluster gene promoter has been mapped.

In this Aim I intend to identify protein interactors and map genomic regions bound by Nup98 fusions, as further discussed in Aims 1.1 and 1.2.

#### **AIM 1.1: Determining protein interaction networks for leukemogenic Nup98 fusions.**

Proteins are usually the ‘active’ molecules in a cell, the effectors of the diseased phenotype. Understanding their interaction with cellular factors is paramount for the comprehension of the phenotype they generate, the mechanistic causes of the disease. Protein interaction information can be obtained by compiling interaction data available in the literature, using computational prediction (homologue interactions, putative protein domains interaction, etc.) or using unbiased experimental mapping strategies.<sup>52</sup> In the case of Nup98 fusion proteins, very few protein interactions have been reported in the literature and they do not contain all appropriate controls. Computational predictions for these chimeric proteins would have to be based on indirect data for the 2 full-length proteins forming the chimera, defeating the purpose of identifying differential interactions arising in the fusion proteins and increasing the likelihood of false interactions. Given these restrictions, experimental mapping of Nup98 fusions protein interaction networks seems to be the most suited methodology.

Experimental mapping of protein interaction networks is currently achieved by using two main methodologies: binary protein interaction mapping (yeast two hybrid assays) and protein complexes characterization (affinity purification followed by mass spectrometry).<sup>53</sup> For this Aim, protein complex characterization is the best approach. Y2H assays are not recommended for transcriptional transactivator, which is hypothesized to be the case for Nup98 fusion proteins, and they are performed in an entirely different cell type, which is potentially detrimental since oncogenic Nup98 fusions have been exclusively detected in hematopoietic cells. Affinity purification, on the other hand, can be performed under near physiological conditions in the cell of interest maintaining protein modifications. It can quantitate the protein interactions detected, and its protocol can be combined with other approaches to increase the information obtained (such as the use of cross-linkers or tandem purifications).

In this Aim I will immunoprecipitate GFP tagged Nup98 fusions expressed in HSCs for 72h. Using a protocol adapted from Hubner,<sup>54</sup> non-denaturing cell lysis and protein solubilization will render protein complexes accessible for purification by an antibody against GFP covalently bound to magnetic bead columns. An automated liquid-handling platform will be used for the affinity purification, decreasing experiment variability and time. Purified protein complexes will be trypsin digested directly on columns and eluted peptides submitted to high-resolution liquid chromatography tandem mass spectrometry (LC-MS/MS) for identification (Fig. 7). Since all

experiments will be done in triplicate, label-free quantification of the MS data is also possible (MaxQuant),<sup>55,56</sup> distinguishing specific binders from background proteins as well as providing an estimate of the complex stoichiometry.

Protein interactions detected can be compared to public protein interaction databases<sup>57</sup> to aid in the protein interaction (PPI) network design. A few different analyses can be performed to extract biological insights from these interaction networks. Evaluating each network separately, we will likely notice that several of the nodes (representing a protein) have only a few edges (protein interactions), and a few of the nodes will be highly connected (hubs), holding the network together. Identification of these hubs can indicate important players in the cell; perturbation of these hub proteins by Nup98 fusions can have effects that propagate through the interaction network, making them good drug targets. Searching the network for clusters of highly interconnected nodes, we can identify different protein complexes present. Nodes in these complexes will usually have the same cellular localization and related functions. These protein complexes are usually enriched for specific biological processes and can be mapped into known cellular pathways; connections between complexes can reveal unknown biological relationships that may be important in disease development (Fig. 8 shows examples of PPI networks).

Comparison between the different networks obtained can be even more informative. Extracting the differential interactions between Nup98 fusion proteins and full-length Nup98 can identify and quantify perturbed and novel interactions, providing clues to how their propagation through the network may contribute to the disease phenotype. Comparing the interaction network of different Nup98 fusions can identify common processes affected, as well as alterations specific to groups of translocations such as: AML vs. T-ALL causing translocations, HD vs. PHD vs. CC domains containing fusions, chromatin interacting vs. other fusions, etc. Combining individual and comparative network analysis with ontology and pathway integration, we will begin to understand the effect NUP98 translocations have in changing the cellular landscape.

These protein interaction networks will be complemented with chromatin interaction networks in Aim 1.2, and will serve as the scaffold for mapping gene expression data in Aim 2.

#### AIM 1.2: Mapping genome-wide Nup98 fusions interaction with chromatin.

As described above, over 75% of known leukemic Nup98 fusions can potentially interact with chromatin and/or DNA. In contrast, remarkably little is known about the targets of these interactions. A few Nup98 fusion chromatin immunoprecipitation experiments followed by PCR amplification of genes of interest have revealed a role for this interaction in regulating gene expression. NUP98-HOXA9 fusions interact with the KBTBD10 gene promoter in K562 cells, increasing its expression level.<sup>58</sup> In mouse hematopoietic progenitors NUP98-KDM5A fusions bind the promoters of HOXA6-A10, this stops polycomb complex proteins from silencing them, maintains H3K4me3 and acetylated histones in their promoter, and leads to increased expression of these genes.<sup>34</sup> Similarly, NUP98-NSD1 binds the promoters of HOXA7 and HOXA9, leading to histone acetylation (p300 recruitment by Nup98's N-terminal) and H3K36me3 (NSD1 C-terminal function) that increase gene expression.<sup>35</sup>

In this Aim I will map genome-wide interactions of Nup98 fusions with chromatin. Determining the target sequences interacting with Nup98 fusions can help identify the altered regulatory inputs and link them to the transcriptional outputs observed in these translocations. Currently the main methodology for mapping protein-DNA interactions is chromatin immunoprecipitation (ChIP), in which the protein of interest (Nup98 fusions) is cross-linked to chromatin and purified from the cell of interest, reversal of cross-linking allows purification of

bound DNA fragments that can be identified. Identification of the DNA sequences purified can be achieved by microarray hybridization or DNA deep sequencing, with the second being more advantageous due to increased sensitivity, specificity and genome-wide breadth. Recently, an improved protocol coupling ChIP-seq and exonuclease digestion of DNA not cross-linked to purified proteins allowed single nucleotide resolution of interacting sequences, increasing sensitivity and accuracy while decreasing background (Fig. 9).<sup>59</sup> I will use this protocol to identify chromatin regions bound to GFP tagged Nup98 fusions expressed in HSCs for 72h.

The result of the experiment above will be a large number of short DNA sequences. These are aligned to the genome and converted into the number of reads present in each genomic position. After adjusting results to remove background (non-specific sequences), regions enriched in sequence reads above a set threshold (peak calling) are ranked according to their statistical significance. This genomic map of regions binding Nup98 fusions can then be analyzed to identify genes bound by Nup98 fusions, regulatory elements (promoters, enhancers, repressors, insulators) enriched, and sequence specificity/preference of the interaction (Fig. 10 exemplifies an insert from a ChIP generated genome map).

Integrating data from other groups, for example from different hematopoietic cells or other leukemia cells, I will also compare these results to several other genomic tracks; such as: regions bound by other transcription factors or chromatin modifying enzymes, CpG islands and epigenetic markers, nucleosome positioning, etc. Unsupervised data integration can be used in an unbiased search for emerging patterns, hypothesizing from my preliminary results for example:

- Nup98 fusions may interact with promoters of genes rich in CpG islands that are not methylated in HSCs or LSCs but become methylated in more differentiated cells (such as HOX genes); this may also coincide with changes in histone post-translational modifications present.
- There may be combinatorial interaction of Nup98 fusions and transcription factors such as Jun, Myc, CEBP, HoxA9/Meis1 in a set of genes enriched for myeloid cell differentiation and self-renewal regulation.

Integrating this data with the protein interaction network defined in Aim 1.1 will also produce interesting insights, such as:

- Transcriptional regulators interacting with Nup98 fusions may bind to and regulate the same genes that interact with fusions.
- Gene-sets emerging from pattern searches may be similarly regulated and encode proteins belonging to the same complex or affecting similar functions in the cell.
- Chromatin modifying enzymes interacting with Nup98 fusions may also show an increase in their epigenetic marks in genes bound by these fusions.

The results from Aim 1 will provide an overview of the place occupied by these Nup98 fusions within the cell. I expect that several exciting hypotheses and unexpected connections will become apparent after careful examination of this landscape. In order to further the understanding of the functional role these fusions have within this cellular network, in keeping with this project's initial hypothesis that they may act as rogue transcriptional regulators, in Aim 2 I will explore the effects of nup98 translocations in the cells transcriptome.

## **AIM 2: Effects of NUP98 translocations in the transcriptome of HSCs.**

The effect of some Nup98 fusions in gene expression has previously been explored through microarray experiments. However, broader comparisons between fusions are still lacking, and some have not been studied in detail. Differing experimental conditions present in previously

reported microarray studies of Nup98 fusions limits their direct comparability hampering the identification of common effects. Since most fusions have been shown to affect gene expression in the cell, and given the common changes identified for a few of the Nup98 fusions in the preliminary results of this project, it seems imperative to extend this evaluation to the remaining known NUP98 translocations.

Analyzing the transcriptome, all transcripts present in the cell, can provide invaluable information explaining the disease phenotype observed as a consequence of genomic rearrangements. Transcriptome analysis can be used to assess changes in gene expression by quantification of mRNA species present, but it can also provide information on the presence of mutant transcripts, splice isoform abundance, RNA editing events, preferential allele expression, non-coding RNA abundance, etc. Changes in gene expression are usually studied using two methods, microarray hybridization or deep sequencing of RNA molecules converted to cDNA. RNA-seq presents several advantages when compared to microarray experiments: it is more reproducible, it has lower background, the dynamic range of detection is larger, and estimated mRNA levels correspond better to protein levels in the cell.<sup>60</sup> In addition, more information can be extracted from RNA-seq experiments, such as all the described above in transcriptome analysis.

In this Aim I will evaluate the transcriptome of HSCs expressing different NUP98 translocations. I will purify the total RNA pool from HSCs expressing different GFP tagged Nup98 fusions for 72h. This sample will be depleted of ribosomal RNAs and submitted for RNA-seq (Fig. 11). As in the ChIP-seq experiment described in Aim 1.2, the results of this experiment will be a large number of short cDNA sequences, which will be used in several different analyses.

After they are properly filtered and mapped into the genome sequence, read counts can be used to quantify RNA levels present in the cell. Since the number of reads mapped to a specific RNA are not only a function of its abundance but also of its length, they can be converted into reads per kilobase per million, generating a value directly comparable for genes within the sample. Normalization of this value can also allow inter-study comparisons of gene expression levels. Furthering this analysis with a focus on exon spanning reads and read count in alternatively spliced exons, different isoform abundances can also be quantified. Merging this data with that previously obtained in Aim 1 will help clarify the mechanisms by which Nup98 fusions affect gene expression, for example:

- Genes presenting a change in splice isoform abundance may be interacting with a Nup98 fusion that can directly (such as Nup98-NSD1) or indirectly (recruiting a protein interactor to the gene locus) cause epigenetic changes known to affect splicing, like H3K36me3.<sup>61</sup>
- Genes bound by Nup98 fusions at their regulatory regions, regulated by interacting transcription factors, can have the effects this novel interaction has on their transcription clarified by the RNA-seq results.

The integration of these results will also generate a regulatory network of gene expression (Fig. 12 shows an example of the construction of a gene regulatory network). For instance, Nup98 fusions may bind the locus of a transcription factor gene and recruit interacting histone acetyltransferases (CBP/p300 for example) leading to an increase in its expression level, this type of change would be in the first level of the regulatory network. Increased levels of this transcription factor promote transcription of other genes and miRNAs; these genes will also have increased expression detected but will not be bound by Nup98 fusions, representing a second level of the network. The miRNA with increased expression will lead to reduction in the levels

of its target mRNAs, representing the third level of the regulatory network. Mapping all the data obtained into this regulatory network will allow the identification of key regulatory changes caused by NUP98 translocations as well as the consequences spread by them in a ripple effect through the network, leading to the disease phenotype observed. These key regulators affected would likely be optimum targets for future drug interventions and the effects of these drugs could be modeled through the network in order to infer their effect on the disease.

### **Concluding remarks:**

The project described here aims at improving our understanding of how NUP98 translocations lead to leukemia. Based on previous literature and the preliminary results presented I hypothesized that these Nup98 fusions act as rogue transcriptional regulators, affecting mainly cell differentiation and self-renewal. The directed approach used in experimental design and data integration as well as the limited amount of data available on these genomic aberrations indicates that a more comprehensive approach is still needed. In the experiments described here I will broaden the understanding of how these fusions fit in the cellular landscape and how they lead to the disease phenotype. Generating an integrated differential disease network mapping the protein-protein and protein-chromatin interactions of these fusions as well as their effects in the cellular transcriptome will produce a scaffold from which to tease out important affected pathways and identify important therapeutic interventions.

One of the strengths of this type of unbiased genome-wide experiment is that it has no previous assumptions or expected results intrinsically attached to it. Therefore, all results are informative whether or not they fit into an initial hypothesis. Network analysis can identify key nodes, subnetwork modules and the relation between them, allowing the inference of functional principles and mechanisms underlying cancer development. It is important to remember, however, that this type of approach is not an end in itself; it must be used to generate novel testable hypothesis that must be experimentally validated in more directed experiments.

### **Future directions:**

Due to the broad-based, unbiased characteristics of the experiments proposed here, it is difficult to theorize on results to be obtained. Independent of the findings ensuing from this project, the most important expected result is the generation of several new directly testable hypotheses. Any future directions will focus on more targeted approaches to test the hypotheses generated, be it in wet-lab or in silico experiments. The importance of any such project is to ultimately create a positive impact for patients suffering from the disease studied, be it through better understanding of its mechanisms, improved diagnostic or prognostic assessment, or generation of novel therapeutic interventions. A widespread initial study, like the one proposed in this project, has potential to provide significant contribution to all these areas of discovery.

The network level understanding of a disease is an invaluable resource, as new knowledge is created it can be integrated into this network, facilitating its understanding in a larger context of the disease; with data integration this network can be almost infinitely expanded making it more robust and detailed. Therefore, it seems likely that the future of cancer biology will greatly lean on systems biology and the creation of tumor biology networks as models of this disease.

## **References:**

- 1 Canadian Cancer Society., Public Health Agency of Canada., Statistics Canada. Canadian cancer statistics 2011 featuring colorectal cancer. In: *Canadian cancer statistics*,. Toronto, Ont.: Canadian Cancer Society. 2011; 1 electronic text (132 p.).
- 2 Vardiman JW. The World Health Organization (WHO) classification of tumors of the hematopoietic and lymphoid tissues: an overview with emphasis on the myeloid neoplasms. *Chem Biol Interact* 2010; **184**: 16-20.
- 3 Chen J, Odenike O, Rowley JD. Leukaemogenesis: more than mutant genes. *Nat Rev Cancer* 2010; **10**: 23-36.
- 4 Passegue E, Jamieson CH, Ailles LE *et al*. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proc Natl Acad Sci U S A* 2003; **100 Suppl 1**: 11842-9.
- 5 Mohrin M, Bourke E, Alexander D *et al*. Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell* 2010; **7**: 174-85.
- 6 Zhang Y, Rowley JD. Chromatin structural elements and chromosomal translocations in leukemia. *DNA Repair (Amst)* 2006; **5**: 1282-97.
- 7 Zhang Y, Gostissa M, Hildebrand DG *et al*. The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Adv Immunol* 2010; **106**: 93-133.
- 8 **Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer.** In: (Mitelman F JBaMF, ed), Vol. 2011. <http://cgap.nci.nih.gov/login.ezproxy.library.ualberta.ca/Chromosomes/Mitelman>. 2011.
- 9 Bachas C, Schuurhuis GJ, Hollink IH *et al*. High-frequency type I/II mutational shifts between diagnosis and relapse are associated with outcome in pediatric AML: implications for personalized medicine. *Blood* 2010; **116**: 2752-8.
- 10 Scandura JM, Boccuni P, Cammenga J *et al*. Transcription factor fusions in acute leukemia: variations on a theme. *Oncogene* 2002; **21**: 3422-44.
- 11 Cedar H, Bergman Y. Epigenetics of haematopoietic cell development. *Nat Rev Immunol* 2011; **11**: 478-88.
- 12 He H, Hua X, Yan J. Epigenetic regulations in hematopoietic Hox code. *Oncogene* 2011; **30**: 379-88.
- 13 Xu S, Powers M. Nuclear pore proteins and cancer. *Semin Cell Dev Biol* 2009; **20**: 620-30.
- 14 Chen ST, Lee JC. An inflammatory myofibroblastic tumor in liver with ALK and RANBP2 gene rearrangement: combination of distinct morphologic, immunohistochemical, and genetic features. *Hum Pathol* 2008; **39**: 1854-8.
- 15 Köhler A, Hurt E. Gene regulation by nucleoporins and links to cancer. *Mol Cell* 2010; **38**: 6-15.
- 16 Ma Z, Hill DA, Collins MH *et al*. Fusion of ALK to the Ran-binding protein 2 (RANBP2) gene in inflammatory myofibroblastic tumor. *Genes Chromosomes Cancer* 2003; **37**: 98-105.

- 17 De Keersmaecker K, Rocnik JL, Bernad R *et al.* Kinase activation and transformation by NUP214-ABL1 is dependent on the context of the nuclear pore. *Mol Cell* 2008; **31**: 134-42.
- 18 Van Vlierberghe P, van Grotel M, Tchinda J *et al.* The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. *Blood* 2008; **111**: 4668-80.
- 19 Ageberg M, Drott K, Olofsson T *et al.* Identification of a novel and myeloid specific role of the leukemia-associated fusion protein DEK-NUP214 leading to increased protein synthesis. *Genes Chromosomes Cancer* 2008; **47**: 276-87.
- 20 Gamble MJ, Fisher RP. SET and PARP1 remove DEK from chromatin to permit access by the transcription machinery. *Nat Struct Mol Biol* 2007; **14**: 548-55.
- 21 Gough SM, Slape CI, Aplan PD. NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. *Blood* 2011; **118**: 6247-57.
- 22 Iwamoto M, Asakawa H, Hiraoka Y *et al.* Nucleoporin Nup98: a gatekeeper in the eukaryotic kingdoms. *Genes Cells* 2010; **15**: 661-9.
- 23 Oka M, Asally M, Yasuda Y *et al.* The mobile FG nucleoporin Nup98 is a cofactor for Crm1-dependent protein export. *Mol Biol Cell* 2010; **21**: 1885-96.
- 24 Blevins M, Smith A, Phillips E *et al.* Complex formation among the RNA export proteins Nup98, Rae1/Gle2, and TAP. *J Biol Chem* 2003; **278**: 20979-88.
- 25 Griffis E, Altan N, Lippincott-Schwartz J *et al.* Nup98 is a mobile nucleoporin with transcription-dependent dynamics. *Mol Biol Cell* 2002; **13**: 1282-97.
- 26 Kalverda B, Pickersgill H, Shloma V *et al.* Nucleoporins directly stimulate expression of developmental and cell-cycle genes inside the nucleoplasm. *Cell* 2010; **140**: 360-71.
- 27 Kasper L, Brindle P, Schnabel C *et al.* CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol Cell Biol* 1999; **19**: 764-76.
- 28 Bai X, Gu B, Yin T *et al.* Trans-repressive effect of NUP98-PMX1 on PMX1-regulated c-FOS gene through recruitment of histone deacetylase 1 by FG repeats. *Cancer Res* 2006; **66**: 4584-90.
- 29 Romana SP, Radford-Weiss I, Ben Abdelali R *et al.* NUP98 rearrangements in hematopoietic malignancies: a study of the Groupe Francophone de Cytogénétique Hématologique. *Leukemia* 2006; **20**: 696-706.
- 30 Moore MA, Chung KY, Plasilova M *et al.* NUP98 dysregulation in myeloid leukemogenesis. *Ann NY Acad Sci* 2007; **1106**: 114-42.
- 31 Nakamura T, Largaespada D, Lee M *et al.* Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nat Genet* 1996; **12**: 154-8.
- 32 Borrow J, Shearman AM, Stanton VP *et al.* The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nat Genet* 1996; **12**: 159-67.
- 33 Hussey D, Dobrovic A. Recurrent coiled-coil motifs in NUP98 fusion partners provide a clue to leukemogenesis. *Blood* 2002; **99**: 1097-8.
- 34 Wang G, Song J, Wang Z *et al.* Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* 2009; **459**: 847-51.
- 35 Wang GG, Cai L, Pasillas MP *et al.* NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat Cell Biol* 2007; **9**: 804-12.

- 36 Yung E, Sekulovic S, Argiropoulos B *et al.* Delineating domains and functions of NUP98 contributing to the leukemogenic activity of NUP98-HOX fusions. *Leuk Res* 2011; **35**: 545-50.
- 37 Palmqvist L, Pineault N, Wasslavik C *et al.* Candidate genes for expansion and transformation of hematopoietic stem cells by NUP98-HOX fusion genes. *PLoS One* 2007; **2**: e768.
- 38 Jankovic D, Gorelo P, Liu T *et al.* Leukemogenic mechanisms and targets of a NUP98/HHEX fusion in acute myeloid leukemia. *Blood* 2008; **111**: 5672-82.
- 39 Huang dW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**: 1-13.
- 40 Gotea V, Ovcharenko I. DiRE: identifying distant regulatory elements of co-expressed genes. *Nucleic Acids Res* 2008; **36**: W133-9.
- 41 Szklarczyk D, Franceschini A, Kuhn M *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011; **39**: D561-8.
- 42 Owen C, Fitzgibbon J, Paschka P. The clinical relevance of Wilms Tumour 1 (WT1) gene mutations in acute leukaemia. *Hematol Oncol* 2010; **28**: 13-9.
- 43 Yang L, Han Y, Suarez Saiz F *et al.* A tumor suppressor and oncogene: the WT1 story. *Leukemia* 2007; **21**: 868-76.
- 44 Pabst T, Mueller BU. Complexity of CEBPA dysregulation in human acute myeloid leukemia. *Clin Cancer Res* 2009; **15**: 5303-7.
- 45 Luo H, Li Q, O'Neal J *et al.* c-Myc rapidly induces acute myeloid leukemia in mice without evidence of lymphoma-associated antiapoptotic mutations. *Blood* 2005; **106**: 2452-61.
- 46 Suzuki H, Forrest AR, van Nimwegen E *et al.* The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet* 2009; **41**: 553-62.
- 47 Branciamore S, Chen ZX, Riggs AD *et al.* CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors. *Proc Natl Acad Sci U S A* 2010; **107**: 15485-90.
- 48 5. Hematopoietic Stem Cells. In: *Stem Cell Information*, Vol. 2012. Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services. 2011.
- 49 Tse W, Bunting KD. The expanding tool kit for hematopoietic stem cell research. *Methods Mol Biol* 2008; **430**: 3-18.
- 50 Uchida N, Hsieh MM, Hayakawa J *et al.* Optimal conditions for lentiviral transduction of engrafting human CD34+ cells. *Gene Ther* 2011; **18**: 1078-86.
- 51 Verhoeven E, Cosset FL. Hematopoietic stem cell targeting with surface-engineered lentiviral vectors. *Cold Spring Harb Protoc* 2009; **2009**: pdb.prot5276.
- 52 Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease. *Cell* 2011; **144**: 986-98.
- 53 Nibbe RK, Chowdhury SA, Koyutürk M *et al.* Protein-protein interaction networks and subnetworks in the biology of disease. *Wiley Interdiscip Rev Syst Biol Med* 2011; **3**: 357-67.
- 54 Hubner NC, Bird AW, Cox J *et al.* Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *J Cell Biol* 2010; **189**: 739-54.

- 55 Mueller LN, Rinner O, Schmidt A *et al.* SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* 2007; **7**: 3470-80.
- 56 Mortensen P, Gouw JW, Olsen JV *et al.* MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J Proteome Res* 2010; **9**: 393-403.
- 57 Zhang GL, DeLuca DS, Brusic V. Database resources for proteomics-based analysis of cancer. *Methods Mol Biol* 2011; **723**: 349-64.
- 58 Yassin ER, Sarma NJ, Abdul-Nabi AM *et al.* Dissection of the transformation of primary human hematopoietic cells by the oncogene NUP98-HOXA9. *PLoS One* 2009; **4**: e6719.
- 59 Rhee HS, Pugh BF. Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. *Cell* 2011; **147**: 1408-19.
- 60 Fu X, Fu N, Guo S *et al.* Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* 2009; **10**: 161.
- 61 Luco RF, Pan Q, Tominaga K *et al.* Regulation of alternative splicing by histone modifications. *Science* 2010; **327**: 996-1000.
- 62 Emig D, Salomonis N, Baumbach J *et al.* AltAnalyze and DomainGraph: analyzing and visualizing exon expression data. *Nucleic Acids Res* 2010; **38**: W755-62.
- 63 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**: 118-27.
- 64 Kent WJ, Sugnet CW, Furey TS *et al.* The human genome browser at UCSC. *Genome Res* 2002; **12**: 996-1006.
- 65 Singh GB, Kramer JA, Krawetz SA. Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res* 1997; **25**: 1419-25.
- 66 Gehlenborg N, O'Donoghue SI, Baliga NS *et al.* Visualization of omics data for systems biology. *Nat Methods* 2010; **7**: S56-68.
- 67 Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet* 2010; **11**: 476-86.
- 68 Martin JA, Wang Z. Next-generation transcriptome assembly. *Nat Rev Genet* 2011; **12**: 671-82.
- 69 Cheng C, Yan KK, Hwang W *et al.* Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. *PLoS Comput Biol* 2011; **7**: e1002190.

Table I: Nup98 fusion genes characteristics overview.

Nup98 fusion genes			Total	AML	ALL
Interaction	DNA/Chromatin binding		77.78%	95.24%	16.67%
	Transcription regulator		51.85%	66.67%	0.00%
Protein domains	Homeobox		35.71%	40.00%	0.00%
	Coiled-coil		62.96%	52.38%	100.00%
	Zinc finger PHD		17.86%	20.00%	0.00%
Biological Process	Embryonic regionalization		59.26%	57.14%	66.67%
	Cell differentiation		18.52%	23.81%	0.00%
	Regulation of cell proliferation		22.22%	19.05%	33.33%
	Transcription		51.85%	47.62%	66.67%
	Chromatin modification		14.81%	9.52%	33.33%
Lineage	Myeloid		77.78%		
	Lymphoid			22.22%	

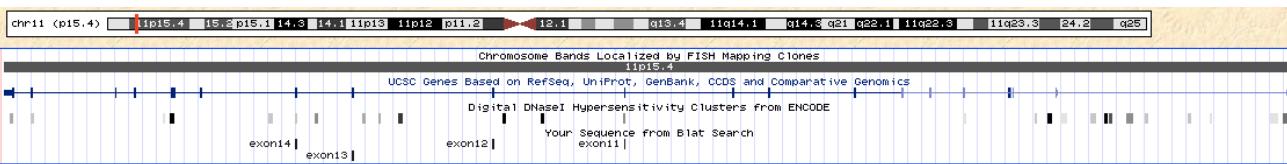
Table II: Transcription factors with TFBS enriched in genes with deregulated expression in the presence of nup98 fusions.

Common TFs	Symbol	Gene ID	Occurrence		Importance	
			up reg genes	down reg genes	up reg genes	down reg genes
CEBP	cebpα	12606	8.82%	4.66%	0.01544	0.0459
	cepbβ	12608				
CMAF	maf	17132	4.41%	1.43%	0.0024	0.00896
CRX	crx	12951	2.94%	1.08%	0.00037	0.00329
DEC	Bhlhe40	20893	7.35%	2.87%	0.07077	0.00621
E2A	tcf3	21423	8.82%	7.17%	0.02206	0.00341
E2F1DP1	e2f1	13555	5.88%	2.15%	0.06655	0.00324
	tfdp1	21781				
E2F1DP1RB	e2f1	13555	5.88%	2.87%	0.03318	0.01098
	tfdp1	21781				
	rb1	19645				
E2F4DP1	e2f4	104394	5.88%	2.87%	0.00147	0.00229
	tfdp1	21781				
FOXJ2	foxj2	60611	7.35%	1.79%	0.13281	0.01915
FOXO4	foxo4	54601	7.35%	1.08%	0.00184	0.00184
GC	CpG islands		16.18%	19.35%	0.03808	0.13548
GCM	gcm2	107889	7.35%	4.30%	0.06147	0.03602
	gcm1	14531				
GRE	glucocorticoid res. elem.		1.47%	1.08%	0.0136	0.00197
HFH1	foxq1	15220	2.94%	0.36%	0.01454	0.00205
HIF1	hif1a	15251	11.76%	9.68%	0.02061	0.04876
HNF1	hnf1a	21405	8.82%	2.51%	0.03529	0.01356
	hnf1b	21410				
IK1	ikzf1	22778	4.41%	2.15%	0.02509	0.00403
LXR	Nr1h2	22260	7.35%	2.15%	0.07146	0.01984
	nr1h3	22259				
MYB	myb	17863	7.35%	2.51%	0.02941	0.00044
MYC	myc	17869	1.47%	2.87%	0.01418	0.02151
MYOGENIN	myog	17928	7.35%	6.09%	0.05285	0.00515
NERF	elf2	69257	8.82%	5.02%	0.09816	0.01944
NFKAPPAB50	nfkb1	18033	7.35%	2.51%	0.16176	0.00659
NFKB	rela	19697	8.82%	6.81%	0.02105	0.06719
NFY	nfia	18027	13.24%	4.66%	0.03174	0.00453
	nfib	18028				
	nfic	18029				
	nfix	18032				
P53	trp53	22059	8.82%	1.43%	0.14366	0.00409
PAX	pax1	18503	5.88%	2.51%	0.04559	0.01687
PAX9	pax9	18511	4.41%	2.87%	0.01665	0.01093
RSRFC4	mef2a	17258	2.94%	0.72%	0.00126	0.00905
SREBP	srebf1	20787	5.88%	2.87%	0.03676	0.0145
	srebf2	20788				
STAT	stat1	20846	7.35%	2.51%	0.0432	0.01443
TAL1	tal1	21349	5.88%	2.15%	0.06324	0.00225
TCF4	tcf4	21413	2.94%	2.15%	0.04403	0.01667
TEL2	telo2	71718	1.47%	2.15%	0.00729	0.01726
TST1	pou3f1	18991	1.47%	1.08%	0.01556	0.0121
USF2	usf2	22282	1.47%	2.87%	0.00023	0.02222
WT1	wt1	22431	14.71%	6.45%	0.03808	0.03055
ZF5	zfp161	22666	27.94%	23.30%	0.08507	0.13499
OCT4	pou5f1	18999	2.94%	1.08%	0.01838	0.00753

Table III: GFP tagged constructs to be transduced into HSCs.

Tag	N-terminal	C-terminal
GFP	NUP98 (GLFG domain)	ADD3
		CCDC28A
		DDX10
		HHEX
		HMGB3
		HOXA11
		HOXA13
		HOXA9
		HOXC11
		HOXC13
		HOXD11
		HOXD13
		IQCG
		KDM5A
		LNP1
		MLL
		NSD1
		PHF23
		PRRX1
		PRRX2
		PSIP1
		RAP1GDS1
		RARG
		SETBP1
		TOP1
		TOP2B
		WHSC1L1
		--
	NUP98 (Full length)	--
	--	--

A



B

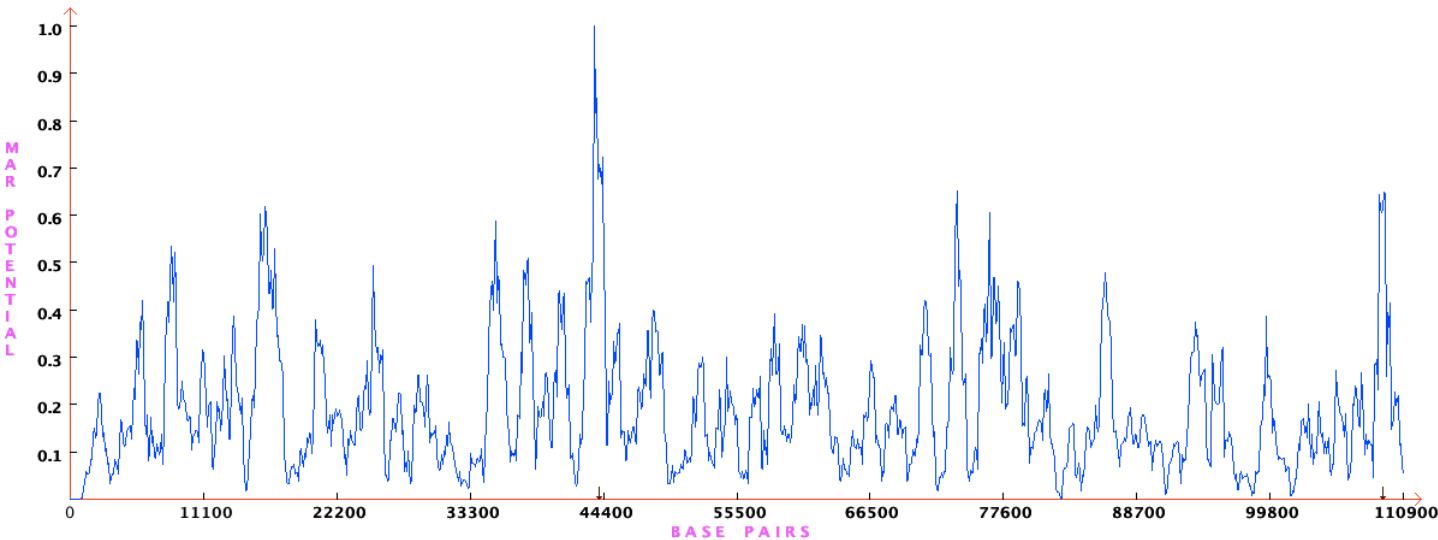
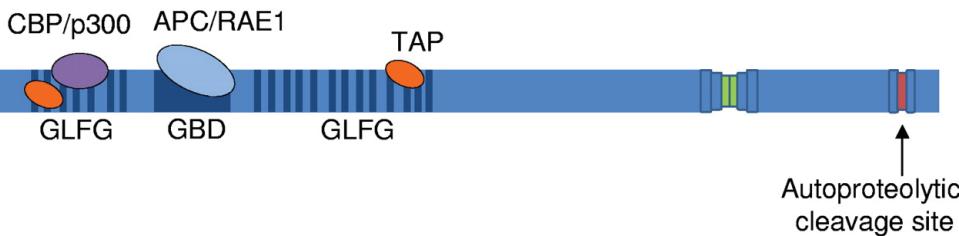


Figure 1: NUP98 gene chromatin structure. A – UCSC genome browser<sup>64</sup> tracks displaying NUP98 gene structure (most commonly translocated region, exons 11 to 14, marked) and DNase I hypersensitivity sites. B – S/MAR enrichment on NUP98 gene, defined by MARfinder.<sup>65</sup>

A



B

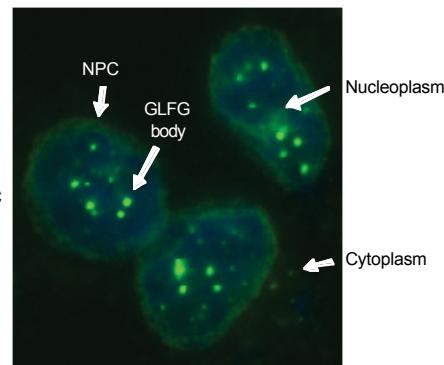


Figure 2: A – Nup98 protein domains and known interactions<sup>21</sup> B – DNA stained with DAPI (blue) and GFP-Nup98 (green) expressed in 293T cells visualized by immunofluorescence microscopy.

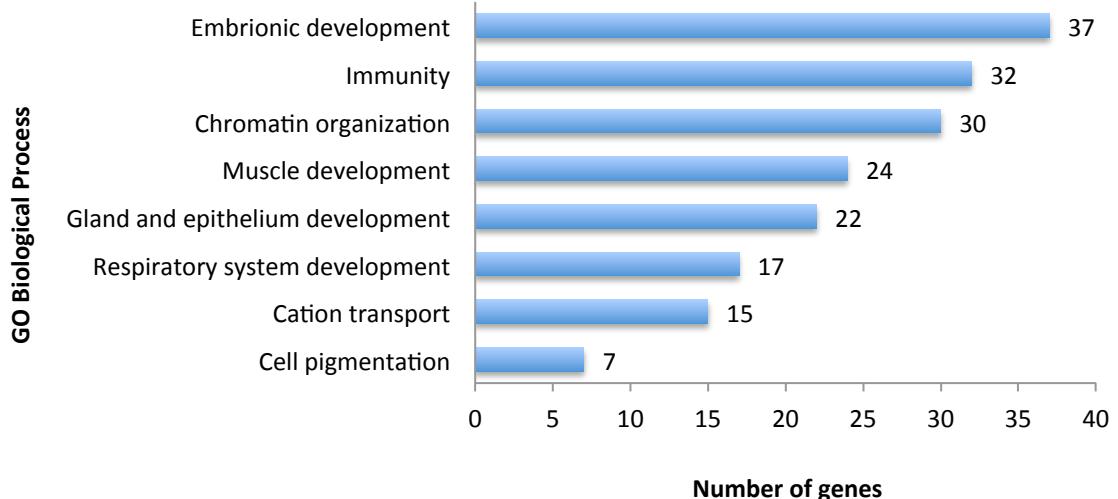


Figure 3: GO annotation biological process enrichment for all genes differentially expressed in the presence of Nup98 fusion proteins.

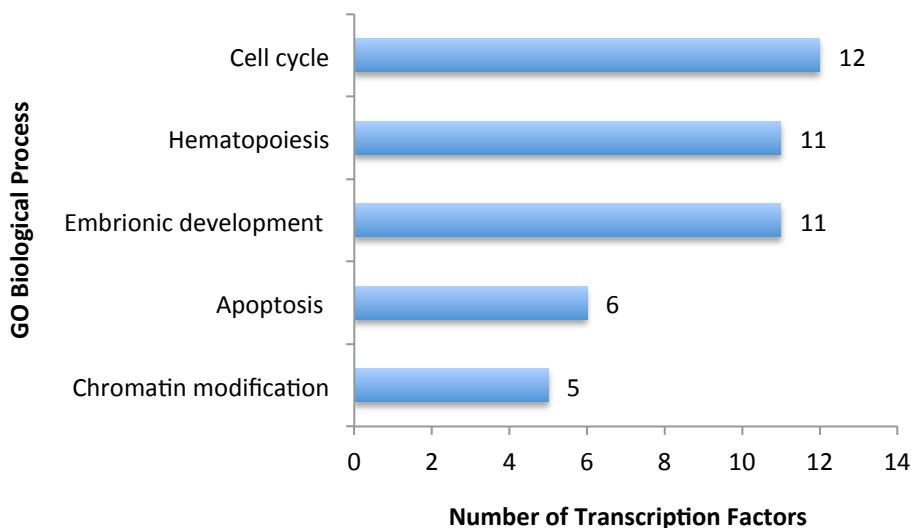


Figure 4: GO annotation biological process enrichment for all TFs with TFBS enriched in genes differentially expressed in the presence of Nup98 fusion proteins.

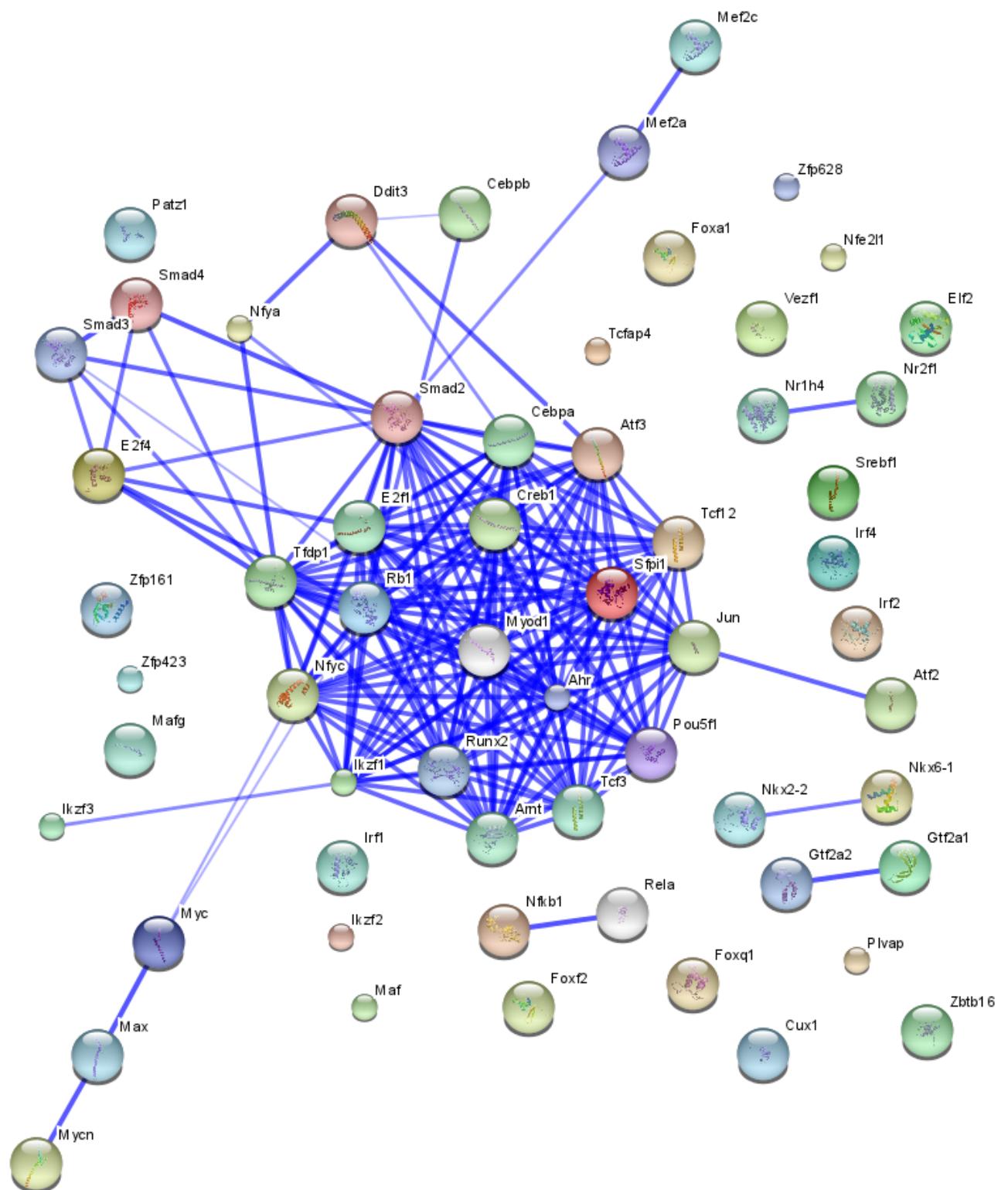


Figure 5: Protein-protein interaction network of TFs with TFBS enriched in genes differentially expressed in Nup98 translocations.

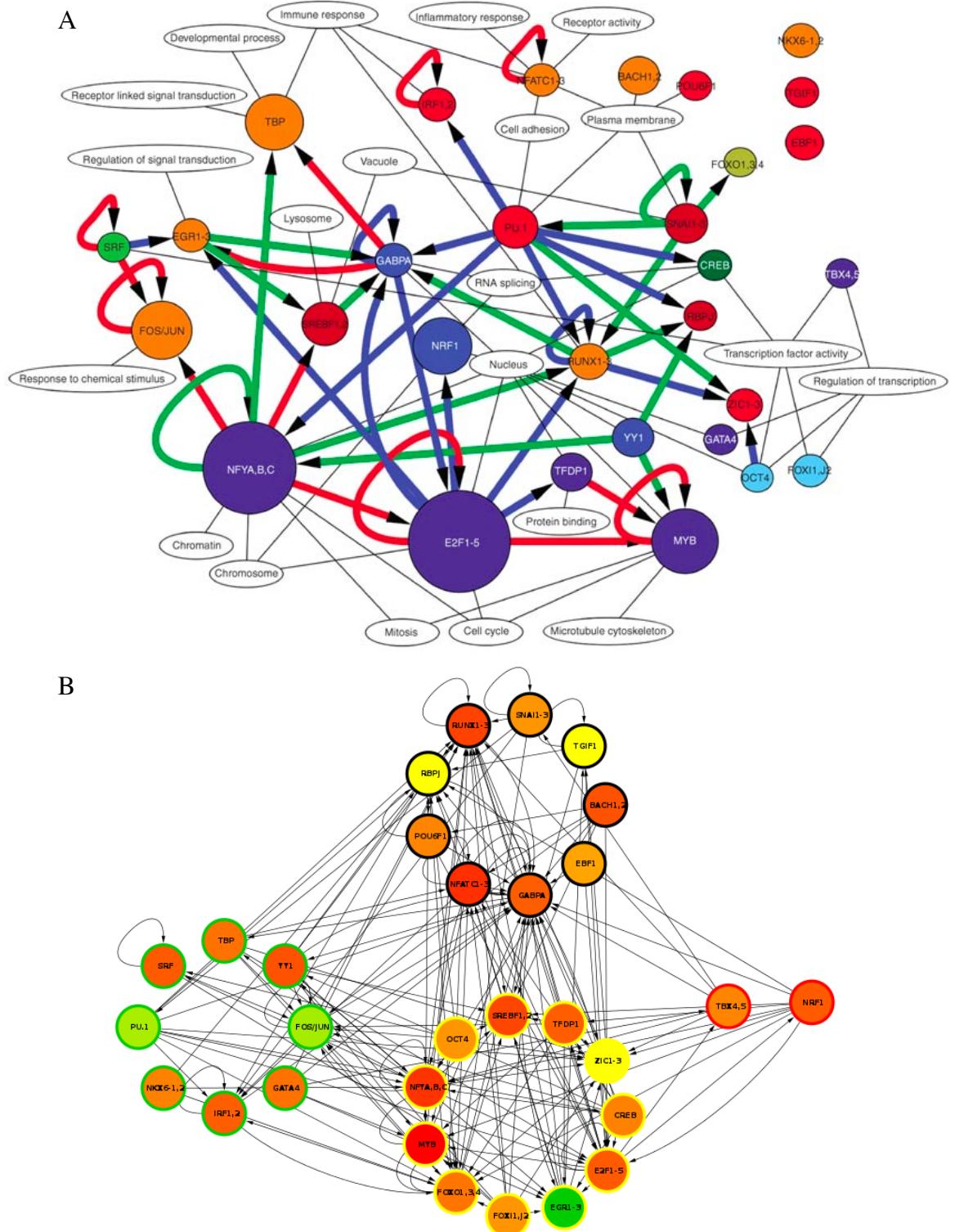


Figure 6: A - The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line.<sup>46</sup> B – Superposition of results from Nup98 fusion microarrays into the network from A, node color indicates expression level of TF (red – decreased, green – increased) and node border indicates enrichment of TFBS in groups of genes with increased (green) or decreased (red) expression (yellow border - TFBS present in up and down regulated genes).

## Label-free approach

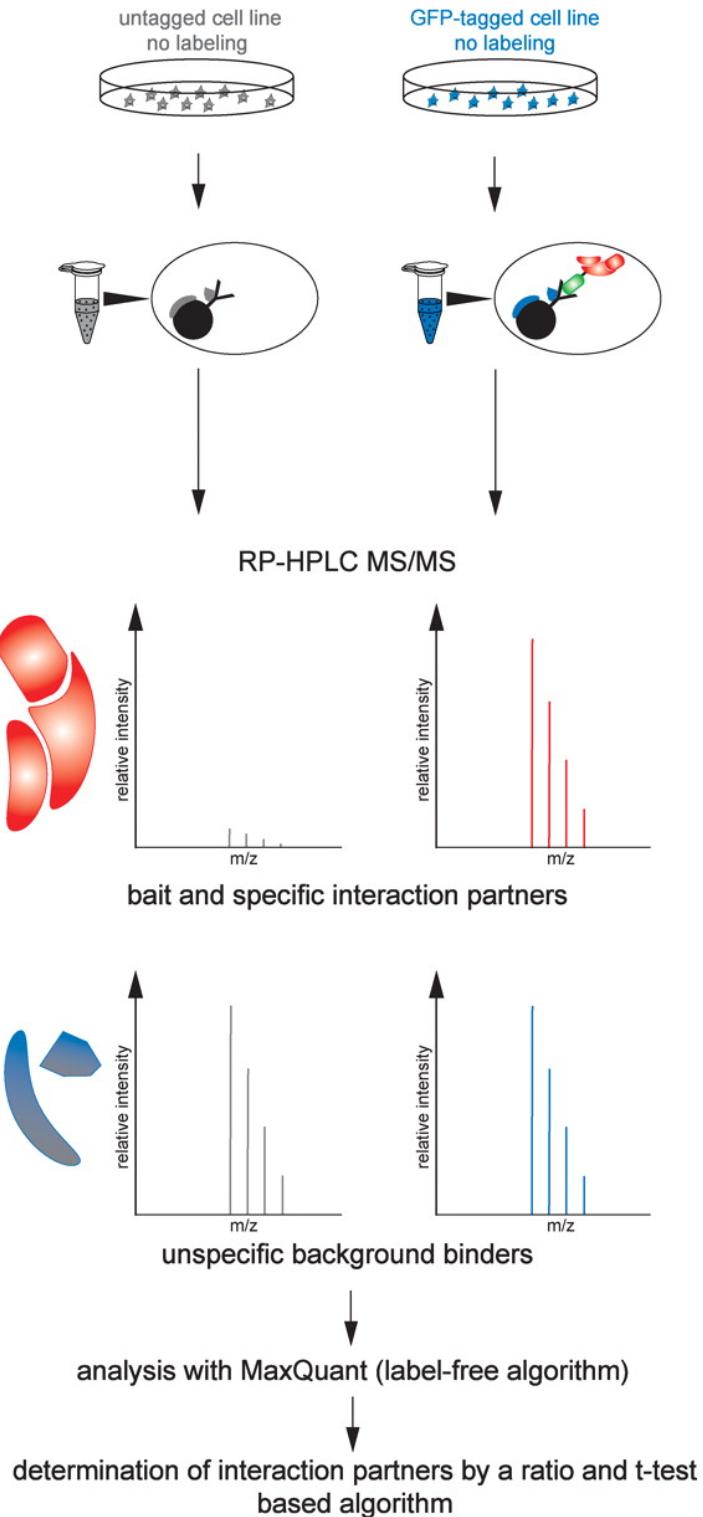


Figure 7: Protocol overview for affinity purification and label-free mass spectrometry quantification of protein complexes.<sup>54</sup>

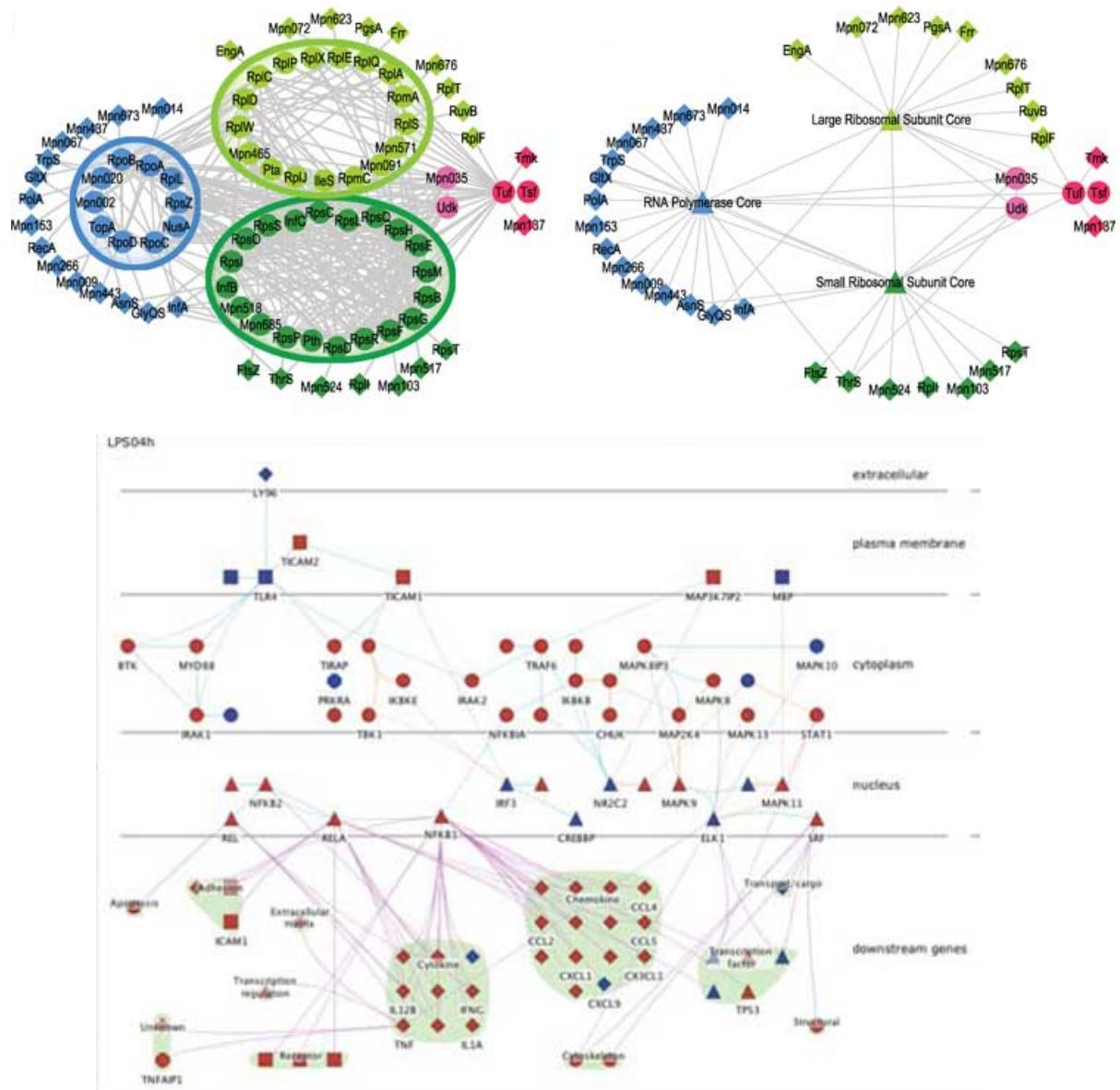


Figure 8: PPI network examples.<sup>66</sup> Exemplifying biological process enrichment, important network hubs and protein complexes distributed according to cellular compartments.

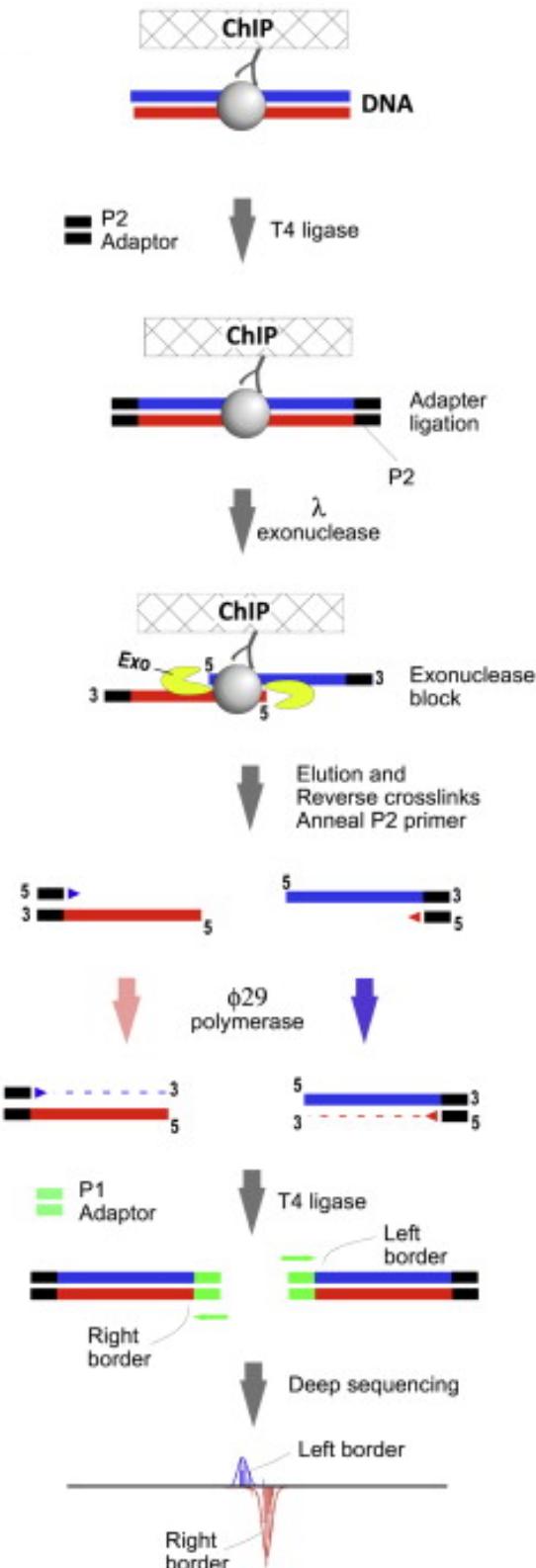
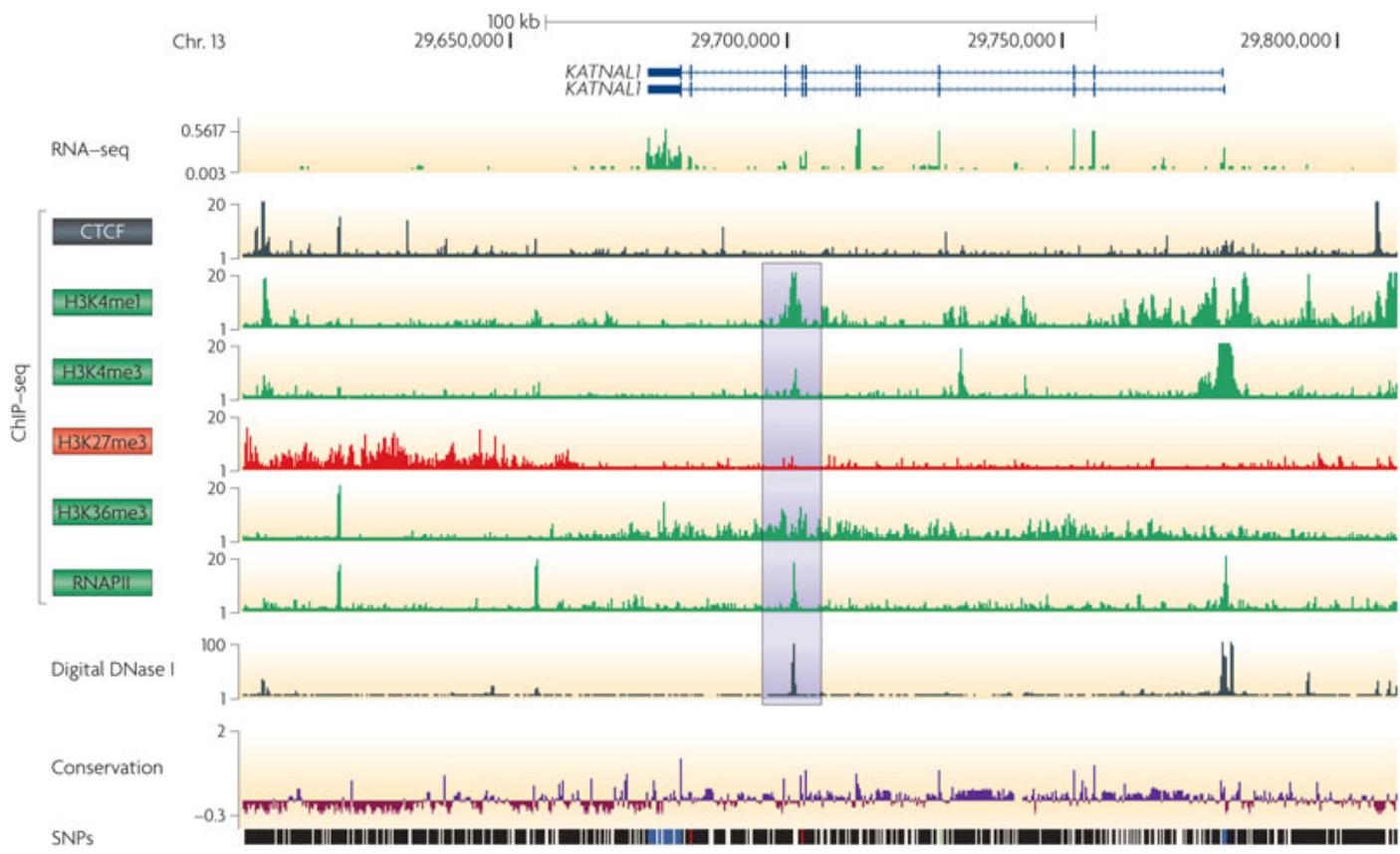


Figure 9: Protocol overview of ChIP-exo method followed by deep sequencing.<sup>59</sup>



Nature Reviews | Genetics

Figure 10: ChIP-seq results in a genome map example visualized in UCSC genome browser.<sup>67</sup>

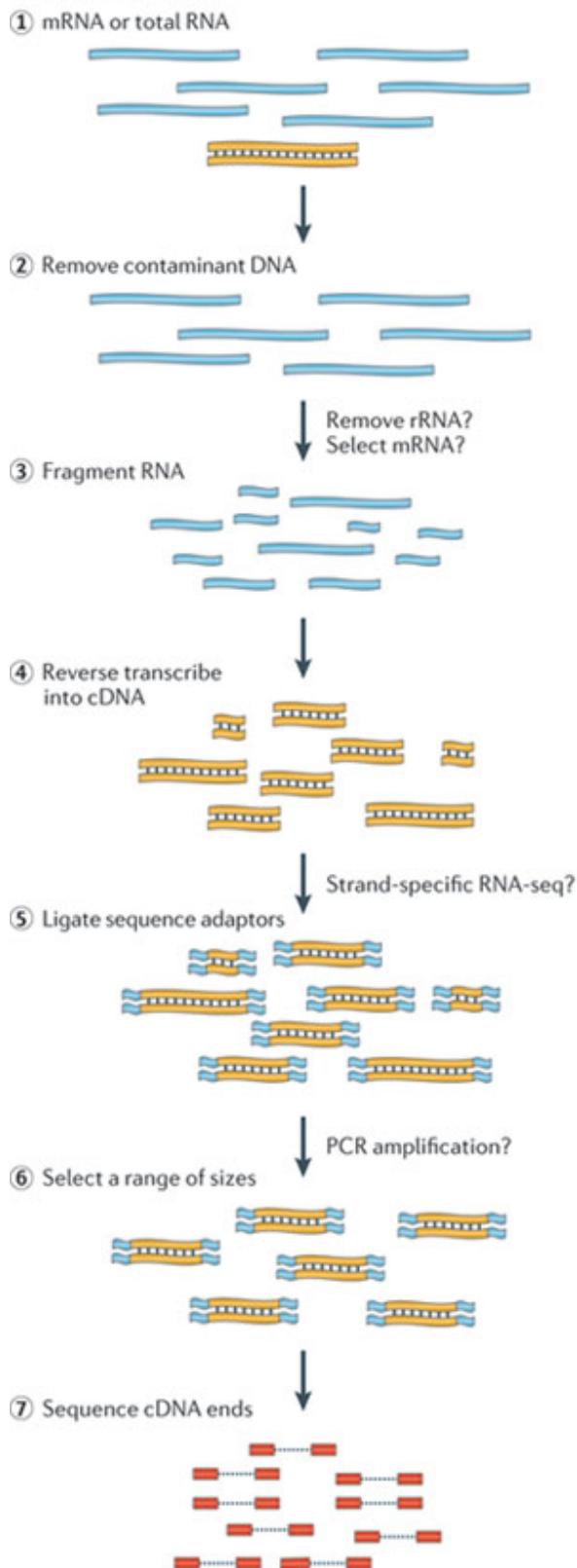
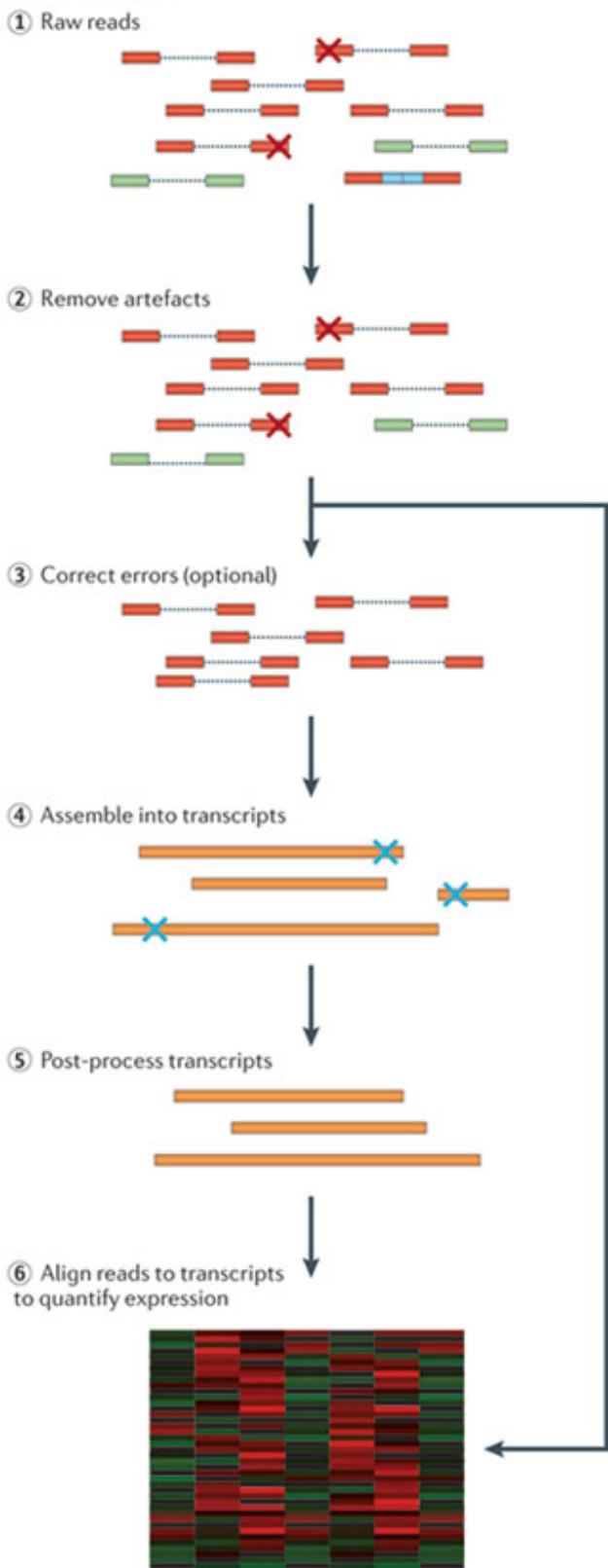
**a Data generation****b Data analysis**

Figure 11: Next generation RNA-seq protocol overview.<sup>68</sup>

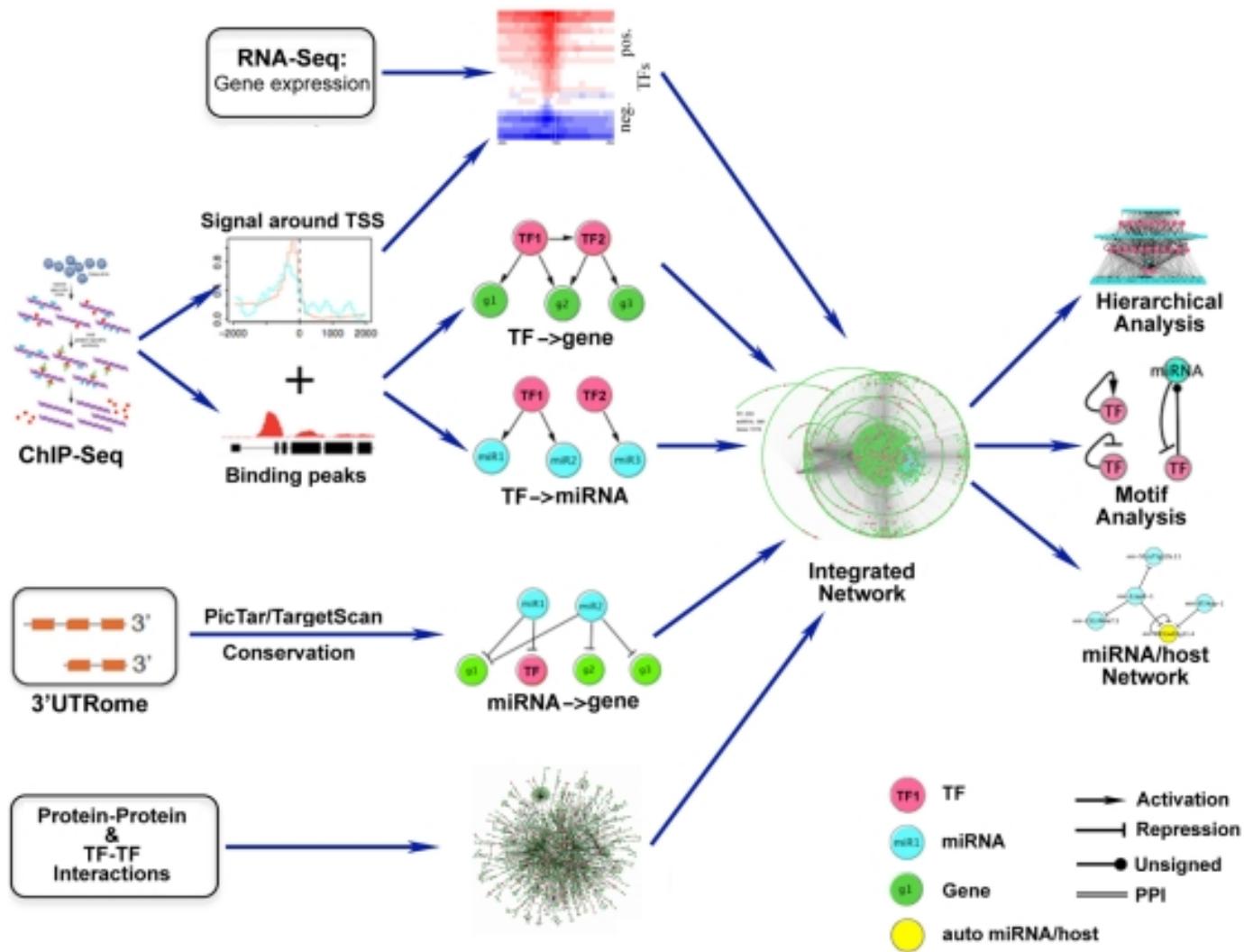


Figure 12: Data collection, integration and analysis for construction of a gene expression regulatory network.<sup>69</sup>

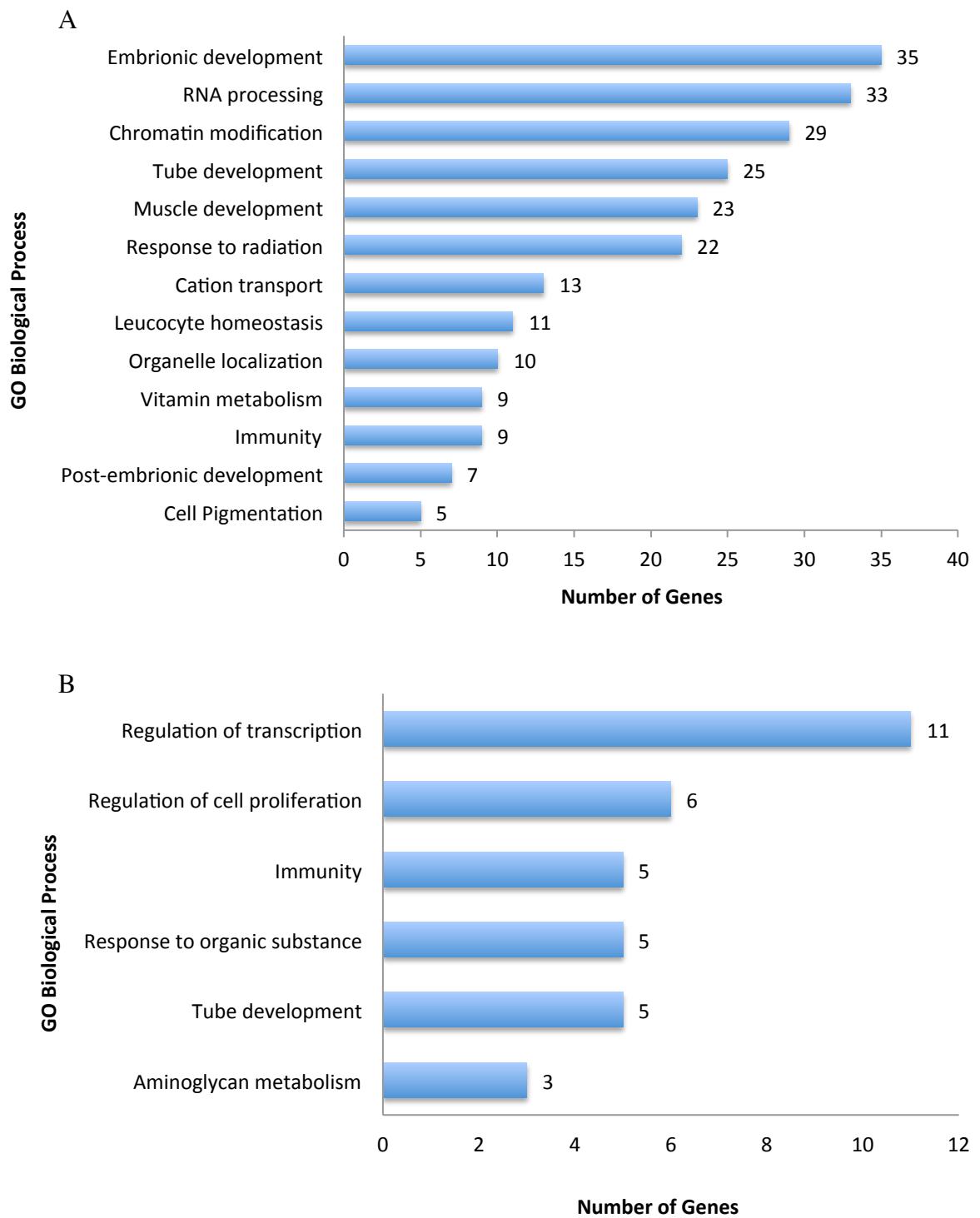
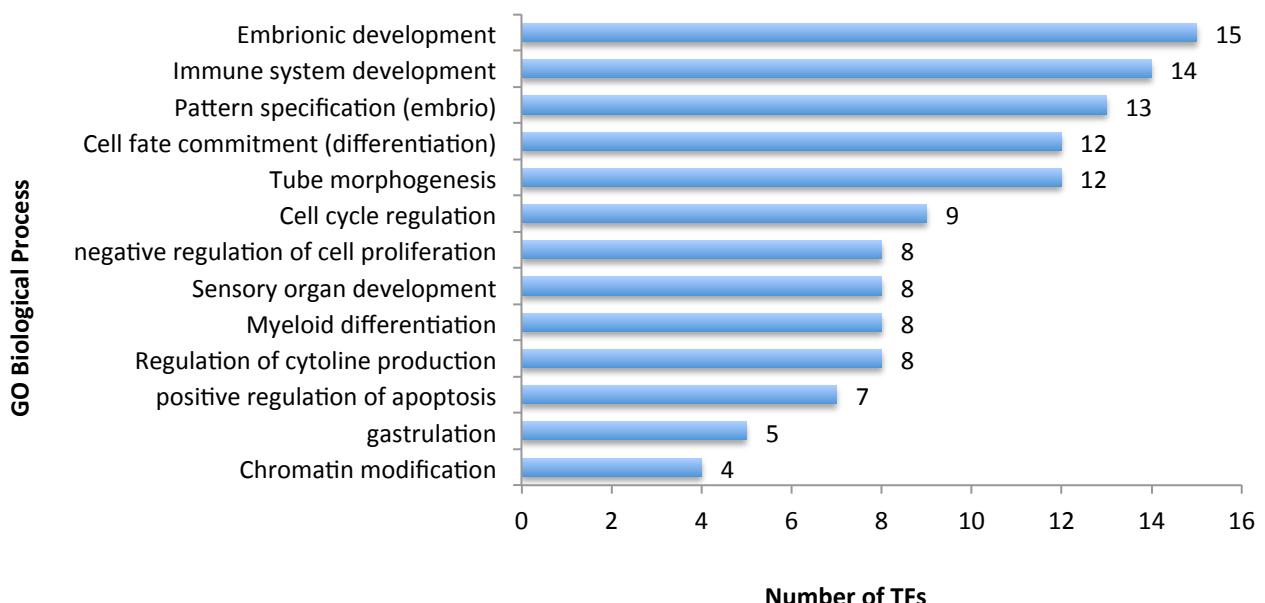


Figure S1: GO annotation biological process enrichment for genes with expression down-regulated (A) or up-regulated (B) in the presence of Nup98 fusion proteins.

A



B

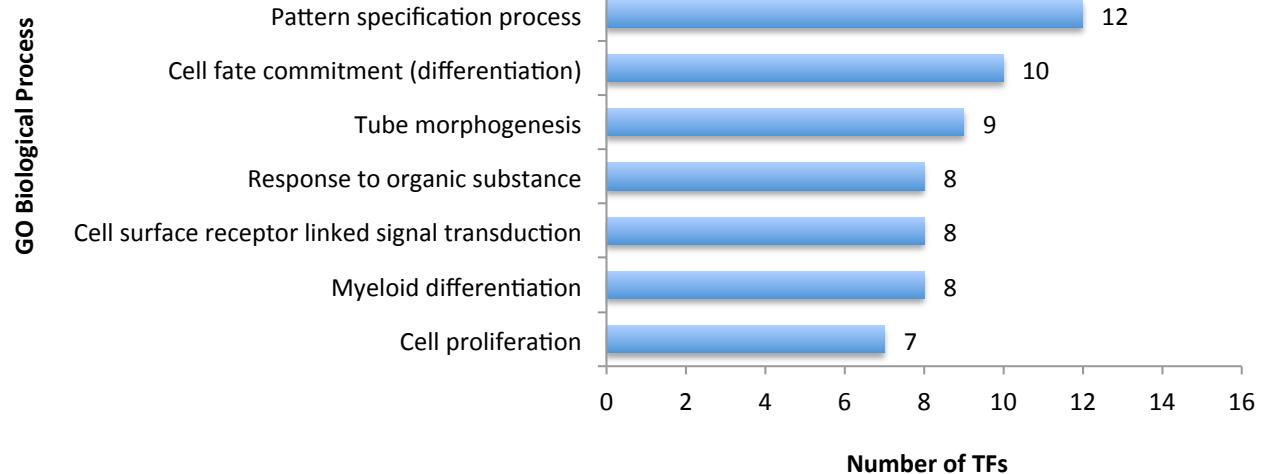


Figure S2: GO annotation biological process enrichment for TFs with TFBS enriched in genes with expression up-regulated (A) or down-regulated (B) in the presence of Nup98 fusion proteins.

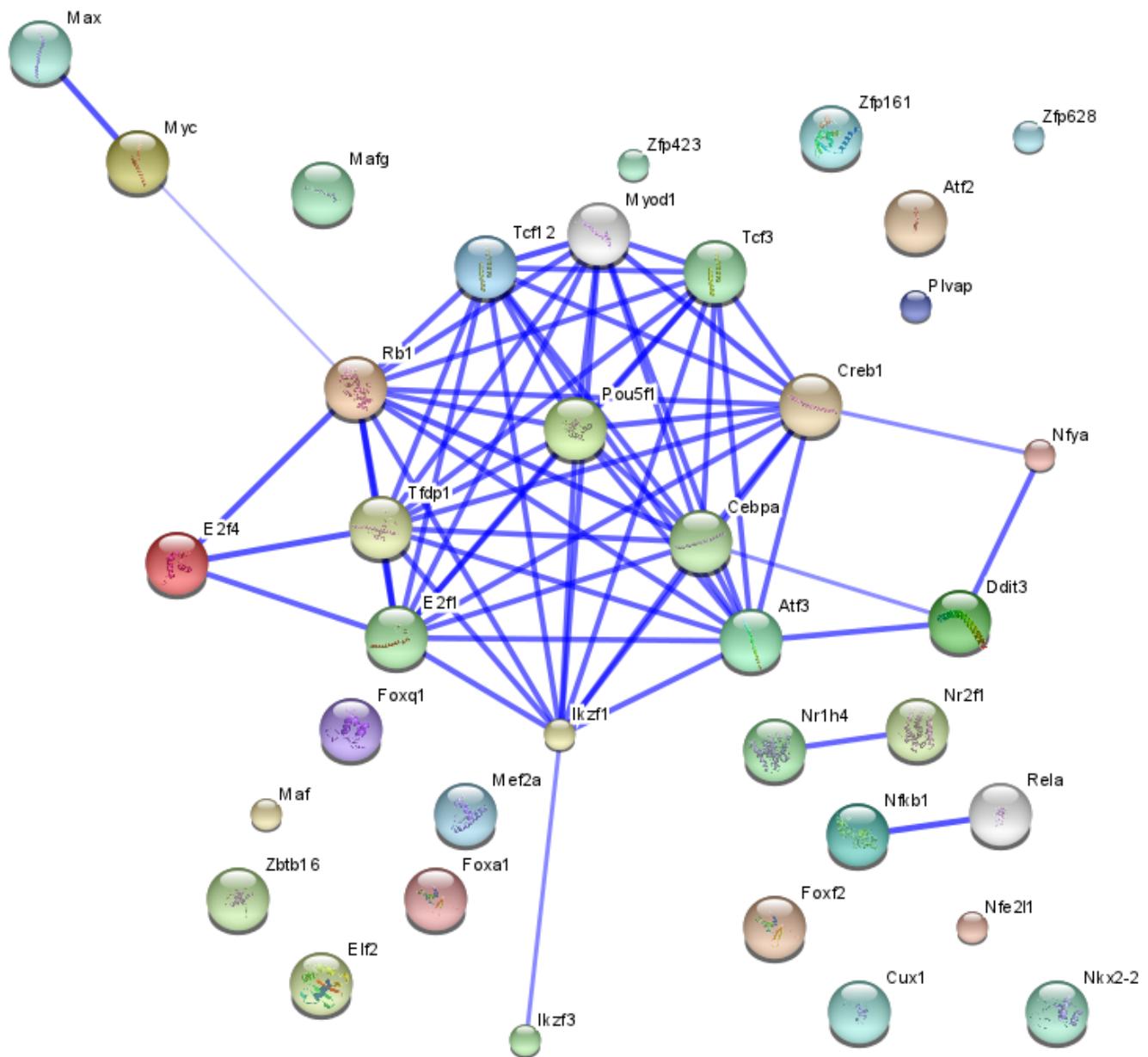


Figure S3: Protein-protein interaction network of TFs with TFBS enriched in genes with down-regulated expression in Nup98 translocations.

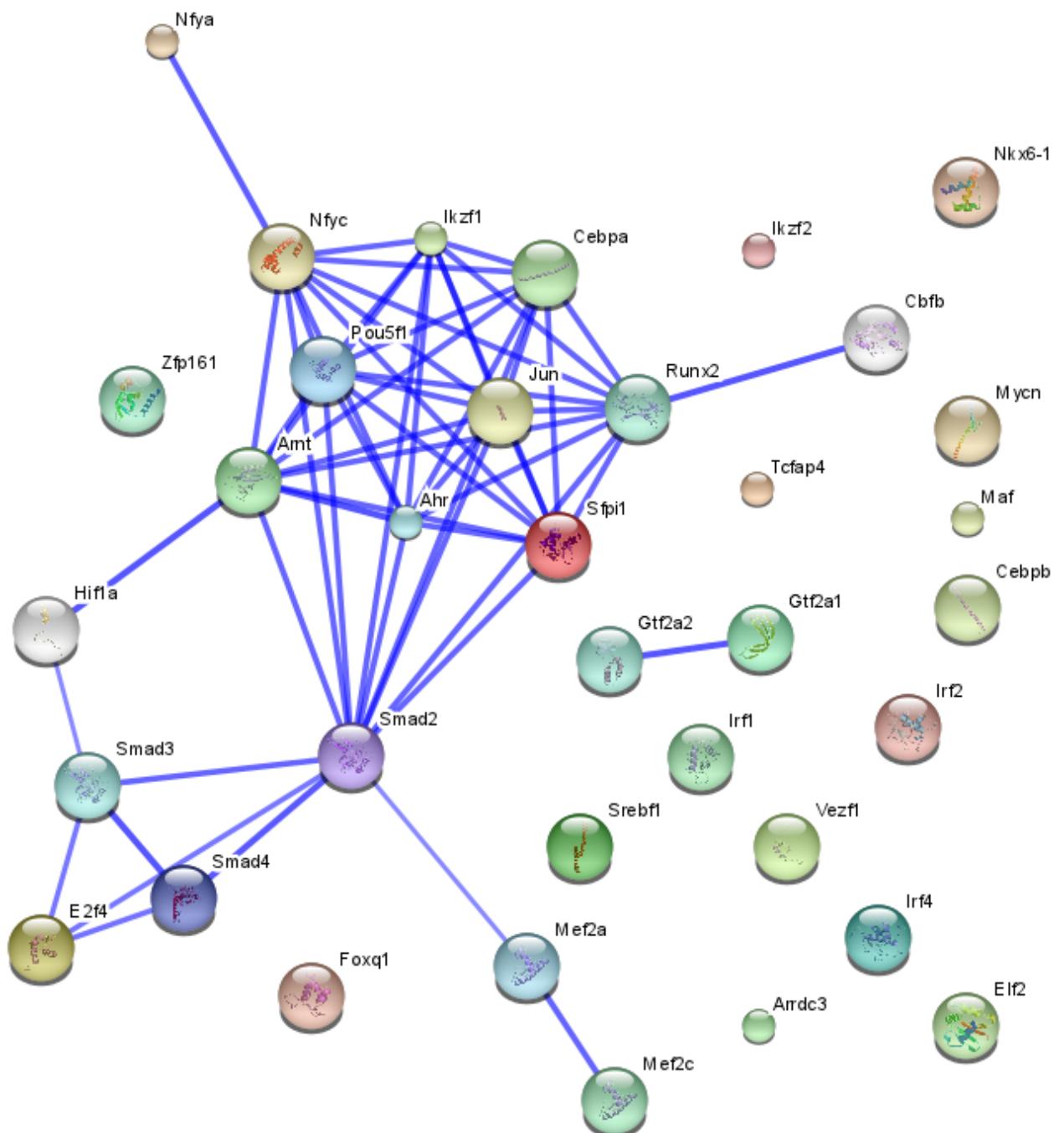


Figure S4: Protein-protein interaction network of TFs with TFBS enriched in genes with up-regulated expression in Nup98 translocations.