Journal of
Computer Science & Systems Biology

Vallabhu et al., J Comput Sci Syst Biol 2014, 7:6
DOI: 10.4172/jcsb.1000156

Research Article Open Access

A Systems Biology View of the Spliceosome Component Phf5a in Relation to Estrogen and Cancer

Rishu Vallabhu, Eva Falck and Angelica Lindlöf*

Systems Biology Research Centre, University of Skövde, Box 408, 541 28 Skövde, Sweden

Abstract

Cancer is a broad term for a wide spectrum of diseases and which involves the alteration in expression levels of several hundreds of genes. As such, the study of the disease from a systems biology point of view becomes rational, as the properties of a system as a whole may be very different from the properties of its individual components. However, understanding a network at the systems level not only requires knowledge about the components of the network, but also the interactions between them.

Here, a systems biology view of the rat PHD finger protein 5A (*Phf5a*) gene was attempted; a gene previously identified as aberrantly expressed in estrogen dependent endometrial adenocarcinoma tumors from both rat and human. Phf5ais a highly conserved cysteine rich (C4HC3) zinc finger and such proteins predominantly have a role in chromatin mediated transcriptional regulation. Moreover, PHF5A is a component of the macromolecular complex spliceosome that takes part in pre-mRNA splicing and spliceosome component coding genes have previously been shown to be implicated in various cancer types and suggested to potentially be novel antitumor drugs.

To derive a systems biology view, in this study, a weighted gene network was inferred from a list of genes having correlated expression profiles to *Phf5a* as nodes, and common transcription factors and microRNAs regulating these genes together with annotation about biological process ontology term(s) and pathway(s) as edge weights. In the inferred network a higher weight indicates more annotation shared between two genes and, hence, the network facilitates the identification of closely interacting genes with *Phf5a*. The results show that highly weighted edges connect *Phf5a* to other spliceosome components, but also to genes involved in the metabolism of proteins, proteasome and DNA replication, repair and recombination. The results also link *Phf5a*to the Myc/Rb/E2F pathway, one of the central pathways associated with cancer. The proposed method for inferring a weighted gene network can easily be applied to other genes and diseases.

Keywords: Cancer; Estrogen; Spliceosome; Phf5a; Systems biology; Weighted gene network

Introduction

Cancer is a broad term for a wide spectrum of diseases where the cells have gained capability to divide uncontrollably, by overcoming fundamental regulatory mechanisms controlling cell division. In a multistep process, cells become malignant by acquiring several genetic mutations that ultimately alter various molecular pathways and which subsequently lead to the development of proliferating cells [1]. Several hundreds of genes have been listed to be implicated in cancer and mutations in genes controlling the cell cycle, apoptosis and angiogenesis have been shown to be important in the progression of the disease [2]. As there are many genes and pathways involved, the study of the disease from a systems biology point of view becomes rational [1,3]. This view intends to fit genes/proteins into a system, rather than studying each gene/protein in isolation, based on the observation that the properties of a system as a whole may be very different from the properties of its individual components. However, understanding a network at the systems level not only requires knowledge about the components of the network, but also the interactions between them. Moreover, an important property of biological networks is degeneracy, which is the capability of structurally different elements (seen as nodes in the network) to perform the same function [4,5]; degeneracy keeps the biological system flexible and adds robustness to it. However, there is commonly a minimal set of genes that are essential for the system to survive [4,6]. Interestingly, essential genes have previously been proven to be potential drug targets and suggested to be considered in cancer therapy as well [7-10]. Essential genes tend to be evolutionarily more conserved than non-essential genes as they accomplish basic cellular functions.

The PHD finger-like domain protein 5a (PHF5A) is a highly conserved cysteine rich (C4HC3) zinc finger and such proteins predominantly have a role in chromatin mediated transcriptional regulation [11-14]. Moreover, PHF5A is a component of the subunit Splicing factor 3b (SF3b) [15], which in turn is a component of the U2 small nuclear riboproteins (snRNA) complex-an important part of the spliceosomal machinery. The macromolecular complex *spliceosome* takes part in pre-mRNA splicing and this complex comprises the components U1, U2, U5 and U4/U6 snRNAs. The U2 snRNA complex, of which PHF5A is a component of, is involved in the two first steps of the splicing process [16,17]. Pre-mRNA splicing involves removal of introns from pre-mRNA to produce a mature mRNA. In eukaryotes, alternative splicing of pre-mRNAs is a major factor for the diversity of proteins and functional complexity, and is indispensable for the expression of essential genes [18,19]. High-throughput sequencing studies have shown that 92-94% of human multiexon genes undergo alternative splicing.

Spliceosome component coding genes have previously been shown

***Corresponding author:** Angelica Lindlöf, University of Skövde, Systems Biology Research Centre, Skövde, Sweden, Tel: +46500448349; Fax: +46500448000; E-mail: angelica.lindlof@his.se

Received July 24, 2014; **Accepted** August 22, 2014; **Published** August 24, 2014

Citation: Vallabhu R, Falck E, Lindlöf A (2014) A Systems Biology View of the Spliceosome Component *Phf5a* in Relation to Estrogen and Cancer. J Comput Sci Syst Biol 7: 193-202. doi:10.4172/jcsb.1000156

to be over-expressed in lung, breast and ovarian cancers [20], and mutations in genes coding for the spliceosome proteins SF3B1, U2FA1 and SFRS2 have been found to be associated with myelodysplastic syndromes (MDSs) in humans, which are chronic neoplasms of hematopoietic stem cells that often progress to acute myeloid leukemia (AML) [21]. Interestingly, the chemicals pladienolide derivatives and Spliceostatin A have been shown to display antitumor activity by binding to the SF3b complex and thereby inhibiting the spliceosome, which results in impaired splicing and altered gene expression patterns [22,23]. *Phf5a*/*PHF5A* itself has previously been studied on gene-level by conventional means, with the aim to characterize the gene and its protein product. For example, it has been identified to be essential for the formation and maintenance of glioblastoma multiforme (GBM) [24], an aggressive malignant brain tumor, and also suggested to act as a transcription factor or co-factor in the up-regulation of the Gap junction alpha 1 (*Gja1*) in the presence of estrogen in rat [11,14]. *Gja1* is a connexin shown to be down-regulated in cancer cells and, moreover, connexins have previously been shown to act as tumor suppressors [25,26]. Falck and Karin-Levan (2013) previously found *Phf5a*/*PHF5A*to be aberrantly expressed in estrogen dependent endometrial adenocarcinoma (EAC) tumors from rat and human type I tumors [27]. Additionally, homologs of *Phf5a* in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* have been identified as critical genes in pre-mRNA splicing and cell cycle regulation in these species, as cells lacking this gene showed an arrest in the spliceosome assembly and failed to go through the cell cycle, respectively [11,28]. In vertebrates PHF5A is 100% identical at the amino acid level and in multicellular organisms the degree of DNA sequence similarity is over 80%. Genes that have a high evolutionary conservation are commonly retained for their functional importance, as they are required to accomplish basic cellular functions [29].

Due to a high genetic heterogeneity in human, the study of complex diseases such as cancer may be difficult to perform on human samples. Therefore, as a complement to studies in human, model organisms have previously been commonly used. The rat model provides a good choice since this species has similarities in pathogenesis and histopathological properties to those of human and has therefore been extensively used in the study of various cancer types [30,31]. For example, the database Array Express [32] lists several hundreds of experimental studies related to cancer in this species.

In this study we aimed to develop a systems biology view of the rat *Phf5a* in relation to estrogen, since previouslya strong correlation of *Phf5a* expression to malignant samples of estrogen dependent EAC in BDII rats had been identified [27]. This was accomplished by choosing published microarray studies from experiments in rats related to estrogen. Also, narrowing in on a particular focus resulted in a reduced number of suitable data sets and thereby the workload during data analysis. In this study, six different microarray studies were finally included, which is still a substantial number that plausibly can provide important information.

Microarrays measure the expression of thousands of genes simultaneously and are therefore suitable for co-expression analysis, since a list of highly correlated genes can easily be generated from the data by using a correlation test [33,34]. However, co-expression networks alone cannot reveal how correlated genes might be co-regulated or associated by participating in the same pathway or biological process, for example. Therefore, annotation about microRNAs (miRNAs) and transcription factors (TFs) need to be added, as these are main classes of gene regulatory mechanisms [35]. Moreover, annotation in form of ontology term(s), pathway(s) and protein interaction(s) is also valuable in the characterization of genes. Finally, integration of knowledge from different data sources would seem complex unless it is presented in a comprehensible manner [36,37]. Therefore, in this study, a network model was used to integrate the data and visualize it, by inferring a weighted network where the genes are represented as nodes and annotation in form of TF and miRNA regulation, protein interactions, biological ontology terms and pathways as an additive weighted edge, where a higher weight means more annotation shared between a pair of genes. By identifying and analyzing central genes, i.e., those connected with highest weighted edges in the network, a number of interesting clues were revealed. For example, we conclude that *Phf5a* is possibly a target of Myc, a TF that has a prominent role in the control of DNA replication and which mutated form has been shown to be implicated in various cancer types [38,39]. We can also link *Phf5a* to the Rb/E2F pathway, by being a target of Myc and E2f1 TFs that regulate this pathway. The Rb/E2F pathway is critical in the initiation of DNA replication and the cell cycle, and is commonly disrupted in various cancer types [40]. The expression pattern of *Phf5a* is also correlated to ribosomal components and this could be attributed by the fact that Myc controls the expression of ribosomal components [41]. We also identified a number of miRNAs that potentially target *Phf5a* and these could also be used as a strategy to slow down tumor progression.

Results

Generation of a systems biology view

Generation of a systems biology view requires the integration of knowledge from various data sources and a comprehensible presentation of the integrated data [36,37]; to accomplish this number of analysis steps were implemented (Figure 1). In this study we used a reverse engineering model to construct a gene network that integrated the knowledge obtained from the various data sources. The model starts with choosing a set of suitable microarrays based on the condition in

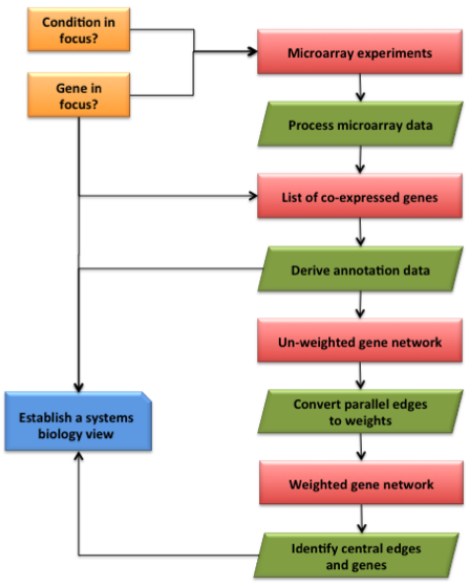

Figure 1: Systems biology view: Illustrates the flowchart for deriving a systems biology view of the gene in focus, by selecting appropriate microarray experiments, extracting correlated genes and annotation data for these genes, deriving gene networks based on the annotation data, and finally establishing a systems biology view from identified central edges and genes.

Exper.	References	Platform	Tissue	Species/Sex	Rat model	Treatment	Corr. genes (pos./neg.)
E-GEOD-13003	[92]	SWEGENE Rat 70mer oligonucleotide array	Endometrium, cervix and uterus	Rattus norvegicus/Female	BDII	-	90/6
E-GEOD-13319	[93]	A-AFFY-43	Uterine leiomyoma	Rattus norvegicus/Females	Eker	-	27/1
E-MEXP-999	[94]	A-AFFY-25	Uterus	Rattus norvegicus/Female	Charles River VAF plus	10 µg/kg of ethinyl estradiol	1/1
E-TOXM-20	[95]	A-AFFY-25	Uterus and ovaries	Rattus norvegicus/Female	Sprague-Dawley	0.1/1/10 µg/kg/day of ethinyl estradiol for 4 days	25/11
E-GEOD-24672	[96]	A-AFFY-43	Testes	Rattus norvegicus/Male	LBNF1	Irradiation with acyline and flutamide for a period of 2/4 weeks in doses of 30-110 picograms/ml	1/2
E-GEOD-40713	Unpublished	A-AFFY-43	Mammary gland	Rattus norvegicus/Male, female	Sprague-Dawley	0.1/1/10 µg/kg/day of ethinyl estradiol for 11 days	130/8

Table 1: Microarray experiments: The following microarray experiments were used in this study to derive correlated genes to *Phf5a*. Exper., reference in ArrayExpress; Ref., reference to published paper; Platform, microarray platform used in the experiment; Tissue, which tissue(s) were used in the experiment; Species/Sex, which rat species and sex that was used in the experiment; Rat model, which rat model was used in the experiment; Treatment, type of treatment applied to the rats; Corr. genes, number of correlated genes to *Phf5a*.

focus, together with which processing tools should be used to analyze the microarray data. Thereafter, a list of correlated genes to the gene under study is derived from the microarray data. The list of correlated genes is subsequently submitted to various gene discovery databases/tools to derive ontology and pathway annotation, TF and miRNA binding as well as protein interactions. The compiled annotation is first used to generate an un-weighted network, where the nodes represent the correlated genes and the edges any shared annotation between a pair of genes. For example, if two genes are regulated by the same TF than this is represented by an edge in the network. Thereafter, the un-weighted network is converted to a weighted network by counting all edges shared between each pair of nodes and replacing these edges by a single edge with a weight, where the weight represents all annotation shared by the two genes. For example, if two genes in the un-weighted network are connected by three edges (representing regulation by the same TF(s), miRNA(s), and/or gene ontology terms, etc.), then in the weighted network the weight of the single edge will be three. Here, the most important genes, so called central genes, are the ones connected by edges with the highest weights. Using the weighted network to identify edges with high weights simplifies the task of discerning central genes and constructing a systems biology view of these genes centering on the gene in focus.

Generation of un-weighted network

In total data from six different microarray studies were collected from ArrayExpress [32] based on the condition in focus, i.e., the expression of rat *Phf5a* in relation to estrogen (Table 1). The experiments were either conducted on estrogen sensitive tissues or rats treated with estrogen. The microarray data was pre-processed using various packages in R statistical language, in order to derive expression profiles for all genes in each experiment.

Pearson correlation (PC) test was applied and all genes having a correlated expression profile (PC ≥ |0.7|) to *Phf5a*'s expression profile in at least one experiment were included in subsequent analyses. The cutoff for PC was based on the number of correlated genes that were derived; a cutoff of 0.8 resulted in very few genes (and for some experiments in no correlated genes) and a cutoff of 0.6 resulted for some experiments in a very high number of genes (several thousands). Using a cutoff of 0.7, in total 303 correlated genes (~1% of all genes in the pool) were derived from the six different experiments, of which 274 and 29 were positively and negatively correlated, respectively. Interestingly, there were no overlaps of correlated genes between the experiments and, additionally, the number of correlated genes from each experiment varied substantially (Table 1). Most number of correlated genes was derived from experiment E-GEOD-40173, with 130 positively and 8 negatively correlated genes, and least from E-MEXP-999 with only one positively and one negatively correlated gene, respectively. In E-GEOD-40173 doses of 0.1/1/10 µg/kg/day of ethinyl estradiol were used for 11 days and tissues used were mammary glands, whereas in E-MEXP-999 a single dose of 10 µg/kg of ethinyl estradiol was used and tissues were collected from the uterus.

The list of correlated genes (including *Phf5a*) was submitted to the Database for Annotation, Visualization, and Integrated Discovery (DAVID) [42], to derive Gene Ontology Biological Process terms (GO_BP_FAT) [43] associated with these genes. In total 277 (91%) of the correlated genes could be mapped to an official gene symbol in DAVID and 160 (53%) of them could be mapped to at least one GO_BP_FAT term. Subsequently, by setting a gene cutoff value ≥ 15 (i.e., terms for which at least 15 of the correlated genes were annotated with) and a p-value ≤ 0.05, 10 significant GO_BP_FAT terms were retrieved, of which 88 (29%) of the correlated genes were annotated with. Most number of genes (25 of the correlated genes; 8%) was annotated with response to organic substance. However,the only term *Phf5a* was annotated with was positive regulation of macromolecule metabolic process. As we intended to develop a network centered on *Phf5a,* we aimed to find more GO_BP_FAT terms in common between this gene and the correlated genes. Therefore, a second round of GO_BP_FAT terms were obtained, but where the gene cut off limit was decreased to 10 and 5, respectively (the p-value was retained on the same level). Using a gene cut off 10 increased the number of significant GO_BP_FAT terms to 30, but still *Phf5a* was only annotated with the term *positive regulation of macromolecule metabolic process*. Using a gene cut off 5 resulted in a list of 78 significant GO_BP_FAT terms to which 159 (52%) of the correlated genes were annotated with. In this case, six terms were retrieved for *Phf5a*: positive regulation of macromolecule metabolic process, mRNA metabolic process, RNA splicing, nuclear mRNA splicing via spliceosome, *RNA* splicing *via* transesterification, and *RNA* splicing *via* transesterification with bulged adenosine as nucleophile. Another 26 (9%) of the correlated genes were also annotated with these six terms and this information was used as edges in the un-weighted network; an edge between two genes represented a common GO_BP_FAT term.

From DAVID, a list of overrepresented KEGG pathways [44] was

also obtained. Using a gene cutoff value ≥ 5 and *p*-value ≤ 0.05 resulted in seven different pathways and in total 56 (18.5%) of the correlated genes were annotated with at least one of these pathways. However, the only pathway retrieved for *Phf5a*was the Spliceosome (KEGG: rno03040), since there is currently no other evidence of Phf5a/PHF5A participating in any other pathway(s). Another 11 (4%) of the correlated genes were also annotated with this pathway and this information was used as edges in the un-weighted network; an edge between two genes represented participation in the same pathway.

The list of correlated genes was submitted to the database Chip Enrichment Analysis (ChEA) [45], to obtain a list of TF(s) binding to these genes. The database contains genome-wide DNA TF binding site data from ChIP-chip, ChIP-seq, ChIP-PET, and DamID experiments derived for rat, mouse and human. However, when using rat binding site data solely, there was binding site information for only 42 (16%) of the correlated genes and there was no information for *Phf5a*. Therefore, data for all three species (i.e., rat, mouse and human) was used instead, by selecting all species as filtering criteria and a *p*-value ≤ 0.05. It has previously been demonstrated that there is a high conservation in binding sites when the function is also conserved, and since Phf5a has a high conservation across these three species it seemed reasonable to use information from all three species [46]. This resulted in 210 (69%) of the correlated genes having binding site information for at least one of 338 TFs. The list of TFs was subsequently filtered by removing those with no binding site data for *Phf5a*, and which reduced the number of TFs to 29 and the number of correlated genes to 194 (64%). The information about TF binding site(s) was used as edges in the un-weighted network; an edge between two genes represented regulation by the same TF.

From the miRNA database miRWalk [47]a list of validated and predicted (*p*-value ≤ 0.01) miRNAs binding to any of the correlated genes was obtained. This list was subsequently filtered to exclude miRNAs not binding to *Phf5a*, which resulted in total 37 miRNAs that bound to 136 (45%) of the correlated genes. Of these miRNAs, 10 were predicted/validated to bind to the coding region of *Phf5a*, 24 to the 3'UTR region, and 3 to both of these regions; there were no miRNAs predicted/validated to bind to the 5' UTR region. All 37 miRNAs were used as edges in the network, where an edge between two genes represented regulation by the same miRNA.

From the Search Tool for the Retrieval of Interacting Genes/Protein (String) database [48] information about protein-protein interactions for *Phf5a*were obtained. Using a low confidence score of 0.150 and the active prediction "Experiments" 10 predicted functional partners to *Phf5a* were obtained. However, none of the correlated genes was represented among these partners and therefore no annotation about protein-protein interactions was included in the network.

The final un-weighted network included 252 (83%) of the correlated genes and these were linked by 33,620 edges; hence, some of the correlated genes (17%) did not have an association to *Phf5a* other than expression correlation. The majority of the edges were represented by TF binding sites (76%), followed by miRNAs (22%), GO_BP_FAT terms (1%), and KEGG pathways (0.2%). Moreover, the genes stilled from all experiments and, hence, no experiment was filtered out by this method.

Generation of weighted network and analysis of central genes

From the un-weighted network a weighted network was inferred, by converting all edges between each pair of genes in the un-weighted network to a single edge with an additive weight, i.e., the weight represented the sum of all edges between each pair of genes in the un-weighted network. This procedure reduced the number of edges to 15,827. The distribution of the weights showed that they ranged from 1-24 and the majority of the edges had very low weights; 96.6% of the edges had a weight ≤ 5 and only 3% had a weight ≥ 10 (Figure 2). In order to identify the most important interactions, edges with the highest weights were extracted and a separate network of the nodes that were connected by these edges was inferred (Figure 3). Here, edges with weights ≥ 15 were arbitrarily chosen to be "central edges", which reduced the network to 19 nodes (referred to as "central genes") and which represents 6% of all genes in the weighted network. These genes were connected by 18 central edges and which represents 0.1% of all weighted edges. Additionally, some of the central genes were also interconnected by edges with intermediate weights (i.e., weights ≥ 10; see dotted lines in (Figure 3). The majority of the central genes were positively correlated to *Phf5a*; in fact only one of them was negatively correlated (Figure 2). The central genes were correlated to *Phf5a* in four of the microarray experiments: E-GSE-40173 (8 of them; 44%), E-GEOD-13319 (5 of them; 28%), E-GEOD-13003 (4 of them; 22%), and E-TOXM-20 (1 of them; 6%).

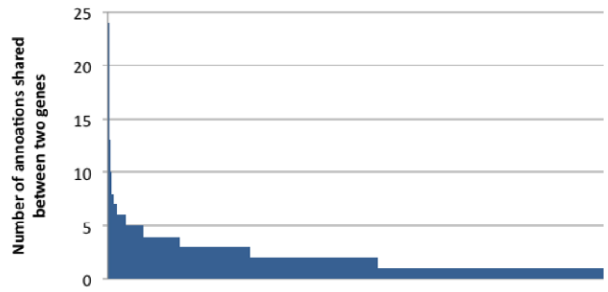

Figure 2: Distribution of number of annotations: The figure shows the distribution of the number of annotations shared between two correlated genes. There are 15,827 gene pairs among the correlated genes and the x-axis shows gene pair number sorted on number of annotations shared between a pair of gene (from highest to lowest). The y-axis shows the number of annotations shared between two genes. Most number of annotations shared between two genes is 24 and least number of annotations is 1 (which means they only share correlation in expression profiles, but no other annotation).

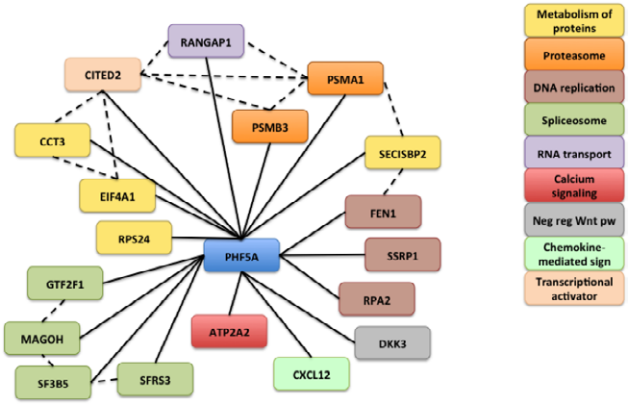

Figure 3: Network of central genes: On the left in this figure the central genes are indicated and how they are linked in the weighted network. Solid lines indicate that two genes share ≥ 15 annotations and dotted lines that two genes share ≥ 10 annotations, but <15 annotations. On the right in the figure the coloring is explained and which was based on literature searches.

A literature search on the central genes revealed a rather diverse set of molecular functions, but the majority of the genes could be related to the biological processes *Spliceosome*, Metabolism of proteins, *DNA* replication, repair and recombination and Proteasome (Figure 3). Four of the central genes, besides *Phf5a*, could be linked to the Spliceosome, of which two were components of the spliceosome complex: Splicing factor 3b subunit 5 (*Sf3b5*) and Serine/Arginine-rich splicing factor 3 (*Sfrs3*) [16,49]. The other two were a General transcription factor IIF (*Gtf2f1*) and a Mago-Nashi homolog (*Magoh*). Gtf2f1/GTF2F1function as a general transcription initiation factor that binds to RNA polymerase II and helps to recruit the initiation complex [50], whereas Magoh/MAGOH is a component of the exon junction complex (EJC) that bind to splice junction sites on mRNAs [51]. The genes *Sf3b5*, *Sfrs3* and *Magoh* were correlated to *Phf5a* in the experiment E-GSE-40173, whereas *Gtf2f1* in E-GEOD-13319.

Three of the genes that could be related to *DNA* replication, repair and recombination were correlated to *Phf5a* in the experiment E-GSE-13319: the Flap structure-specific endonuclease 1 (Fen1), Replication protein A2 (Rpa2) and Structure specific recognition protein 1 (Ssrp1). Fen1/FEN1 is a multifunctional nuclease involved in DNA repair by cleaving the 5'-overhanging flap structure and process the 5'-end of downstream Okazaki fragments [52]. Previous studies have shown that FEN1 can directly interact with estrogen receptor-alpha (ERα) and influence estrogen-responsive gene expression, and that *FEN1* itself is regulated by estrogen [53]. Rpa2/RPA2 is a subunit of the Replication Protein A (RpA), which is essential for chromosomal DNA replication and critical for cell cycle checkpoint activation, and is hyperphosphorylated in response to DNA damage [54,55]. The gene has been shown to be regulated by E2F1, a TF that is regulated by ERα [56]. Ssrp1/SSRP1 is a component of the FACT complex, which is a general chromatin factor that acts to reorganize nucleosomes and shown to be strongly associated with poorly differentiated aggressive cancers [57].

Two of the genes correlated to *Phf5a* in the experiment E-GSE-40173 were Proteasome subunits Alpha type-1 (*Psma1*) and Alpha type-2 (*Psma2*), which are components of the core 20S proteasome [58]. The proteasome's main function is to degrade unneeded or damaged proteins and is essential for many cellular processes, such as cell cycle control, gene expression regulation and tumor growth. Moreover, proteasome activity has previously been shown to increase in the presence of estrogen in murine microglial cells [59]. Another two genes were correlated in this experiment and which could be related to Metabolism of proteins, the Chaperonin Containing TCP1 Subunit 3 (*Cct3*) and Ribosomal protein S4 (*Rps24*). Cct3/CCT3 is a chaperonin that assists the folding of proteins in an ATP-dependent manner and Rps24/RPS24 is a component of the 40S subunit in the ribosome, which catalyzes protein synthesis [60,61].

However, the gene with most annotations in common with *Phf5a* was *Cbp/P300*-Interacting Transactivator with *Glu/Asp-Rich* Carboxy-Terminal (*Cited2*) and which was correlated to *Phf5a* in the experiment E-GEOD-13003. Cited2/CITED2 has previously been shown to, amongst others, act as a transcriptional co-activator of the p300/CBP-mediated transcription complex and enhance estrogen-dependent transactivation mediated by estrogen receptors [62].

For the central genes the edge weights were mainly represented by TFs (56%) and miRNAs (36%), and only to a small extent GO_BP_FAT terms (7%) and KEGG pathways (1%). In total 29 TFs and 37 miRNAs were validated/predicted to regulate the central genes. Moreover, 11 (38%) of the these TFs had been listed as a member of one or several cancer pathways in KEGG (i.e., Myc, E2f1/E2F1, Fli1/FLI1, Runx1/RUNX1, Myc-n, Pparg/PPARG, Ccnd1/CCND1, Ppard/PPARD, Spi1/SPI1, Srf/SRF and Esr1/ESR1) and four (14%) of them as a member of the cell cycle pathway (i.e., Myc, E2f1/E2F1, Ccnd1/CCND1 and E2f4/E2F4). Top TFs, i.e., those that regulated most of the central genes, were Myc (95% of the central genes), E2f1 (79% of the central genes), Fli1 (74% of the central genes), Runx1 (74% of the central genes) and Hnf4a (58% of the central genes), and top miRNAs were rno-miR-207 (47% of the central genes) and rno-miR-129-5p (42% of the central genes).

Myc (c-*Myc*) is a well-studied oncogene; previous studies have shown that cells lacking *Myc* cannot grow and cells over-expressing *Myc* have an increased proliferation rate [39,63]. The gene has been shown to be estrogen-induced and, moreover, being rapidly induced by estrogen in estrogen receptor (ER)-positive breast cancer cells. When Myc is bound to the estrogen receptor it causes activation of cyclin dependent kinases (Cdk2 and Cdk4/6), which together with cyclins act as a complex to drive the progression of the cell cycle. The expression of cyclins is cell cycle specific, but these proteins also have a role in transcriptional regulation [64]. Moreover, according to information in ChEA [45], the transcriptional regulators *Ccnd1/CCND1*, *Fli1/FLI1*, *E2f1/E2F1* and *E2f4/E2F4* are regulated by Myc (Figure 4). *Ccnd1/CCND1* has also been shown to be induced by estrogen, but not, however, by induced *Myc* expression, indicating that other response elements in the promoter region of *Ccnd1/CCND1* are required for its induction [65-67]. For example, Sabbah et al. (1999) showed that a cAMP response element, besides estrogen, was critical for the induction of *Ccnd1/CCND1* [68]. Moreover, this gene has also been shown to interact with members of the retinoblastoma protein (Rb) family [69]. Interestingly, E2f1/E2F1 and E2f4/E2F4 have been shown to be inactive when bound to Rb proteins, but activated when released upon phosphorylation of Rb by cyclins (such as Ccnd1/CCND1) and cyclin dependent kinase complexes [70]. Additionally, E2F TFs are important regulators of genes required for cell cycle progression [71].

Fli1/FLI1 is a proto-oncogene that has previously been demonstrated to undergo translocations in Ewing sarcoma and acute myeloid leukemia (AML) cases [72]. According to information in

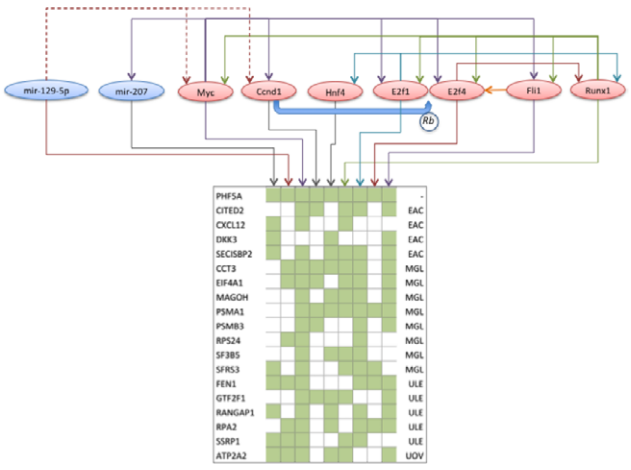

Figure 4: Regulation of central genes: The figure shows TFs and miRNAs predicted/validated to regulate Phf5a and the central genes. On top of figure, miRNAs and TFs are indicated with blue and red circles, respectively. Arrows indicate a predicted/validated regulation of the central genes listed in the box at the bottom of the figure. Green squares indicate a regulation by either a miRNA or TF. The different experiments included in this study are also indicated to the right of the box: E-GEOD-13003 (EAC), E-GSE-40173 (MGL), E-GEOD-13319 (ULE) and E-TOXM-20 (UOV).

ChEA, this gene is regulated by Myc and Runx1/RUNX1, amongst others, and itself regulates the TF *E2f4/E2F4* (Figure 4). Runx1/RUNX1 is a transcriptional activator for various genes having a role in hematopoiesis and has been established as a tumor suppressor in AML [73,74]. Moreover, Runx1/RUNX1 can act to promote G_1-S cell transition via its transactivation domain and is a transcriptional activator of *Cyclin D3*, another cyclin involved in the cell cycle [75]. *Runx1/RUNX1* itself is regulated by E2f1/E2F1 and E2f4/E2F4, amongst others, and reported to regulate *Myc*, *Fli1/FLI1* and *E2f4/E2F4* (Figure 4).

Hnf4a/HNF4A is a gene required for the development of the kidney, liver and intestine [76]. The DNA binding ability of this protein is related to its phosphorylation status, as only its phosphorylated form can bind to DNA. The basic functions of Hnf4a/HNF4A include regulation of genes involved in amino acid metabolism, lipid and bile acid synthesis. Hnf4a/HNF4A has been identified as an important gene in hepatocyte differentiation and the loss of Hnf4a has been associated with hepatocellular carcinoma in mouse [77,78]. Interestingly, overexpression of Hnf4a has been shown to block carcinogenesis and metastasis in a rat model of hepatocellular carcinoma [79]. According to ChEA [45], Hnf4a/HNFA is regulated by E2f1/E2F1, but itself does not regulate any of the transcriptional regulators previously mentioned (Figure 4).

According to information in miRWalk [47], mir-207 is predicted to target nine of the central genes, but, interestingly, none of the TFs previously mentioned (Figure 4). The expression level of mir-207 has previously been shown to be up-regulated by Myc in mouse mammary tumors and be down-regulated in estrogen-treated mice [80,81]. Mir-129-5p, on the other hand, is predicted to target eight of the central genes, but, similar to mir-207, none of the TFs previously mentioned. However, mir-129-5p has been shown to be a target of the *APC* gene in human, a gene that has previously been shown to repress the expression levels of *CCND1* and *Myc*; a down-regulation of mir-129-5p in human Hep-2 cells led to an increase of *APC* expression and which correlated with lower expression levels of *CCND1* and *Myc* in these cells [82]. Hence, an indirect connection between mir-129-5p and *Ccnd1/CCND1* and *Myc* is a possibility. Moreover, an overexpression of mir-129-5p in gastric cancer cells as well as in E10 lung epithelial cells was shown to result in significant G_1 phase arrest. Mir-129-5p has also been shown to target *CDK6*, a kinase involved in G_1-S transition in the cell cycle, as an over-expression of the miRNA resulted in inhibition of *CDK6* [83,84]. However, *Cdk6* was not among the correlated genes derived in this study.

Discussion

PHF5A is a highly conserved zinc finger protein and such proteins commonly participate in fundamental mechanisms of gene expression, e.g., as TFs and mediators of protein-protein interactions, but they can also have more specific functions, such as participating in cell growth regulation and differentiation [85]. PHF5A has previously been shown to act as a transcriptional regulator and also be involved in pre-mRNA splicing, by being a component of the U2 snRNP complex in the spliceosome machinery [16,18,19]. Moreover, essential genes tend to be evolutionarily more conserved than non-essential genes as they accomplish basic cellular functions and on the DNA level *PHF5A* has a sequence similarity over 80% in multicellular organisms. Since essential genes participate in basic cellular functions they have proven to be potential drug targets and could be considered for cancer therapy as well [7,8]. Additionally, spliceosome components are highly interesting since any obstruction in pre-mRNA splicing would halt the expression of cell cycle genes and as dividing cells require a tightly regulated expression of many essential genes, this would lead to interference in cell division and ultimately cell death [10,18]. Furthermore, several spliceosome component-coding genes have previously been shown to be over-expressed in lung, breast and ovarian cancers, implicating their role in cancer progression [20]. For example, *Phf5a/PHF5A* has been identified as essential for the formation and maintenance of glioblastoma multiform (GBM), an aggressive malignant primary brain tumor [24] and be aberrantly expressed in estrogen dependent EAC tumors from both rat and human [27]. The facts given above made it interesting to analyze *Phf5a/PHF5A* from a systems biology view. Since the study of complex diseases, such as cancer, can be difficult to perform on human samples due to a high genetic heterogeneity, the use of rat as a model organism has been shown to be a good complement; rat has similarities in both pathogenesis and histopathological properties to those of human. Hence, we choose to derive a systems biology view of the rat *Phf5a* gene.

In order to analyze the rat *Phf5a* from systems biology view, we utilized the wealth of publicly available microarray studies for rat and derived genes correlated in their expression levels to *Phf5a* in the different experiments as the basis. However, due to a large number of available data sets, we choose to focus on studies related to estrogen, since *Phf5a* had previously been linked to malignant samples of EAC in rats. The results showed that the number of correlated genes varied between the datasets and, interestingly, there were no overlaps among the correlated genes derived from the different experiments (Table 1). From the experiment E-GEOD-40173 in which estrogen stimulation response was measured in mammary glands after estrogen treatment for 11 days, 130 positively and 8 negatively correlated genes were derived, and from the experiment E-GEOD-13003 that used samples of EAC, 90 positively and 6 negatively correlated genes were derived. These two experiments comprised 77% of the correlated genes, i.e., the majority of the correlated genes were derived from either breast tissue continuously treated with estrogen or estrogen dependent endometrial adenocarcinoma tissue, which indicates the relation to estrogen dependence since in both cases the underlying mechanism would have been stimulation of estrogen sensitive tissues by estrogen. In the experiment E-GEOD-24672 a longer duration (4 weeks) of estrogen exposure was used than in E-GEOD-40173 (11 days), but only three correlated genes were derived from E-GEOD-24672. However, the tissue used in E-GEOD-24672 was testis, which has poor estrogen sensitivity due to a low expression of the estrogen receptor. Consequently, we hypothesize that this was the main underlying reason for a few number of correlated genes derived from E-GEOD-24672.

Subsequently, we collected annotation for *Phf5a* from various sources as well as for the correlated genes and derived a weighted gene network based on this information. The idea behind generating a weighted network was that stronger interactions would be represented by higher edge weights, whereas weaker and relatively unimportant interactions would be represented by lower edge weights. For example, genes sharing a high number of TFs are presumed to have a high expression correlation [86]. However, TFs alone do not regulate the expression of genes, since miRNAs have recently been shown to have a vital impact as well, and genes sharing both TFs and miRNAs have a greater probability of forming various activation loops, e.g., feed forward, auto regulatory, negative feedback loops, etc. [87]. When principal classes of gene regulators (such as TFs and miRNAs) are combined with previous established interaction data, the identification of closely interacting (central) genes becomes easier [88]. From the

weighted network it was possible to identify several central genes (by setting an arbitrary threshold on the weight) that shared a large number of annotations with *Phf5a*. Subsequently, from the reduced weighted network compromising only the central genes, we could identify top TFs and miRNAs, i.e., TFs and miRNAs regulating a large number of the central genes (by arbitrarily selecting the number of TFs and miRNAs to further analyze).

The reduced weighted network clearly showed that the TF Myc has a central role regarding the regulation of *Phf5a* and the central genes; the TF was reported to regulate 95% of the central genes. *Myc* is a well-known onco-gene that is erratically expressed in about 70% of all human cancers [39,63,89]. Overexpression of *Myc* causes overexpression of E2F TFs, which are inactive when bound to Rb proteins and activated when released upon phosphorylation of Rb by cyclin and cyclin dependent kinase complexes. The Rb/E2F pathway is one of the central pathways associated with cancer; it regulates the initiation of DNA replication and is disrupted in almost all human cancers [40,90]. Rb proteins and E2F TFs oppose each other in actions, and both of them are associated with G_1-S phase transition; the interaction of E2F family members with Rb proteins is a key event in proper cell cycling. *Phf5a* is reported to be regulated by both E2f1 and E2f4, and E2f1 regulate as well the spliceosome components *Magoh*, *Sf3b5* and *Sfrs3* that were correlated to *Phf5a*. This establishes a link between the spliceosome and the Myc/Rb/E2F pathway. Moreover, as previously described, both Rb proteins, E2F TFs as well as Runx1/RUNX1 is associated with G_1-S phase transition in the cell cycle. Additionally, mir-129-5p, one of the miRNAs identified in this study to target Phf5a/PHF5A, has been shown to also target *CDK6*, a kinase involved in G_1-S transition in the cell cycle. This indicates an interesting link between the spliceosome and the cell cycle.

However, *Myc* or *E2F* TFs were not represented among the correlated genes derived, indicating that other regulators as well are affecting the expression of these genes. Moreover, the lack of overlap among the correlated genes derived from the different experiments further support this observation. The annotation analysis of the central genes revealed clues to other regulators that might influence the expression of these genes. For example, a number of miRNAs were validated/predicted to target many of the central genes and, in more specific, miR-207 and mir-129-5p were predicted to target more than 40% of them. Interestingly, the expression level of mir-207 has previously been shown to be up-regulated by Myc in mouse mammary tumors and be down-regulated in estrogen-treated mice [80,81], and mir-129-5p to be up-regulated by *APC*, a gene that is down-regulated by both *CCND1* and *MYC* [82]. These miRNAs are new interesting regulators that potentially influence the expression of *Phf5a*. Targeting these miRNAs and others identified in this study can also be a strategy to slow down tumor progression, which should be further investigated. Moreover, experimentally establishing the spliceosome as target for both Myc and E2F TFs will strengthen the case of spliceosome inhibitors to be used in the treatment of cancer, as the Myc/Rb/E2F pathway is one of the most important pathways in tumor progression and the spliceosome is indispensable for proper expression of essential genes.

Materials and Methods

Microarray data and analysis

The following microarray data sets were downloaded from ArrayExpress [32]: E-GEOD-13003, E-GEOD-13319, E-MEXP-999, E-TOXM-20, E-GEOD-24672 and E-GEOD-40713. Annotation regarding platform, tissue, species/sex, rat model and treatment used in each experiment can be found in Table 1, as well as reference to publication. R statistical language was used to carry out the microarray pre-processing. The E-GEOD-13003 is a two-dye data with Cy3 (Cyanine 3) and Cy5 (Cyanine 5) columns. The Genepix files for this experiment were downloaded and pre-processed with the packages Marray and Array Quality available in Bioconductor [91]. These microarrays were within-array normalized using the function normalizes within arrays and the method Loess and background corrected with the method minimum. Thereafter, they were between-array normalized with the function normalize between arrays and the method Rquantile, because the red channel in this experiment represented the RNA reference.For the experiments E-GEOD-13319, E-MEXP-999 and E-GEOD-40713 the pre-processed data was directly downloaded from ArrayExpress. For the experiments E-TOXM-20 and E-GEOD-24672 the cell-files were downloaded and pre-processed using the *Affy* package and the *MAS5* function for normalization.

Extraction of correlated genes

Pearson correlation (PC) test was applied on all experiments separately and for each experiment genes with a correlated expression profile to *Phf5a*'s expression profile were derived. A user-defined function was designed in R to conduct the correlation tests and filter out all genes having a PC $\geq |0.7|$ compared to *Phf5a*.

Extraction of gene annotations

The Affymetrix IDs for the correlated genes were submitted to the Database for Annotation, Visualization and Integrated Discovery (DAVID) [42], to get official gene symbols. This was done for all experiments except E-GEOD-13003, since for this experiment the platform used was SWEGENE Rat 70mer oligonucleotide array V1.0 and therefore the data did not have Affymetrix IDs. Instead, gene IDs for the SWEGENE array was downloaded from Gene Expression Omnibus (GEO), which was then subjected to DAVID analysis. The Functional Annotation Tool in DAVID was used to derive over-represented Gene Ontology Biological Process terms and pathways from the Kyoto Encyclopedia of Genes and Genomes pathway database (KEGG) [44], using a p-value ≤ 0.05 and a gene cutoff value ≥ 5.

The gene symbols of the correlated genes were submitted to the database Chip Enrichment Analysis (ChEA) [45] and a list of transcription factors binding to any of the correlated genes was downloaded. This list was subsequently filtered to only retain those transcription factors that also bind to*Phf5a* and having a *p*-value ≤ 0.05.

The gene symbols of the correlated genes were submitted to the database miRWalk [47] to obtain a list of miRNAs predicted (p-value ≤ 0.01) or validated to bind to the genes. This list was subsequently filtered to exclude miRNAs not binding to *Phf5a*.

Protein interactions for PHF5A were obtained from the Search Tool for the Retrieval of Interacting Genes/Protein (String) database [48], using the active prediction "Experiments" and a confidence score of 0.150.

Generation of gene networks

Gene networks were generated in R statistical language using correlated genes as nodes and extracted gene annotations as edges; for two genes to be interacting they must either be predicted/validated to be regulated by the same transcription factor or miRNA, or annotated with the same biological process ontology term, have a direct protein interaction or be a member of the same pathway. The gene networks

were created using the *igraph* package available in R. Initially, an un-weighted network was created, using the function *graph.edgelist* from *igraph,* which was then converted to a weighted network using the functions *graph.adjacency* and *get.adjacency* from *igraph*. Parallel interactions between nodes were converted to a single edge with a weight, where the weight indicated the number of annotations shared between two genes. From the weighted network central genes were identified, by setting an arbitrary threshold on the edge weight and removing genes having a weight lower than the this threshold.

Authors' contributions

RV and planned the data analyses together, RV and AL did the majority of the data processing and analysis, RV, AL and EF contributed in the interpretation of the data. RV and AL wrote the draft of the manuscript and all authors participated in the editing and improvement of the manuscript. AL coordinated the work. All authors read and approved the manuscript and AL gave the final approval of the version to be published.

References

1. Laubenbacher R, Hower V, Jarrah A, Torti SV, Shulaev V, et al. (2009) A systems biology view of cancer. Biochim Biophys Acta 1796: 129-139.
2. Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. Nat Med 10: 789-799.
3. Trewavas A, Francois Jacob (2006) A brief history of systems biology. Every object that biology studies is a system of systems, Plant Cell 18 2420-2430.
4. Whitacre JM (2010) Degeneracy: a link between evolvability, robustness and complexity in biological systems. Theor Biol Med Model 7: 6.
5. Tononi GO, Sporns, Edelman GM (1999) Measures of degeneracy and redundancy in biological networks. Proc Natl Acad Sci USA 96: 3257-3262.
6. Juhas M, Eberl L, Church GM (2012) Essential genes as antimicrobial targets and cornerstones of synthetic biology. Trends Biotechnol 30: 601-607.
7. Conde-Pueyo N (2009) Human synthetic lethal inference as potential anti-cancer target gene detection. BMC Syst Biol 3: 116.
8. Dotsch A (2010) Evolutionary conservation of essential and highly expressed genes in Pseudomonas aeruginosa. BMC Genomics 11: 234.
9. Luo B (2008) Highly parallel identification of essential genes in cancer cells. Proc Natl Acad Sci 105: 20380-20385.
10. van Alphen RJ (2009) The spliceosome as target for anticancer treatment. Br J Cancer 100: 228-232.
11. Oltra E (2004) A novel RING-finger-like protein Ini1 is essential for cell cycle progression in fission yeast. J Cell Sci 117: 967-974.
12. Halbach T, Scheer N, Werr W (2000) Transcriptional activation by the PHD finger is inhibited through an adjacent leucine zipper that binds 14-3-3 proteins. Nucleic Acids Res 28: 3542-3550.
13. Aasland R, Gibson TJ, Stewart AF (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci 20: 56-59.
14. Trappe R (2002) Identification and characterization of a novel murine multigene family containing a PHD-finger-like motif. Biochem Biophys Res Commun 293: 816-826.
15. Gao Y (2013) Comparison of splicing factor 3b inhibitors in human cells. Chembiochem 14: 49-52.
16. Golas MM (2003) Molecular architecture of the multiprotein splicing factor SF3b. Science 300: 980-984.
17. Chen W, MooreMJ (2014) The spliceosome: disorder and dynamics defined. Curr Opin Struct Biol 24: 141-149.
18. Chen L, Tovar-Corona JM, Urrutia AO (2012) Alternative splicing: a potential source of functional innovation in the eukaryotic genome. Int J Evol Biol 596274.
19. Carvalho RF, Feijao CV, Duque P (2013)On the physiological significance of alternative splicing events in higher plants. Protoplasma 250: 639-650.
20. Quidville V (2013) Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. Cancer Res 73: 2247-2258.
21. Makishima H (2012) Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood 119: 3203-3210.
22. Kaida D (2007) Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol 3: 576-583.
23. Kotake Y (2007) Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol 3: 570-575.
24. Hubert CG (2013) Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. Genes Dev 27: 1032-1045.
25. Schlemmer SR, Novotny DB, Kaufman DG (1999) Changes in connexin 43 protein expression in human endometrial carcinoma. Exp Mol Pathol 67: 150-163.
26. McLachlan E (2006) Connexins act as tumor suppressors in three-dimensional mammary cell organoids by regulating differentiation and angiogenesis. Cancer Res 66: 9886-9894.
27. Falck E, Klinga-Levan K (2013) Expression patterns of Phf5a/PHF5A and Gja1/GJA1 in rat and human endometrial cancer. Cancer Cell Int 13: 43.
28. Wang Q, Rymond BC (2003) Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. Mol Cell Biol 23: 7339-7349.
29. Ahituv N (2007) Deletion of ultraconserved elements yields viable mice. PLoS Biol 5: e234.
30. Vollmer G (2003) Endometrial cancer: experimental models useful for studies on molecular aspects of endometrial cancer and carcinogenesis. Endocr Relat Cancer 10: 23-42.
31. Samuelson E (2009) Molecular classification of spontaneous endometrial adenocarcinomas in BDII rats. Endocr Relat Cancer 16: 99-111.
32. Parkinson H (2007) Array Express--a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 35: D747-750.
33. Lee HK (2004) Coexpression analysis of human genes across many microarray data sets. Genome Res 14: 1085-1094.
34. Hu R (2009) Detecting intergene correlation changes in microarray analysis: a new approach to gene selection. BMC Bioinformatics 10: 20.
35. Chen K, Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet 8: 93-103.
36. Fordyce P, Ingolia N (2011) Integrating systems biology data to yield functional genomics insights. Genome Biol 12: 302.
37. Gehlenborg N, Nitin S Baliga , Alexander Goesmann, Matthew A Hibbs, Hiroaki Kitano, et al. (2010) Visualization of omics data for systems biology. Nat Methods 7: S56-S68.
38. Bretones G, Delgado MD, Leon J (2014) Myc and cell cycle control. Biochim Biophys Acta
39. Albihn A, Johnsen JI, Henriksson MA (2010) MYC in oncogenesis and as a target for cancer therapies. Adv Cancer Res 107: 163-224.
40. Nevins JR (2001) The Rb/E2F pathway and cancer. Hum Mol Genet 10: 699-703.
41. vanRiggelen J, Yetil A, Felsher DW (2010) MYC as a regulator of ribosome biogenesis and protein synthesis. Nat Rev Cancer 10: 301-309.
42. Dennis G Jr (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4: P3.
43. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, et al. (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32: D258-261.
44. Ogata H (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27: 29-34.
45. Lachmann A (2010) ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics 26: 2438-2444.
46. Hemberg M, Kreiman G (2011) Conservation of transcription factor binding events predicts gene expression across species. Nucleic Acids Res 39: 7092-7102.

47. Dweep H, Sticht C, Pandey P, Gretz N (2011) miRWalk--database: prediction of possible miRNA binding sites by walking the genes of three genomes. J Biomed Inform 44: 839-847.

48. vonMering C, Jensen LJ, Snel B, Hooper SD, Krupp M, et al. (2005) STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res 33: D433-437.

49. Zahler AM, Lane WS, Stolk JA, Roth MB (1992) SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev 6: 837-847.

50. Buratowski S, Zhou H (1993) Functional domains of transcription factor TFIIB. Proc Natl Acad Sci USA 90: 5633-5637.

51. Kataoka N, Diem MD, Kim VN, Yong J, Dreyfuss G, et al. (2001) Magoh, a human homolog of Drosophila mago nashi protein, is a component of the splicing-dependent exon-exon junction complex. EMBO J 20: 6424-6433.

52. Lieber MR (1997) The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. Bioessays 19: 233-240.

53. Schultz-Norton JR (2007) The deoxyribonucleic acid repair protein flap endonuclease-1 modulates estrogen-responsive gene expression. Mol Endocrinol 21: 1569-1580.

54. Din S, Brill SJ, Fairman MP, Stillman B, et al. (1990) Cell-cycle-regulated phosphorylation of DNA replication factor A from human and yeast cells. Genes Dev 4: 968-977.

55. Zernik-Kobak M,Vasunia K, Connelly M, Anderson CW, Dixon K, et al. (1997) Sites of UV-induced phosphorylation of the p34 subunit of replication protein A from HeLa cells. J Biol Chem 272: 23896-23904.

56. Kalma Y (2001) Expression analysis using DNA microarrays demonstrates that E2F-1 up-regulates expression of DNA replication genes including replication protein A2. Oncogene 20: 1379-1387.

57. Orphanides G, LeRoy G, Chang CH, Luse DS, Reinberg D (1998) FACT, a factor that facilitates transcript elongation through nucleosomes. Cell 92: 105-116.

58. Wu X, Zhao SH, Yu M, Zhu ZM, Wang H, et al. (2005) Physical mapping of four porcine 20S proteasome core complex genes (PSMA1, PSMA2, PSMA3 and PSMA6). Cytogenet Genome Res 108: 363.

59. Reed JL, Dimayuga FO, Davies LM, Keller JN, Bruce-Keller AJ (2004) Estrogen increases proteasome activity in murine microglial cells. Neurosci Lett 367: 60-65.

60. Walkley NA, Demaine AG, Malik AN (1996) Cloning, structure and mRNA expression of human Cctg, which encodes the chaperonin subunit CCT gamma. Biochem J 313: 381-389.

61. Brown SJ (1990) A cDNA encoding human ribosomal protein S24. Gene 91: 293-296.

62. Lau WM, Doucet M, Huang D, Weber KL, Kominsky SL (2013) CITED2 modulates estrogen receptor transcriptional activity in breast cancer cells. Biochem Biophys Res Commun 437: 261-266.

63. Levens D (2002) Disentangling the MYC web. Proc Natl Acad Sci USA 99: 5757-5759.

64. Campaner S, Doni M, Hydbring P, Verrecchia A, Bianchi L, et al. (2010) Cdk2 suppresses cellular senescence induced by the c-myc oncogene. Nat Cell Biol 12: 54-59.

65. Spyratos F, Andrieu C, Vidaud D, Briffod M, Vidaud M, et al. (2000) CCND1 mRNA overexpression is highly related to estrogen receptor positivity but not to proliferative markers in primary breast cancer. Int J Biol Markers 15: 210-214.

66. Nakamura Y, Felizola SJ, Kurotaki Y, Fujishima F, McNamara KM, et al. (2013) Cyclin D1 (CCND1) expression is involved in estrogen receptor beta (ERβ) in human prostate cancer. Prostate 73: 590-595.

67. Cicatiello L, Addeo R, Sasso A, Altucci L, Petrizzi VB, et al. (2004) Estrogens and progesterone promote persistent CCND1 gene activation during G1 by inducing transcriptional derepression via c-Jun/c-Fos/estrogen receptor (progesterone receptor) complex assembly to a distal regulatory element and recruitment of cyclin D1 to its own gene promoter. Mol Cell Biol 24: 7260-7274.

68. Sabbah M, Courilleau D, Mester J, Redeuilh G (1999) Estrogen induction of the cyclin D1 promoter: involvement of a cAMP response-like element. Proc Natl Acad Sci U S A 96: 11217-11222.

69. Dowdy SF, Hinds PW, Louie K, Reed SI, Arnold A, et al. (1993) Physical interaction of the retinoblastoma protein with human D cyclins. Cell 73: 499-511.

70. Takahashi Y, Rayman JB, Dynlacht BD (2000) Analysis of promoter binding by the E2F and pRB families in vivo: distinct E2F proteins mediate activation and repression. Genes Dev 14: 804-816.

71. Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, et al. (2002) E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev 16: 245-256.

72. Truong AH, Ben-David Y (2000) The role of Fli-1 in normal cell function and malignant transformation. Oncogene 19: 6482-6489.

73. Janes KA (2011) RUNX and its understudied role in breast cancer. Cell Cycle 10: 3461-3465.

74. Falini B, Nicoletti I, Martelli MF, Mecucci C (2007) Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. Blood 109: 874-885.

75. Bernardin-Fried F, Kummalue T, Leijen S, Collector MI, Ravid K, et al. (2004) AML/RUNX increases during G1 to S cell cycle progression independent of cytokine-dependent phosphorylation and induces cyclin D3 gene expression. J Biol Chem 279: 15678-15687.

76. Grigo K, Wirsing A, Lucas B, Klein-Hitpass L, Ryffel GU (2008) HNF4 alpha orchestrates a set of 14 genes to down-regulate cell proliferation in kidney cells. Biol Chem 389: 179-187.

77. Bonzo JA, Ferry CH, Matsubara T, Kim JH, Gonzalez FJ (2012) Suppression of hepatocyte proliferation by hepatocyte nuclear factor 4Î± in adult mice. J Biol Chem 287: 7345-7356.

78. Watt AJ, Garrison WD, Duncan SA (2003) HNF4: a central regulator of hepatocyte differentiation and function. Hepatology 37: 1249-1253.

79. Ning BF, Ding J, Yin C, Zhong W, Wu K, et al. (2010) Hepatocyte nuclear factor 4 alpha suppresses the development of hepatocellular carcinoma. Cancer Res 70: 7640-7651.

80. Sun Y, Wu J, Wu SH, Thakur A, Bollig A, et al. (2009) Expression profile of microRNAs in c-Myc induced mouse mammary tumors. Breast Cancer Res Treat 118: 185-196.

81. Dai R, et al. (2008) Suppression of LPS-induced Interferon-gamma and nitric oxide in splenic lymphocytes by select estrogen-regulated microRNAs: a novel mechanism of immune modulation. Blood 112: 4591-4597.

82. Li M, Tian L, Wang L, Yao H, Zhang J, et al. (2013) Down-regulation of miR-129-5p inhibits growth and induces apoptosis in laryngeal squamous cell carcinoma by targeting APC. PLoS One 8: e77829.

83. Yu X, Song H, Xia T, Han S, Xiao B, et al. (2013) Growth inhibitory effects of three miR-129 family members on gastric cancer. Gene 532: 87-93.

84. Wu J, Qian J, Li C, Kwok L, Cheng F, et al. (2010) miR-129 regulates cell proliferation by downregulating Cdk6 expression. Cell Cycle 9: 1809-1818.

85. Matthews JM, Sunde M (2002) Zinc fingers--folds for many occasions. IUBMB Life 54: 351-355.

86. Marco A, Konikoff C, Karr TL, Kumar S (2009) Relationship between gene co-expression and sharing of transcription factor binding sites in Drosophila melanogaster. Bioinformatics 25: 2473-2477.

87. Arora S, Rana R, Chhabra A, Jaiswal A, Rani V (2013) miRNA-transcription factor interactions: a combinatorial regulation of gene expression. Mol Genet Genomics 288: 77-87.

88. Wu JH, Sun YJ, Hsieh PH, Shieh GS (2013) Inferring coregulation of transcription factors and microRNAs in breast cancer. Gene 518: 139-144.

89. Grandinetti KB, David G (2008) Sin3B: an essential regulator of chromatin modifications at E2F target promoters during cell cycle withdrawal. Cell Cycle 7: 1550-1554.

90. Knudsen ES, Wang JY (2010) Targeting the RB-pathway in cancer therapy. Clin Cancer Res 16: 1094-1099.

91. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.

92. Karlsson S, Olsson B, Klinga-Levan K (2009) Gene expression profiling

predicts a three-gene expression signature of endometrial adenocarcinoma in a rat model. Cancer Cell Int 9: 12.

93. Crabtree JS, Jelinsky SA, Harris HA, Choe SE, Cotreau MM, et al. (2009) Comparison of human and rat uterine leiomyomata: identification of a dysregulated mammalian target of rapamycin pathway. Cancer Res 69: 6171-6178.

94. Naciff JM, Overmann GJ, Torontali SM, Carr GJ, Khambatta ZS, et al. (2007) Uterine temporal response to acute exposure to 17alpha-ethinyl estradiol in the immature rat. Toxicol Sci 97: 467-490.

95. Naciff JM, Torontali SM, Overmann GI, Carr GJ, Tiesman JP, et al. (2005) Evaluation of the gene expression changes induced by 17-alpha-ethynyl estradiol in the immature uterus/ovaries of the rat using high density oligonucleotide arrays. Birth Defects Res B Dev Reprod Toxicol 74: 164-184.

96. Zhou W, Bolden-Tiller OU, Shao SH, Weng CC, Shetty G, et al. (2011) Estrogen-regulated genes in rat testes and their relationship to recovery of spermatogenesis after irradiation. Biol Reprod 85: 823-833.