

Gene expression is regulated biology essay

[Science](#), [Biology](#)



**ASSIGN
BUSTER**

UNIVERSITY OF ESSEX SCHOOL OF BIOLOGICAL SCIENCES FINAL YEAR

PROJECT, B. Sc. Biomedical Science By: Alexandru C. Calciu Supervisor: John

D. Norton Date: 20. 03. 2013 Word counts: Abstract - 223 Introduction -

1377 Methods, Results and Discussion - 4409

LIST OF CONTENTS

Contents

LIST OF ABBREVIATIONS

AML - acute myeloid leukaemia ARACNe - Algorithm for the Reconstruction of

Accurate Cellular Networks ARE - AU-rich element ARED - AU-rich element

containing mRNA database AUBP - AU-binding protein BIND - Biomolecular

Interaction Network Database CN-AML - cytogenetically normal acute myeloid

leukaemia ELAV - embryonic lethal abnormal vision HuR - human antigen

RHzf - haematopoietic zing finger MILE - Microarray Innovations

in LEukemia RRM - RNA recognition motif UTR - untranslated region

Abstract

When discussing the mechanisms through which gene expression is regulated, one of the key points that needs to be addressed is post-transcriptional regulation. Over the past few decades, this mechanism that involves AREs in the 3' UTR of mRNAs together with the AU-binding proteins that affect the mRNA stability, has been involved in numerous types of carcinogenetic and leukaemogenetic processes. In this study, the relation between AUBP genes and various types of cancer was analysed to determine if there is any correlation. Moreover, this study focused on HuR/ELAVL1 and analysed its interactions with other genes, based on data obtained from AML

patients. With the help of this collected data as well as data on gene interactions from the BIND database, results have shown that HuR/ELAVL1 participates in a complex network of interactions with other genes, both positively and negatively correlated. It was concluded that even though the studied AUBP was not found to be directly involved in AML through regulating the stability of targeted genes, it is possible that HuR/ELAVL1 plays a role in the process of leukaemogenesis through genes such as SKI, ZBTB16, STAT3, PRKCI, SELL, SMAD1, WEE1, NFE2, JUNB and BACH1. The suggested mechanism behind this involves genes that are directly regulated by HuR/ELAVL1 and in turn, act as a central regulator for other genes through protein-protein or protein-DNA interactions.

Introduction

ARE's and post-transcriptional regulation In 1970, Francis Crick published what is probably one of the most well-known journal articles in the field of biology, the Central Dogma of Molecular Biology. In this article, Crick hypothesised that protein synthesis is the result of two closely related and intertwined processes that take place inside the cell. Today, these processes are known as transcription and translation. Simply put, transcription is the process through which a strand of DNA is copied into a strand of mRNA, while translation is the process that uses that strand of mRNA to synthesise the final product. These are the two mechanisms that provide the basis for what is today referred to as gene expression. Gene expression is a highly complex process that undergoes regulation at various checkpoints to ensure the normal growth and development of eukaryotic cells. This regulation is achieved through a mechanism that can act at the level of DNA through

differential transcription, or at the level of protein through selective degradation. However, the regulation of gene expression has also been observed to act at the level of RNA, through changes in translational efficiency as well as changes in the stability of mRNA. (Brennan and Steitz, 2001) One such category of elements that act at the level of RNA and regulate the process of gene expression is represented by AU-rich elements (AREs). According to Barreau et al. (2005), AREs are elements of 50 to 150 nucleotides, located in the 3' untranslated region (UTR) of many mRNAs. These elements are rich in adenosine and uridine bases and are most known for their capability of degrading host mRNA through a deadenylation-dependent mechanism. This characterisation is supported by von Roretz and Gallouzi (2008) who add that AREs often include an AUUUA pentamer repeat and are one of "the most widely studied" amongst elements that alter mRNA stability. It is then useful to look back in time to when AREs were first described by Miller et al. (1984). Their experiment consisted of comparing the mRNA of v-fos, an oncogene from the FBJ murine osteosarcoma virus and the mRNA of c-fos, the protooncogene counterpart of v-fos. This comparison revealed significant differences located in the 3' UTR of their respective mRNAs, differences which were manifested after translation occurred. Today, the c-fos 3' UTR is known to contain an ARE. A classification of AREs has been made by Chen and Shyu (1995), discriminating between these elements on the basis of sequence and decay characteristics. According to this classification, class I includes the previously mentioned c-fos ARE and contains elements with one to three copies of the AUUUA pentamer, found embedded within U-rich regions. Class II is represented by AREs which

contain two copies of an UUAUUUA(U/A)(U/A) nonamer that overlap, and finally, class III is represented by AREs which completely lack the characteristic AUUUA pentamer, but instead mediate mRNA degradation through U-rich sequences. It is impossible to talk about AREs and how they influence mRNA stability in eukaryotic cells without also referring to the role of AU-binding proteins (AUBPs). As characterised by Barreau et al. (2005), these proteins recognise and bind to the afore-mentioned AREs and together act towards regulating mRNA stability. The role of these proteins becomes of crucial importance when considering their possible interactions with proto-oncogenes and cytokines, where changes in mRNA stability may lead to a loss of regulation and the induction of carcinogenesis. So far, at least 14 distinct AUBPs have been identified, most of which are associated with the degradation of their target mRNA. However, it is important to mention the two special categories of AUBPs, namely those that act towards stabilising their target mRNA (HuR/ELAVL1) and those which can either stabilise or degrade their target mRNA (AUF-1). (von Roretz and Gallouzi, 2008 ; Schott and Stoecklin, 2010 ; Brennan and Steitz, 2001)HuR/ELAVL1 and its involvement in carcinogenesisHuR was initially discovered by Ma et al. (1996) who described it as a distant member of the embryonic lethal abnormal vision (ELAV) family of RNA-binding proteins. HuR has been found to contain three RNA recognition motifs: RRM1, RRM2 and RRM3. The first two of these motifs play a role in binding and mediating ARE recognition, while the third motif is believed to maintain the stability of the RNA-protein complex by interacting with and binding to the poly (A) tail. (Yuan et al., 2010)Presently, HuR is one of the most well researched and documented

proteins that bind to AREs. The interest in this protein can be attributed to the fact that instead of enhancing degradation like most other AUBPs, HuR stabilises its target mRNA. Two possible mechanisms presented by Brennan and Steitz (2001) have been suggested for this stabilising effect. In the first mechanism, HuR is thought to act by shielding the body of its target mRNA from degradation, instead of simply slowing the rate of deadenylation. The second mechanism suggests that when HuR is over-expressed, it isolates other factors required for degradation and slows down the process. The second mechanism is also supported by Yuan et al. (2010) who suggest that by forming oligomers on its target mRNAs, HuR can possibly block the binding of other, destabilising AUBPs. According to a recent study conducted by Lebedeva et al. (2011), HuR was found to be an abundant protein that is able to identify and interact with up to 4874 genes. Bearing this in mind, it is of no surprise that Cho et al. (2011) suggest the implication of HuR in biological processes such as cell proliferation, immune and stress response, differentiation and carcinogenesis. The effects that HuR has on these processes have been the main point of multiple recent studies. Nakamura et al. (2011) have focused on the interaction between HuR and the RNA-binding protein haematopoietic zinc finger (Hzf), concluding that these two proteins play a key role in the positive post-transcriptional regulation of p53. Donahue et al. (2011) have instead focused on the effects that HuR has on survivin, a protein that is associated with a poor outcome in multiple malignancies if over-expressed. In this study they have shown that HuR binds to a 288 bp fragment in the 3' UTR region of survivin mRNA. When HuR was over-expressed, the levels of survivin were found to decrease, a result that is

inconsistent with the stabilising role that HuR is thought to have.

Interestingly, the mechanism that was suggested by the authors in order to explain this is that HuR over-expression leads to an over-expression of p53, a known negative transcriptional regulator which then affects the overall levels of survivin. This study is a good illustration of the extensive and complex pathways that surround the interaction between AREs and AUBPs. From a clinical perspective, one of the most important aspects of HuR's ability to influence the stability of mRNAs is its role in carcinogenesis. According to Audic and Hartley (2004), HuR has been found to be implicated in the up-regulation of several growth factors, linking it to the promotion of certain malignant brain tumours and colon cancers. It has also been reported by Topisirovic et al. (2006) that the HuR-mediated over-expression of eIF4E plays a major role in acute myeloid leukaemia as well as several cases of head and neck squamous cell carcinoma. Furthermore, HuR has also been found to be over-expressed in both the chronic and blast crisis phase of chronic myeloid leukaemia, with a reported increase during the transition of the condition from the chronic phase to the blast crisis phase. (Radich et al., 2006 in Baou et al., 2011)

Aims and objectives

The aim of this study is to determine if there is any significant correlation between the expression levels of AUBP genes and several types of cancer. Furthermore, this study focuses on HuR/ELAVL1 and identifies its mRNA targets and the genes whose expression it subsequently regulates. This is achieved by scanning a series of acute myeloid leukaemia patients and identifying any genes whose expression is positively or negatively correlated with the expression of HuR/ELAVL1, as well as determining if these genes are involved in any types

of malignancies. Finally, this study aims to create a map of the previously identified target genes, as well as other first-neighbour genes that have been found to interact with the studied AUBP and its targets.

Methods

Eight AUBPs were selected for this study: TIAL1, ELAVL1, ZFP36, ZFP36L1, KHSRP, NCL, ZFP36L2 and HNRNPD. Each of these proteins was individually scanned by using the Prognoscan database for all types of cancer, with no data postprocessing, no selected platform and no selected endpoint as parameters. Hits were then recorded for every cancer type where Prognoscan indicated a significant P value, ensuring hits are not double-counted when belonging to the same dataset. All hits were then summarised in a table showing the type of cancer, number of hits and whether they were up-regulated or down-regulated. This table served as the basis on which another colour-coded table was created, showing all eight AUBPs and all the types of cancer used by Prognoscan. In this table, red was used to show over-expression in poor prognosis and blue was used to show under-expression in poor prognosis. Darker shades of these colours were used when the difference between over-expressed and under-expressed hits was greater than 1 and lighter shades of colour were used when the difference between over-expressed and under-expressed hits was 1. If the number of over-expressed hits was equal to the number of under-expressed hits for a certain type of cancer, the cell was left blank. For the next part of the study, 3 microarray datasets were used. The first of these datasets, DS191, belongs to the MILE series (GSE13204) and consists of 191 acute myeloid leukaemia patients with defined cytogenetic abnormalities. The second dataset, DS345,

also belongs to the MILE series and consists of 345 acute myeloid leukaemia patients with no cytogenetic abnormalities. The third dataset, DS251, is part of an independent study (GSE15434) that involved 251 acute myeloid leukaemia patients with no cytogenetic abnormalities. The ARACNe analysis was performed by using modules from the GenePattern online tool. Initially, the ARACNe module was used with a gene expression matrix for the DS191 dataset in '. gct' format as the dataset file, 201726_at as the hub gene identifier for ELAVL1 and a p value of 1e-8. Upon completion, the '. adj' file was pipelined into the Cytoscape module of GenePattern, re-uploading the gene expression matrix as the dataset file. In the Cytoscape Viewer, the file was exported by using File> Export> Edge Attributes and checking only the 'MI Score' field. The resulting file was then modified and re-formatted as a '. gct' file according to the GenePattern Userguide. The SelectFeaturesRows module was used next with the newly formatted '. gct' file as the input filename and a list of all AffyIDs in the dataset compiled using the ARED database as the list filename. The resulting '. gct' file was then pipelined into the MergeColumns module, using a new file containing all the AffyID probesets in the original data matrix and their corresponding gene symbols. In the MergeColumns module, the new AffyID probeset file was used as input filename1 and the previous '. gct' file was used as input filename2. The resulting merged '. gct' file was then pipelined into the GeneNeighbors module, using the gene expression matrix as data filename, 201726_at as gene accession and the total number of probes in the dataset as num neighbours. The resulting '. odf' file was used to extract probesets with a score < 1.0 as positively correlated and probesets with a score > 1.0 as

negatively correlated with the hub probe. The SelectFeaturesRows module was then used with the '. gct' file from the MergeColumns module as input filename and the lists of positively and negatively correlated probes used separately, as list filename. In order to eliminate duplicate genes from the two resulting lists, a comparison was performed and a list of overlapping genes was created. This list was then used to record the MI values of duplicate genes and delete the entry with the lower MI value from the appropriate list (positively or negatively correlated). The edited lists were saved as '. gct' files. Finally, in order to remove gene symbols that may have occurred in both lists, the CollapseDataset module was used separately with the positively correlated and negatively correlated edited '. gct' files as dataset file and an Affymetrix U133 Plus2 chip platform as the chip platform, and the resulting collapsed '. gct' file was saved. This ARACNe analysis was performed in turn for each of the three datasets, following the above-written protocols. The VennDiagram module was then used separately with the resulting positively correlated files and negatively correlated files to determine any overlap between the three datasets. The genes that were determined to be common to all three datasets were used to compile two adjacency matrices in '. sif' format for use with the Cytoscape module of geWorkbench. The initial gene matrix was formatted as a '. txt' file and uploaded as a ' Tab-Delimited Data Matrix' together with the Affymetrix U133 Plus2 chip annotation file in a '. csv' format using ' Affymetrix 3' Expression' as the annotation file type. Finally, the previously generated '. sif' files were loaded, with nodes represented by gene names and the ' Restrict to genes present in microarray' option selected. The resulting

network graphs of the hub AUBP and its interactions with positively/negatively correlated targets were exported using a '. pdf' file format. An additional scan of the resulting correlated gene lists was performed in Prognoscan, exclusively for blood cancers. Hits were counted in the results of this scan as previously mentioned, this time only focusing on results for acute myeloid leukaemia (AML). Two lists were created, one for positively correlated targets and one for negatively correlated targets with the number of hits and expression pattern. These lists were then used as the basis for a literature-wide search with the help of the online GeneCards database, and notes were made on description and involvement in pathology for every gene that returned at least one hit from the Prognoscan search. Due to space-related limitations, these tables can be found on the CD accompanying this report. The previously generated '. sif' files were now converted to a '. csv' format by deleting the first two columns and adding the hub gene, ELAVL1, at the top of the list. A new project was created in geWorkbench, using the same parameters as previously described, but uploading the '. csv' files through ' Load By Symbol'. Furthermore, the Cellular Network Knowledge Base module was used to download information from the BIND database and merge it individually with the positively and negatively correlated gene lists. The generated network graphs were then edited through the Import> Attribute from table option to show a yellow border for nodes representing AUBP targets and a red or blue colour for nodes that have been found to be over-expressed or under-expressed in cases of AML through the Prognoscan search. The final network graphs were also exported in the same manner as the previous ones, under the form of a

‘. pdf’ file. Statistical analysis was performed on the ARACNe analysis for each of the three datasets at four distinct stages. The first statistical test was performed to determine whether the ARACNe-inferred target list contains a higher proportion of ARED genes than the dataset. For this purpose, the proportion of ARED genes in the entire dataset was determined. The total number of unique genes in the ARACNe output was determined by using the CollapseDataset module from GenePattern, using the MI score ‘. gct’ file as dataset file and the Affymetrix U133 Plus2 chip platform as chip platform. These numbers were then used when performing a two-tailed Chi-square test with Yates’ correction, and the results were tabulated for all three datasets. The second statistical test was performed to determine if the number of positively and negatively correlated genes from the ARACNe output is significantly different from the number of genes in the entire dataset. The total number of such genes in the dataset was determined from the previously generated ‘. odf’ file after removing duplicate entries, while the total number of such genes in the ARACNe output was determined by observing the overlap between the positively/negatively correlated genes in the dataset and those from the ‘. gct’ file generated from the recently ran CollapseDataset module. The ratio of positively correlated genes over negatively correlated genes was determined separately for the entire dataset and the genes from the ARACNe output. These ratios were then subjected to a two-tailed Chi square test with Yates’ correction to determine statistical significance. The third statistical test was performed to determine if there is any under or over- representation of ARED hits amongst the positively and negatively correlated ARACNe genes. For this, the number of

ARED hits in the entire dataset was determined by using a pre-compiled list of ARED gene symbols with the positively/negatively correlated gene lists and the number of ARED ARACNe hits taken from the input data used for the VennDiagram module. A two-tailed Fisher's exact test was then used separately for the positively correlated targets and the negatively correlated targets to determine statistical significance. The final statistical test performed on the generated data was to determine if the overlap shown by the previously generated Venn diagrams is statistically significant. For this test, the 'R Bioconductor' program was used to perform the hypergeometric distribution version of Fisher's exact test. This was done by using the command: $n_A = 100$; $n_B = 200$; $n_C = 1025$; $n_{A_B} = 501$ -phyper(n_{A_B} , n_B , $n_C - n_B$, n_A) In the above example, n_A represents the number of genes in the first set, n_B represents the number of genes in the second set, n_C represents the number of total positively/negatively correlated ARED genes in the entire dataset and n_{A_B} represents the number of overlapping genes. This command was executed for all pairings between the three datasets.

Results

Prognoscan Analysis The first step in analysing the 8 selected AU-binding proteins (AUBPs) was to use the Prognoscan database and observe how they are involved in various types of cancer. Table 1 provides a summary of the results obtained from this analysis, where darker shades of colour were used to signal cases in which positively or negatively correlated hits predominated, with lighter shades used to signal cases in which the number of positively and negatively correlated hits was more balanced.

Table 1 – Colour-coded analysis showing the expression pattern of the eight studied AUBPs and their involvement in a selection of cancer types; blue shows under-expression and red shows over-expression, with light shades indicating a moderate degree of confidence while dark shades indicate a high degree of confidence

It is worth noting that the type of cancer that has produced the highest amount of hits is represented by breast cancer, while renal cell carcinoma and glioblastoma did not produce any hits. When focusing on the AUBP of interest in this study, ELAVL1, it is important to take into account the fact that 10 out of the total of 11 hits are characterised by over-expression, as well as the fact that when analysing the expression levels of ELAVL1 in the context of AML, no hits were recorded. ARACNe Analysis After the full ARACNe analysis was performed for all 3 datasets (DS191, DS345, DS251), the results were tabulated and analysed for significance. The values that can be seen in Table 2 provide information as to whether the results obtained from the ARACNe analysis contain a higher or lower proportion of ARED hits than the number of hits to be expected by chance. Results show that for both DS191 and DS345, the number of targets inferred by ARACNe that are also ARED hits is significantly higher than the number expected by chance, while for DS251 there is no significant enrichment of the ARACNe results with ARED hits.

Table 2 – Statistical analysis of the frequency of ARED genes in the ARACNe output list**ARACNe genes****ARED hits****ARED hits/ARACNe genes (%)****P value*****DS_191**

2581620240.0022

DS_345

3547833230.0052

DS_251

2623575220.4030*Note: proportion of ARED genes in entire dataset is

1641/6128 = 21%

Positively Correlated**Negatively Correlated****ARACNe Positively Correlated****ARACNe Negatively Correlated****Dataset Ratio****ARACNe Ratio****P Value****DS_191**

5242323219089301, 622, 050, 0001

DS_345

53063235266312641, 642, 10, 0001

DS_251

50833474182710671, 461, 710, 0004

Table 3 – Statistical analysis of the frequency of ARACNe genes in positively and negatively correlated gene lists

Aside from analysing the frequency of ARED hits, it is also important to see if the ratio of positively correlated hits and negatively correlated hits is the same for the ARACNe targets before ARED filtering than the ratio that would be expected by chance. By examining the data presented in Table 3, we can see that for all three datasets, the ratio of positively and negatively correlated targets was determined to be significantly higher in the case of ARACNe inferred targets than in the case of dataset targets, showing an enrichment with positively correlated targets in the case of ARACNe targets.

Table 4 – Statistical analysis of the frequency of ARED genes in both positively and negatively correlated ARACNe output lists**POS****Frequency of ARED hits in dataset****Frequency of ARED hits in ARACNe output****P Value****DS_191**

19%19%1

DS_345

20%18%0, 1187

DS_251

17%14%0, 0006

NEG**Frequency of ARED hits in dataset****Frequency of ARED hits in ARACNe output****P Value****DS_191**

26%27%0, 735

DS_345

25%27%0, 3067

DS_251

26%30%0, 2145 Since the previous statistical tests performed on ARACNe results showed that the ARACNe targets are enriched for ARED hits, it is useful to determine if this is also true when considering positively and negatively correlated targets separately (Table 4). Results show that, interestingly, in the case of positively correlated targets, the only dataset to show a significant difference in the frequency of ARED hits between targets belonging to the entire dataset and targets belonging to the ARACNe output is DS251. In the case of negatively correlated targets, none of the 3 datasets exhibited significant differences in the studied frequencies.

Table 5 - The P values as calculated for each overlap between the 3 datasets, considering positively correlated targets

P Value

DS191 \cap DS345

0

DS191 \cap DS251

0

DS345 \cap DS251

0

DS191 \cap DS345 \cap DS251

0

Fig. 1 - Venn diagram showing the overlap between the positively correlated genes of the three datasets

Table 6 - The P values as calculated for each overlap between the 3 datasets, considering negatively correlated targets

P Value

DS191 \cap DS345

0

DS191 \cap DS251

0

DS345 \cap DS251

0

DS191 \cap DS345 \cap DS251

0

Fig. 2 - Venn diagram showing the overlap between the negatively correlated genes of the three datasets

Having used three different datasets, it is important to determine if there are any statistically significant overlaps between them. The results of this can be seen in Figure 1 for the positively correlated genes and in Figure 2 for the negatively correlated genes. One important thing to notice is that in both cases, DS345 has more overlapping targets with DS191 than it has with DS251. It can also be seen from the figures that the number of targets overlapping between DS191 and DS251 is in both cases, very small. By analysing the overlap with the help of " R Bioconductor", the P value was determined for all overlaps (Tables 5 and 6), showing that in all cases, the overlap between datasets is statistically significant. Cytoscape visualisation

The main aim of the study is that of essentially ' mapping' the immediate interactions that take place between ELAVL1 and other genes. Bearing that in mind, two network graphs were computed by using Cytoscape, showing the genes inferred by the ARACNe analysis and common to all 3 datasets that were identified to be either positively (Figure 3) or negatively (Figure 4) correlated with the AUBP of interest. Two further networks were generated with the help of the BIND database, showing protein-protein and protein-DNA interactions between ELAVL1 and ARACNe

targets, as well as several additional nodes determined to be closely related by the BIND database. The two initial networks were then merged with these newly generated networks, resulting in two more extensive networks for both positively (Figure 5) and negatively (Figure 7) correlated targets. When examining these network graphs, it can be seen that there are several so-called ‘hub genes’ that have appeared. One very interesting element that can be observed from these graphs is the fact that a high proportion of these hub genes are also known AUBP targets, some of them even being involved in leukaemogenesis.

Fig. 3 – Network graph showing the interactions between the target AUBP, ELAVL1, and positively correlated ARACNe-inferred genes common to DS191, DS345 and DS251

Fig. 4 – Network graph showing the interactions between the target AUBP, ELAVL1, and negatively correlated ARACNe-inferred genes common to DS191, DS345 and DS251

Fig. 5 – Network graph of nodes showing the protein-protein (purple) and protein-DNA (pink) interactions involving the target AUBP, positively correlated targets and other genes
a. c. d. b.

Fig. 6 – Selected sections obtained from the network graph of positively correlated targets, focusing on a. SMAD2, b. HDAC1, c. PRKCI and d. RALBP4

Fig. 7 – Network graph of nodes showing the protein-protein (purple) and protein-DNA (pink) interactions involving the target AUBP, negatively correlated targets and other genes
a. b.

Fig. 8 – Selected sections obtained from the network graph of negatively correlated targets, focusing on a. SMURF1 and b. MAFG

In order to analyse the network graphs in more detail, snapshots were taken of several ‘hub genes’ that are AUBP targets and involved in the process of leukaemogenesis (Figure 6 for positively correlated targets, Figure 8 for negatively correlated targets). To further investigate these newly found hub genes, an analysis was performed using Genecards and peer-reviewed literature. The summary of this analysis can be seen in Table 7 for the positively correlated genes and in Table 8 for the negatively correlated genes. In addition to being over/under-expressed in AML, it can be seen that most of these hub genes are also involved in a wide variety of carcinogenic processes.

Table 7 – The expression pattern, function and pathology of the previously selected genes from the network graph of positively correlated targets (data taken from CD)

Table 8 – The expression pattern, function and pathology of the previously selected genes from the network graph of negatively correlated targets (data taken from CD)

Discussion

The expression pattern of AUBPs When examining the results of the PrognosScan analysis summarised in Table 1, several observations can be made. Focusing on the types of malignancies involved, it can be seen that breast cancer has produced the highest number of hits. While one explanation for this is that there is a high proportion of AUBP-regulated AREs in the mRNA of tumour suppressor genes specific to breast cancer, it is also

important to take into account that the number of datasets available for interrogation in the Prognoscan database is significantly higher for breast cancer than it is for the other types of malignancies. At the opposite end, renal cell carcinoma has produced no hits, a result that leads to the assumption that AUBP regulation plays a very limited role, if at all, in the mRNAs involved in this pathway. It is important to mention that a limitation of the methods used in this study is represented by the disproportionate number of datasets that are available for each type of malignancy in the Prognoscan database, a limitation that may have an impact on the accuracy of the reported results. Examining Table 1 from an AUBP-centric point of view, we can see that the protein that has produced the highest number of hits is ZFP36L1 with hits for three out of the five types of blood cancers studied. This suggests that ZFP36L1 may play an important role in the regulation of erythroid development, a fact supported by the findings of Vignudelli et al. (2010). Amongst all of the hits produced by this AUBP, we can see that the negatively-correlated ones predominate, a statement which is also true for the other member of this family, ZFP36L2. This finding is consistent with the anti-inflammatory and anti-cancer effects of this family of proteins, as reported by Sanduja et al. (2011). When shifting focus to ELAVL1, it can be seen that there is an overwhelmingly higher proportion of positively-correlated hits in rapport to negatively-correlated hits. This suggests that the main mode of action of ELAVL1 is that of stabilising its target mRNAs, a mechanism widely acknowledged in related literature (Brennan and Steitz, 2001). In the case of breast cancer, results indicate that the over-expression of ELAVL1 is associated with a poor prognosis. This

result is consistent with the findings of Yuan et al. (2010) who have reported that the role of this AUBP is pivotal in breast carcinogenesis and tumour progression. One surprising result of this analysis is that no hits were reported for ELAVL1 in the case of blood cancers (with the exception of multiple myeloma), despite having been previously reported to have an important role in the development of normal haematopoiesis by Baou et al. (2011). Interpreting results from the ARACNe analysis

When studying the results of the ARACNe analysis and trying to determine their statistical significance, it is important to take into consideration that two of the datasets contain AML patients that have been determined to be cytogenetically normal (DS345 and DS251), while the other dataset (DS191) contains AML patients that have been determined to have cytogenetical abnormalities. It is also of importance to highlight that DS191 and DS345 both belong to the MILE series, while DS251 has been obtained from an independent study. Considering the above, it can be assumed that the ARACNe results for datasets 345 and 251 would exhibit certain similarities. Both datasets that belong to the MILE series show a higher proportion of AREs amongst their ARACNe-inferred targets (Table 2), suggesting a possible connection between the stabilising effect of HuR/ELAVL1 on ARE containing mRNAs and the development of both cytogenetically normal AML (CN-AML) and AML with cytogenetic abnormalities. These results, however, are not consistent with those obtained from the DS251 dataset, which does not show a greater proportion of AREs among ARACNe-inferred targets. One reason for which this inconsistency is observed can be attributed to differences between the two populations of patients selected for these studies, or

differences in the experimental techniques used to gather information and compile the genetic make-up of these patients. All of the studied datasets have shown a significantly ($P < 0.05$) higher ratio of positively correlated genes over negatively correlated genes in the case of ARACNe-inferred targets than the ratio expected by chance (Table 3). This finding suggests that the ARACNe analysis is biased towards positively-correlated genes when identifying targets, a fact that can be linked to HuR/ELAVL1's well-documented property of stabilising targets and increasing protein expression. In Table 4, we can see that when considering the frequencies of AREs between the entire dataset and ARACNe-inferred results, separately for positively correlated and negatively correlated targets, most datasets show no significant difference between the two ratios. However, one exception is represented by DS251 in the case of positively correlated targets, which shows a significant ($P < 0.05$) decrease in the frequency of AREs among ARACNe-inferred targets. This finding is once again inconsistent with the trend exhibited by the other dataset containing CN-AML patients, and could once again be explained by possible differences in the selected patients or experimental procedures. Analysing dataset overlap It is important to determine if there is any overlap between the three datasets that have so far been analysed separately. Figures 1 and 2 provide a graphic visualisation of this overlap, while the data presented in Tables 5 and 6 tells us that all the overlaps presented in the afore-mentioned figures are highly significant from a statistical point of view ($P < 0.05$). Having determined that, it is interesting to analyse this overlap further by considering the patient categories which these datasets represent. Since DS345 and DS251 both

represent CN-AML patients, it would be expected that the overlap between these two datasets is greater than the overlap between DS345 and DS191 when looking at both positively and negatively correlated targets. However, as it can be seen from the respective figures, the overlap between the two datasets belonging to the same study consists of more genes than the overlap between studies consisting of patients with the same type of AML. Considering this result alongside all the other observations made so far in this study, it can be hypothesised that the activity of HuR/ELAVL1 does not discriminate between the two sub-types of acute myeloid leukaemia.

Interactions between HuR/ELAVL1 and 'first neighbour' genes
The generated network graphs present in Figures 5 and 7 provide an overview of the interactions between ELAVL1, the targets it regulates (yellow border) and 'first neighbour' genes as suggested by the BIND database. These network graphs can provide valuable insight into the complex underlying mechanisms through which the studied AUBP can influence the process of leukaemogenesis. One focus point of this study is represented by nodes that have been determined to be AUBP targets as well as associated with a poor prognosis in AML when they were over-expressed (red) or under-expressed (blue). A few of these nodes have emerged as key 'hub genes', connected to ELAVL1 while also playing a central role in interactions with other proteins. For the positively correlated network graph, four such 'hub genes' have been selected for a more in-depth analysis (Figure 6). The first one of these, SMAD2, shows multiple protein-protein interactions with neighbouring genes. According to Table 7, SMAD2 appears to be widely involved in various carcinogenetic processes, especially when under-expressed. However,

according to the results recorded from Prognoscan, SMAD2 is over-expressed in AML. This apparent contradiction is merely an indication of the complexity of interactions at play, and in order to be able to understand these mechanisms, it is required to look past the first layer of genes involved. What can be observed from this network graph is that ELAVL1 positively regulates the expression of SMAD2, and, indirectly, also regulates the genes linked to SMAD2. This suggests a possible mechanism for the involvement of ELAVL1 in the cancer types outlined in the initial Prognoscan result (Table 1) such as breast cancer through PAK1 (Bostner et al., 2010) or even in the development of leukaemia through SKI (Ueki et al., 2008). The second selected 'hub gene', HDAC1 (Figure 6b), can be found over-expressed in several cancer types (Table 7), showing consistency with the stabilising role of ELAVL1. What makes HDAC1 a remarkable 'hub gene' is its interactions with other genes such as ZBTB16, reported to be extensively involved in promyelocytic leukaemia (Kang et al., 2003) and the process of leukaemogenesis (Dong et al., 1996). STAT3 has also been reported to be involved in leukaemogenesis (Hillion et al., 2008) and BCL3 has been involved in chronic lymphocytic leukaemia (Schlette et al., 2005). The third highlighted gene, PRKCI (Figure 6c), has been involved in a wide range of malignancies such as breast cancer and lung cancer, offering a mechanism for the involvement of ELAVL1 in these (Table 7). However, PRKCI also interacts with genes such as PAWR, reported to be under-expressed in chronic lymphocytic leukaemia (Kukoc-Zivojnov et al., 2004) or SELL, also reported to be involved in chronic lymphocytic leukaemia (Csanaky et al., 1994). The final highlighted gene for positively correlated targets is RALBP1

(Figure 6d), a gene that unlike the previously selected ones, has been found to be under-expressed in AML. RALBP1 has also been reported to play a role in breast and lung cancers (Table 7), offering yet another possible mechanism through which ELAVL1 could be linked with these types of cancer. ELAVL1 could also have a role in carcinogenesis through indirectly regulating RAC1, a gene reported to promote transcription in colorectal carcinoma (Matos and Jordan, 2006), and also reported to be mutated in brain cancers such as astrocytomas (Hwang et al., 2004). When examining the 'hub genes' selected from the negatively correlated network graph (Figure 7) and their involvement in pathology (Table 8), it is surprising to see that they have not been directly linked to any type of malignancy. This suggests that the genes which interact with these two nodes must be responsible for the involvement of ELAVL1 in carcinogenesis, through indirect regulation. It then comes as no surprise that several genes that exhibit a protein-protein interaction with SMURF1 (Figure 8a) have been linked to various types of malignancies such as SMAD1 to lymphoma (Munoz et al., 2004) or WEE1, to leukaemogenesis (Lin et al., 2006). The other selected 'hub gene', MAFG (Figure 8b), has been reported to be under-expressed in AML following the PrognosScan analysis, and also exhibits several protein-protein interactions with genes such as NFE2, involved in erythroleukaemia (Forsberg et al., 2000), JUNB, reported to be inactivated in chronic myeloid leukaemia (Yang et al., 2003) and BACH1, reported to also be involved in erythroleukaemia (Tahara et al., 2004).

Conclusion

This study has confirmed the involvement of AUBP genes in various types of carcinogenetic and leukaemogenetic processes. Moreover, it has been shown that HuR/ELAVL1 does not discriminate between cases of CN-AML and cases of AML with defined cytogenetic abnormalities when considering the expression patterns of genes that are positively and negatively correlated with the studied AUBP. Finally, network graphs of interactions between ELAVL1, the target genes it regulates and other 'first neighbour' genes have been generated, illustrating the complex interactions that take place in this system. While the genes that are regulated by HuR/ELAVL1 have not been found to be implicated in the process of leukaemogenesis, when analysing secondary genes that are indirectly regulated by HuR/ELAVL1, extensive roles in leukaemogenesis have been found for SKI, ZBTB16, STAT3, PRKCI, SELL, SMAD1, WEE1, NFE2, JUNB and BACH1, indicating a possible mechanism through which HuR/ELAVL1 can be involved in the aforementioned types of blood cancer.

Proposal for future work

In this study, network graphs were generated that essentially mapped the interactions that take place between HuR/ELAVL1 and the genes it regulates, as well as other 'first neighbour' genes. While this map has provided a basis for suggesting possible mechanisms regarding the involvement of this AUBP in the processes of carcinogenesis and leukaemogenesis, a very useful further study could involve analysing this graph in much greater detail and expanding it to go beyond 'first neighbour' genes and map the complex interactions in which this AUBP is involved. Furthermore, this expansion of

the presented network graphs can also be done for other AUBPs by using the data obtained from other similar studies that have each focused on a different AUBP. After generating these expanded network graphs, the resulting graphs can then be used to research and propose possible mechanisms of interaction between these genes and their implication in various types of cancer, possibly leading to novel anti-cancer targets for future drug development. The proposed study can be undertaken by initially performing a GeneCard search to characterise every gene that can be found in the original network graphs. In order to determine further interactions between these genes and other genes, the Human Protein Reference Database can be used to analyse their involvement in several well-known pathways. After generating a new and expanded network graph in Cytoscape for each of the original eight AUBPs, a literature-wide search can be performed for the network's newly added genes to determine their degree of involvement in carcinogenesis and/or leukaemogenesis. Once this has been completed, the authors could then focus on describing possible mechanisms of interaction between the original AUBP and these newly discovered genes. This could lead to enhancing our understanding of the genes affected by AUBPs and how these can be involved in cancer, information that could provide pharmaceutical laboratories with novel pathways and interactions that can be targeted for the development of new anti-cancer drugs.