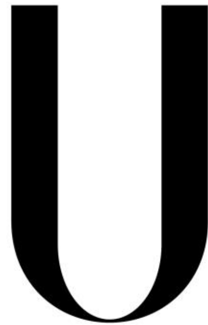
University of Lisboa
Faculty of Sciences

Department of Plant Biology



LISBOA

UNIVERSIDADE
DE LISBOA

PROGRESS REPORT
YEAR 2
(1 January 2016 – 14 October 2016)

Analysis of the translatome by ribosome profiling in colorectal cancer

Joana Filipa Pires Silva

Biological Systems: Functional and Integrative Genomics
PhD Programme in Biology/Biochemistry

PhD in Biology *(specialty: Systems Biology)*
SFRH/BD/106081/2015

Supervised by Prof. Luísa Romão and Prof. Augusto Luchessi

Contents

Page

1. Abstract **3**

2. State of the Art **4**

3. Objectives **6**

4. Summary of Activities **7**

Objective 1 – *Analysis of the translatome* of normal (NCM460) and CRC (HCT116) cell lines by ribosome profiling 7
Objective 2 – *Study of the main biological function and cellular pathways of the proteins differentially expressed in normal and CRC cells* 8
Objective 3 – *Study of the non-coding mRNA sequences (AltORFs in 5'UTR and/or 3'UTR) and non-annotated ORFs that appear differentially translated in normal and CRC cells* 11

5. Publications and Conference Communications **12**

5.1. Communications at scientific conferences 12

6. Attendance at Scientific Meetings, Courses and Training Visits **13**

6.1. Conferences 13
6.2. Courses and Training Visits 13

7. Future Work and Timeline **14**

8. References **17**

1. Abstract

Colorectal cancer (CRC) has a high incidence and mortality rates worldwide [1]. CRC carcinogenesis is a continuous accumulation of genetic alterations with concomitant variations in gene expression profiles [2]. To study the variations of gene expression profiles involved in cancer progression, genome-wide analyses have so far focused on the abundance of mRNA as measured either by microarray or RNA sequencing [3,4]. However, neither approach provides information on protein synthesis, which is the true end-point of gene expression [3-5]. Ribosome profiling (Ribo-Seq) emerges to monitor *in vivo* translation, providing global and quantitative measurements of translation by deep sequencing of ribosome-protected mRNA fragments (RPFs) [5,6]. This technique has revealed unexpected complexity in translation, including the presence of ribosomes outside of classical protein-coding regions of the transcriptome [3].

The main goal of this project is to determine the changes between the translatome of CRC and normal colorectal cells and their role in CRC tumorigenesis. For that, we aim to analyze ribosome profiling data already available for the CRC cell line HCT116, and eventually data from non-neoplasic colorectal cells (if available). Gene ontology and network interaction analysis of the differentially translated mRNAs will elucidate the main molecular pathways through which the corresponding proteins are involved in CRC progression. Furthermore, we aim to analyze the function of translatable small open reading frames (sORFs), such as the upstream ORFs (uORFs), and/or the corresponding encoded peptides in the regulation of CRC progression.

We have performed a computational analysis of HCT116 Ribo-Seq data to detect potential translatable uORFs. For that we are currently determining the number of RPFs in the 5'UTR of transcripts. Meanwhile, and based on previously published data about the prediction/detection of translatable alternative ORFs (AltORFs) in CRC cells [7], *ABCE1*, *ABCF1*, *ABCF2* and *ABCF3* mRNAs were chosen for further studies. To analyze their mRNA expression levels, we performed semi-quantitative RT-PCR analysis using RNA from HCT116, Caco-2 and SW480 CRC cells, as well as from non-neoplasic colorectal NCM460 cells. Our results show that these mRNAs are down-regulated in HCT116 cells comparing to their expression in the other three cell lines. In addition, we have been involved in mapping, by circular rapid amplification of cDNA ends (cRACE), cloning and sequencing, the exact 5' end of the *ABCE1* 5'UTR. After getting this information, we will clone this 5'UTRs in a reporter construct that will allow us to test the *ABCE1* uORF potential function in CRC progression.

2. State of the Art

Despite the significant advances in diagnosis and treatment of cancer, this disease still appears as a leading cause of death worldwide [8]. Colorectal cancer (CRC) arises as one of the most harmful cancers being, as estimated in 2012, the third most common and the fourth more deadly cancer worldwide [1]. At early stages, CRC patients can reach cure with surgery and systemic therapies, however, in advanced stages of the disease, the 5-year survival rate drops significantly to around 10% compared to the 90% obtained in early stages [9,10]. Several are the therapeutic approaches currently available for CRC, but the problem of possible recurrent or advanced cases still leads to a high difficulty in reaching the cure in CRC patients [11,12]. The tumorigenesis process in CRC cells is based in a continuously accumulation of genetic alterations leading to changes in the overall gene expression profiles. This creates the need for genome wide analyses of gene expression to reveal the cellular pathways involved in the different stages of carcinogenesis, and then translate this information to clinical application [2]. Gene expression profiles of CRC cells have been widely carried out by microarray [11,13-15] and RNA deep sequencing (RNA-Seq) analyses [10,16]. In a high-throughput genome wide analysis of 276 samples of cancer and normal colorectal samples, a pool of 32 recurrently mutated genes was identified, among them the expected *APC*, *TP53*, *SMAD4*, *PIK3CA* and *KRAS*. Integration of data about mutations, copy number and mRNA expression changes gave the ability to understand how several cellular pathways are deregulated in CRC, including the WNT, MAPK, PI3K, TGF-β and p53 pathways [17]. However, the clinical application of the so far identified molecular markers in prognosis and treatment of CRC continue to be poor, reinforcing the need for the discovery of new molecular markers [2].

Gene expression studies based on translation analyzes give answers about the regulatory changes in protein expression patterns adopted by a cell in a constantly changing environment due to several stimuli [5,18]. Gene expression regulation is controlled at the levels of transcription, mRNA processing and stability and protein translation [19]. It has been shown that post-transcriptional regulation of gene expression is modulated by features present in non-coding regions of mRNAs, both downstream and/or upstream leader regions, and can be used to rapidly change the gene expression profiles in the presence of different stimuli [18-20]. Translation is also largely regulated at the initiation level, as well as at elongation and termination levels [18]. However, most translational regulation is exerted at the translation initiation step, where the initiation codon is identified and decoded. Nevertheless, when translation is inhibited by cell stress, alternative mechanisms of translation initiation act to maintain the synthesis of certain proteins required either for the stress response or to aid recovery from the stress stimulus. In many cases, features in the 5' untranslated regions (5'UTRs) of these mRNAs such as upstream open reading frames (uORFs) are important for translational control. About 15% of the human mRNAs present uORFs within their 5'UTR [21]. The uORFs are conspicuously common in certain classes of genes, such as oncogenes and genes involved in the control of cell growth and differentiation, making easy to understand that mutations or polymorphisms in uORFs can be associated with the development of several diseases, including cancer [21-23]. For a uORF to function as a translational regulatory element, its initiation codon must be recognized, at least at certain times, by the scanning

40S ribosomal subunit and associated initiation factors. When the uORF recognition is regulated by a so-called leaky scanning mechanism, ribosomes either scan through the upstream AUG codon or recognize it, initiating translation. In the case the uORF is recognized and translated by a scanning ribosome, multiple alternative fates are available to the ribosome. One option is for the ribosome to remain associated with the mRNA, continue scanning, and reinitiate further downstream, at either a proximal or distal AUG codon. The potential of a ribosome to reinitiate further downstream, as well as the site at which it reinitiates, can vary depending on both trans-acting factors and the structure of the mRNA [24]. Another option for the ribosome is to stall during either the elongation or termination phase of uORF translation, creating a blockage to additional ribosome scanning. Previous reports indicate that ribosome stalling is mediated by the structure of the peptide encoded by the uORF [22]. uORFs can also affect gene expression by altering mRNA stability. The similarity in the cistronic organization of a uORF-containing mRNA to that of an mRNA containing a nonsense mutation has suggested the potential of a uORF-bearing mRNA to trigger nonsense-mediated mRNA decay (NMD). Indeed, it has been shown that NMD functions to control the physiologic levels of transcripts bearing uORFs [25,26].

Deep sequencing technologies and microarray studies are focusing in the study of the abundance and content of mRNA in the cell, however the challenge holds with the identification of mRNAs that are in fact being synthesized into proteins [3-5]. In fact, the correlation between transcript and protein abundance is underestimated, as it does not account for the translational regulation mechanisms, thus limiting and masking the analysis of gene expression [5]. Moreover, a large percentage of the transcriptome was mistakenly thought to have no coding properties [3,4]. Ribosome profiling is a new technique with high accurate determination and quantification of *in vivo* translation with single-nucleotide resolution based in deep sequencing of mRNA fragments covered with translating ribosomes to map the exact position of ribosomes in the transcripts [3,6]. The advent of ribosome profiling led to the identification of translation beyond the known annotated coding sequences. In fact, ribosome profiling data of the supposedly non-coding fragments demonstrate hallmarks of eukaryotic translation [3]. Furthermore, this novel approach can be informative about other start sites relative to the annotated canonical start codon leading to alternative ORFs (AltORFs) overlapping or not the main protein-coding ORF. In addition, it can also demonstrate translation of uORFs in the 5'UTR, as well as that the 3'UTR can be in fact translated [4,27]. Also, long noncoding RNAs (lncRNAs) emerge by ribosome profiling as coding sequences and being associated with crucial biological functions [4,5]. The example of ribosome profiling studies carried out in mouse embryonic stem cells (mESCs) well expose the complexity of the mammalian proteome. This work revealed unexpected translation products like N-terminal extentions, truncated products, as well as uORFs initiated at canonical and non-canonical start codons [28]. Furthermore, ribosome profiling has been widely applied in several cancer cell lines, such as the kidney cancer cell line HEK293 [29,30] and the cervical cancer cell line HeLa [30,31], among others. Interestingly, ribosome profiling of the CRC cell line HCT116 gave information about new proteins not yet annotated and possibly originated from N-terminal extensions or uORFs [32]. Thus, a more reliable understanding of the translatome of a cell is of most importance to understand the molecular and cellular mechanisms underlying normal and disease conditions [7].

3. Objectives

This doctoral project has five main objectives, namely:

1) Analysis of the translatome of normal (NCM460) and CRC (HCT116) cell lines by ribosome profiling. Translation levels of the main ORF or AltORFs will be quantified and the ones differentially translated will be further studied (objectives 2 to 4).

2) Study of the main biological function and cellular pathways of the proteins differentially expressed in normal and CRC cells:

   2.1) bioinformatics and gene ontology analysis.

   2.2) experimental analyses to confirm the relevance of such proteins in CRC tumorigenesis process.

3) Study of the non-coding mRNA sequences (AltORFs in 5'UTR and/or 3'UTR) and non-annotated ORFs that appear differentially translated in normal and CRC cells:

   3.1) determination of the relevance of such regions in translational regulation.

   3.2) analysis of the biological function of small peptides encoded by AltORFs (uORFs and/or others) and their relevance in CRC progression.

4) Validation of the obtained results in primary cell cultures of CRC and in normal colon mucosa.

5) Thesis writing and manuscript(s) preparation.

The work carried out towards each of these objectives is described in the following section according to each of these objectives.

4. Summary of Activities

Objective 1 – *Analysis of the translatome of normal (NCM460) and CRC (HCT116) cell lines by ribosome profiling.*

To answer our question regarding the differential translational patterns between CRC and normal colorectal cells and to identify potential proteins and/or uORFs encoded peptides associated to the tumorigenesis process, we thought in implementing a ribosome profiling procedure. To get this information we would need to perform this technique, for instance in different colorectal cell lines. However, ribosome profiling is an expensive and time-consuming methodology and thus we cannot have the results and their extensive analysis in a desirable time scale for this PhD. As there is already available Ribo-Seq data for the CRC HCT116 cell line, the analysis of this data can be helpful to circumvent this issue. However, the lack of Ribo-Seq data for non-neoplasic colorectal cells complicates this analysis. Nevertheless, we are now focused in determining the number of ribosome-protected mRNA fragments (RPFs) in the 5'UTR of transcripts, which is potentially associated with the presence of uORFs, using the ribosome profiling data from HCT116 cells.

This task is being performed in collaboration with the BioSYS PhD student Hugo Santos from Prof. Margarida Gama-Carvalho Group. The Ribo-Seq dataset from HCT116 cells was downloaded from NCBI Geo Datasets with the accession number GSE58207. The computational analysis has been performed as described by Crappé and co-workers [32]. Briefly, first, the 3' adaptor sequence from the sequenced reads is trimmed using the STAR internal clipping function. Second, these reads are mapped to STAR indices to discard potential interferents like rRNA and tRNA molecules. Third, the reads are aligned to the human genome (assembly GRCh37 from Ensembl) using STAR program and only the uniquely mapped reads are considered for further analysis.

To search for the existence of putative translatable uORFs we are conducting our analysis in terms of determining the levels of ribosome occupancy (i.e. number of RPFs) at: (i) 5'UTRs, and (ii) corresponding main coding sequences (CDSs). The RPFs distribution is determined using R/Bioconductor. Transcripts bearing ribosomes at their 5'UTRs have currently been grouped based in their 5'UTR length (to minimize any feature length bias in the RPFs count). The same holds for the CDS sequence RPFs measurements. Therefore, a correlation between RPFs accumulation in the 5'UTR and CDS of a transcript can point some translational regulation by the translatable 5'UTR. Thus, the average number of RPFs for the 5'UTR and CDS per transcript will be determined and translation efficiency represented as the ratio of 5'UTR/CDS. A high ratio can then be associated to an enrichment of RPFs in the 5'UTRs of a transcript compared to the corresponding CDS. Thus, from this result we can speculate that possibly this 5'UTR carries a uORF with translational regulatory function. The transcripts with 5'UTRs showing a significant ribosome occupancy, will be investigated in the next tasks.

Objective 2 - *Study of the main biological function and cellular pathways of the proteins differentially expressed in normal and CRC cells*.

Here we aim to characterize the proteins encoded by the previously selected transcripts bearing a higher level of translation in their 5'UTRs (consistent with the existence of uORFs) in terms of their biological function and its association to CRC tumorigenesis process. As we are now setting all the conditions for a proper analysis of Ribo-Seq data in HCT116 cells, we do not have yet a set of putative targets for further study. However, in the meantime and based in an initial search in the literature we found a study performed by Vanderperre and co-workers [7] which predicts/detects translatable AltORFs in a panel of colorectal cancer cell lines. Based in this information, we chose the mRNAs of *ABCE1* and *ABCF1*, *ABCF2* and *ABCF3* for further studies. The corresponding proteins belong to the ATP-binding cassette (ABC) protein family, subfamily E and F, respectively. The association of ABC proteins to cancer is usually determined by its function as drug transporters, being seen as efflux carriers of anticancer drugs which leads to drug resistance. However, more than just simple transporters, this family of proteins has been linked to different stages of cancer development (initiation, progression and acquisition of invasive and metastatic features), pointing out that these proteins are critical targets in the success of cancer treatment [33].

To understand the biological functions of these four proteins, we have performed bioinformatics and gene ontology analysis using different curated databases (Objective 2.1). First, we wanted to access the expression levels of these proteins in CRC cells compared to the normal colorectal tissues. For that, we used The Human Protein Atlas database. Our results show that ABCE1 protein is highly expressed in a panel of colorectal cancers. ABCE1 protein is also overexpressed in non-neoplasic colorectal cells, which reveals that this protein plays major roles in the cell. On the other hand, the three ABCF proteins show reduced expression in CRC cells as compared to ABCE1. Moreover, ABCF1 and ABCF3 exhibit higher levels of protein expression in normal tissues than in CRC cells. The study by Hlavata *et al.* (2012) [33] shows a correlation between transcript and protein levels for ABCE1, being up-regulated in CRC cancer tissues. Moreover, the three ABCF mRNAs levels do not change between cancer and normal tissues.

To disclose the molecular functions and general biological process of those four ABC proteins we performed a gene ontology analysis. For this analysis it was used the Gene Ontology Consortium and Quick GO (EMBL-EBI). As shown in Table 1, the term 'translation' is consistently highlighted for the ABC proteins in study. In fact, for each one of these proteins it is discriminated some role in the translation machinery, as in translation initiation, elongation or even in ribosome recycling, as associated to ABCE1 [27]. Additionally, corroborating these results, String analysis (Figure 1) shows that ABCE1, ABCF1 and ABCF2 interacts with several eukaryotic initiation factors (eIFs), that are major players in translation. Thus, having in mind that the translation patterns are known to be deregulated in cancer, the association of these proteins to the translation machinery might indicate their involvement in cancer.

Table 1 – Bioinformatics characterization of ABCF1, ABCF2, ABCF3 and ABCE1 proteins. Gene Ontology Consortium (http://pantherdb.org/) and Quick GO from EMBL-EBI (http://www.ebi.ac.uk/QuickGO/) were accessed to obtain information about the molecular function, general biological process and localization of those proteins. *Young *et al.* (2015) [27]

Gene ID	Protein ID	Molecular function	Biological process	Localization
ABCF1	**Q8NE71**	• Transporter activity • **Translation elongation factor** • **Translation initiation** • ATPase activity • ATP/protein/**RNA and ribosome binding**	• **Positive regulation of translation** • Transport • Ribosome biogenesis • **Inflamatory response**	• Nucleus • Cytoplasm • Ribosome
ABCF2	**Q9UG63**	• Transporter activity • **Translation elongation factor** • ATPase activity • ATP binding	• Transport	• Mitochondrial envelope
ABCF3	**Q9NUQ8**	• Transporter activity • **Translation elongation factor** • ATPase activity • ATP/protein binding	• Defense response to virus	• Membrane
ABCE1	**P61221**	• ATPase activity • ATP/protein binding • Ribonuclease inhibitor activity (ex. RNaseL) • **Ribossome recycling***	• Negative regulation of catalytic activity • Response to vírus • **RNA turnover** • Transport	• Cytopalsm • Mitochondrial matrix

We have been analysing by semi-quantitative reverse transcription-PCR (RT-PCR) the levels of each ABC mRNA in colorectal cell lines of interest: three CRC cell lines (HCT116, Caco-2, SW480) and a non-neoplasic colorectal cell line NCM460. As shown in Figure 2, the four ABCs are generally down-regulated in HCT116 cells when compared to the non-neoplasic cell line and the two other CRC cell lines Caco-2 and SW480. Moreover, for *ABCF1 and ABCF2* an expression increase was detected in SW480 cells (a prone to be metastatic cell line), as well as for *ABCE1* but in a lesser extent. For *ABCF1* it was shown a higher transcript level in Caco-2 cells (a tumor cell line with primary-like features). For *ABCF3* a high expression level is noticed in the non-neoplasic NCM460 cells. This results need to be confirmed by reverse transcription-quantitative PCR (RT-qPCR). Moreover, as soon as we have specific antibodies, we will analyse the protein levels of these four transcripts (Objective 2.2).

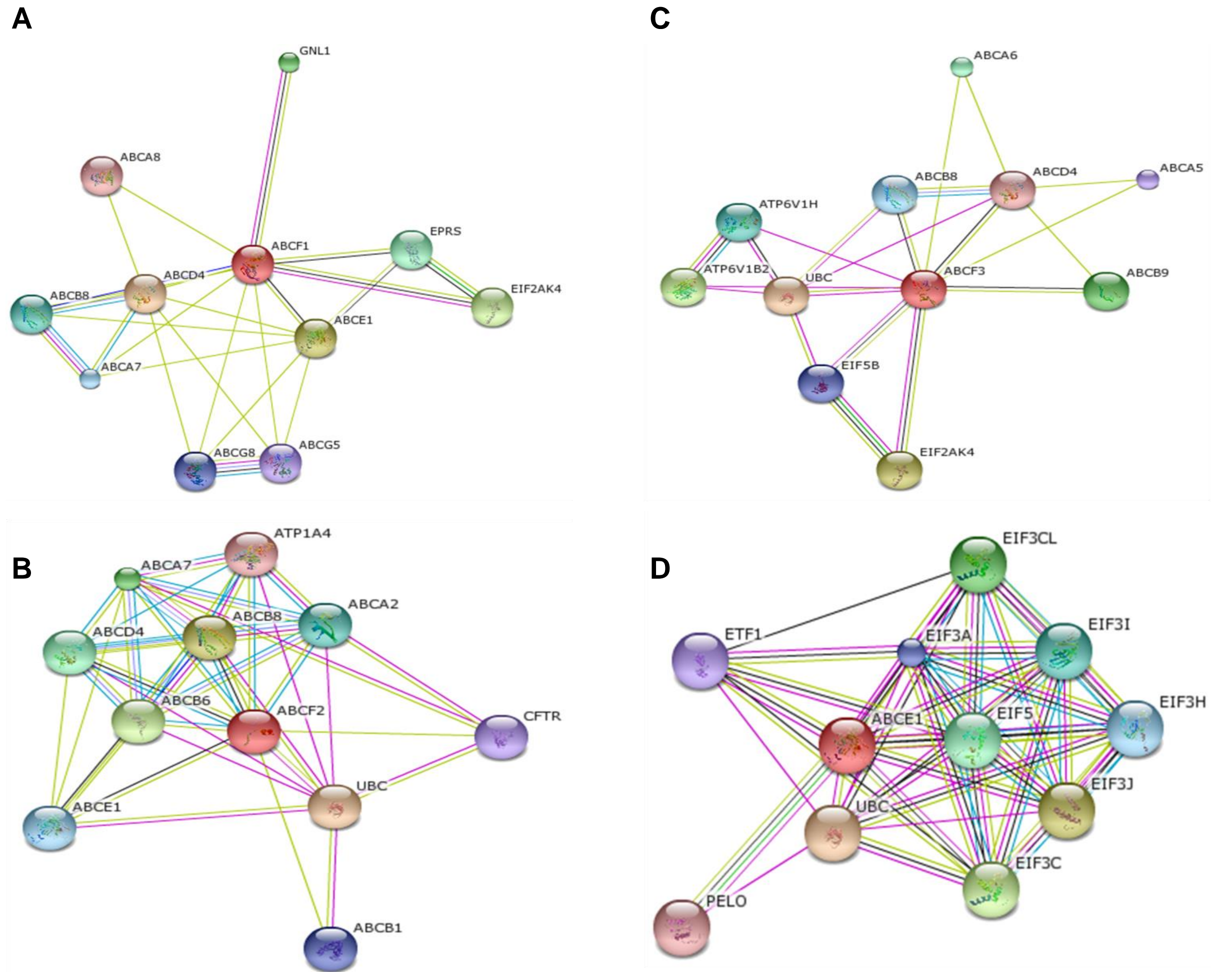


Figure 1 – Protein-protein interactions of ABCF1 (A), ABCF2 (B), ABCF3 (C) and ABCE1 (D) using String database.

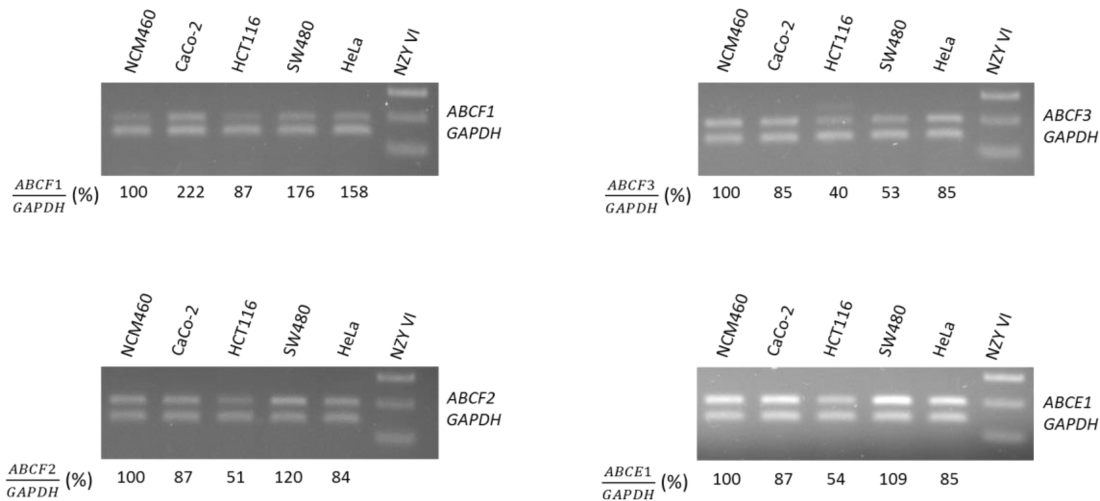


Figure 2 – Semi-quantitative RT-PCR of the transcripts *ABCF1*, *ABCF2*, *ABCF3* and *ABCE1* in non-neoplasic colorectal (NCM460) and CRC (HCT116, Caco-2 and SW480) cell lines. GAPDH is used as an internal control. The ratio ABC/GAPDH (%) presented is the ratio of each cell line normalized to the one in NCM460 cells. NZY VI – molecular marker VI from NZYTech.

Objective 3 - *Study of the non-coding mRNA sequences (AltORFs in 5'UTR and/or 3'UTR) and non-annotated ORFs that appear differentially translated in normal and CRC cells.*

To study if the potential uORFs identified are translational regulators of the corresponding main coding sequence (Objective 3.1), we intent to construct reporter plasmids carrying the cDNA sequence of the 5'UTR containing the uORF(s) of interest fused to the *Firefly* luciferase (Fluc) ORF. To clone the 5'UTR of each one of the ABC transcripts we first need to characterize their true 5' end. For that we have been using the circular rapid amplification of cDNA ends (cRACE). This technique, as described by McGrath [34], is based in circularized RNA used for cDNA synthesis in order to perform the PCR reactions from which the end products will be further cloned and sequenced. Briefly, this protocol has four major steps: RNA decapping, RNA circularization, cDNA synthesis and nested PCR [34], making the protocol setup for cRACE very challenging. After several optimizations and applying an adapted protocol from Ruiz de los Mozos *et al.*, 2013 [36], we got the first sequencing results and these that do not provide information about the 5'UTR sequence of *ABCE1*, as indicated by performing a nucleotide basic local alignment with BLAST. This result can be possibly explained by the substitution of the decapping agent from Tobacco Acid Pyrophosphatase (TAP) to RNA 5' Pyrophosphohydrolase (RppH), since the decapping properties of RppH are not so well characterized as for TAP. To solve this problem, we will sequence genomic DNA, as well as cDNA, to identify the complete sequence of *ABCE1* 5'UTR.

Despite this setback, we tried to proceed to the cloning of *ABCE1* 5'UTR based in the sequence annotations in RefSeq (NCBI) and Ensembl databases. The 5'UTR of *ABCE1* has different 5' ends depending on the database used; for instance, the 5'UTR annotated in Ensembl is 72 nts longer at the 5' end of the 5'UTR compared to the 5'UTR in RefSeq. The cloning procedure is based in synthesis by overlap extension (SOEing) PCRs using as template the cDNA from NCM160 cells. In this technique, the first SOEing (SI) PCR allows the amplification of the full 5'UTR of interest. Thus after performing the SI PCR we do not have any amplification product for the 5'UTR of *ABCE1*. This can be explained by the existence of a different 5' end for this transcript than the one annotated in the above mentioned databases. This explanation arrives from a previous sequencing reaction with a forward primer at an internal position in the 5'UTR of *ABCE1* that shows that the 3' end of the 5'UTR is consistent with what is annotated in the databases.

5. Publications and Conference Communications

5.1. Communications at scientific conferences

1. *Joana Silva, Augusto Luchessi, Luísa Romão* (2016) *Analysis of the translatome by ribosome profiling in colorectal cancer*. Poster presented at the *Advanced Lecture Course on Systems Biology*, Innsbruck (Austria). February 28 - March 5. *From Molecules to Functional Phenotypes* **PS2-28**, Abs 62, p. 102.

2. *Joana Silva, Hugo Santos, Margarida Gama-Carvalho, Luísa Romão* (2016) *Analysis of the translatome by ribosome profiling in colorectal cancer*. Poster presented at the *Ciência 2016 – Science and Technology in Portugal Meeting*, Lisbon (Portugal). 4-6 July. Virtual presentation.

6. Attendance at Scientific Meetings, Courses and Training Visits

6.1. Conferences

I participated in the following conferences during this year:

1. **Ciência 2016 – Science and Technology in Portugal Meeting**. Lisbon (Portugal). 4-6 July.

6.2. Courses and Training Visits

I participated in the following courses and training visits during this year:

1. *Advanced Lecture Course on Systems Biology*, organised by *SYSBIO 2016 Course Office*, which took place in the Hotel Grauer Bär, Innsbruck (Austria). February 28 – March 5.

2. *NGS Data Analysis, RNAseq, ChIPseq*, organized by *The Gulbenkian Training Programme in Bioinformatics*, which took place in the Instituto Gulbenkian de Ciência, Lisbon (Portugal). March 29 – April 1.

7. Future Work and Timeline

Within the next year of doctoral work, we propose:

1) Objective 1 – *Analysis of the translatome of normal (NCM460) and CRC (HCT116) cell lines by ribosome profiling.*

   In addition to the analysis of HCT116 Ribo-Seq data already in course, we intent to apply the same computational analysis to a normal control of non-neoplasic colorectal cells, when the Ribo-Seq data is available.

2) Objective 2 – *Study of the main biological function and cellular pathways of the proteins differentially expressed in normal and CRC cells.*

   Objective 2.1 – *Bioinformatics and gene ontology analysis.*

   The same bioinformatics and gene ontology analyzes will be performed to promising targets selected from the computational analysis carried out in the Objective 1. Other bioinformatics tools will be further included in this analysis as they will be relevant for characterization of the proteins of interest.

   Objective 2.2 – *Experimental analyzes to confirm the relevance of such proteins in CRC tumorigenesis process*

   To follow with the experimental validation of the results achieved from the previous bioinformatics and gene ontology analyzes for the ABC proteins and other possible targets, we will perform: (ii) Western blot (WB) to obtain the levels of protein expression in the previous used cell lines; (iii) immunofluorescence (IF) to understand the subcellular localization; and (iv) immunoprecipitation (IP) assays to confirm protein-protein interactions with implementing a silencing approach by RNA interference (RNAi) or overexpression assays.

   Furthermore, to evaluate the role of such proteins in CRC tumorigenesis process we will promote their overexpression or knockdown and access the effects at the levels of: (i) cell proliferation by cell counting via MTT assays and cell cycle progression by flow cytometry analysis with propidium iodide staining; (ii) cell survival by the analysis of programmed cell death (apoptosis) evasion with TUNEL labeling protocols; and (iii) migratory and invasive proprieties via wound healing (evaluated by time lapse microscopy) and matrigel assays in CRC monolayers [37]. Additionally, if appropriated, we can also expose the colorectal cell lines to different cell stimuli, such as hypoxia and starvation.

3) Objective 3 – *Study of the non-coding mRNA sequences (AltORFs in 5'UTR and/or 3'UTR) and non-annotated ORFs that appear differentially translated in normal and CRC cells.*

Objective 3.1 – *Determination of the relevance of such regions in translational regulation.*

To study the role of uORFs as regulators of the translation process, we will construct reporter plasmids carrying the cDNA sequence of the 5'UTR containing the uORF(s) of interest fused to the *Firefly* luciferase (Fluc) ORF. The CRC cell lines (HCT116, Caco-2 and SW480) and non-neoplasic NCM460 cells will be transiently co-transfect with this construct and with another expressing *Renilla* luciferase (Rluc) ORF, as an internal control. Luciferase activity will be measured by luminometry assays and the mRNA levels quantified by RT-qPCR to normalize the results and obtain the translation efficiencies, as previously described in Barbosa and Romão [35] and Onofre and co-workers [27]. The results will be compared to those obtained using constructs with a non-functional uORF. For that, a site directed mutagenic process will be carry out in the initiation codon (AUG or others) of the uORF(s) and their analysis will be performed as above.

Objective 3.2 – *Analysis of the biological function of small peptides encoded by AltORFs (uORFs and/or others) and their relevance in CRC progression.*

To analyze if the uORFs of interest can in fact be translated, we will clone reporter plasmids in which the uORF sequence is fused in frame with the FLuc ORF. A site-directed mutagenic approach will be applied to mutate the stop codon of the uORF and the start codon of the FLuc ORF, thus translation will only occur if ribosomes initiate translation at the uORF initiation codon. The detection of an N-terminally extended luciferase protein will be accessed by WB.

To understand if the translational regulation is dependent on the peptide sequence of the uORF, we will use constructs such as the ones described in the Objective 3.1, but modified by shifting the reading frame of the uORF to get a different encoded amino acidic sequence. These constructs will be transiently transfected in the colorectal cell lines and luciferase activity and mRNA levels will be measured as in the Objective 3.1. The role of these uORFs encoded peptides in the CRC tumorigenesis will be evaluated as described in the Objective 2.2 and using plasmid constructs as the ones in the Objective 3.1.

4) Objective 4 – *Validation of the obtained results in primary cell cultures of CRC and in normal colon mucosa.*

If possible, considering potential time and cost constrains in the course of this doctoral plan, we intend to validate the better hits (translatable uORFs with a biological role in CRC cells) in mice and CRC tumors. For that we will use transgenic mice expressing the uORF of interest in which we will establish CRC xenografts. The expression levels of the uORF encoded peptide will be measured and its effects in CRC xenografts will be assessed at the levels of cell survival and proliferation as described in the Objective 2.2. Moreover, the correlation between the levels of proteins and/or the uORFs encoded peptides with the CRC stages will be determined by statistical analysis. Significant correlations will be analyzed, if possible, in terms of quantification of the abundance of such proteins and/or small peptides in CRC tumors. Thus, the knowledge acquired from these studies will help

in the identification of possible diagnostic, prognostic and/or therapeutic markers in CRC.

5) Thesis writing and manuscript(s) preparation.

Year		2015						2016						2017						2018					
Objectives / Months		January February	March April	May June	July August	September October	November December	January February	March April	May June	July August	September October	November December	January February	March April	May June	July August	September October	November December	January February	March April	May June	July August	September October	November December
Objective 1							■	■	■	■	■	■	■												
Objective 2	Objective 2.1								■	■	■	■	■	■											
	Objective 2.2								■	■	■	■	■	■	■	■	■								
Objective 3	Objective 3.1									■	■	■	■	■	■	■									
	Objective 3.2													■	■	■	■	■	■	■	■	■			
	Objective 4																■	■	■	■	■	■	■		
	Objective 5															■	■	■	■	■	■	■	■	■	■

Figure 3 - Timeline of the objectives proposed to study the translatome of colorectal cancer cells by ribosome profiling.

8. References

[1] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* **136**: E359-386.

[2] Kheirelseid EAH, Miller N, Chang KH, Nugent M, Kerin MJ (2013). Clinical applications of gene expression in colorectal cancer. *J Gastrointest Oncol* **4**: 144-157.

[3] Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJS, Jackson SE, Wills MR, Weissman JS (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep* **8**: 1365-1379.

[4] Jackson R, Standart N (2015). The awesome power of ribosome profiling. *RNA* **21**: 652-654.

[5] Ingolia NT (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* **15**: 205-213.

[6] Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**: 218-223.

[7] Vanderperre B, Lucier JF, Bissonnette C, Motard J, Tremblay G, Vanderperre S, Wisztorski M, Salzet M, Boisvert FM, Roucou X (2013). Direct Detection of Alternative Open Reading Frames Translation Products in Human Significantly Expands the Proteome. *PLoS One* **8**: e70698.

[8] Siegel RL, Miller KD, Jemal A (2015). Cancer Statistics, 2015. *CA Cancer J Clin* **65**: 5-29.

[9] Siegel RL, Naishadham D, Jemal A (2013). Cancer Statistics, 2013. *CA Cancer J Clin* **63**: 11-30.

[10] Xiao WH, Qu XL, Li XM, Sun YL, Zhao HX, Wang S, Zhou X (2015). Identification of commonly dysregulated genes in colorectal cancer by integrating analysis of RNA-Seq data and qRT-PCR validation. *Cancer Gene Ther* **22**: 278-284.

[11] Kitahara O, Furukawa Y, Tanaka T, Kihara C, Ono K, Yanagawa R, Nita ME, Takagi T, Nakamura Y, Tsunoda T (2001). Alterations of Gene Expression during Colorectal Carcinogenesis Revealed by cDNA Microarrays after Laser-Capture Microdissection of Tumor Tissues and Normal Epithelia. *Cancer Res* **61**: 3544-3549.

[12] ACS - American Cancer Society (2014). Learn About Cancer, Colon/Rectum Cancer, Detailed Guide, Treatment of colon cancer by stage. http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-treating-by-stage-colon (Last visit: 24.9.2015)

[13] Bianchini M, Levy E, Zucchini C, Pinski V, Macagno C, Sanctis P, Valvassori L, Carinci P, Mordoh J (2006). Comparative study of gene expression by cDNA microarray in human colorectal cancer tissues and normal mucosa. *Int J Oncol* **29**: 83-94.

[14] Pei H, Zhu H, Zeng S, Li Y, Yang H, Shen L, Chen J, Zeng L, Fan J, Li X, Gong Y, Shen H (2007). Proteome Analysis and Tissue Microarray for Profiling Protein Markers Associated with Lymph Node Metastasis in Colorectal Cancer. *J Proteome Res* **6**: 2495-2501.

[15] Kim IJ, Lim SB, Kang HC, Chang HJ, Ahn SA, Park HW, Jang SG, Park JH, Kim DY, Jung KH, Choi HS, Jeong SY, Sohn DK, Kim DW, Park JG (2007). Microarray Gene Expression Profiling for Predicting Complete Response to Preoperative Chemoradiotherapy in Patients with Advanced Rectal Cancer. *Dis Colon Rectum* **50**: 1342-1353.

[16] Nome T, Thomassen GOS, Bruun J, Ahlquist T, Bakken AC, Hoff AM, Rognum T, Nesbakken A, Lorenz S, Sun J, Barros-Silva JD, Lind GE, Myklebost O, Teixeira MR, Meza-Zepeda LA, Lothe RA, Skotheim RI (2013). Common Fusion Transcripts Identified in Colorectal Cancer Cell Lines by High-Throughput RNA Sequencing. *Transl Oncol* **6**: 546-553.

[17] The Cancer Genome Atlas Network (2012). Comprehensivemolecular characterization of human colon and rectal cancer. *Nature* **487**: 330-337.

[18] Barbosa C, Peixeiro I, Romão L (2013). Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLoS Genet* **9**: e1003529.

[19] Calvo SE, Pagliarini DJ, Mootha VK (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Nati Acad Sci U S A* **106**: 7507-7512.

[20] Arribere JA, Gilbert WV (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res* **23**: 977-987.

[21] Wethmar K, Smink JJ, Leutz A (2010). Upstream open reading frames: molecular switches in (patho)physiology. *Bioessays* **32**: 885-893.

[22] Morris DR, Geballe AP (2000). Upstream Open Reading Frames as Regulators of mRNA Translation. *Mol Cell Biol* **20**: 8635-8642.

[23] Onofre C, Tomé F, Barbosa C, Silva AL, Romão L (2015). Expression of human hemojuvelin is tightly regulated by two upstream open reading frames that respond to iron overload in hepatic cells. *Moll Cell Biol* **35**: 1376-1389.

[24] Poyry TAA, Kaminski A, Jackson RJ (2004). What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame?. *Genes Dev* **18**: 62-75.

[25] Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genetics* **36**: 1073-1078.

[26] McGlincy NJ, Tan LY, Paul N, Zavolan M, Lilley KS, Smith CWJ (2010). Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. *BMC Genomics* 11: 565-583.

[27] Young DJ, Guydosh NR, Zhang F, Hinnebusch AG, Green R (2015). Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3'UTRs In Vivo. *Cell* **162**: 872-884.

[28] Ingolia NT, Lareau LF, Weissman JS (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity of Mammalian Proteomes. *Cell* **147**: 789-802.

[29] Gao X, Wan J, Liu B, Ma M, Shen B, Qian SB (2015). Quantitative profiling of initiation ribosomes *in vivo*. *Nat Methods* **12**: 147-153.

[30] Liu B, Han Y, Qian SB (2013). Cotranslational Response to Proteotoxic Stress by Elongation Pausing of Ribosomes. *Mol Cell* **49**: 453-463.

[31] Guo H, Ingolia NT, Weissman JS, Bartel DP (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**: 835-840.

[32] Crappé J, Ndah E, Koch A, Steyaert S, Gawron D, De Keulenaer S, De Meester E, De Meyer T, Van Criekinge W, Van Damme P, Menschaert G (2013). PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. *Nucleic Acids Res* **43**: e29.

[33] Hlavata I, Mohelnikova-Duchonova B, Vaclavikova R, Liska V, Pitule P, Novak P, Bruha J, Vycital O, Holubec L, Treska V, Vodicka P, Soucek P (2012). The role of ABC transporters in progression and clinical outcome of colorectal cancer. *Mutagenesis* **27**: 187-196.

[34] McGrath PT (2011). Characterizing cDNA Ends by Circular RACE. *Methods Mol Biol* **772**: 257-265.

[35] Barbosa C, Romão L (2014). Translation of the human erythropoietin transcript is regulated by an upstream open reading frame in response to hypoxia. *RNA* **20**: 594-608.

[36] Ruiz de los Mozos I, Vergara-Irigaray M, Segura V, Villanueva M, Bitarte N, Saramago M, Domingues S, Arraiano CM, Fechter P, Romby P, Valle J, Solano C, Lasa I, Toledo-Arana A (2013). Base Pairing Interaction between 5'- and 3'-UTRs Controls icaR mRNA Translation in *Staphylococcus aureus*. *PLoS Genet* **9**: e1004001.

[37] Moniz S, Martinho O, Pinto F, Sousa B, Loureiro C, Oliveira MJ, Moita LF, Honavar M, Pinheiro C, Pires M, Lopes, JM, Jones C, Costello JF, Paredes J, Reis RM, Jordan P (2013). Loss of WNK2 expression by promoter gene methylation occurs in adult gliomas and triggers Rac1-mediated tumour cell invasiveness. *Hum Mol Genet* **22**: 84-95.