**Universita’ degli studi di Parma**

Dipartimento di Scienze Chimiche, della vita e della Sostenibilità Ambientale

Dottorato in Biotecnologie e Bioscienze XXXII ciclo

**Relazione attività di ricerca – terzo anno**

**Dottoranda:** Simona Cantarella

**Tutor:** Prof. Giorgio Dieci

**Coordinatore:** Prof. Simone Ottonello

**Relazione attivita’ di ricerca**

**Research project:** Exploring the potential of *Alu* RNAs as novel epigenetic players and molecular biomarkers in cancer biology

**Introduction**

*Alu* retrotransposons are repetitive sequences that constitute the 10% of the human genome. These elements have often been referred to as “junk DNA” for their apparent lack of a biological function and the tight epigenetic silencing in normal cell conditions. However, evidence such as *Alu* upregulation in response to Adenovirus 5 (Ad5) infection suggests their involvement in cell functions. We hypothesized that Adenovirus early e1a protein might have a role in *Alu* activation. Indeed, it is very well established that e1a induces global chromatin rearrangements of the host genome through interactions with the tumor suppressor Rb and the histone acetyltransferase p300. e1a is also capable to bind the Swi/Snf related helicase p400 protein, but the functional role of this interaction has not been clarified yet.

The aim of this PhD project is to dissect the molecular mechanism of *Alu* dysregulation in response to Ad5 infection and to investigate new possible roles of *Alu* RNAs in gene expression.

During the first two years of my PhD we performed *Alu* transcriptome profiling in e1a-infected human fibroblasts, revealing a 6-fold average induction of *Alu* elements compared to mock infected cells. Further, the expression of *AluSp* sequence was analyzed in cells infected with e1a mutants that are not capable to bind Rb, p300 or p400: from RT-qPCR data we could verify that e1a-p400 binding mutant is the least effective in increasing *AluSp* expression.

In the meanwhile, possible new roles of *Alu* RNAs were analyzed in fibroblasts and HeLa cells that stably overexpress the sequences *AluSq2* or *AluSx*. Stable cell lines were generated through lentiviral infection, and gene expression was analyzed through RNA-sequencing. The upregulation of genes belonging to cell cycle pathways could be detected in human fibroblasts, while no effect was observed in HeLa cells.

**Results**

The last year of my PhD focuses on a deeper analysis of *Alu* expression profile data and on the experimental validation of transcriptome analysis of fibroblasts overexpressing *AluSq2* or *AluSx*.

*Alu* transcriptome data analysis revealed the expression of 1265 *Alu* sequences in e1a and mock infected cells (overexpressed *Alu*s), while 778 sequences were detected in e1a cells but not in mock samples (activated *Alu*s). Interestingly, the overexpressed *Alu*s showed a higher read counts value than the activated *Alu*s. This suggests that those *Alu* sequences that are expressed at a basal level in mock samples are also efficiently expressed in response to e1a infection.

In order to confirm the hypothesized mechanism of e1a-p400 involvement in the activation of *Alu* sequences, the expression of the not yet investigated element *AluSc* was analysed in cells infected with e1a mutants that are not capable to bind Rb, p300 or p400. RT-qPCR data confirmed the trend towards a lower expression of *AluSc* in e1a-p400 binding mutant. Next, we asked whether this effect is specific to *Alu* elements or generally involves RNA Polymerase III genes. Therefore, we tested the expression of 7SL and U6 genes in wild type e1a and e1a-p400 binding mutant, observing no changes in gene expression. This result was also confirmed by RNA-seq data analysis.

RT-qPCR data should also be confirmed at a genome-wide level, therefore we asked the construction of total RNA-libraries of fibroblasts infected with e1a binding mutants. RNA-seq libraries were synthetized by collaborators at The University of California, Los Angeles (UCLA), and sequencing was performed by Fulgent Genetics (Temple City, California). RNA-seq data are currently under analysis.

Lastly, we asked whether the induction of *Alu* sequences in response to e1a could be influenced by the expression of nearby coding genes. Using the online software GREAT, we analysed the presence of expressed genes within 100 kb from the centre of an expressed *Alu* body. The analyses show a significant enrichment of *Alu* expressed sequences nearby expressed coding genes, and this enrichment increases when considering the genomic association between Differentially Expressed (DE) coding genes and DE *Alu*s.

Overall, the data suggest that *Alu* overexpression in response to e1a infection could depend on the existence of a basal permissive transcription state, which could be influenced by the epigenetic landscape of nearby active coding genes.

It is not clear what is the functional activation of *Alu* sequences in response to virus infection. To this purpose, the coding transcriptome of fibroblasts that stably overexpress *AluSq2* or *AluSx* was analysed in more detail using the Ingenuity Pathway Analysis software (IPA, Qiagen). A significant enrichment of pathways belonging to cell cycle was confirmed in *Alu*s-overexpressing cells, but not in control samples. We also performed upstream regulator analysis to identify potential molecules that could modulate the expression of the detected DE genes, revealing FOXM1, CDKN1A and YAP1 as possible candidates. Moreover, in order to unravel the mechanism underlying *Alu* RNA modulation of cell transcriptome, we analysed the possible localization of *Alu* elements within the 3’ UTR of Differentially Expressed genes. We found an enrichment of *Alu* sequences in the 3’ UTR of DE genes detected in *AluSq2* overexpressing cells, suggesting the involvement of an intermolecular base-pairing between *Alu* sequences in the modulation of gene expression. Lastly, miRNA targeting of *Alus* and DE genes was investigated. Interrogating the databases miRBase and miRDB, we found that *AluSx* hosts target sequences for two miRNAs, which in turn were found to potentially target three DE genes. Interesting, the gene PRR11 belongs to cell cycle pathway and is found upregulated in our RNA-seq dataset.

Lastly, cell cycle modulation induced by *Alu* RNA was experimentally confirmed by flow cytometry. Cell synchronization was obtained through serum withdrawal for 48h, and cell cycle was afterwards released adding fresh serum-added medium for 24h. FACS analyses were performed through Propidium Iodide staining, and cells overexpressing *AluSq2* and *AluSx* were found significantly enriched in the S-phase, compared to cells overexpressing a control sequence or an empty vector.

The results presented here are published in Cantarella et al, 2019 {Cantarella, 2019 #38}.