Karin Fleischhanderl; Martina Fondi

Experimental validation of candidates of tissue-specific and CpG-island-mediated alternative polyadenylation in mouse

108 - Biotechnologie

Abstract

Keywords:

Alternative polyadenylation, CpG islands, DNA methylation, Epigenetic, Intragenic promoters

Einleitung

Introduction

The definition of epigenetics has evolved over time and is now typically understood as mitotically or meiotically heritable changes in gene expression that are not caused by changes in the DNA sequence (Holliday, R. 1994). We know that genes are not all active at the same time. Several epigenetic mechanisms including DNA methylation and histone modifications are used to regulate gene expression in cells. Gene expression is the process of transcription of information encoded in DNA into RNA, before translation into protein (Bentley, D.L. 2014). DNA methylation modifies the function of DNA and typically acts to repress gene transcription if located in a gene promoter. Different cells have different methylation patterns, contributing to transcript diversity.

It is still unclear how approximately 25,000 mammalian genes give rise to the nearly 250,000 observable transcript isoforms. By analysing RNA-seq data from mouse tissues, a potential mechanism was identified that could significantly contribute to the complexity of the mammalian transcriptome. The transcriptome is partly generated through alternative splicing and polyadenylation (poly(A)). These processes can be influenced by intragenic CpG islands (CGIs) through poly(A) site selection regulation. Alternative splicing is a co-transcriptional event that aims at increasing transcript diversity. About 95% of multi-exonic genes undergo alternative splicing and generate at least two different transcript isoforms by differential exon inclusion (Pan, Q. *et al.* 2008).

The mouse and human genomes contain nearly the same number of CGIs. CGIs can be separated into CGIs associated with annotated transcription start sites (TSS), as they coincide with promoters of annotated genes, and represent about 50%. The remaining half is split into CGIs either within gene bodies (intragenic) or between gene bodies (intergenic) (Deaton, A.M. / Bird, A. 2011). The intragenic and intergenic CGIs are also called 'orphan' CGIs. There is evidence that about 40% of 'orphan' CGIs are associated with transcriptional initiation and represent novel promoters, as the majority recruits RNA Polymerase II (RNAPII) (Illingworth, R.S. et al. 2011). Many orphan CGIs are active promoters in a tissue-specific manner (Deaton, A.M. / Bird, A. 2011).

Based on the CpG density and DNA methylation state, the genome can be divided into two different categories: The bulk of the genome, which is CpG-poor and predominantly methylated (~80%), and CpG islands (CGIs). Through alternative polyadenylation, more than one transcript isoform can arise depending on multiple poly(A) sites, as present in about 70% of human genes. (Derti, A. et al. 2012). CpG islands were considered as a characteristic of housekeeping genes, but it is now apparent that CGIs are also utilised as promoters in tissue-specific genes (Blackledge, N.P. / Klose, R. 2014). CGIs are about 1000bp long, GC-rich, CpG-rich and predominantly non-methylated compared to an almost completely methylated CpG-poor genomic landscape. As mentioned, 70% of annotated gene promoters are associated with CGIs and almost all CGIs are sites of transcription initiation.

Currently, it is generally understood, that the choice of poly(A) sites is related to tissue type and developmental stage (Tian, B. 2013). One method of alternative poly(A) site control is through trans-acting processing factors, that are dependent on cell-type specific activity. One example are the immunoglobulin genes (Edwalds-Gilbert, G. 1997). However, there is evidence shown at two imprinted genes H13 (Wood, A.J. et al. 2008) and Herc3 (Cowley, M. 2012), that epigenetic modifications can act in *cis* to regulate poly(A) site selection.

Professor Oakey's group has identified a novel murine imprinted locus, located on mouse chromosome 2 which contains two protein-coding genes: *H13* and *Mcts2*. *Mcts2* is a protein-coding retrogene located within the fourth intron of the multi-exonic gene *H13* ('host gene'). A CGI (gDMRs) within *H13* is differentially methylated between the maternal and paternal alleles and includes the promoter for *Mcts2*. The 'host gene' *H13* generates multiple transcripts differing at the 3' ends by using alternative poly(A) sites. The polyadenylation sites are used in an allele-specific manner depending on the DNA methylation and promoter activity state of the CpG island promoter of *Mcts2*. The *Mcts2* promoter is unmethylated and thus active on the paternal allele where truncated *H13* transcripts arise, terminating at polyadenylation sites upstream of *Mcts2*. On the maternal allele, the *Mcts2* promoter is methylated and thus silenced and inactive, which allows the utilisation of downstream *H13* polyadenylation sites. There are several poly(A) sites located within the *H13* gene, generating at least five different transcripts. Three isoforms arise from the maternal allele (H13a, H13b, and H13c) and two from the paternal allele (H13d and H13e). H13d and H13e are only generated if *Mcts2* is expressed (Wood, A.J. et al. 2008), indicating that transcription from an internal site could be responsible for transcripts terminating upstream of the CGI.

Aims of the study

Hypothesising that the mechanism of poly(A) site selection/alternative polyadenylation may operate genome-wide in a tissue-specific manner, 30 primary mouse tissues were analysed leading to the identification of about 1,700 gene loci with preliminary evidence supporting the hypothesis. At these loci, the tissue-specific activity of intragenic CGIs is correlated with changes in pre-mRNA processing and

specifically how the transcript is spliced and/or polyadenylated. Therefore I hypothesised that Intragenic CGIs are regulated by DNA methylation and their activity is associated with an increase of premature polyadenylation of the host gene. Based on this hypothesis I tried to answer the following two research questions:

(1) Can the RNA-seq-based observations be recapitulated by qRT-PCR assays applied to a subset of 10 chosen candidate loci?

(2) Is a differentially methylated CGI the epigenetic mechanism which regulates the CGI activity?

Methods

Next Generation Sequencing short read data only provide limited direct evidence for the hypothesised mechanism. Therefore, transcriptional activity and DNA methylation were explored and characterised at 10 representative candidate loci in detail. The objective of the project was to select 10 candidate loci, based on inspection of the RNA-seq data and other data on the UCSC genome browser and carry out qRT-PCR to quantify transcript abundance at the selected loci. This provided data to answer my first research question.

Furthermore, bisulfite PCR & direct Sanger sequencing, followed by cloning & Sanger sequencing of the intragenic CGI at those loci was carried out. The results were compared between high-activity and low-activity CGI transcriptional states. This provided data to answer my second research question.

Results and Discussion

The in-depth analysis resulted in an ambiguous pattern compared to the imprinted genes. The characterised loci can be roughly divided into two different groups. They were separated into loci without a difference in upstream terminating transcripts and loci with differences in upstream terminating transcripts. Three loci (Nacc2, Zadh2 and Hoxa) differed in terms of upstream terminating transcripts between at least two tissues. These tissues varied regarding the DNA methylation of the CGI. However, the methylation varied in a broad range (Figure 1B). At another three loci (Elf2, Adck2-Ndufb2 and Brf1), the tissues showed no difference in the amount of upstream terminating transcripts. Interestingly, except for one tissue pair, at these loci, all CGIs revealed unmethylated CpG islands (Figure 1A). For two loci in each group, the datasets were incomplete and therefore, it was assumed they behave similarly.

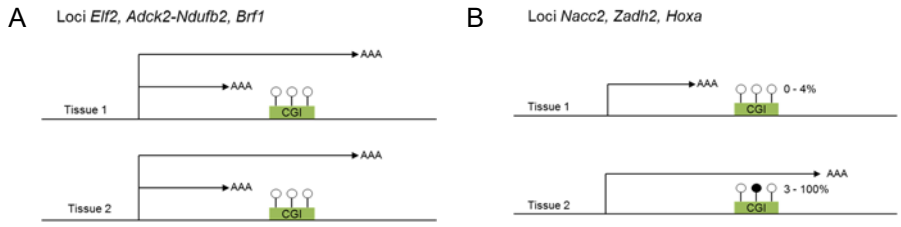

Figure 1: 6 out of 10 characterised candidate loci. The green bar indicates CGI, empty circles show unmethylated CpGs, filled circles display methylated CpGs. AAA indicates transcript termination.

According to our hypothesis, intragenic CGIs are regulated by DNA methylation and their activity is associated with an increase of premature polyadenylation of the host gene. We asked two questions: i) if the RNA-seq-based observations can be recapitulated by qRT-PCR on a subset of 10 chosen candidate loci; ii) if differential methylation regulates the activity of CGIs. The first question cannot be

answered with a yes or no. As discussed above, some loci were consistent with the RNA-seq data, whereas some loci differed. The same applies to the second question. There were only four tissue pairs where a differentially methylated CGI had an effect on the amount of upstream terminating transcripts. Further experiments are required to investigate the influence of intragenic promoter on gene expression.

The overall pattern after extensive experiments was not as expected and frequently not completely consistent with the RNA-seq data. The data suggest an overall trend towards increased transcriptional elongation through (more) methylated CGIs, but that trend is by no means definitive. Furthermore, it is crucial to match the mouse strain, tissues, and developmental stages. The optimisation of the experiments would also include working with single isolated cell types or cell lines. However, in spite of the unexpected results, this project shed new light on the role of intragenic CGI methylation in relation to alternative polyadenylation, as a mechanism of tissue-specific gene expression regulation. Therefore, future work is necessary to understand the complex mechanism of epigenetic marks influencing gene expression. Future work will be conducted to examine the effect of intragenic promoters on alternative polyadenylation with two different approaches. Additionally, a *Mcts2* knock-in and a knock-out construct will be generated. The expectation is that more transcripts will terminate downstream if the CpG island of *Mcts2* is removed. A second approach is a relocation of an active *Mcts2* (knock-in) in an intron of *Fam13c*. *Fam13c* is the targeted gene and located at chromosome 10 with similar features to H13, but lacks intragenic CGIs and it does not exhibit imprinted expression. Both approaches will help to investigate the influence of intragenic promoter on gene expression.

References

Holliday, R. Epigenetics: an overview. *Developmental Genetics* **15,** 453–457 (1994).

Wood, A.J. *et al.* Regulation of alternative polyadenylation by genomic imprinting. *Genes & Development* **22,** 1141–1146 (2008).

Tian, B., Hu, J., Zhang, H. & Lutz, C.S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Research* **33,** 201–212 (2005).

Derti, A. *et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Research* **22,** 1173–1183 (2012).

Tian, B. & Manley, J.L. Alternative cleavage and polyadenylation: the long and short of it. *Trends in Biochemical Sciences* **38,** 312–320 (2013).

Edwalds-Gilbert, G., Veraldi, K.L. & Milcarek, C. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Research* **25,** 2547–2561 (1997).

Cowley, M., Wood, A.J., Böhm, S., Schulz, R. & Oakey, R.J. Epigenetic control of alternative mRNA processing at the imprinted Herc3/Nap1l5 locus. *Nucleic Acids Research* **40,** 8917–8926 (2012).

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* **40,** 1413–1415 (2008).

Deaton, A.M. *et al.* Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Research* **21,** 1074–1086 (2011).

Illingworth, R.S. *et al.* Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genetics* **6,** e1001134 (2010).

Deaton, A.M. & Bird, A. CpG islands and the regulation of transcription. *Genes & Development* **25,** 1010–1022 (2011).

Blackledge, N.P. & Klose, R. CpG island chromatin. *Epigenetics* **6,** 147–152 (2014).