



Impact of Alternative Promoter Usage in Normal and Disease Gene Regulation

Anders Kvist

Home Department: Department of Cell and Molecular Biology, Division for Developmental Biology, Lund University

External Lab: Swegene cDNA Microarray Resource Centre, Department of Oncology, Lund University

Doctor's Degree: 2001-07-31, Department of Ecology, Lund University

Project Description: The results from large scale EST and full length mRNA sequencing projects and of directed studies of specific genes have shown that many human and mouse genes use alternative promoters in their regulation. Alternative promoter usage can affect gene expression in diverse ways. Different promoters may direct production of different mRNA isoforms, directly through different transcription start sites or indirectly by promoter-directed exon inclusion. The resulting transcripts may encode different protein isoforms or may differ only in their 5' untranslated regions, affecting mRNA stability and translation efficiency. Other genes use promoters of different strength to direct tissue specific expression. Tight regulation is essential for accurate gene function, and loss of this control may have serious phenotypic effects. There are several examples of disease associated alternative promoter usage, for example promoters specifically activated in a number of cancer forms.

The current project aims (i) to identify genes regulated by alternative promoters in the human and mouse genomes and to characterize these genes with regard to function and evolutionary history, (ii) to investigate disease association of these genes using computational approaches, and (iii) to find tissue and disease association of alternative promoter usage, using microarray hybridization.

Impact of alternative promoter usage in normal and disease gene regulation

Swegene Postdoc Program Annual report 2004

Anders Kvist

Home department: Patrik Medstrand lab, Section for Developmental Biology, Dept of Cell & Molecular Biology, Lund University.

External lab: Åke Borg lab, cDNA Microarray Centre, Dept of Oncology, Lund University

Summary

The first year of my Swegene postdoc has been scientifically stimulating and productive. Our project has gained attention from biotech companies and collaborations have been initiated with Operon Biotechnologies Inc and are underway with a second company. My main work has focused on the bioinformatic identification of alternative first exons and promoter regions as outlined in my research plan. This work has been done in the Medstrand lab. The results of the analysis have allowed us to design promoter-specific oligonucleotide-microarrays and my postdoc project is now entering its second phase, where the bioinformatic tools and databases developed, and the novel microarrays, are applied to investigate the role of alternative promoters in normal and disease gene regulation. This part of the project is on-going and is possible due to close collaboration between the Medstrand and Borg labs. We have also started new collaborative projects within the field of regulation of gene expression by non-coding RNAs. I have become involved in projects studying the expression microRNAs and antisense RNAs by combining bioinformatic analyses with expression studies using both microarrays and large scale cloning (of microRNAs). This first year has given me more than I had hoped for, both in terms of scientific results, the development and broadening of my own skills as a bioinformatician and biologist and the network of contacts I have developed. I am very grateful for the opportunity the Swegene postdoc program has given me.

Introduction

Background

Mammalian genes are often regulated by alternative promoters active in different tissues or at different developmental stages. Different promoters may direct production of different mRNA isoforms, directly through different transcription start sites or indirectly by promoter-directed exon inclusion. The resulting transcripts may encode different protein isoforms or may differ only in their 5' untranslated regions, affecting mRNA stability and translation efficiency. Promoters may also differ in strength to direct different levels of expression. Tight regulation is essential for accurate gene function, and loss of this control may have serious phenotypic effects. There are several examples of disease associated alternative promoter usage, for example promoters specifically activated in a number of cancer forms.

Objective

The objective of my project is to systematically analyze the human genome to investigate the impact of multiple promoter usage in normal and disease gene regulation. Novel experimental and computational approaches and a combination of bioinformatics, comparative genomics and microarray expression analysis are applied to address these questions. Elucidating the transcriptional regulatory network is a challenge of the post-genomic era. Understanding how genetic variation and genetic changes affect gene regulation may provide a basis for explaining susceptibility to some complex genetic diseases and other human disorders.

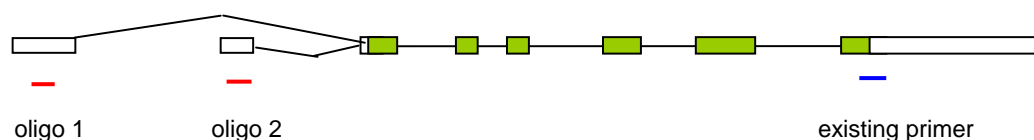
Specific aims

The project aims (i) to identify genes regulated by alternative promoters in the human and mouse genomes and to characterize these genes with regard to function and evolutionary history, (ii) to investigate disease association of these genes using computational approaches, and (iii) to find tissue- and disease- association of alternative promoter usage, using microarray hybridization.

Results of the past year

Main project

- 1) *Bioinformatic identification of alternative first exons and promoter regions (Medstrand lab)*. Only a small fraction of all human promoters is known, primarily due to incomplete data on true transcriptional starts sites (which define the upstream region as the promoter). Starting from from all publicly available full-length and partial human mRNA sequences (including ESTs), I have identified alternative promoter regions, many of them novel, for a large fraction of human genes. A recent review estimates that 18 % of human genes have alternative promoters (Landry JR et al. 2003. Trends Genet 19:640-8). Our results put this figure over 30%. I am currently working on a paper describing the characteristics and genomic organization of alternative 5' exons and promoters of human genes.
- 2) *Design of promoter-specific oligonucleotide-microarrays (Medstrand and Borg labs)*. The results of the bioinformatic analysis have allowed us to design entirely novel oligonucleotide-microarray slide sets specific for transcripts derived from unique promoters. Oligos have been designed to be specific to the 5' exon regions uniquely identifying alternative promoter usage (see fig. below, red bars indicate promoter-specific oligos)



To limit expenses we had to reduce the initial number of oligos to be synthesized and spotted onto the arrays. Therefore we have restricted the analysis to three sets of genes: (1) genes that use a retroelement sequence as physiological promoter, (2) genes potentially related to schizophrenia / bipolar disorders (based on microarray studies), and (3) genes identified from gene expression data as being differentially expressed in estrogen receptor positive (ER+) and negative (ER-) breast cancers.

The three sets of genes correspond to three planned applications of the microarray slides in research projects by the collaborating partners.

Our oligo set can be used to assess the impact of alternative promoters in gene regulatory functions – a critical component in tissue-specific gene regulation which is largely unexplored. It will also allow us to specifically investigate if certain promoter regions are aberrantly regulated in disease tissue, e.g. breast cancer and brain tissue from patients affected by schizophrenia / bipolar disorders.

The microarray design project is done in cooperation with Operon Biotechnologies Inc. Operon contributes to the project both financially and with their experience and expertise in oligo design.

Side tracks

- 3) *Bioinformatic identification of alternative splicing.* As part of the process of identifying alternative promoters, I have had to develop a method to reliably identify alternative spliced transcript variants of human genes. A database of alternative splice variants, exons and introns is being constructed based on the results of this analysis.
- 4) *Antisense transcription.* I have collaborated with Carlos Rovira (Borg lab) on a project to use cDNA microarrays to identify antisense transcription of human genes. My part of the project has been to mine evidence of antisense transcription from public databases to verify the results of the microarray pilot experiments. As part of the process of identifying alternative splicing, I will also identify antisense transcripts and these will form part of the database described above.
- 5) *Cloning and identification of short RNAs.* I have collaborated with Carlos Rovira on a project to clone and identify short RNAs. As part of this project I have supervised (together with Carlos Rovira and Patrik Medstrand) the bioinformatic part of a masters thesis project (examensarbete) by Helena Persson. Helena cloned and identified short RNAs in the BT-474 breast cancer cell line. Several types of short, non-coding RNAs that regulate gene expression are known. We found evidence of several types of one class of small RNAs, the microRNAs, in the BT-474 cell line. MicroRNAs are produced as approximately 22 nucleotide fragments from longer doublestranded precursors that are cleaved by the RNase III endonuclease Dicer. We also identified other putative products of Dicer cleavage and are currently conducting experiments to verify these findings. One paper about short RNAs in the BT-474 cell line will be submitted soon.
- 6) *Design of microRNA microarrays.* In another collaboration between the Medstrand and Borg labs we are designing a novel oligonucleotide-microarray slide set for microRNA expression profiling. A biotech company has expressed interest in this project and a collaboration is under development. We expect to run the first hybridizations on these slides this autumn.
- 7) *Design of a novel phylogenetic clustering algorithm.* Together with Daniel Svenback (PhD student, Medstrand lab) and Patrik Medstrand, we have developed a new phylogenetic clustering algorithm. The clustering algorithm allows for variable evolutionary rates within clades and uses a criterion combining low variance within clusters with good separation between clusters to find a stable clustering. The aim has been to produce a clustering as close as possible to that which would have been found by human curation. A paper is being written about this algorithm.

Plans for the next year

Main project

1) *Conservation of alternative promoter usage between mouse and human.*

Mammalian genomes are highly conserved and it is now generally thought that gene regulation is the main factor responsible for the diversity of form and function between species. Promoter regions play a key role in the regulation of gene expression and our results indicate that alternative promoter usage may be an important mechanism in this regulation. Therefore, a timely and important question to investigate is the conservation of promoter regions between mammalian species. The conservation of alternative promoters between orthologous genes in mouse and human will shed light on how promoter usage has evolved and diversified in mammalian evolution. Questions to investigate include differences in promoter conservation between constitutive and alternative promoters and between recently diverged gene families and those of older origin. In combination with expression profiling (e.g. from EST data), the evolution of tissue and developmental stage specific promoter usage can also be addressed.

2) *Experimental validation of alternative promoter regions.* The promoter-specific microarrays we have designed will be used to validate the results of our bioinformatic analysis of alternative promoters. We will run hybridizations with a selection of normal tissue samples and cell lines. Selected genes will be analyzed using standard procedure (Q-PCR, transfection experiments etc) by us and in collaboration.

3) *Expression profiling of promoter usage in breast cancer and schizophrenia / bipolar disorders.* We will use our promoter-specific microarrays to analyse promoter usage differences in normal and pathogenic conditions. Microarray slides will be simultaneously hybridized with randomly labeled mRNA extracted from different tissue samples and from a combination of reference samples. Promoter usage differences indicating aberrant promoter regulation in pathogenic conditions will be further investigated and confirmed using standard procedures. The set of breast cancer related genes represented on the microarray slides include many ER regulated genes (containing EREs in their promoter). Several genes known to be regulated by ER are still expressed at high levels in ER- tumors. The transcription of these genes could be driven by other transcription factors acting on alternative promoters. Identifying these genes could be a good starting point to study and identify the presence of new transcription factors of importance in development of the aggressive ER- breast cancers.

4) *Database of alternative promoters and alternative splicing.* I am in the process of building a database of alternative splicing, alternative promoters and antisense transcripts of human genes. This database will allow convenient access and crossreferences to a number of gene related resources. It will also allow easy extraction of sequences unique to alternative splice forms, specific exons or splices and extraction of promoter sequences and intronic sequences for further analyses. A future plan is to develop a web interface to this database.

Side tracks

5) *MicroRNA profiling of breast cancer samples using microRNA microarrays.* During this autumn we will run pilot experiments with our microRNA microarrays and if all goes well we will apply them to profile the expression of microRNAs in breast cancer.