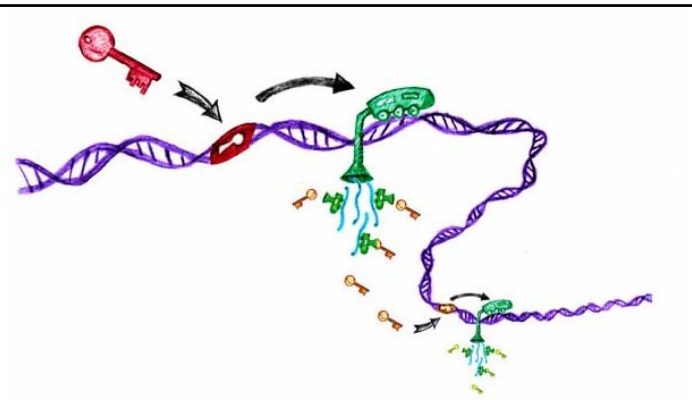


Promoter sequence analysis of genes, differentially expressed in sheep, following a nematode parasite challenge

Shivashankar Nagaraj
CSIRO Livestock Industries

Poster 69

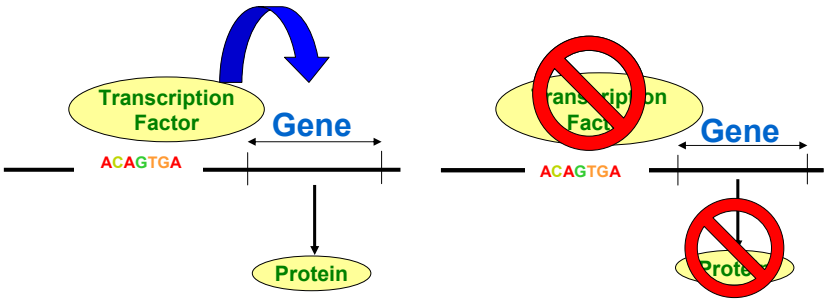
CSIRO

Outline

- Background
- Previous experimental work
- Goals
- Results
- Future directions

Regulation of Transcription

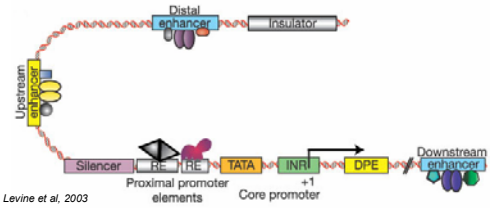
- Regulatory proteins, binding to the up-stream region of a gene act as either promoters or suppressors of transcription



- Regulation of transcription is time, cell and tissue specific

Promoter architecture

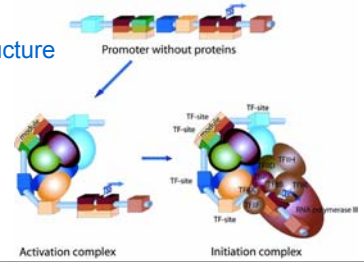
- Organization of promoter motifs represents a "footprint" of the transcriptional regulatory mechanisms
- Complex transcriptional control modules



Levine et al, 2003

Active promoters have a unique 3-dimensional structure

Changing the order or spacing of transcription factor binding sites (TFBS) can change the overall structure of the promoter and thus affect transcription



Werner et al, 2003

Previous work

Available online at www.sciencedirect.com
ScienceDirect
International Journal for Parasitology 38 (2008) 431–442
www.elsevier.com/locate/ijpara

Gastrointestinal nematode challenge induces some conserved gene expression changes in the gut mucosa of genetically resistant sheep

Aaron Ingham [a,*], Antonio Reverter [a], Ross Windon [b], Peter Hunt [b], Moira Menzies [a]

[a] CSIRO Livestock Industries, Queensland Bioscience Precinct, 306 Carmody Road, St. Lucia, Qld 4067, Australia
[b] CSIRO Livestock Industries, Armidale, NSW, Australia

Received 22 May 2007; received in revised form 18 July 2007; accepted 19 July 2007

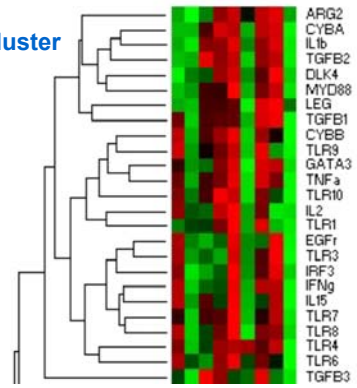
Abstract

Sheep have a varying ability to resist infection with gastrointestinal nematodes. This ability is due in part to genetic differences that exist between individuals. In order to define these differences we have used real-time PCR to quantify gene expression responses in the gut mucosal surface of genetically resistant and susceptible sheep, following a nematode challenge. Expression profiles were determined in response to two different nematode species, *Haemonchus contortus* and *Trichostrongylus colubriformis*, and in divergent sheep originating from two different genetic backgrounds. Results show that the response generated differs between resistant and susceptible animals and is further impacted by the origin of the sheep and nematode species used for challenge. However, some conserved features of a response mounted by a resistant or a susceptible animal were identified. Genes found to be more abundantly expressed in resistant animals include markers of an early inflammatory response, several Toll-like receptors (*TLR2, 4, 9*) and free radical producing genes (*DUOX1* and *NOS2A*). Conversely, genes differentiating susceptible animals indicate a prolonged response and development of a chronic inflammatory state, characterised by elevated expression of members of the NF-κβ signalling pathway (*IKBKB* and *NFKBIA*) together with delayed expression of regulatory markers such as *IL2RA* (CD25), *IL10* and *TGFβ2*. While multiple nematode response pathways were identified, the identification of conserved aspects of the response which associate with resistance provides evidence that alternative nematode control strategies, such as breeding for resistant animals, may be feasible.

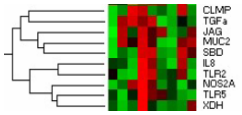
Data

- **Sets of differentially expressed (DE) genes in resistant and susceptible animals following nematode challenge**

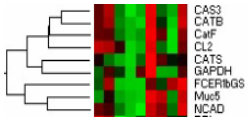
TLR cluster



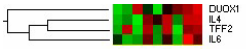
XDH cluster



Cathepsin Cluster



DUOX Cluster

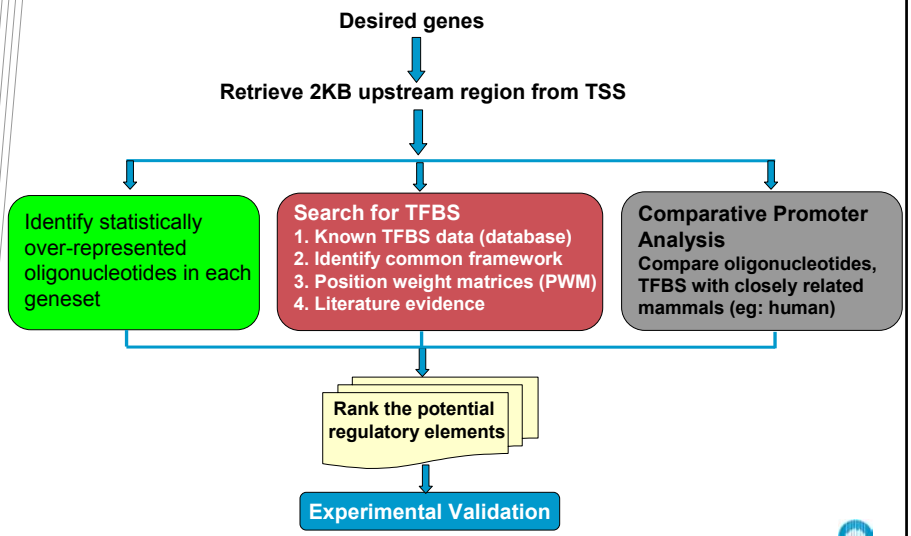


Co-expression → **Transcriptional co-regulation** → **Common regulatory elements**

Goals

- To conduct a comprehensive analysis of promoter sequences for DE genes
- To identify potential TF binding sites involved in regulation
  * To find common motifs shared by the majority of TFs
  * Are there any global regulators (TFs) that turns on DE genes in sheep?
- To benchmark currently available methods and establish a broad protocol for regulatory sequence analysis
- To extensively use human promoter sequence data (well studied) using comparative genomics approaches
- To apply this protocol to other related projects at CSIRO as a starting point to understand transcription regulation

Promoter sequence analysis - schema



Over-represented consensus binding sites discovered in 22 Toll Like Receptor (TLR) pathway genes

Motif 1: TCAGAAA

P-value : 3.5e-05

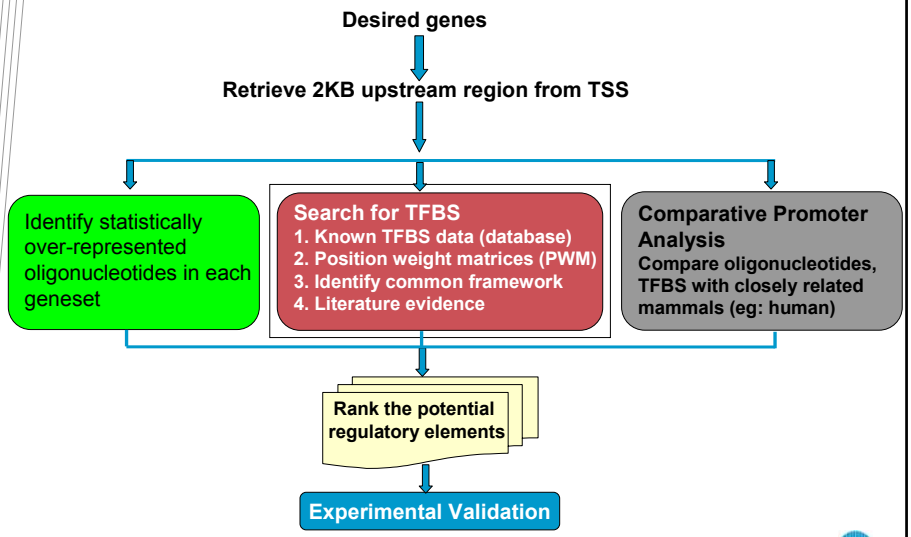
Motif 2: AGAGAAA

P-value : 3.5e-06

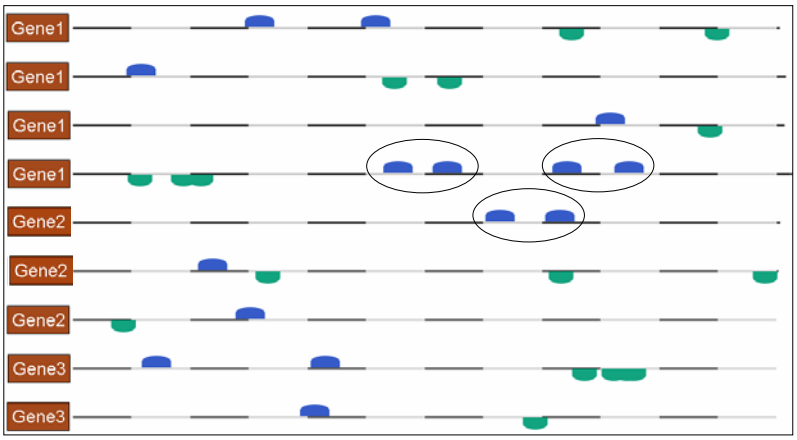
Motif 3: GGGAGGA

P-value : 2.1e-05

Promoter sequence analysis - schema



Common patterns (frameworks) of TFBS identified in promoters from a set of three differentially expressed genes in TLR pathway



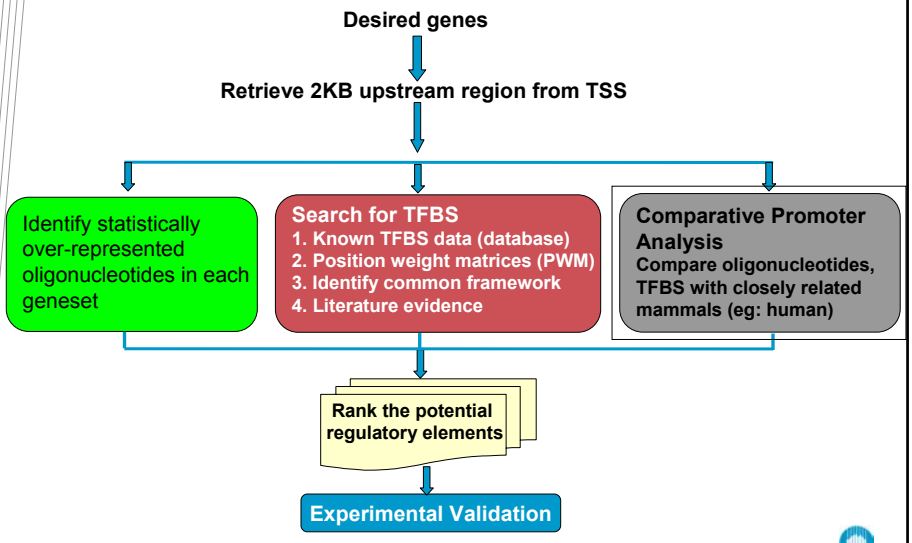
TFBS common to all sequences in a set of three promoters from TLR pathway

Selected hits from matrix search for transcription factor binding sites

Query Gene	Position (strand)	Sequence	Transcription Factor Name
Resistant: TLR2	150 (-)	aCACTTga	Nkx2-5 (homeobox gene)
Resistant: DUOX1	948 (-)	acaAACAAac	FOXD3(fork head)
Susceptible: IKBKB	686 (-)	atGGAAAttcc c	NF-kappaB
Susceptible: TGFB2	121 (+)	gatAACGGtc	v-Myb

Literature evidence in human ⇨

Promoter sequence analysis - schema



Promoter region– conserved in mammalian species

List of TFs

V$AHRR V$NKXH V$HOXF V$STAT O$VTBP V$ETSF V$EVI1 V$DICE V$EGRF V$P53F V$THAP V$SP1F
V$MYT1 V$FKHD V$GKLF V$BRNF V$PIT1 V$HNF1 V$SORY V$DMRT V$E2FF V$MYBL V$GATA V$NOLF
V$XBBF

Common TF matches located in aligned regions

TFBS1

```
alignment position      1801......  1811......  1821......  1831.....
Region_8_Human 1653 CCTGTACCTC ATCTACTCCC AGGTCCTCTT CAAGGGCCAA
Region_8_Rhesus 1653 CCTGTACCTC ATCTACTCCC AGGTCCTCTT CAAGGGCCAA
Region_8_Mouse  1546 GTTGTACCTT GTCTACTCCC AGGTTCTCTT CAAGGGACAA
Region_8_Rat    1486 GCTGTACCTT ATCTACTCCC AGGTTCTCTT CAAGGGACAA
Region_8_Cow    1653 GCTTTACCTC ATCTACTCAC AGGTCCTCTT CAGGGGCCAA
                     ********** ********** ********** **********
                     ********** ********** ********** **********
```

```
alignment position      2051......  2061......  2071......  2081......  2091.....
Region_8_Human 1903 CGGCCCGACT ATCTCGACTT TGCCGAGTCT GGGCAGGTCT ACTTTGGGAT
Region_8_Rhesus 1903 CTGCCCGACT ATCTCGACTT TGCCGAGTCT GGGCAGGTCT ACTTTGGGAT
Region_8_Mouse  1793 CTGCCCAAGT ACTTAGACTT TGCGGAGTCC GGGCAGGTCT ACTTTGGAGT
Region_8_Rat    1733 CTGCCCAAGT ACTTAGACat caCGGAGTCC GGGCAGGTCT ACTTTGGAGT
Region_8_Cow    1903 CTGCCGGACT ACCTGGACTA TGCCGAGTCT GGGCAGGTCT ACTTTGGGAT
                     ********** ********** ********** ********** **********
                     ********** ********** ********** ********** **********
```

TFBS2

TFBS3

```
alignment position      2751......  2761......  2771......  2781.....
Region_8_Human 2528 CAGACATGTT TTCCGTGAAA ACGGAGCTGA ACAATAG---
Region_8_Rhesus 2541 CAGACATGTT TTCTGTGAAA ACGGAGCTGA ACAATAG---
Region_8_Mouse  2394 CAGACATGTT TTCTGTGAAA ACGGAGCTGA GCTGtcc---
Region_8_Rat    2377 CCAACATGTT TTCTGTGAAA ACGGAGCTAA ACTaccagct
Region_8_Cow    2526 CAGATGTGTT TTCTGTGAAA ACGGAGCTGA ACTGCAG---
                     ********** ********** ********** ********
                     ********** ********** ********** ********
```

Conclusions

- We believe that the promoter motifs identified in this study have regulatory potential
- Binding of transcription factors to these motifs might explain the differential expression observed following nematode challenge
- Functional variation in these motifs is therefore likely to contribute to an individual's ability to resist infection

Future directions

Computational

- To establish a roadmap for regulatory sequence analysis based on gene expression data

Experimental validation for initial results

- *In-vitro* experiments such as Electrophoretic Mobility Shift Assay (EMSA)
- *In-vivo* experiments (ChIP-Seq experiments)
- The second set of RT-PCR data spanning 102 genes

Acknowledgments

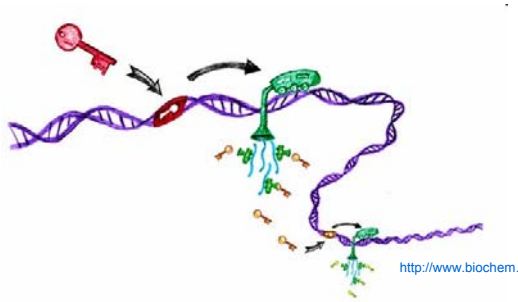
Antonio Reverter
CSIRO Livestock Industries

Aaron Ingham
CSIRO Livestock Industries

- OCE Post-doctoral fellowship
- SheepGenomics

Transcription Factors (TF) - Factsheet

- Found in all living organisms
- Approximately 10% of genes in the mammalian genome code for TFs
- Have affinity for short, degenerate DNA sequences (5-15 bp)
- Contain one or more DNA binding domains (DBDs)
- Mutated TF genes have been shown to cause numerous diseases (Eg: Haemophilia B Leyden)
- Potential Targets for several therapeutic drugs (Eg: breast cancer)



http://www.biochem.arizona.edu/

 Butt et al 1995, Latchman et al 2000

Challenges

- **Ovine genome incomplete, therefore we have chooses bovine as starting point**
- **Several physical binding sites comes up in predictions as a physical TFBS is found every 10 to 15 bps throughout any mammalian genome**
- **A single isolated TF binding site carries no function**
- **TFs work through complexes which are represented on sequence level through sets of TF binding sites in certain distance relationship and orientation ->promoter frameworks**
- **Therefore, we have to decipher gene anatomy landmarks , rather than just promoter motifs**
- **So wee are talking promoter model/ framework and not individual TFBS**

Why are we doing what we are doing ?

- Promoter features provide clues to gene function that are not obvious from the protein sequence alone
- The organizational features of promoters derived from promoter sequences contain information about the spatial and temporal 'functional context' of expression.
- Genes having similar expression patterns contain common motifs in their promoter regions. Thus, a common set of TFs is likely to control these genes
- Understanding the gene regulation in livestock species is still in primitive stages , so Starting point .. **

(Vilo *et al.*, 2000, Klok *et al.* 2002)

Different tissues and families of genes differentially use distinct types of promoters.

Tamoxifen – drug for breast cancer

Gene Cluster	Number of genes per cluster	Sequence	observed occurrences	expected occurrences	P-value
TLR Cluster	25	TCAGAAA	25	9.81	3.5e-05
XDH cluster	10	AGAGAAA	16	3.88	3.3e-06
XDH cluster	10	ccgcgc	18	5.29	1.1e-05
Cathepsin cluster	9	cccccg	17	4.42	4.2e-06