Comparative Sequence Analysis of the Cytokine Gene Cluster on Human Chromosome 5q31
Identification of a potent regulator for IL-4, IL-5 and IL-13

rVISTA
Identifying CNSs involved in transcriptional regulation
Clustering of Transcription Factor Binding Sites
Functional SNPs in Non-coding DNA

Scanning BACs for regulatory elements

1 Mb Human 5q31 Cytokine Gene Cluster
23 genes

Interleukins: Not homologous
Similar 3D structure
Functionally related
Several co-expressed in Th2 Cells

Conserved Noncoding Sequences (CNSs)
Conserved Human/Mouse Sequences in 1 MB Region

270 Transcribed

411

141 Non-Transcribed

Introns 59%

< 1 k from gene 8%

> 1 kb from gene 33%

> 40 bp and >90%

> 60 bp and >80%

> 70 bp and >70%

VISTA: http://www-gsd.lbl.gov/vista/
1 Mb Human 5q31 Cytokine Gene Cluster

>80% and >100bp

113 CNSs

**IL-4  IL-13  IL-5  IRF-1  GMCSF  IL-3**

15 highly conserved non-coding elements were analyzed by zoo PCR.

Most of the elements (red) were highly conserved in mice and human and at least in one other vertebrate.
CNS-1 is present in multiple vertebrates

1. Cow (86%)  Dog (81%)  Rabbit (73%)
2. Genomic position conserved in dog and baboon
3. Single copy in human genome

2 Hypersensitive Sites Map on CNS-1

Mast-cell specific enhancer

Enhancer
Breed with Cre Recombinase transgenic mice

Human YAC A94G6

pRS406.CNS-1loxP

A94G6.CNS-1wt

A94G6.CNS-1del
4 CNS-1\textsuperscript{wt} transgenic lines and 4 CNS-1\textsuperscript{del} transgenic lines.
CNS-1 inhibits Mouse IL4 Expression in Human 5q31 Transgenic Mice in naïve T-cells

Hu IL-4 and IL-13 Expression is reduced in CNS-1 del mice

- TH2
  - hiL-4
  - hiL-13
- TH1
  - 50% reduction
  - 70% reduction

- IL-4
  - Δ CR1
  - IFNγ
Human IL-5 Expression is also reduced in CNS-1\textsuperscript{del} mice

Summary

- By performing human/mouse comparative studies we have identified numerous highly conserved noncoding sequences that are present in multiple organisms.

- CNS-1 is a potent regulator of IL-4, IL-5 and IL13

- Conserved Noncoding sequences are biologically important
Identify conserved noncoding sequences involved in transcriptional regulation in a high-throughput manner

Potential Functions of Conserved Sequences

Chicken β-globins Region
Bell AC. et al, Science 291(5503), 2001
Transcription Factors involved in Th-2 cell specific cytokine expression

Distribution of GATA-3 Transcription factors across 12 kb region
182 Sites (15 Sites/kb)
Regulatory VISTA (rVISTA)

1. Identify transcription factor binding sites for each sequence using library of matrices (TRANSFAC)
2. Identify aligned sites using VISTA
3. Identify conserved sites using dynamic shifting window

<table>
<thead>
<tr>
<th>Species</th>
<th>Ikaros-2</th>
<th>NFAT</th>
<th>Ikaros-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>TGATTCTGAGCAGGAAAGGGCCCATGACCAAGCATTATGAAATCCAGGAACAAATTTTCTACTTTACACCTTCCTTCCTCCCTCTCTCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>TGATTCTGAGCAGGAAAGGGCCCATGACCAAGCATTATGAAATCCAGGAACAAATTTTCTACTTTACACCTTCCTTCCTCCCTCTCTCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dog</td>
<td>TGATTCTGAGCAGGAAAGGGCCCATGACCAAGCATTATGAAATCCAGGAACAAATTTTCTACCTTTACACCTTCCTTCCTCCCTCTCTCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cow</td>
<td>TGATTCTGAGCAGGAAAGGGCCCATGACCAAGCATTATGAAATCCAGGAACAAATTTTCTACTTTACACCTTCCTTCCTCCCTCTCTCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit</td>
<td>TGATTCTGAGCAGGAAAGGGCCCATGACCAAGCATTATGAAATCCAGGAACAAATTTTCTACTTTACACCTTCCTTCCTCCCTCTCTCT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

20 bp dynamic shifting window
>80% ID
<table>
<thead>
<tr>
<th>Transcription Factor</th>
<th>Total Sites</th>
<th>Conserved</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATA-3</td>
<td>36286</td>
<td>1393</td>
<td>3.8</td>
</tr>
<tr>
<td>Ikaros</td>
<td>15376</td>
<td>657</td>
<td>4.3</td>
</tr>
<tr>
<td>NFK-B</td>
<td>571</td>
<td>35</td>
<td>6.1</td>
</tr>
<tr>
<td>AP-1</td>
<td>32386</td>
<td>1612</td>
<td>5.0</td>
</tr>
<tr>
<td>STAT</td>
<td>1511</td>
<td>51</td>
<td>3.4</td>
</tr>
<tr>
<td>YY1</td>
<td>17564</td>
<td>505</td>
<td>2.9</td>
</tr>
<tr>
<td>SP1</td>
<td>9023</td>
<td>383</td>
<td>4.2</td>
</tr>
<tr>
<td>NF1</td>
<td>9984</td>
<td>350</td>
<td>3.5</td>
</tr>
<tr>
<td>NFY</td>
<td>4580</td>
<td>92</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Total ~4.0 %
Clustering of Transcription Factor Binding Sites
Transcription Factor Binding Sites

1. **In vivo crosslinking**
2. **DNA shearing**
3. **Immunoprecipitate chromatin with transcription factor specific antibody**
4. **Hybridize precipitated DNA to DNA chip**
5. **PCR immunoprecipitated DNA**
GATA-3 Responsive Genes

GATA-3:
Th2 cell specific expression
(a/t)GATA(a/g)

IL-4  GM-CSF
IL-5  IL-3
IL-13 IRF-1

Predicted GATA-3 sites
Filtering of predicted GATA-3 binding sites

1. Identify all GATA-3 sites
2. Identify aligned sites using VISTA
3. Identify conserved GATA-3 sites
4. Identify conserved GATA-3 pairs

GATA-3 GATA-3

Human: CCTGTTTCCTCTCAGCATTTATCTTGGGCTTCCTGTGACCC
AATCTTCAAGTGACTT
TTATCAAC
GTTATCAGAAGATCAGA
TTGTCTCACCCAAACACGTTTATGGCTCT

Mouse: CCTGTTTCCTCTCAGCATTTATCT-GGGCTTCCTGTGACCC
AATCTTCAAGTTACTT
TTATCAAC
GTTATCAGAACATCACT
TTGTCTTACCCAAACACGTTTATGGCTCT

Dog: CCTGTTTCCTCTCAGCATTTATCTTGGGCTTCCTGCGACCC
AACCTTCAAGTCACTT
TTATCAGT
GTTATCAGAACATCACG
TCGTCTTACCCAAACACCGTTTATGGCTCT

20 bp dynamic shifting window >80% ID

Conserved GATA-3 pairs

<table>
<thead>
<tr>
<th>GDF-9</th>
<th>APRIL</th>
<th>KIF3A</th>
<th>IL-4</th>
<th>IL-13</th>
<th>RAD50</th>
<th>IL-5</th>
<th>IRF-1</th>
<th>OCTN2</th>
<th>OCTN1</th>
<th>RHL</th>
<th>PKH</th>
<th>GM-CSF</th>
<th>IL-3</th>
<th>LACS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
**IL-5 promoter:**
*functional GATA-3 pair*

-80bp

\[\text{tattaggtgtccttatctgtagtaatta}\]

-50bp

Siegel MD. et al. 1995, JBC

---

**Putative Functional SNPs in Noncoding DNA**

Surveying SNPs across highly conserved CNSs

52 CNSs (≥80% Identity; ≥120 bp)
GM-CSF, IL-3 and Septin
~31 kb/individual
17 Africans & African Americans

23 Northern Europeans

Total SNPs 84

<table>
<thead>
<tr>
<th>Nucleotide Diversity</th>
<th>SNPs</th>
<th>$\theta_s (x10^{-4})$</th>
<th>SNPs</th>
<th>$\theta_s (x10^{-4})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding</td>
<td>2</td>
<td>2.47 ± 1.82 3</td>
<td>3</td>
<td>3.45 ± 2.14</td>
</tr>
<tr>
<td>Intron</td>
<td>34</td>
<td>5.66 ± 1.92 22</td>
<td>22</td>
<td>3.41 ± 1.18</td>
</tr>
<tr>
<td>Intergenic</td>
<td>31</td>
<td>5.42 ± 1.86 17</td>
<td>17</td>
<td>2.76 ± 1.00</td>
</tr>
<tr>
<td>UTRs</td>
<td>3</td>
<td>6.64 ± 4.16 2</td>
<td>2</td>
<td>4.12 ± 3.02</td>
</tr>
<tr>
<td>CNS</td>
<td>24</td>
<td>5.20 ± 1.85 18</td>
<td>18</td>
<td>3.63 ± 1.30</td>
</tr>
</tbody>
</table>

Frequency of SNPs in CNSs similar to SNPs in coding sequences

Standard Neutral Model holds for human 5q31 genetic interval

Africans more diverse

Low Minor allele frequencies for Population specific SNPs - rare SNPs
~3 Million SNPs

Which ones to genotype?

Prioritize SNPs in Transcription Factor Binding Sites for association studies

Human Maj
CTTCCCTGTGACCCAATCTTCAAGTGACTTTTATCAACGTTATCAGAAGATCAGATTGTCTCAC

Human Min
CTTCCCTGTGACCCAATCTTCAAGTGACTTTTATAACGTTATCAGAAGATCAGATTGTCTCAC

Mouse
CTTCCCTGGACCCAATCTTCAAGTGACTTTTACAGTGTTATCAGAACATCACGTCGTCTTAC

GATA  Major Allele      Minor Allele
SNP Binding Affinity     Binding Affinity
C/T  1       0.75

14/77 SNPs occurred in transcription factor binding sites
Scanning Human BACs for Regulatory Elements
Gabriela Loots  
ggloots@lbl.gov

**SNP analysis**  
Poulabi Banerjee

**Computational Analysis**  
Inna Dubchak  
Ivan Ovcharenko  
Lior Pachter  
Jody Schwartz  
Edward M. Rubin


**VISTA:** http://www-gsd.lbl.gov/vista/