Pyrosequencing

Alix Groom
• high-throughput CpG methylation analysis platform
• real-time, sequence-based detection and quantification
• % methylation at multiple adjacent CpG sites
• 80-100 bases sequenced per assay
• 2ug DNA analyse 3 assays/regions of interest
• 24 or 96 samples/assays run at a time
DNA extraction
Bisulphite modification
PCR
Single strand template generation
Pyrosequencing
Bisulphite modification

GGTCAGTGAC\textsuperscript{mCG}

\begin{align*}
\text{C} & \xrightarrow{\text{Bisulphite conversion}} \text{mC} \\
\text{U} & \xrightarrow{\text{PCR amplification}} \text{mC} \\
\text{U} & \xrightarrow{\text{Pyrosequencing analysis}} \text{mC}
\end{align*}

GGT\textsuperscript{mCG}

\begin{align*}
\text{TAGTGAT/CG}
\end{align*}
Polymerase chain reaction

Target sequence

primer annealing

extension

copies of target sequence
Single strand generation

- Denature PCR product
- Capture PCR product with streptavidin beads
- Release single stranded DNA
- Anneal sequencing primer
- Biotin labelled PCR product
Pyrosequencing chemistry

DNA polymerase
ATP sulfurylase
Luciferase
Apyrase

5’ T A G T A G G 3’

3’ T A G T A G G 5’

Polymerase

DNA polymerase
ATP sulfurylase
Luciferase
Apyrase

DNA_{(n)} + dNTP

APS
Luciferin

DNA_{(n+1)} + PPI

Polymerase

Light

Time

Luciferase

ATP
Light

Sulfurylase

APS + PPI
ATP

Apyrase

deNTP

dNDP + dNMP + phosphate

ATP

Apyrase

ADP + AMP + phosphate

Nucleotide sequence

G T - A G G
Nucleotide added

G T A G G
Pyrosequencing timeline

Day 1
- **Bisulphite modification**
  - Sample preparation: 15min-1hr
- DNA-reagent incubation: 3hrs

Day 2
- **PCR**
  - PCR set up: 30min-1hr
  - PCR cycles: 1.5hr
  - Agarose gel: 1hr

Day 3
- **Single strand template generation**
  - Sample prep: 30min-1hr
- **Pyrosequencing**
  - Pyrosequencing run: 10min-1.5hr
Pyrosequencing applications

- Gene specific methylation analysis identified target loci through gene expression studies, literature search, methylation arrays etc.

- Global methylation analysis
  - Methylation of repetitive elements
    - LUMA
Case study

Illumina 27K/450K top hit → verify → Pyrosequencing
VeraCode
Sequenom

Check
no SNPs in probe
which DNA strand CpG site is measured
Case study

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest
Case study: promoter region

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest

- genomatix Gene2Promoter software
Case study: promoter region

- **identify if CpG in promoter region**
- **identify CGI within/adjacent to promoter**
- **capture sequence 4000bp flanking CGI**
- **identify TFBM that contain CpG**
- **select CpG of interest**

- genomatix Gene2Promoter software

---

You submitted 1 genes/keywords/phrases:

Looking for **PTPN20B** in Homo sapiens ...

- protein tyrosine phosphatase, non-receptor type 20B
  (PTPN20B / GeneID: 26095 / GXL_228069) on chromosome 10 of human

- Additionally list orthologous genes in output

  - Continue  
  - Reset Form
identify if CpG in promoter region

identify CGI within/adjacent to promoter

capture sequence 4000bp flanking CGI

identify TFBM that contain CpG

select CpG of interest

• genomatix Gene2Promoter software
identify if CpG in promoter region

identify CGI within/adjacent to promoter

capture sequence 4000bp flanking CGI

identify TFBM that contain CpG

select CpG of interest

- genomatix Gene2Promoter software

Case study:

promoter region

...
Case study: CpG Island

- Identify if CpG in promoter region
- Identify CGI within/adjacent to promoter
- Capture sequence 4000bp flanking CGI
- Identify TFBM that contain CpG
- Select CpG of interest

- UCSC Genome Bioinformatics
- CpG Island explorer
Case study: CpG Island

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest

UCSC Genome Bioinformatics
http://genome.ucsc.edu/
Case study: CpG Island

- UCSC Genome Bioinformatics
  http://genome.ucsc.edu/

**Human BLAT Results**

**BLAT Search Results**

<table>
<thead>
<tr>
<th>ACTIONS</th>
<th>QUERY</th>
<th>SCORE</th>
<th>START</th>
<th>END</th>
<th>QSIZE</th>
<th>IDENTITY</th>
<th>CHRO</th>
<th>STRAND</th>
<th>START</th>
<th>END</th>
<th>SPAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>browser details YourSeq</td>
<td>612</td>
<td>1</td>
<td>612</td>
<td>612</td>
<td>100.0%</td>
<td>10</td>
<td>48827823</td>
<td>612</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>608</td>
<td>1</td>
<td>612</td>
<td>612</td>
<td>99.7%</td>
<td>10</td>
<td>46610902</td>
<td>612</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>585</td>
<td>1</td>
<td>612</td>
<td>612</td>
<td>97.9%</td>
<td>10</td>
<td>4934868</td>
<td>592</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>29</td>
<td>267</td>
<td>300</td>
<td>612</td>
<td>96.9%</td>
<td>17</td>
<td>79420091</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>26</td>
<td>568</td>
<td>595</td>
<td>612</td>
<td>100.0%</td>
<td>X</td>
<td>129627074</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>25</td>
<td>133</td>
<td>158</td>
<td>612</td>
<td>100.0%</td>
<td>17</td>
<td>18393131</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>24</td>
<td>131</td>
<td>159</td>
<td>612</td>
<td>96.2%</td>
<td>8</td>
<td>38554758</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>22</td>
<td>445</td>
<td>468</td>
<td>612</td>
<td>87.0%</td>
<td>7</td>
<td>104654081</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>22</td>
<td>552</td>
<td>574</td>
<td>612</td>
<td>100.0%</td>
<td>7</td>
<td>63223715</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>22</td>
<td>552</td>
<td>574</td>
<td>612</td>
<td>100.0%</td>
<td>7</td>
<td>63223715</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>20</td>
<td>122</td>
<td>141</td>
<td>612</td>
<td>100.0%</td>
<td>9</td>
<td>94406683</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>20</td>
<td>264</td>
<td>283</td>
<td>612</td>
<td>100.0%</td>
<td>1</td>
<td>43896606</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browser details YourSeq</td>
<td>20</td>
<td>135</td>
<td>154</td>
<td>612</td>
<td>100.0%</td>
<td>6</td>
<td>166825789</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Case study: CpG Island

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest

- UCSC Genome Bioinformatics
  - http://genome.ucsc.edu/
Case study: CpG Island

- UCSC Genome Bioinformatics http://genome.ucsc.edu/

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest

CpG Island Info

- Position: chr10:48827593-48828126
- Band: 10q11.22
- Genomic Size: 534
- View DNA for this feature (hg19/Human)
- Size: 534
- CpG count: 47
- C count plus G count: 372
- Percentage CpG: 17.6%
- Percentage C or G: 69.7%
- Ratio of observed to expected CpG: 0.79
Case study sequence capture

- UCSC Genome Bioinformatics
  http://genome.ucsc.edu/

- CGI Chr10:48 827 593-48 828 126
- 4000bp downstream 48 823 593
- 4000bp upstream 48 832 126
Case study: sequence capture

- UCSC Genome Bioinformatics
  http://genome.ucsc.edu/

Chr10: 48823593 - 48832126

- Identify if CpG in promoter region
- Identify CGI within/adjacent to promoter
- Capture sequence 4000bp flanking CGI
- Identify TFBM that contain CpG
- Select CpG of interest
Case study

sequence capture

- UCSC Genome Bioinformatics
  http://genome.ucsc.edu/

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest
Case study sequence capture

- UCSC Genome Bioinformatics
  http://genome.ucsc.edu/

Identify if CpG in promoter region

Identify CGI within/adjacent to promoter

Capture sequence 4000bp flanking CGI

Identify TFBM that contain CpG

Select CpG of interest
Case study: sequence capture

- Identify if CpG in promoter region
- Identify CGI within/adjacent to promoter
- Capture sequence 4000bp flanking CGI
- Identify TFBM that contain CpG
- Select CpG of interest

UCSC Genome Bioinformatics
http://genome.ucsc.edu/

Extended DNA Output:

>chr10:48823593-48823126

GCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGTGTCCTTCCGCCTCCAGATTCAGT

Case study sequence capture
Case study

sequence capture

- UCSC Genome Bioinformatics
  http://genome.ucsc.edu/

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest
Case study: transcription factor binding module

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest

- TFBM defined 2+ TFBS in defined order and orientation
- HNF1
- GATA
- TTGTACTAA CGATATGCCATGCTA

- UCSC TFBS: single binding factor information
- JASPAR: http://jaspar.cgb.ki.se
- TransCompel
- Genomatix ModelInspector
Case study: transcription factor binding module

- Identify if CpG in promoter region
- Identify CGI within/adjacent to promoter
- Capture sequence 4000bp flanking CGI
- Identify TFBM that contain CpG
- Select CpG of interest

Genomatix
http://www.genomatix.de/
Case study

transcription factor binding module

identify if CpG in promoter region

identify CGI within/adjacent to promoter

capture sequence 4000bp flanking CGI

identify TFBM that contain CpG

select CpG of interest

Genomatix

http://www.genomatix.de/
Case study: transcription factor binding module

- Identify if CpG in promoter region
- Identify CGI within/adjacent to promoter
- Capture sequence 4000bp flanking CGI
- Identify TFBM that contain CpG
- Select CpG of interest

Genomatix
http://www.genomatix.de/

A total of 23 matches was found in 1 sequences.
### Identify if CpG in promoter region

- Case study: transcription factor binding module

### Identify CGI within/adjacent to promoter

### Capture sequence 4000bp flanking CGI

### Identify TFBM that contain CpG

### Select CpG of interest

## Genomatix

http://www.genomatix.de/

---

### Inspecting sequence PTPN20B_6_shelf (1 - 2000):

**Model: NRSF_GATA_01 (161 - 190 (+))**

<table>
<thead>
<tr>
<th>Matrix element</th>
<th>Str</th>
<th>Sequence</th>
<th>Core sim.</th>
<th>Mat. sim.</th>
<th>Distance to next element</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSN4BF1INFA.01</td>
<td>(+)</td>
<td>CTCAGGCTTGCTGGAATGCTTATT</td>
<td>1.000</td>
<td>0.911</td>
<td>11 bp</td>
</tr>
<tr>
<td>VSN4BF1INFA.02</td>
<td>(+)</td>
<td>CTGAGGATATTACG</td>
<td>1.000</td>
<td>0.911</td>
<td>—</td>
</tr>
</tbody>
</table>

**Model: SMAD_E2F_01 (170 - 149 (-))**

<table>
<thead>
<tr>
<th>Matrix element</th>
<th>Str</th>
<th>Sequence</th>
<th>Core sim.</th>
<th>Mat. sim.</th>
<th>Distance to next element</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSN4SMAD01SMAD</td>
<td>(+)</td>
<td>GGAAGAAGAC</td>
<td>0.750</td>
<td>0.710</td>
<td>8 bp</td>
</tr>
<tr>
<td>VSB2F1E2F14.01</td>
<td>(+)</td>
<td>CTGAGGATATTACG</td>
<td>1.000</td>
<td>0.963</td>
<td>—</td>
</tr>
</tbody>
</table>

**Model: GATA_NKXH_01 (274 - 245 (-))**

<table>
<thead>
<tr>
<th>Matrix element</th>
<th>Str</th>
<th>Sequence</th>
<th>Core sim.</th>
<th>Mat. sim.</th>
<th>Distance to next element</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSN4GATAVSN4GATA.02</td>
<td>(-)</td>
<td>GAAATTTGCTTCAA</td>
<td>1.000</td>
<td>0.929</td>
<td>14 bp</td>
</tr>
<tr>
<td>VSN4GATAVSN4GATA.01</td>
<td>(-)</td>
<td>CTGAGGATATTACG</td>
<td>1.000</td>
<td>0.926</td>
<td>—</td>
</tr>
</tbody>
</table>

**Model: CEBP_NFAT_01 (555 - 582 (+))**

<table>
<thead>
<tr>
<th>Matrix element</th>
<th>Str</th>
<th>Sequence</th>
<th>Core sim.</th>
<th>Mat. sim.</th>
<th>Distance to next element</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSN4CEBP1CEBP</td>
<td>(+)</td>
<td>GAAATTTGCTTCAA</td>
<td>1.000</td>
<td>0.929</td>
<td>14 bp</td>
</tr>
<tr>
<td>VSN4NFAT1NFAT.02</td>
<td>(-)</td>
<td>CTGAGGATATTACG</td>
<td>1.000</td>
<td>0.926</td>
<td>—</td>
</tr>
</tbody>
</table>

**Model: NRCH_NKXH_CEBP_01 (574 - 713 (+))**

<table>
<thead>
<tr>
<th>Matrix element</th>
<th>Str</th>
<th>Sequence</th>
<th>Core sim.</th>
<th>Mat. sim.</th>
<th>Distance to next element</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSN4NRCHVSN4CEBP.01</td>
<td>(-)</td>
<td>CCAAGGAAAGTGGCAAATG</td>
<td>1.000</td>
<td>0.936</td>
<td>58 bp</td>
</tr>
<tr>
<td>VSN4NRCHVSN4CEBP.02</td>
<td>(+)</td>
<td>GTGCAAGTAATTTAATTG</td>
<td>0.000</td>
<td>0.765</td>
<td>58 bp</td>
</tr>
<tr>
<td>VSN4NRCHVSN4CEBP.03</td>
<td>(+)</td>
<td>GTGCAAGTAATTTAATTG</td>
<td>1.000</td>
<td>0.927</td>
<td>—</td>
</tr>
</tbody>
</table>

**Model: HIF1_GATA_01 (850 - 863 (-))**

<table>
<thead>
<tr>
<th>Matrix element</th>
<th>Str</th>
<th>Sequence</th>
<th>Core sim.</th>
<th>Mat. sim.</th>
<th>Distance to next element</th>
</tr>
</thead>
<tbody>
<tr>
<td>VSN4HIF1HIF1.04</td>
<td>(+)</td>
<td>GTTATTACG</td>
<td>1.000</td>
<td>0.988</td>
<td>13 bp</td>
</tr>
<tr>
<td>VSN4GATAVSN4GATA.06</td>
<td>(+)</td>
<td>CTGAGGATATTACG</td>
<td>1.000</td>
<td>0.943</td>
<td>—</td>
</tr>
</tbody>
</table>
identify if CpG in promoter region
identify CGI within/adjacent to promoter
capture sequence 4000bp flanking CGI
identify TFBM that contain CpG
select CpG of interest
Case study

1. Identify if CpG in promoter region
2. Identify CGI within/adjacent to promoter
3. Capture sequence 4000bp flanking CGI
4. Identify TFBM that contain CpG
5. Select CpG of interest

- If analysing specific CpG can capture adjacent CpGs
- Do you want to analyse
  - CGI
  - CpG Shore
  - CpG Shelf
  - CpG Open Sea
Case study

PSQ design software

• Paste in bisulphite modified sequence of interest flanked by ~300bp
• Select target CpG
• Maximum amplicon length ~600bp
• Ensure primers do not cover SNPs or CpGs
Case study: PyroMark CpG assays

- 84,000+ predesigned assays
  - 30,000+ human assays
  - 30,000+ mouse assays
  - 24,000+ rat assays


**Search for and order PyroMark CpG Assays**

Enter one or more search terms (such as Entrez Gene ID, Ensembl Gene ID, RefSeq ID, or gene symbol) to find and order PyroMark CpG Assays.

Search for: **PTPN20B**

<table>
<thead>
<tr>
<th>Name</th>
<th>Details</th>
<th>Cat. no.</th>
<th>List price</th>
<th>Your price</th>
</tr>
</thead>
<tbody>
<tr>
<td>PyroMark CpG Assay (200)</td>
<td>PCR and sequencing primers for Pyrosequencing analysis of gene-specific CpG methylation after DNA bisulfite conversion (200 reactions; tube format)</td>
<td>Varies</td>
<td>£44.50</td>
<td>Log In*</td>
</tr>
<tr>
<td>PyroMark CpG Assay 96 well (50)</td>
<td>PCR and sequencing primers for Pyrosequencing analysis of gene-specific CpG methylation after DNA bisulfite conversion (50 reactions; plate format); minimum order of 24 assays per plate</td>
<td>Varies</td>
<td>£30.00</td>
<td>Log In*</td>
</tr>
<tr>
<td>PyroMark CpG Assay 96 well (200)</td>
<td>PCR and sequencing primers for Pyrosequencing analysis of gene-specific CpG methylation after DNA bisulfite conversion (200 reactions; plate format); minimum order of 24 assays per plate</td>
<td>Varies</td>
<td>£44.50</td>
<td>Log In*</td>
</tr>
</tbody>
</table>

**Select all**

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Species</th>
<th>Entrez Gene ID</th>
<th>siRNAs</th>
<th>shRNA</th>
<th>Real-Time Assays</th>
<th>DNA methylation Assays</th>
<th>Long-Range PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTPN20B</td>
<td>Human</td>
<td>26095</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pttn20b</td>
<td>Rat</td>
<td>306281</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Free delivery on online orders over 300 GBP!
• Level of methylation, accuracy within ~ 5%
  exclude assays with <5% or >95% methylation

• No preferential amplification of unmethylated or methylated DNA

• Bisulphite conversion is completed
Pyrosequencing run

- Run time dependent on sequence length, 10min-1.5hr
Pyrosequencing analysis

![Pyrosequencing Analysis Image]
Pyrosequencing analysis

- Samples run in duplicate are within 5%
- 0% and 100% controls are comparable between plates
- inter/intraplate replicates are comparable
- negative DNA control no signal
• high-throughput CpG methylation analysis platform

• real-time, sequence-based detection and quantification

• % methylation at multiple adjacent CpG sites

• genotyping
References

• Helen E. White, *Clinical Chemistry* 52:6 1005–1013 (2006) Quantitative Analysis of *SRNPN Gene Methylation* by Pyrobequencing as a Diagnostic Test for Prader–Willi Syndrome and Angelman Syndrome

• http://www.pyrosequencing.com/

• http://www.qiagen.com/products/bytechnology/pyrosequencing

• UCSC Genome Bioinformatics http://genome.ucsc.edu/

• Genomatix http://www.genomatix.de/
Pyrosequencing

Alix Groom