Bacterial Methylation and Role in Understanding Disease

Michael Jennings

Pacific Biosciences, West Coast UGM, Stanford, December 3, 2015
Host adapted bacterial pathogens

*Neisseria*
- meningitis, septicemia
- gonorrhea

*Haemophilus*
- respiratory tract, otitis media

*Helicobacter pylori*
- gastritis, cancer

*Streptococcus pneumoniae*
- Pneumonia, meningitis, septicemia

These bacteria express on their surface virulence factors that are essential for successful colonization of the host.
Surface expressed virulence factors.

From: Rosenstein et al., NEJM 2001 344(18):1378-1388.
Phase variation

Phase variation is the reversible, high frequency switching of bacterial surface antigens.
Phase variation via simple tandem repeats

Repeats in coding sequence  Homopolymers eg poly G or poly C
Tetranucleotides eg GATC
Heptanucleotides eg CAAACAA

All changes are reversible
Phase variation via simple tandem repeats

Repeats in coding sequence $\rightarrow$ Alter expression by frame shift mutations.

All changes are reversible
Phase variation - a powerful system for generating phenotypic diversity

3 genes expressing different phase variable virulence factors

Independent switching of these factors can give rise to 8 different phenotypes \( (2^3=8) \) from a single strain.

The genome of *N. meningitidis* strain MC58 is reported to have >50 phase variable genes - a potential repertoire of \( 5 \times 10^{14} \) phenotypes \( (2^{50} = 5 \times 10^{14}) \) from a single strain!
Avogadro and the meningococcus

Considering both phase- \((5 \times 10^{14})\) and antigenic variation (pili only \(\sim 10^7\)), the repertoire of surfaces that can be made by a single organism = \(5 \times 10^{21}\) ! As this is a conservative estimate, Avogadro’s number is not far away…..
Many species containing phase variable Type III restriction modification systems

H. influenzae Rd

N. meningitidis MC58

H. pylori J99

H. pylori AG

M. catarrhalis 23246

M. catarrhalis 23246

P. haemolytica A1

M. hypopneumon

M. hypopneumoniae J

S. thermophilus CNRZ1066

M. pulmonis UAB CTIP

600bp
Restriction-Modification Systems

Composed of pairs of opposing enzyme activities
- Endonuclease (Res)
- Methyltransferase (Mod)
Many phase variable type III restriction modification systems that have lost Res function

H. influenzae Rd

N. meningitidis MC58

H. pylori J99

H. pylori AG

M. catarrhalis 23246

M. catarrhalis 23246

P. haemolytica A1

M. hypopneumon

M. hypopneumoniae J

S. thermophilus CNRZ1066

M. pulmonis UAB CTIP

600bp
This situation has evolved independently at least 11 times – selective advantage?
Phase variable R-M systems

Many genes involved in virulence are phase-variable.

Methylation of DNA can control gene expression and is involved in phase variation mechanisms of specific genes. eg. Dam in *E. coli*.

Could the phase-variable methyltransferase play a role in pathogenicity by mediating phase variation of multiple genes?
Phase variable *H. influenzae* type III mod gene
Nontypeable *Haemophilus influenzae* (NTHi)

Causes upper and lower respiratory tract infections - pneumonia, sinusitis, COPD exacerbations and otitis media.

Otitis media - middle ear infection that occurs when bacteria colonising the nasopharynx enter the middle ear space via the Eustachian tube.
Nontypeable *H. influenzae* modA allele distribution

- **723**: (AGCC)$_{22}$ modA2
- **C486**: (AGCC)$_{22}$ modA4
- **1209**: (AGCC)$_{19}$ modA9
- **R2866**: (AGCC)$_{16}$ modA10
- **477**: (AGCC)$_{14}$ modA5

*mod*
Microarray expression profiling *modA2* strain 723

Repeats “ON”

MOD expressed

Methylation of target sites

Random switching ~1/100

MOD not expressed

No methylation of target sites

Microarray

10 genes upregulated

26 genes upregulated

Many genes required for anaerobic growth
Phase variation of a single gene vs the **phase variable regulon** - “Phasevarion”

Repeats (-/+)

- Mod
Nontypeable *H. influenzae* modA2 in an OM model

![Diagram showing different modA variants and their DNA recognition domains.](Image)
Chinchillas challenged with 723 *modA2* ON vs OFF

**Days:**

+2  +4  +7  +10  +14  +18  +22

Epitympanic tap and nasopharyngeal lavage to determine the load of NTHI

**COHORTS:**

1) NTHI 723 *modA2* (ON)

2) NTHI 723 *modA2* (OFF)

Use histology to assess middle ear pathology
Chinchillas challenged with 723 modA2 ON vs OFF

<table>
<thead>
<tr>
<th>DAY</th>
<th>0</th>
<th>4</th>
<th>7</th>
<th>10</th>
<th>14</th>
<th>18</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>22ON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24OFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Death
ON
OFF
No Data

Nature Communications 2015 6: 7828
Chinchillas challenged with 723 $modA2$ ON vs OFF

The observation that ModA2 OFF challenge switches to ON over time supports the hypothesis of niche adaptation in the middle ear.
All other *NTHi* *modA* alleles tested display gene regulation and virulence phenotypes.
Vaccine-candidate Westerns using OMPs

Nature Communications 2015 6: 7828
Phasevarions based on Type III R-M methyltransferases are present in many host adapted Gram negative bacterial pathogens.
Phasevarion mediated gene switching confirmed in 4 species
Evidence for methylation?
Until recently only a few Mod target sites had been defined.

This has limited studies on the mechanism of gene regulation at the promoter level.
Principle of detecting modified DNA bases during PacBio SMRT sequencing.
SMRT methylome analysis was performed by our collaborators at Pacific Biosciences
# Methylome analysis - ModA12

<table>
<thead>
<tr>
<th>Methyltransferase Specificity</th>
<th>Modified Base</th>
<th>Nm modA12 ON</th>
<th>Nm modA12 KAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-GAATC-3’ 3’-CTAG-5’</td>
<td>m6A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’-GAAAGG-3’ 3’-CTTCC-5’</td>
<td>m6A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5’-ACACC-3’ 3’-TGTGG-5’</td>
<td>m6A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Nontypeable *H. influenzae* modA allele distribution
# Haemophilus methylomes

<table>
<thead>
<tr>
<th>modA allele</th>
<th>NTHi strain</th>
<th>Clinical symptoms</th>
<th>Accession number</th>
<th>Methylation sequence</th>
<th>Systematic name</th>
<th>Number sites in genome</th>
<th>Genome size (bp)</th>
<th>Predicted ORFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>modA2</td>
<td>723</td>
<td>OM</td>
<td>CP007472</td>
<td>5'-CCGA(m6A)-3'</td>
<td>M.Hin723I</td>
<td>2270</td>
<td>1,887,620</td>
<td>1868</td>
</tr>
<tr>
<td>modA4</td>
<td>C486</td>
<td>OM</td>
<td>CP007471</td>
<td>5'-CG(m6A)G-3'</td>
<td>M.HinC486I</td>
<td>6203</td>
<td>1,846,507</td>
<td>1783</td>
</tr>
<tr>
<td>modA5</td>
<td>477</td>
<td>OM</td>
<td>CP007470</td>
<td>5'-AC(m6A)GC-3'</td>
<td>M.Hin477I</td>
<td>2548</td>
<td>1,846,259</td>
<td>1813</td>
</tr>
<tr>
<td>modA9</td>
<td>1209</td>
<td>OM</td>
<td>JMQP01000000</td>
<td>5'-CCTG(m6A)-3'</td>
<td>M.Hin1209I</td>
<td>2504</td>
<td>1,895,979</td>
<td>2247</td>
</tr>
<tr>
<td>modA10</td>
<td>R2866</td>
<td>Blood</td>
<td>CP002277*</td>
<td>5'-CCT(m6A)C-3'</td>
<td>M.Hin2866I</td>
<td>1244</td>
<td>1,932,238</td>
<td>1905</td>
</tr>
</tbody>
</table>
### Haemophilus methylomes

<table>
<thead>
<tr>
<th>modA allele</th>
<th>NTHi strain</th>
<th>Clinical symptoms</th>
<th>Accession number</th>
<th>Methylation sequence</th>
<th>Systematic name</th>
<th>Number sites in genome</th>
<th>Genome size (bp)</th>
<th>Predicted ORFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>modA2</td>
<td>723</td>
<td>OM</td>
<td>CP007472</td>
<td>5'-CCGA(m^6A)-3'</td>
<td>M.Hin723I</td>
<td>2270</td>
<td>1,887,620</td>
<td>1868</td>
</tr>
<tr>
<td>modA4</td>
<td>C486</td>
<td>OM</td>
<td>CP007471</td>
<td>5'-CG(m^6A)G-3'</td>
<td>M.HinC486I</td>
<td>6203</td>
<td>1,846,507</td>
<td>1783</td>
</tr>
<tr>
<td>modA5</td>
<td>477</td>
<td>OM</td>
<td>CP007470</td>
<td>5'-AC(m^6A)GC-3'</td>
<td>M.Hin477I</td>
<td>2548</td>
<td>1,846,259</td>
<td>1813</td>
</tr>
<tr>
<td>modA9</td>
<td>1209</td>
<td>OM</td>
<td>JMQP01000000</td>
<td>5'-CCTG(m^6A)-3'</td>
<td>M.Hin1209I</td>
<td>2504</td>
<td>1,895,979</td>
<td>2247</td>
</tr>
<tr>
<td>modA10</td>
<td>R2866 Blood</td>
<td>Blood</td>
<td>CP002277*</td>
<td>5'-CCT(m^6A)C-3'</td>
<td>M.Hin2866I</td>
<td>1244</td>
<td>1,932,238</td>
<td>1905</td>
</tr>
</tbody>
</table>
How do changes in methylation regulate gene expression?
Mapping methylation sites to promoter regions
What about Gram-positive bacterial pathogens?
Streptococcus pneumoniae

Gram-positive bacterial pathogen.

Causes meningitis, sepsis, pneumonia and otitis media.

Major morbidity and mortality worldwide.
Streptococcus pneumoniae

Are DNA methylation changes driving this switching process in the pneumococcus?
Principle of detecting modified DNA bases during PacBio SMRT sequencing.
Table 1 | RM systems of pneumococcal strain D39.

<table>
<thead>
<tr>
<th>RM system name</th>
<th>Type</th>
<th>ORFs*</th>
<th>Specificity†</th>
<th>TRD</th>
<th>Modified base</th>
<th>Number of sites*</th>
<th>Predicted no. of sites</th>
<th>Difference (%)</th>
<th>Strand specificity (%)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpnD39RF1631P (DpnI)</td>
<td>Type II</td>
<td>SPD_1630-1</td>
<td>5'GATC-3' 3'CTAG-5'</td>
<td>—</td>
<td>—</td>
<td>7,164</td>
<td>7,324</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SpnD39I (SpnD39RF1260P)</td>
<td>Type II</td>
<td>SPD_1259-60</td>
<td>5'TCTAGA-3' 3'AGATCT-5'</td>
<td>—</td>
<td>m6A</td>
<td>644</td>
<td>438</td>
<td>47§</td>
<td>—</td>
</tr>
<tr>
<td>SpnD39II (SpnD39RF1079AP)</td>
<td>Type II</td>
<td>SPD_1079-80</td>
<td>5'TCGAG-3' 3'AGCTG-5'</td>
<td>—</td>
<td>m6A</td>
<td>1,509</td>
<td>1,454</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>SpnD39IIIA (SpnD39RF454P)</td>
<td>Type I</td>
<td>SPD_0450-5</td>
<td>5'CRAAN8CTG-3' 3'GYTTN8GAC-5'</td>
<td>1.1, 2.1</td>
<td>m6A</td>
<td>720</td>
<td>438</td>
<td>64§</td>
<td>66§</td>
</tr>
<tr>
<td>SpnD39IIB</td>
<td>Type I</td>
<td>SPD_0450-5</td>
<td>5'CRAAN8TTC-3' 3'GYTTN8AAG-5'</td>
<td>1.1, 2.2</td>
<td>m6A</td>
<td>1,029</td>
<td>665</td>
<td>55§</td>
<td>64§</td>
</tr>
<tr>
<td>SpnD39IIIC</td>
<td>Type I</td>
<td>SPD_0450-5</td>
<td>5'CACNG8TTC-3' 3'GTGN8AAG-5'</td>
<td>1.2, 2.2</td>
<td>m6A</td>
<td>641</td>
<td>876</td>
<td>-27§</td>
<td>66§</td>
</tr>
<tr>
<td>SpnD39IID</td>
<td>Type I</td>
<td>SPD_0450-5</td>
<td>5'CACNG8CTG-3' 3'GTGN8GAC-5'</td>
<td>1.2, 2.1</td>
<td>m6A</td>
<td>428</td>
<td>577</td>
<td>-26§</td>
<td>67§</td>
</tr>
<tr>
<td>SpnD39IIE</td>
<td>Type I</td>
<td>SPD_0450-5</td>
<td>5'CRAAN8CTT-3' 3'GYTTN8GAA-5'</td>
<td>1.1, 2.3</td>
<td>m6A</td>
<td>1,028</td>
<td>665</td>
<td>55§</td>
<td>63§</td>
</tr>
<tr>
<td>SpnD39IIIFP</td>
<td>Type I</td>
<td>SPD_0450-5</td>
<td>5'CACNG8CTT-3' 3'GTGN8GAA-5'</td>
<td>1.2, 2.3</td>
<td>m6A</td>
<td>796</td>
<td>876</td>
<td>-9§</td>
<td>64§</td>
</tr>
<tr>
<td>SpnD39RF782P</td>
<td>Type I</td>
<td>SPD_0782-4</td>
<td>not active; see below</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SpnD39McrBCP</td>
<td>Type IV</td>
<td>SPD_1108-9</td>
<td>not known</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

* ORFs: Open Reading Frames
† Specificity: Nucleotide specificity of the RM system
‡ Strand specificity: Percentage of sites that are methylated on the same strand as the recognition site
Type I RM system and regulation of gene expression

Each of the 6 alternate methylation patterns results in a distinct gene expression profile (RNA seq).

Some of these genes included well characterized virulence factors such as capsule.
SpnIII system and a role in virulence?

Mice were challenged with wild type D39 via intranasal (carriage model) or intravenous (invasive model) and relative fitness of SpnIII alleles was examined.
In vivo selection for switching in the mouse model

NP

Nature Communications 2014 5: 5055
In vivo selection for switching in the mouse model

BLOOD

Inoculum, black
4 hours, grey
30 hours, white

Nature Communications 2014 5: 5055
A phase variable regulon - “Phasevarion”

- Multiple genes switched in a coordinated fashion resulting in “differentiation” into two (on/off Type III) or 6 (6-way recombination Type I) distinct cell types via methylation changes. Distinct phenotypes in virulence model systems.

- Major implications for host/pathogen interactions and vaccine development.

- SMRT methylome analysis has facilitated rapid determination of target sites - confirming methylation and enabling promoter level studies.
Acknowledgements

**Institute for Glycomics, Griffith University**
Freda Jen, John Atack, Kate Seib, Jayde Gawthorne

**The University of Adelaide**
Prof James Paton

**The Research Institute at Nationwide Children’s Hospital**
Jennifer Edwards, Lauren Bakaletz and team

**Leicester University**
Marco Oggioni

**The University of Iowa,**
Michael Apicella

**Monash University (formerly GU)**
Yogitha Srikhanta

**Pacific Biosciences**
Matthew Boitano Tyson Clark, Jonas Korlach

**Funding:** Australian Research Council, National Health & Medical Research Council (Australia); NIH, Helen C Levitt Visiting Professorship (University of Iowa).