1. Biological Background
○○○○○

2. Data Preparation
○○○○○○○

3. Results
○○○○○○○

4. Conclusion
○○

Predicting exon splicing changes triggered by methylation profiles

33rd TBI Winterseminar in Bled

presented by Florian Heyl

February 15, 2018

Freie Universität Berlin

1.1. Splicing

David P. C, et al. (2013)

Freie Universität Berlin
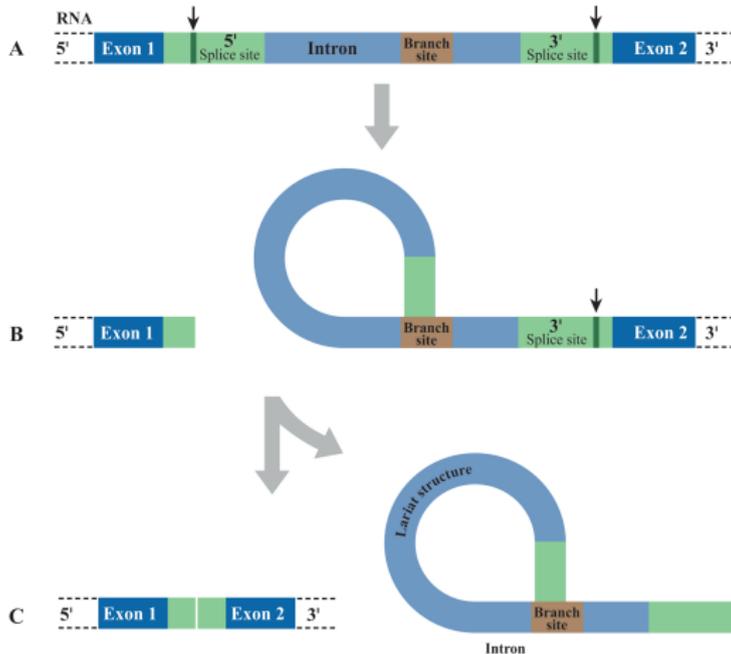
1.2. Alternative Splicing

- 90% of the genes are alternatively spliced
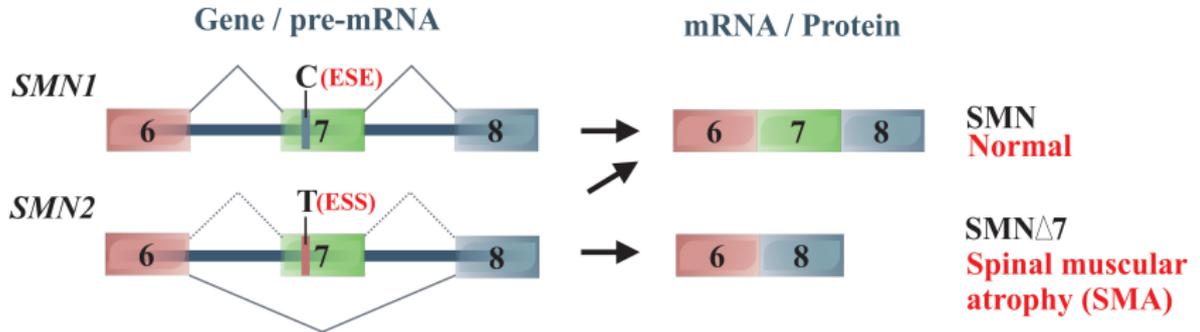
Cooper T. A, et al. (2009)

1.3. Influence of DNA Methylation

- Most prevalent form 5-Methylcytosine (5mC)
- Methylated CpG-islands at the transcription start site induce gene silencing

Does DNA methylation also influences alternative splicing?

Freie Universität Berlin

1.3. Influence of DNA Methylation

- Various proteins linked to DNA methylation and alternative splicing

  - Transcriptional repressor CTCF
  - Multifunctional protein MeCP2
  - Heterochromatin protein HP1

1.3. Influence of DNA Methylation

Predictive Model

- Input: Methylation profiles and target event
- Output: Probability for state of splicing

2.1. Read Counts

1. Biological Background
00000

2. Data Preparation
0●000000

3. Results
0000000

4. Conclusion
00
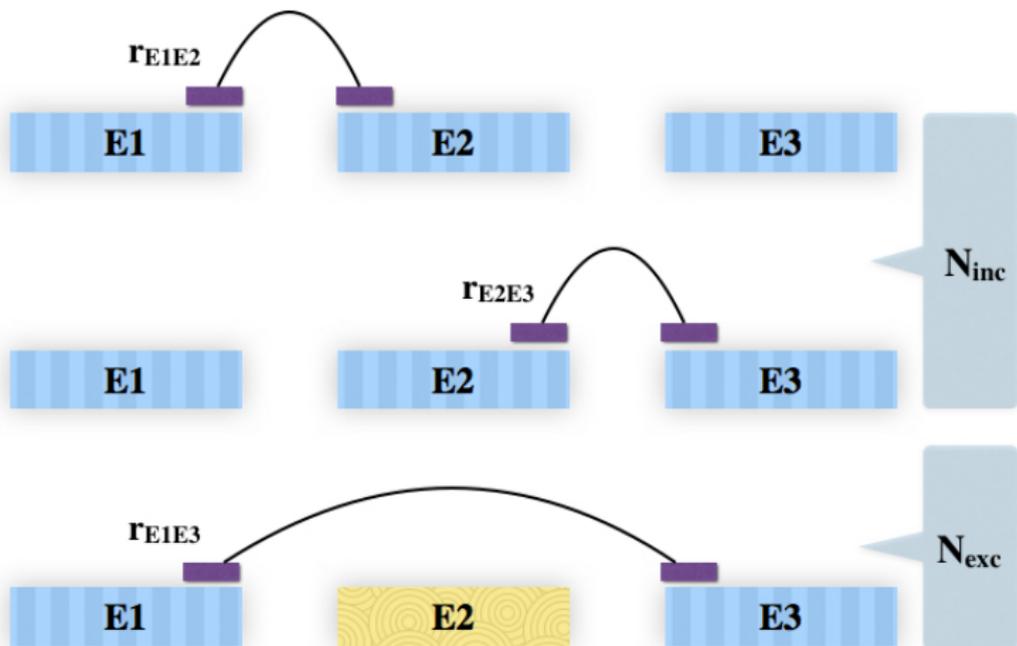
2.1. Read Counts

$$N_{total} = \frac{N_{inc}}{2} + N_{exc}$$

$$\Psi = p(Inc) = \frac{\dfrac{N_{inc}}{2}}{N_{total}}$$

1. Biological Background
ooooo

2. Data Preparation
oo●ooooo

3. Results
ooooooo

4. Conclusion
oo

2.1. Read Counts

$$\text{Beta}(\Psi|10 + \frac{N_{inc}}{2}, 10 + N_{exc}) \qquad \prod_{i=1}^{n} \mu^{x_i} \cdot (1 - \mu)^{1-x_i}$$

$p(\Psi)$

$p(X|\Psi)$

$p(\Psi|X)$

http://mzayat.com/single/143358.html

Freie Universität Berlin
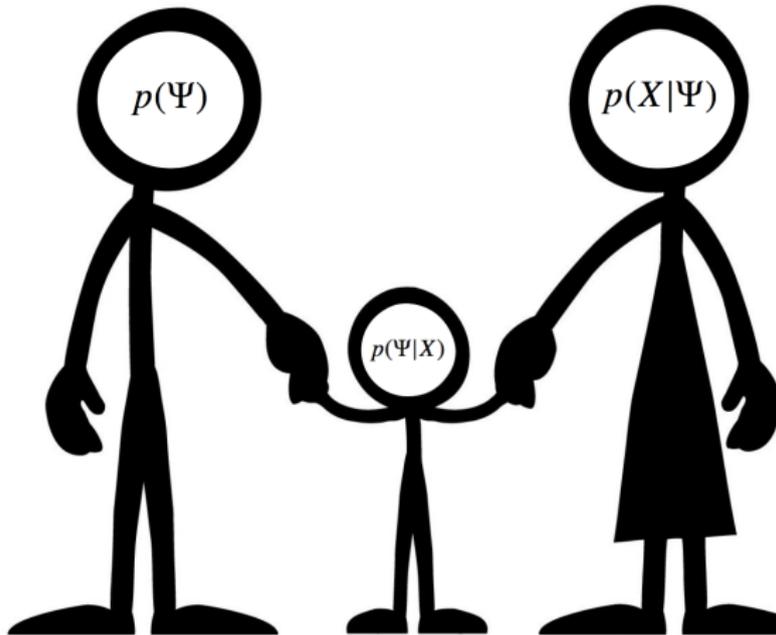
2.1. Read Counts

$$p(\Psi|X) = p(X|\Psi) \cdot p(\Psi)$$
$$\log p(\Psi|X) = \log p(X|\Psi) + \log p(\Psi)$$
$$\frac{d}{d\Psi} \log p(\Psi|X) = \frac{d}{d\Psi} \log p(X|\Psi) + \frac{d}{d\Psi} \log p(\Psi)$$
$$0 = \frac{d}{d\Psi} \log p(X|\Psi) + \frac{d}{d\Psi} \log p(\Psi)$$

Freie Universität Berlin

1. Biological Background
00000

2. Data Preparation
0000●000

3. Results
0000000

4. Conclusion
00

2.2. Get Exons

1. Remove X, Y and M chromosome

Freie Universität Berlin

CSIRO

2.2. Get Exons

1. Remove X, Y and M chromosome
2. Find genes with at least 3 exons

Freie Universität Berlin

1. Biological Background
○○○○○

2. Data Preparation
○○○○●○○○

3. Results
○○○○○○○

4. Conclusion
○○

2.2. Get Exons

1. Remove X, Y and M chromosome
2. Find genes with at least 3 exons
3. Remove short exons (< 50 bp)

Freie Universität Berlin

CSIRO

1. Biological Background
ooooo

2. Data Preparation
oooo●ooo

3. Results
ooooooo

4. Conclusion
oo

2.2. Get Exons

1. Remove X, Y and M chromosome
2. Find genes with at least 3 exons
3. Remove short exons (< 50 bp)
4. Excluded exon < 0.2

Freie Universität Berlin

2.2. Get Exons

1. Remove X, Y and M chromosome
2. Find genes with at least 3 exons
3. Remove short exons (< 50 bp)
4. Excluded exon < 0.2
5. Included exon > 0.8

Freie Universität Berlin

1. Biological Background
00000

2. Data Preparation
00000●00

3. Results
0000000

4. Conclusion
00

2.3. Get Intron Regions

1. Extend 500 bp up- and downstream of exon

2.3. Get Intron Regions

1. Extend 500 bp up- and downstream of exon
2. Remove group (up. intron + exon + do. intron) if introns overlaps with another exon

Freie Universität Berlin

CSIRO

1. Biological Background
ooooo

2. Data Preparation
oooooo●o

3. Results
ooooooo

4. Conclusion
oo

2.4. Methylation Data

2.4. Methylation Data

Krueger F., et al. (2012)

1. Biological Background
○○○○○

2. Data Preparation
○○○○○○○○

3. Results
●○○○○○○

4. Conclusion
○○

3.1. CpG Profiles

Freie Universität Berlin
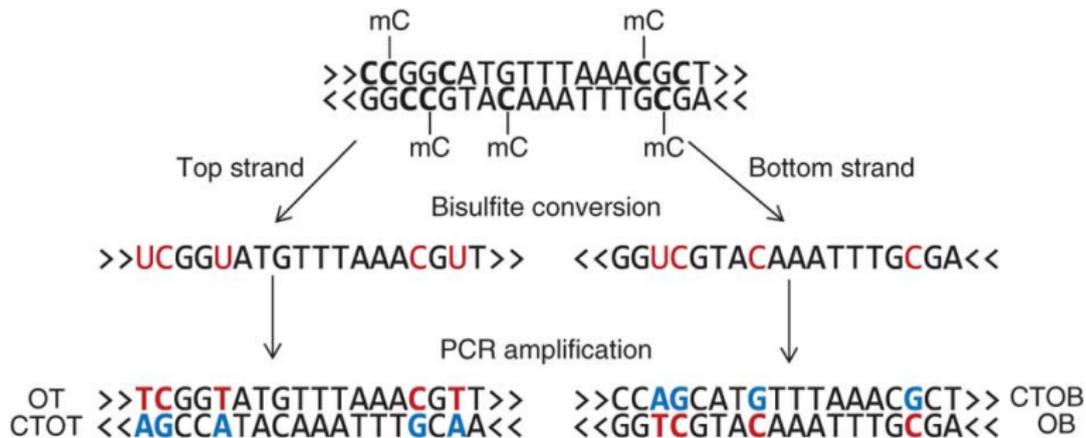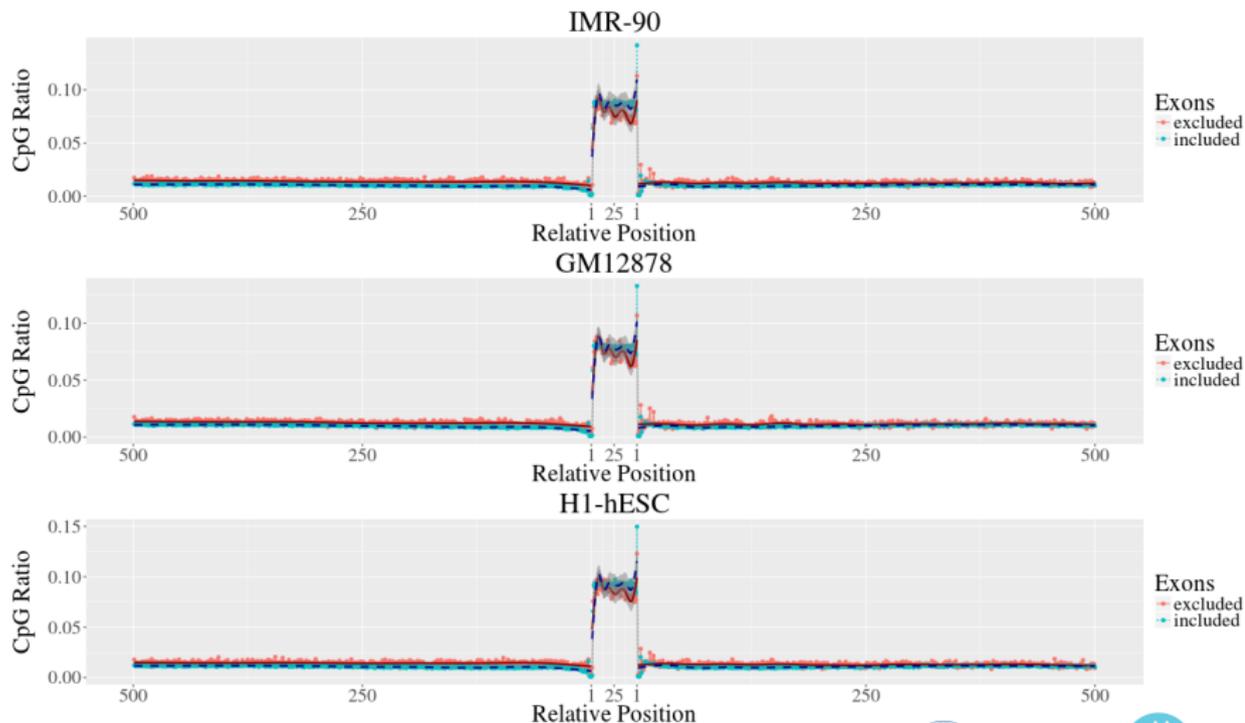
3.2. mCpG/CpG Profile
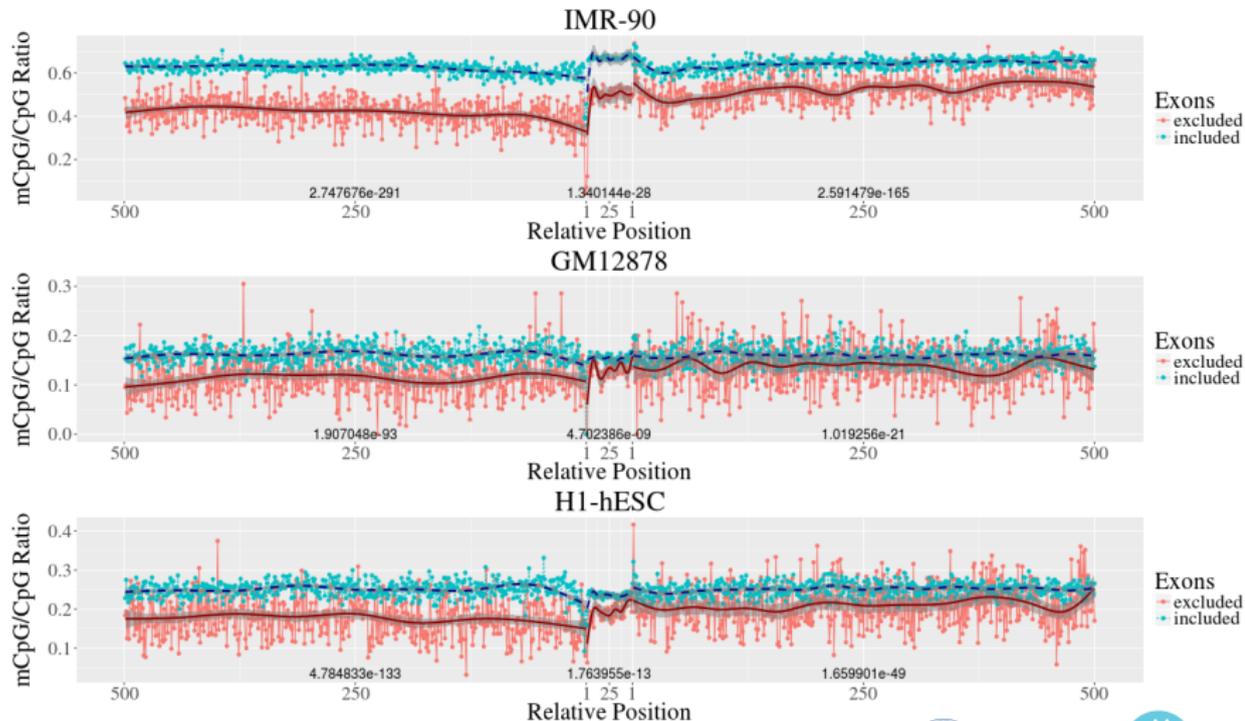
3.3. Gradient Boosting Machine

1. Biological Background
○○○○○

2. Data Preparation
○○○○○○○○

3. Results
○○○●○○○

4. Conclusion
○○
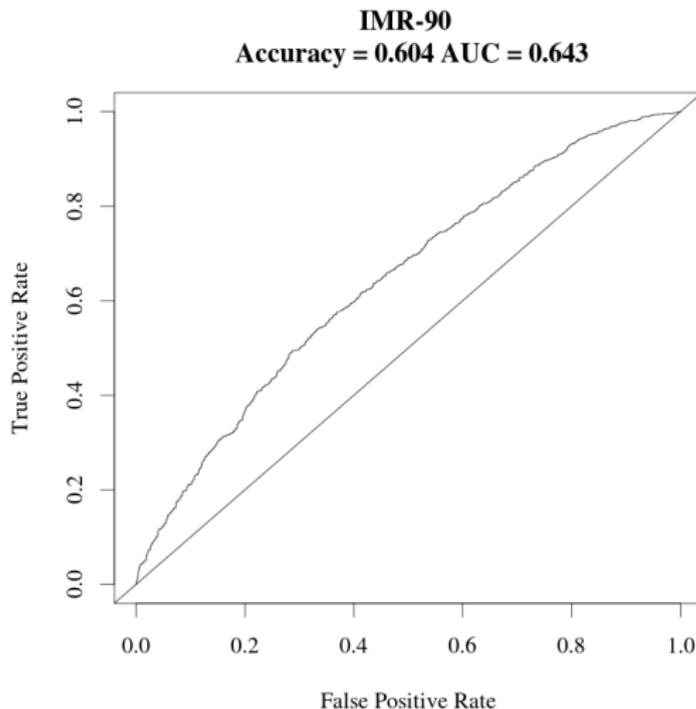
3.4. Why using Deep Learning?

- data x $\rightarrow \phi(x)$
  - $\rightarrow$ problem specific
  - $\rightarrow$ labour-intensive

Angermueller C, et al. (2016)

Freie Universität Berlin

CSIRO

1. Biological Background
○○○○○

2. Data Preparation
○○○○○○○○

3. Results
○○○○●○○

4. Conclusion
○○
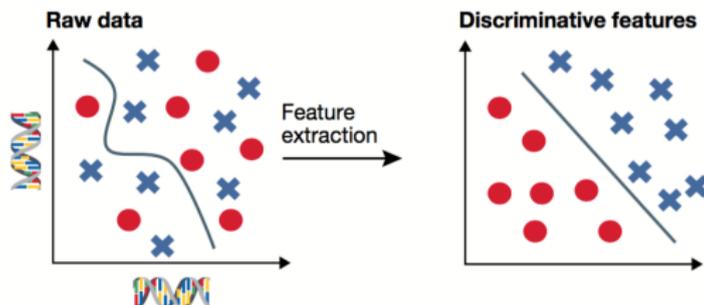
3.5. Architecture of the ANN
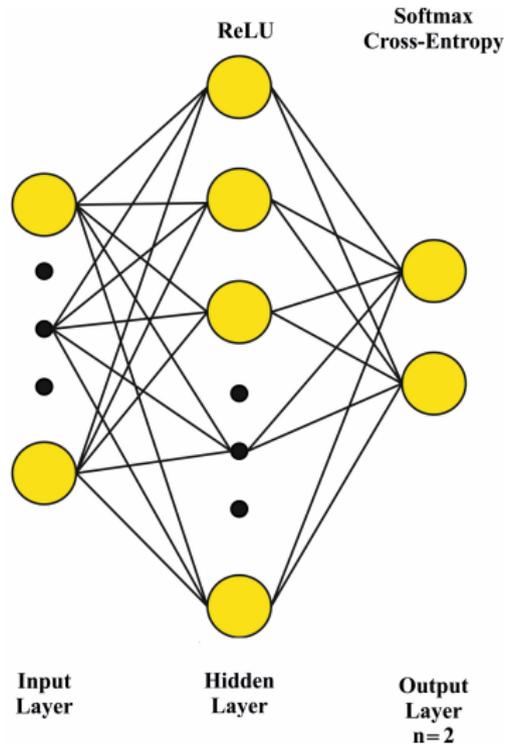
Angermueller C, et al. (2016)

3.6. Performance of the ANN
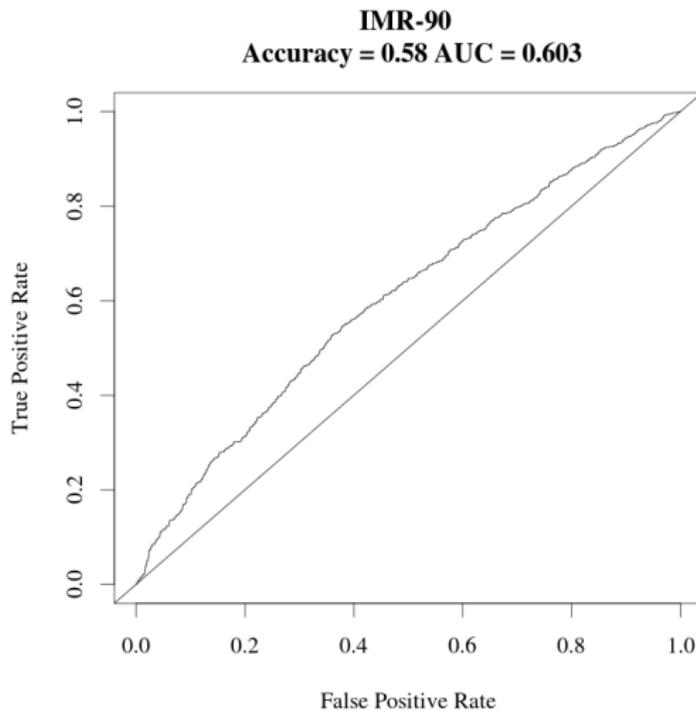
3.7. Convolutional Neural Network
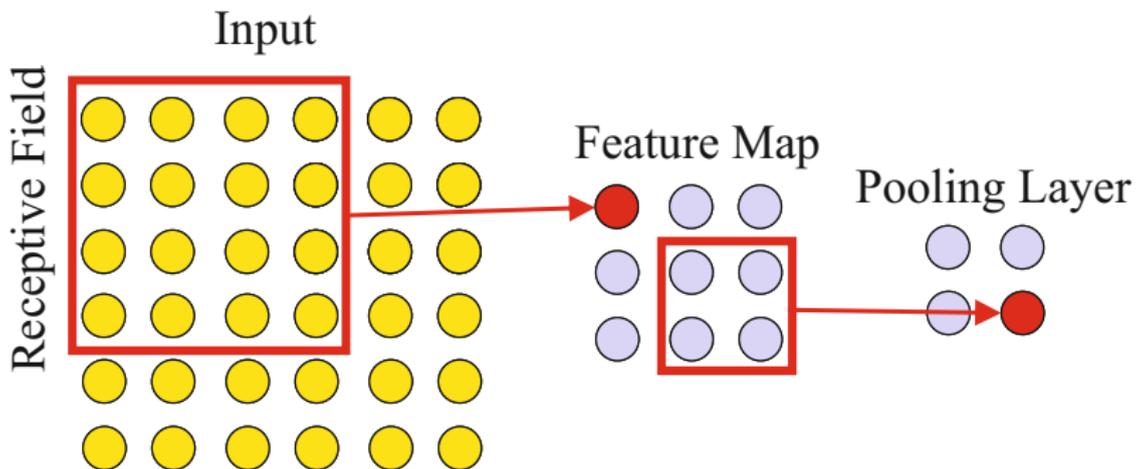
1. Biological Background
○○○○○

2. Data Preparation
○○○○○○○○

3. Results
○○○○○○○

4. Conclusion
●○

4.1. What to Change?

- Use probability of methylation

Freie Universität Berlin

CSIRO

1. Biological Background
ooooo

2. Data Preparation
oooooooo

3. Results
ooooooo

4. Conclusion
●o

4.1. What to Change?

- Use probability of methylation
- Add zeros

Freie Universität Berlin

CSIRO

1. Biological Background
○○○○○

2. Data Preparation
○○○○○○○○

3. Results
○○○○○○○

4. Conclusion
●○

4.1. What to Change?

- Use probability of methylation
- Add zeros
- Apply CNN

Freie Universität Berlin

CSIRO

1. Biological Background
○○○○○

2. Data Preparation
○○○○○○○○

3. Results
○○○○○○○

4. Conclusion
●○

4.1. What to Change?

- Use probability of methylation
- Add zeros
- Apply CNN
- Apply deep learning tricks

Freie Universität Berlin

CSIRO

4.2. Finish

Thanks to:

- Annalisa Marsico (Freie Universität Berlin)
- Denis Bauer

Freie Universität Berlin   CSIRO