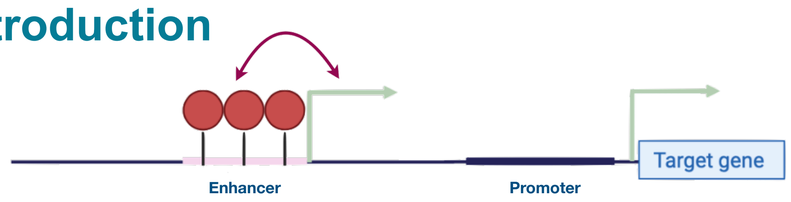
Breast Cancer analysis of interplay between **DNA methylation** and **enhancer activity**

Kim Brunel[1]; Roza Berhanu Lemma[2]; Anthony Mathelier[2]
[1]Magistère Européen de Génétique Paris Master 2
[2]NCMM Computational biology and gene regulation

Introduction



Hypothesis :
Is enhancer activity correlated with CpGs methylation rate?

Breast cancer (BC) is the **most frequent** cancer among women. According to the World Health Organisation, in 2020, there were **2.3 million women diagnosed** with breast cancer and **685 000 deaths** globally.
CpG DNA methylation, a prominent DNA modification accounts for gene expression is **altered in cancer**. Indeed, the righter and eraser methyl proteins are **dynamically regulated.** Moreover, **enhancer and super-enhancer activities** were also previously shown to be **modified in tumor cell types.**

My project aimed to unravel the interplay between DNA methylation and enhancer activities in Breast Cancer.

Strategy

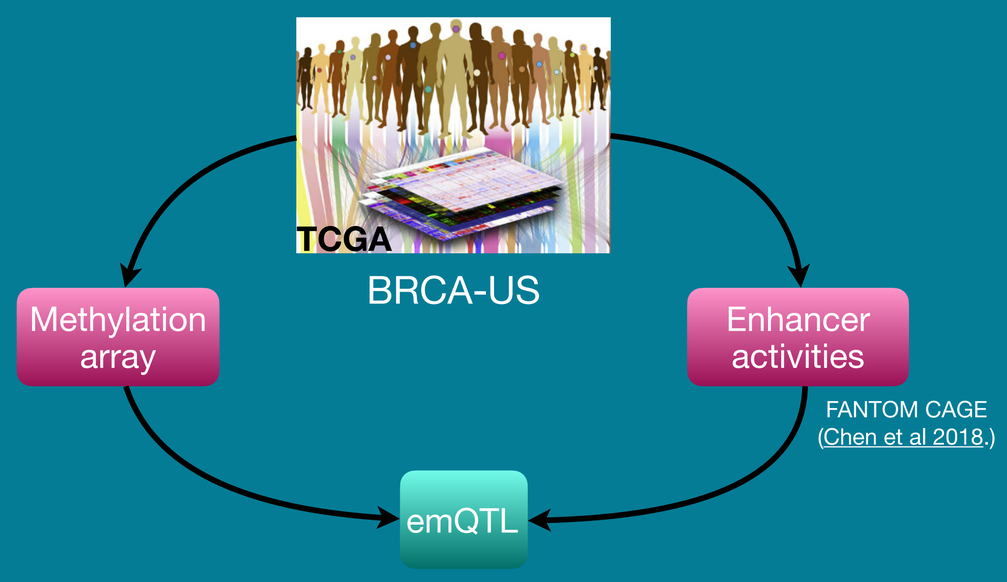
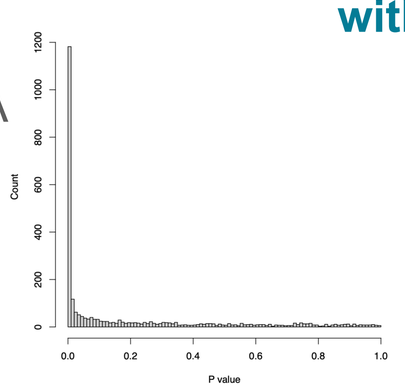


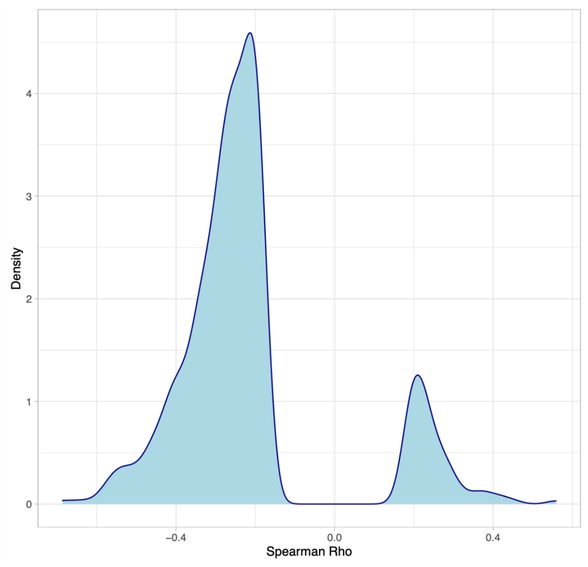
Fig 1 : Two datasets were collected from The Cancer Genome Atlas. The first sets corresponds to the CpGs methylation rate (Methylation array) from BRCA-US datas; the second one (Enhancer activities), collected from FANTOM CAGE method is from BRCA-US datas. These Datas were combined to obtain enhancer methylation quantity trait loci, emQTL pairs. These emQTLs represent the pairs formed by CpGs into enhancers.

CpGs methylation rate is mostly negatively correlated with enhancer activities

A



B

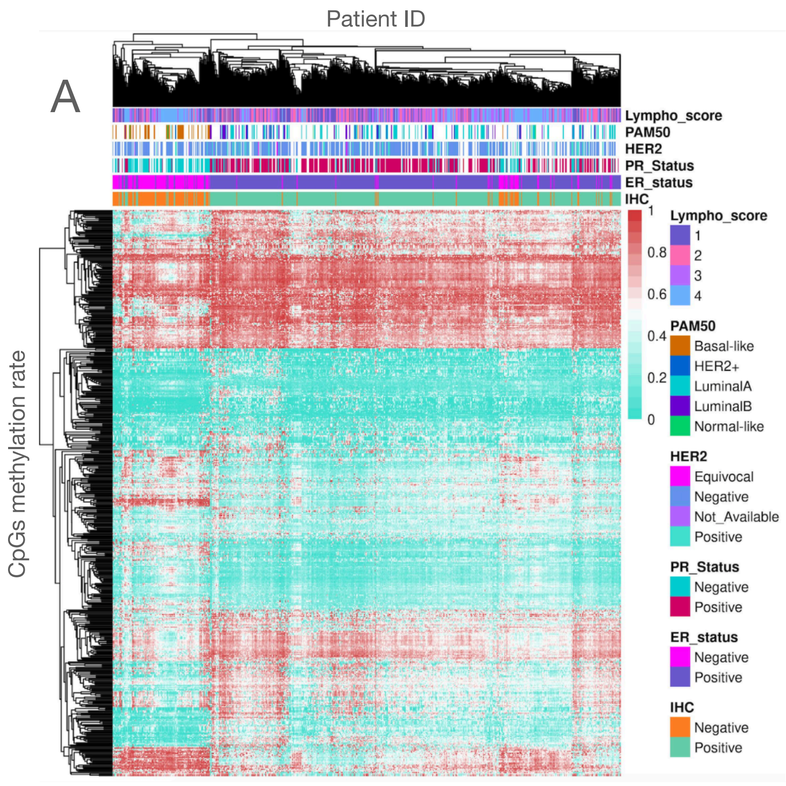


C

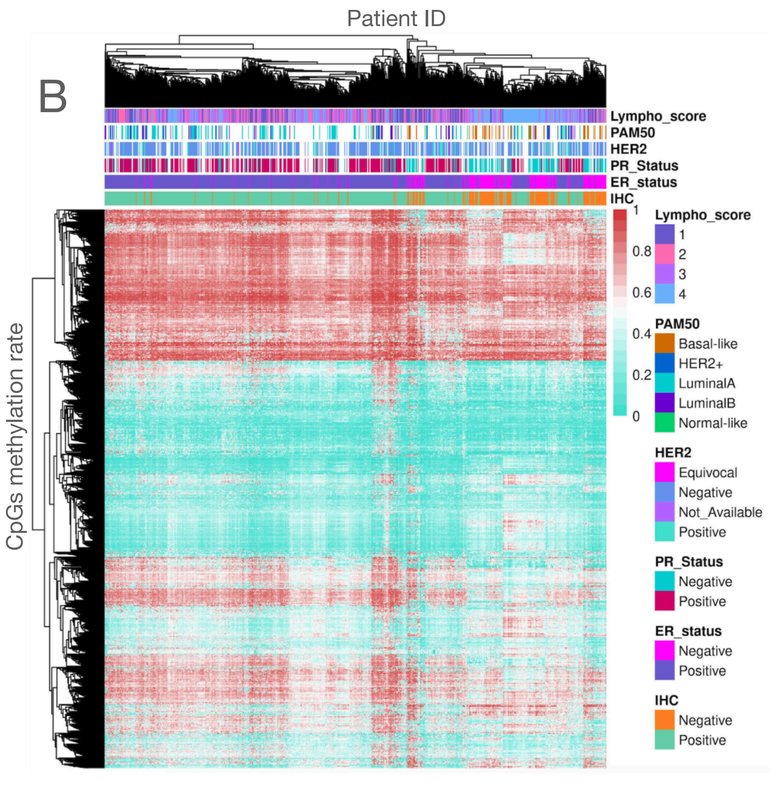
Total enhancer set (from TCGA)	63 480
emQTL (enhancer - CpG paires)	14 120
Enhancer with activity **correlated** to CpG methylation rate	431
Enhancer with activity **positively correlated** to CpG methylation rate	61
Enhancer with activity **negatively correlated** to CpG methylation rate	370
Enhancer with activity **uncorrelated** to CpG methylation rate	1 660
CpG with methylation rate **correlated** to enhancer with activity	555
CpG with methylation rate **uncorrelated** to enhancer with activity	21 112

Fig 2 : (A) P-value histogram before correction. Anti-conservative distribution **(B)** Spearman correlation test. The values were corrected by the Benjamini-Hochberg method. Most of the CpGs presents an anti correlation with the enhancer activities. **(C)** Dataset table.

These enhancer-CpGs negatively correlated pairs are sufficient to rebuilt Breast Cancer identities



CpGs with methylation rate **negatively correlated** to enhancer activity

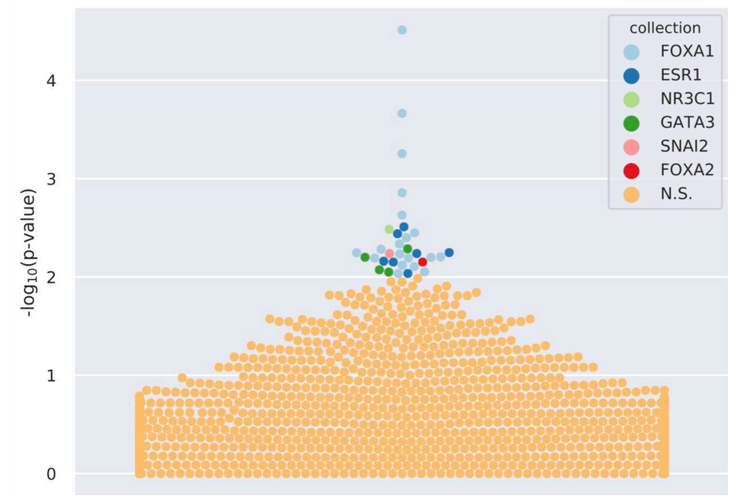


CpGs with methylation rate **uncorrelated** to enhancer activity

Transcription factors enrichment at Enhancer sites.

C

Enhancers with activity **negatively correlated** to CpGs methylation rate



Enhancers with activity **uncorrelated** to CpGs methylation rate)

D

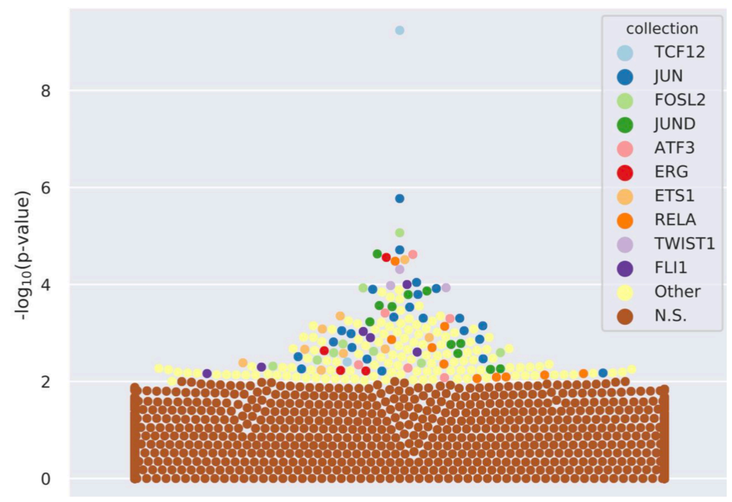
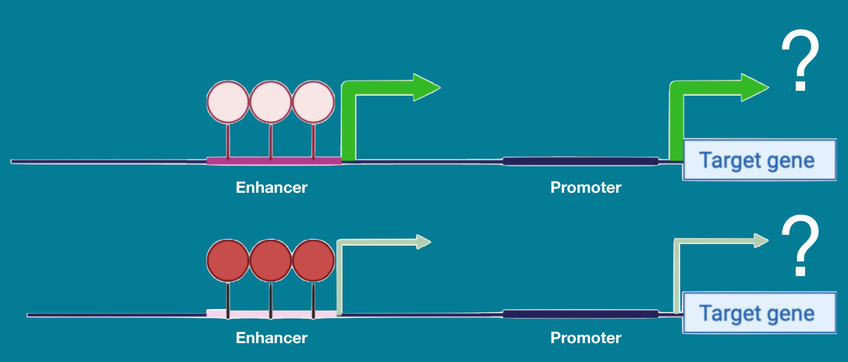


Fig 3 : Hierarchical clustering between CpGs with methylation rate negatively correlated **(A)** and uncorrelated **(B)** to enhancer activity, and the patient ID. The methylation rate of CpGs with methylation rate negatively correlated to enhancer activity is able to cluster the patients, based on their molecular subtypes. The Euclidean Mcquitty method was used to perform the clustering. **(C)** Transcription factors (TFs) enrichment in enhancers with activities negatively correlated to CpGs methylation rate. Unibind analyses were performed. Datas are from human cells lines & tissues (1317 types), combine 9660 ChIP-seq datasets, 841 TFs are implemented. FOXA1 ESR1 GATA3 SNAI2 FOXA2 were found in Unibind analysis and were previously shown to be involved in Breast Cancer; whereas in **(D)** these transcription factors were not found to be enriched.

Working Model & Perspectives



What about other cancers ?
Pan-genomic analyses under going

- These results are based on **correlation tests**, and have to be **verified** by **wet lab analysis**.
- **Next questions:** Can we see the same kind of correlation by looking at the enhancer methylation array and the gene activity associated with the enhancers?
- **Next steps: extend** the analysis to the **other cancer types** (on TCGA) thanks to a **Snakemake code** (the code is almost written)

Informations

Kim Brunel
M2 student - Magistère Européen de Génétique (Paris, France)
brunel.kim3@gmail.com
https://www.linkedin.com/in/kim-brunel-317329164

Roza Berhanu Lemma
Postdoctoral fellow - Oslo University Hospital, NCMM (Oslo, Norway) - Computational biology and gene regulation
r.b.lemma@ncmm.uio.no
https://www.linkedin.com/in/roza-berhanu-lemma-45496a48

Anthony Mathelier
Adjunct Researcher - Oslo University Hospital, NCMM (Oslo, Norway) - Computational biology and gene regulation
anthony.mathelier@ncmm.uio.no
https://www.linkedin.com/in/anthony-mathelier-0b80441a

Université de Paris

MAGISTERE EUROPEEN DE GENETIQUE

EMBL European Molecular Biology Laboratory

NCMM Centre for Molecular Medicine Norway
Nordic EMBL Partnership for Molecular Medicine