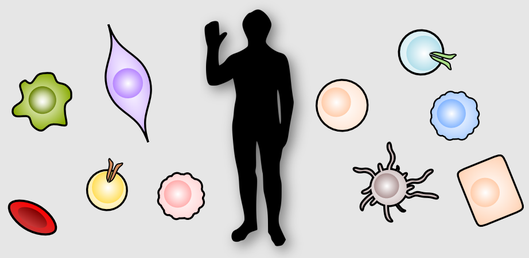
Deep learning reveals the general rules shaping the potential energy landscape of DNA methylation patterns

Dohoon Lee[1], Sun Kim[2,3,4]

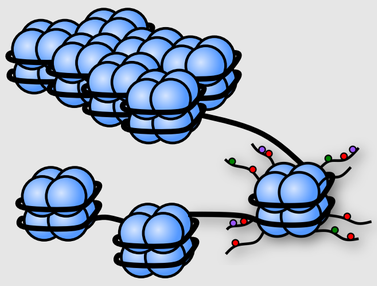
[1]Bioinformatics institute, [2]Interdisciplinary Program in Bioinformatics,
[3]Department of Computer Science and Engineering, and [4]Institute of Engineering Research, Seoul National University, Seoul, Korea

Background

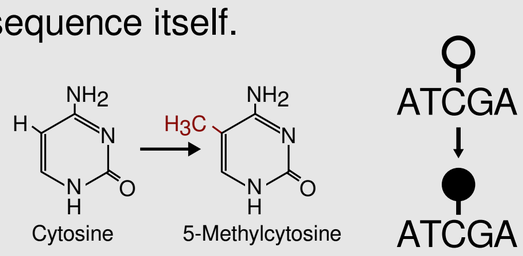


Our body consists of ~30 trillion cells that are morphologically and functionally diverse. These cells originate from a single fertilized egg, and have almost identical genomic DNA sequence.

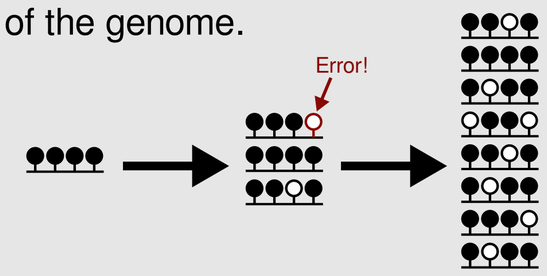
How can such diversity be achieved?



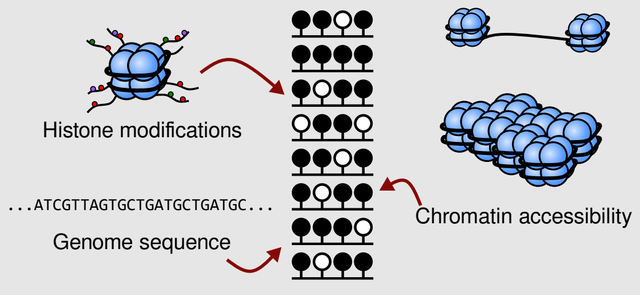
It can be in part explained by ***epigenetic*** modifications, the changes of genome that do not affect DNA sequence itself.



DNA methylation (DNAm) is one of the best studied epigenetic modifications. It denotes the addition of methyl group at CpG dinucleotide of the genome.



Maintenance of DNAm is intrinsically inaccurate. As a result, a population of cells harbor heterogeneous DNAm patterns.



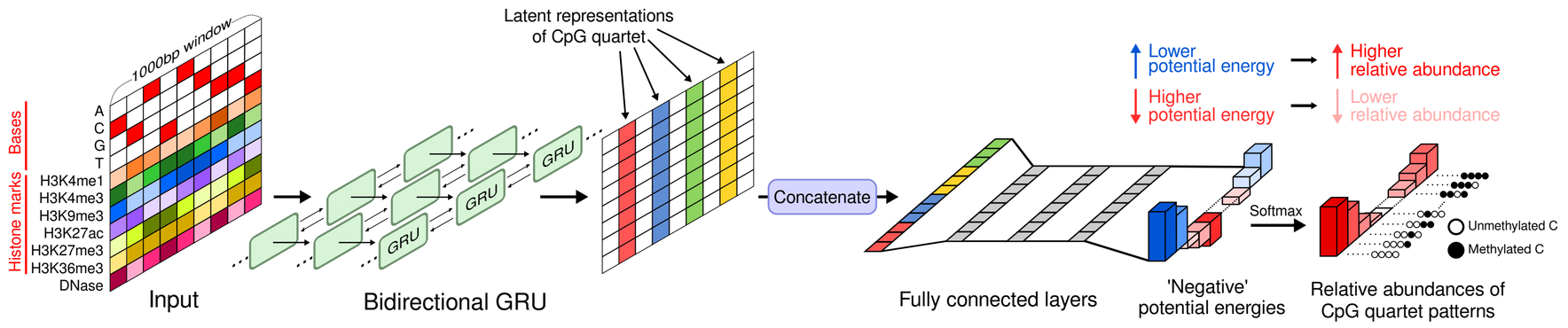
The variability of DNAm is determined by complex interactions between other epi/genomic features such as histone marks, chromatin accessibility and genome sequence itself.

Objective

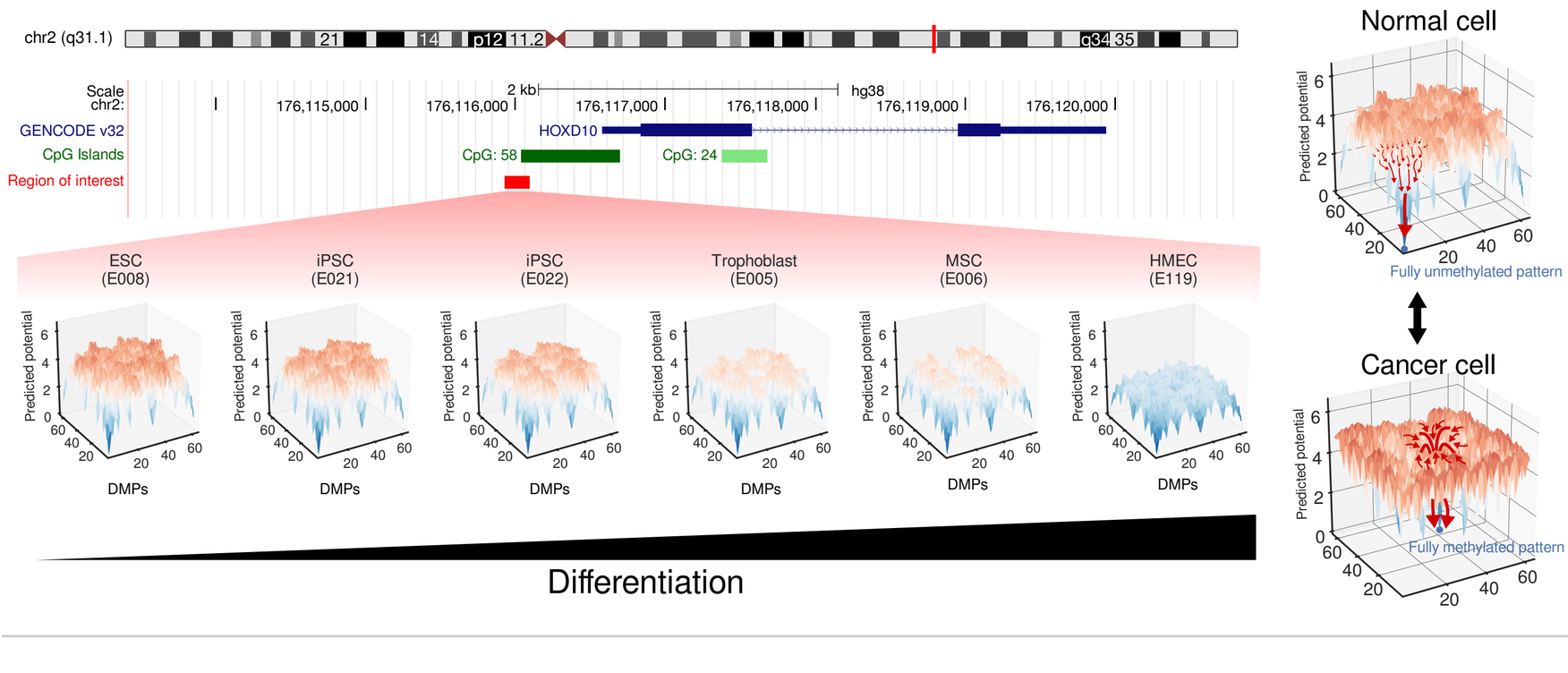
This study is to establish a deep neural network that computationally models the variability of DNAm patterns based on the other epi/genomic features surrounding them. We model the variability of DNAm patterns by **predicting the probability of observing each DNAm pattern**.

Results

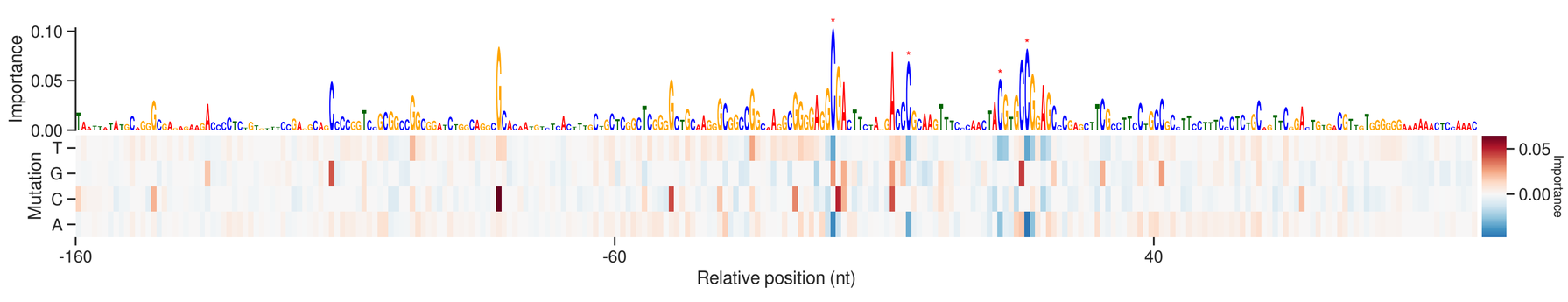
▼ Surprisingly, an extremely simple BiGRU-based integration of epi/genomic features was sufficient to accurately model the regulation of DNAm pattern variability.

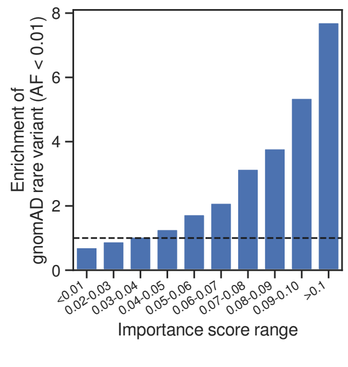


▼ Based on the models, we observed that the DNAm variability is dynamically regulated during differentiation. Furthermore, we revealed remarkable difference of DNAm potential energy landscape between normal and cancer cell.



▼ We can simulate the effect of a mutation on DNAm variability ***in silico***.





◀ Mutations that severely impact DNAm pattern variability are rare in human population, which means that they are evolutionarily deleterious.