



HAL
open science

Nucleosome positioning in yeast: from prediction to synthetic design

Alex Westbrook, Etienne Routhier, Julien Mozziconacci, Romain Koszul,
Jean-Baptiste Boulé

► To cite this version:

Alex Westbrook, Etienne Routhier, Julien Mozziconacci, Romain Koszul, Jean-Baptiste Boulé. Nucleosome positioning in yeast: from prediction to synthetic design. Journées du GDR BIM, Nov 2023, Villeneuve d'Ascq (Lille 3), France. mnhn-04344672

HAL Id: mnhn-04344672

<https://mnhn.hal.science/mnhn-04344672v1>

Submitted on 14 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



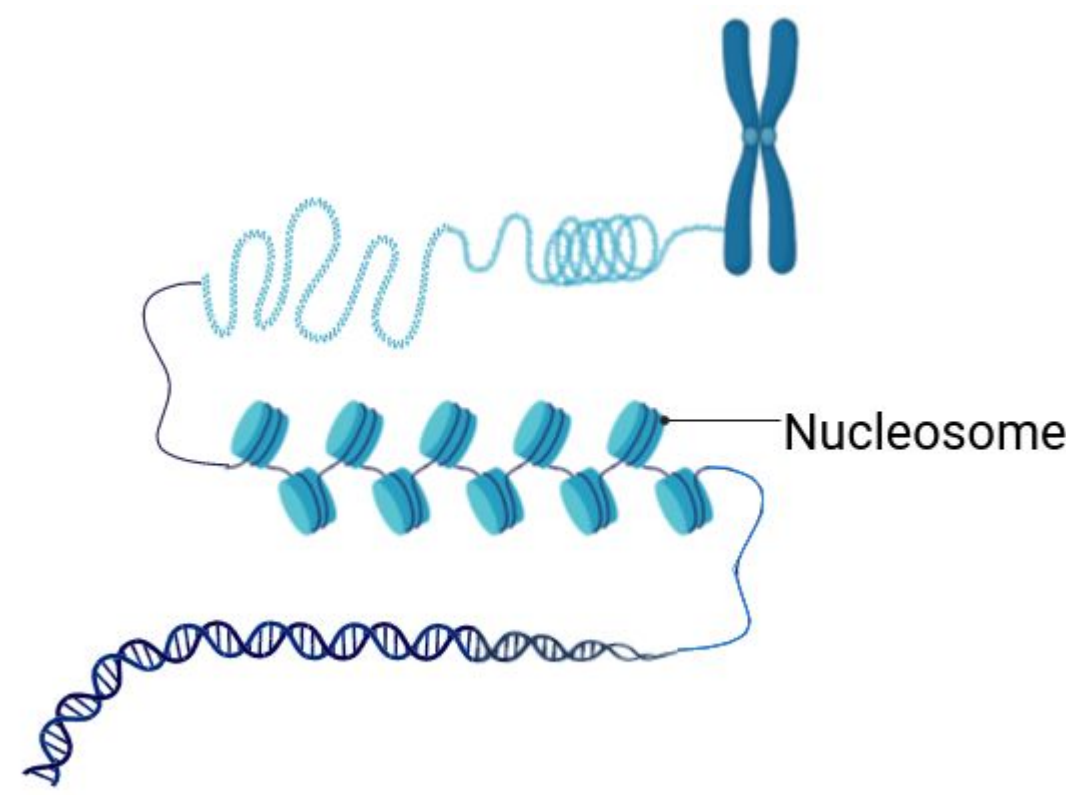
Distributed under a Creative Commons Attribution 4.0 International License

Nucleosome positioning in yeast: from prediction to synthetic design

Alex Westbrook, Etienne Routhier, Julien Mozziconacci

Introduction

Nucleosomes are the primary elements of DNA folding within eukaryotic cells, essential for both **chromosome structure** and **gene regulation**. Nucleosomal tracks are derived experimentally from MNase sequencing. Using **deep neural networks**, these tracks can be predicted from the DNA sequence alone in yeast.

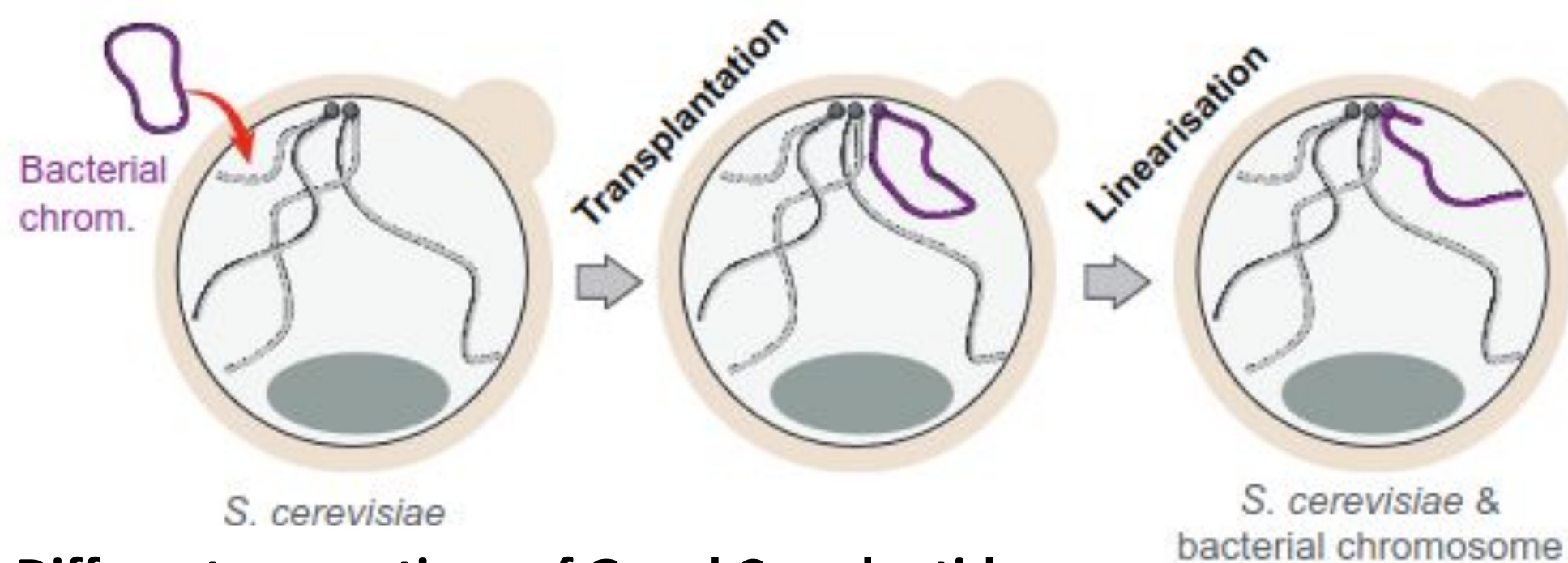


Objectives

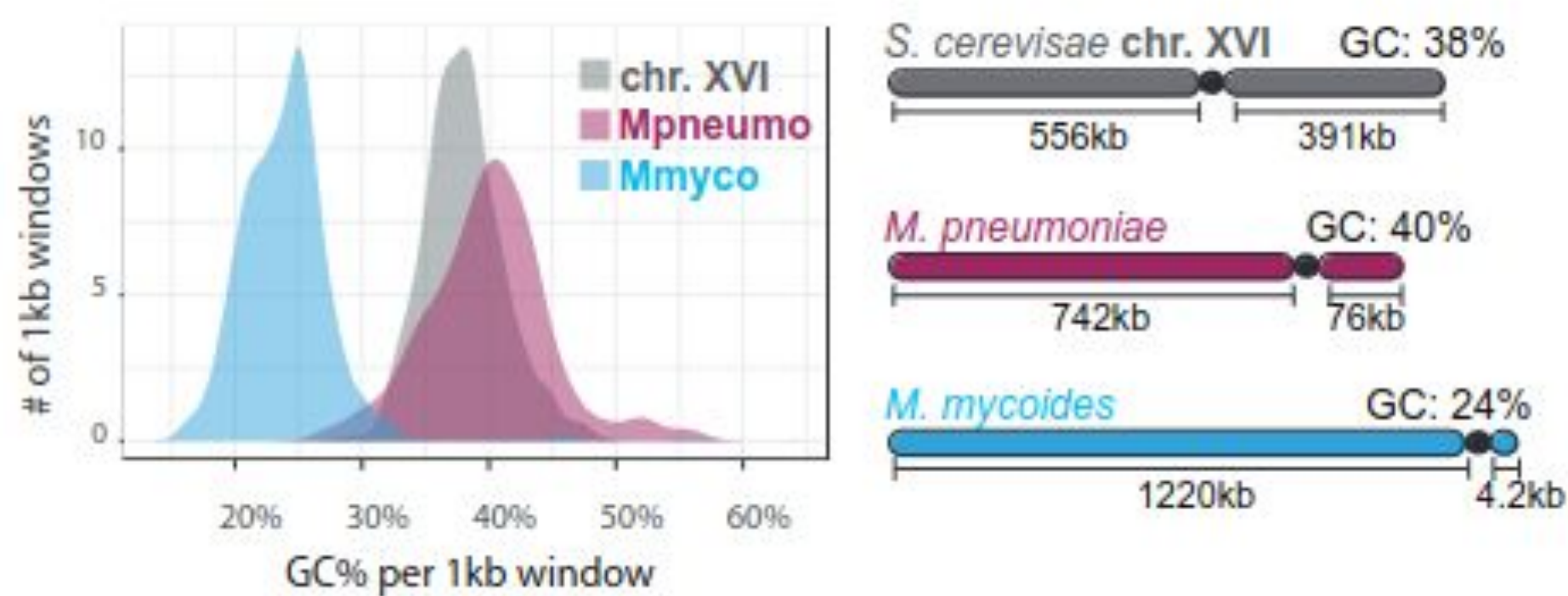
- How well does a yeast model **generalize on exogenous sequences** ?
- How can this model help in **designing new sequences** ?

Exogenous sequences

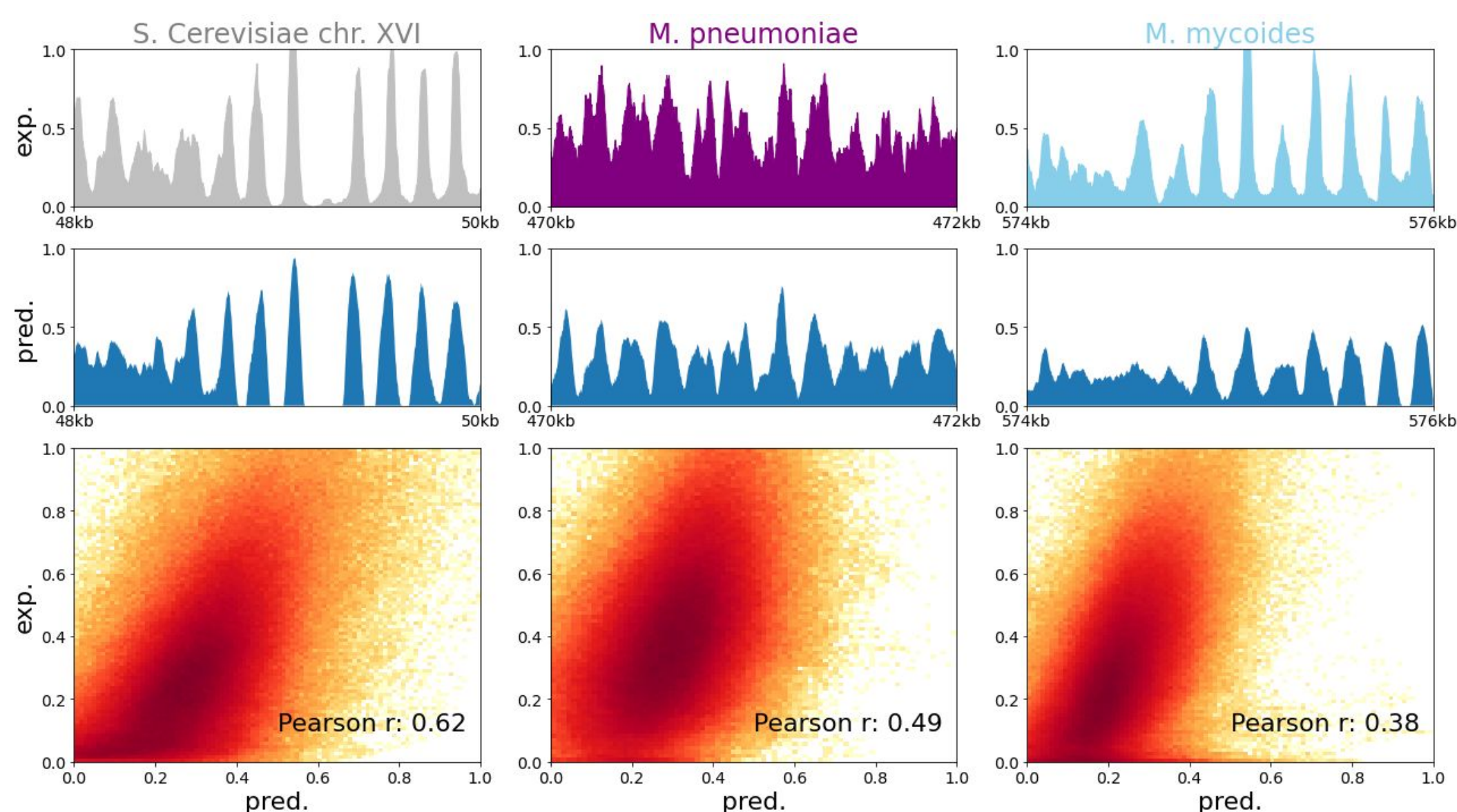
Yeast strains with a bacterial chromosome



Different proportions of G and C nucleotides

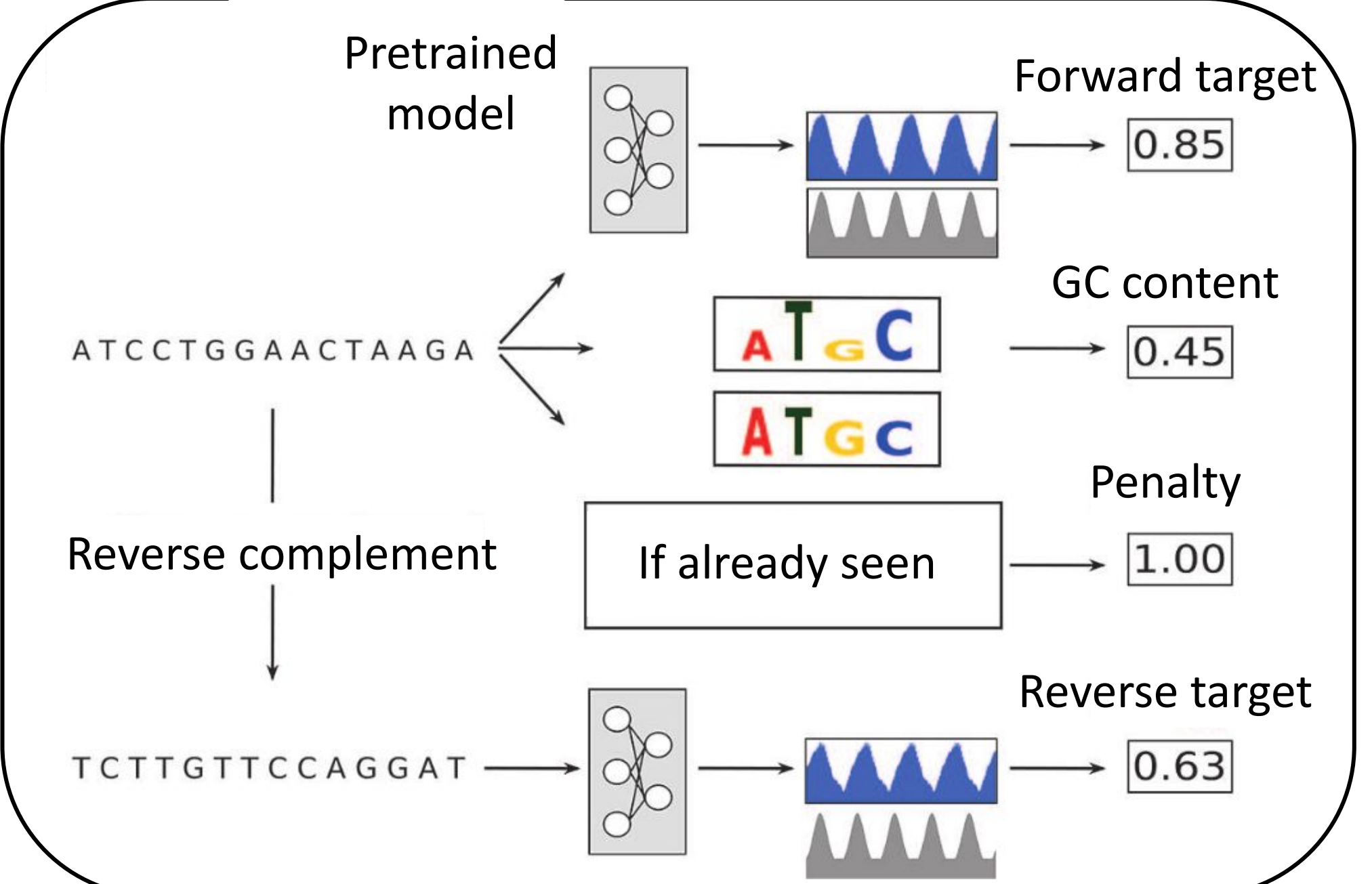
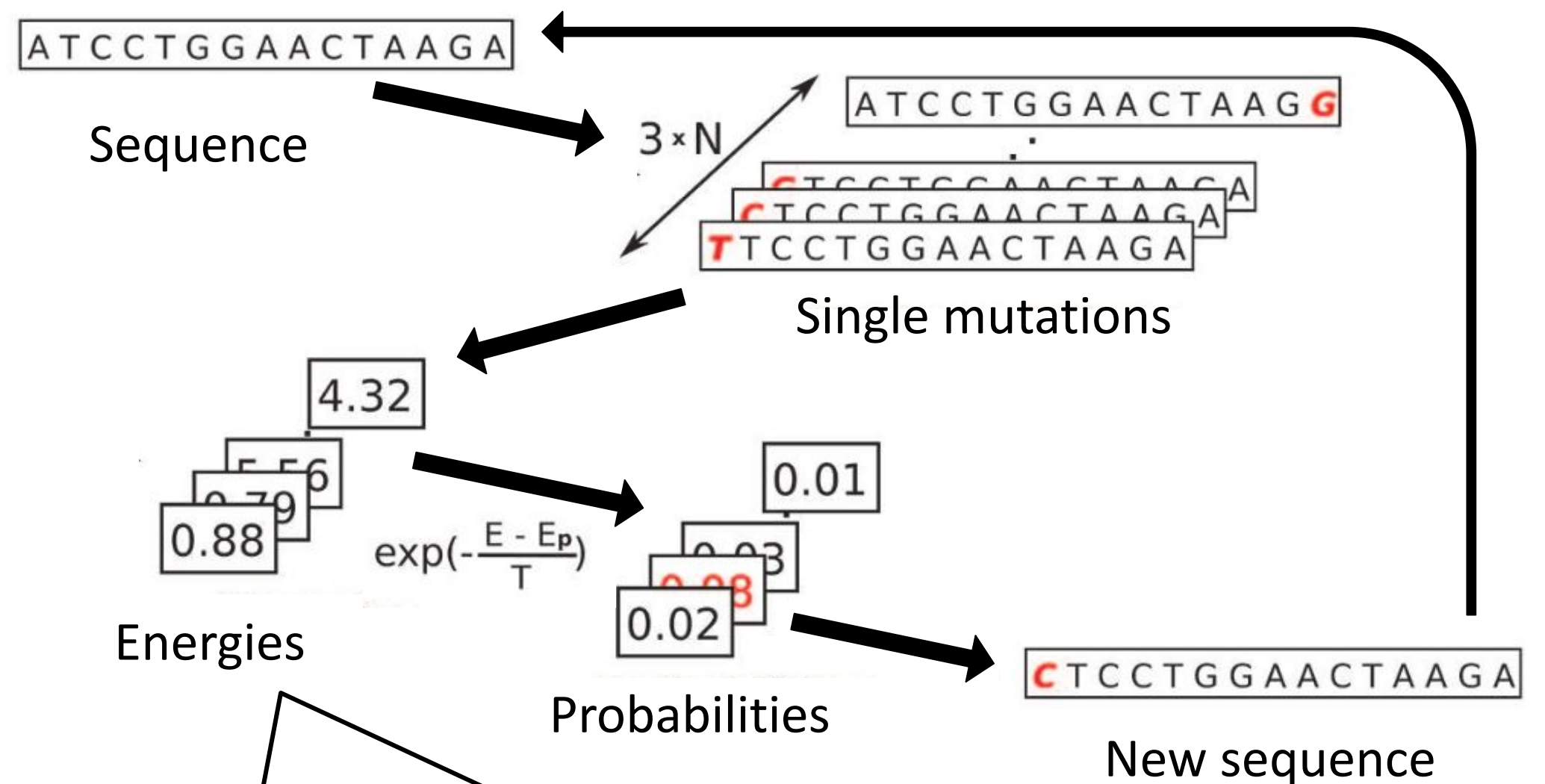


Sample of the predicted and experimental tracks on each sequence, and correlation plots over whole chromosome



Sequence design methodology

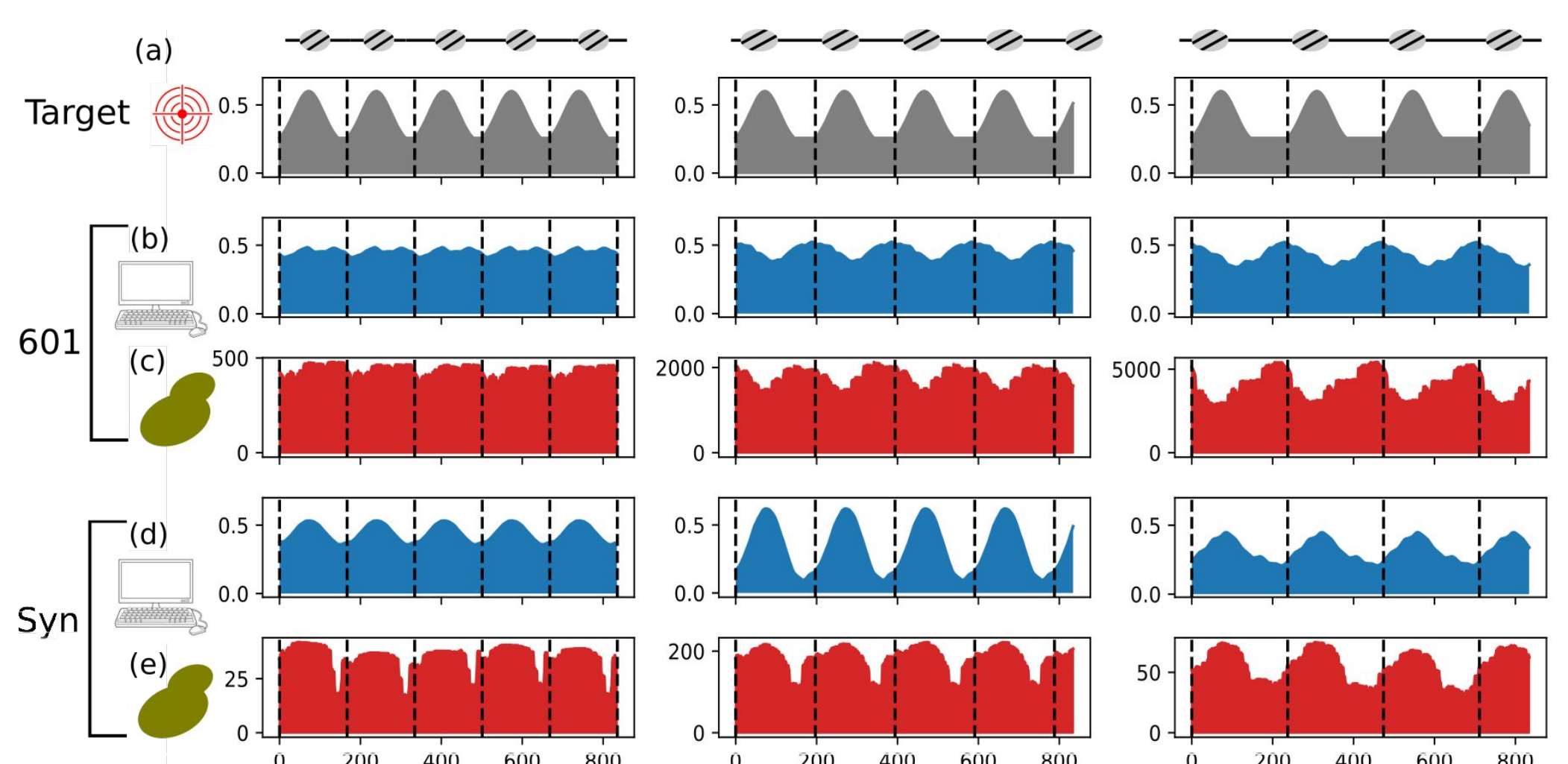
Multi-step mutation selection with model-based energies



Experimental validation

Integration of repeated synthetic sequences in yeast, with **nucleosomal repeat length of 167bp** (naturally occurring in yeast), **197bp** and **237bp** (highest known in eukaryotes)

Comparison to the **601 sequence**, designed in vitro to position nucleosomes preferentially



Conclusion

We show that a nucleosome model trained on yeast chromatin can **accurately predict nucleosome tracks of bacterial chromosomes** inside the yeast nucleus, even when there are notable differences in GC content.

With a computational approach based on deep learning prediction, we designed and validated **regular nucleosome arrays with tailored nucleosomal repeat length** in yeast, much larger than the natural one.

Our next step is to integrate modelling of **polymerase** and **cohesin** binding into our designs, to get better control over the yeast chromatin.



Chapard et al. "Exogenous chromosomes reveal how sequence composition drives chromatin assembly, activity, folding and compartmentalization" *bioRxiv* doi: 10.1101/2022.12.21.520625

Routhier et al. "In silico design of DNA sequences for in vivo nucleosome positioning" *bioRxiv* doi: 10.1101/2023.05.15.540782

Routhier et al. "Genome-wide prediction of DNA mutation effect on nucleosome positions for yeast synthetic genomics" *Genome Research* 2021. 31: 317-326

Contact: Alex Westbrook alex.westbrook@mnhn.fr Julien Mozziconacci julien.mozziconacci@mnhn.fr