A027 · Modeling the Competition Between Transcription Factors and DNA Repair Enzymes for Recognition of DNA Mismatches

Presenter: Anthony Lau — UMass Chan Medical School

Authors: Anthony Lau, Kyle Pinheiro, Wei Zhu, Harshit Sahay, Raluca Gordân

Session: Day 1 · Thu Oct 1, 2026 · 2:15–4:15 PM

DNA mismatches occur frequently throughout the human genome due to either nucleotide misincorporation events that occur during DNA replication or spontaneous deamination of 5-methylcytosine. During DNA repair, the initial step in most DNA repair pathways – including DNA mismatch repair and base excision repair – is the precise recognition of the DNA lesion by the appropriate DNA repair enzyme. Here, we show that transcription factors can directly compete and interfere with the recognition of DNA mismatches by the DNA repair enzymes MutSα and TDG/MBD4, the primary initiators of the DNA mismatch repair and base excision repair pathways, respectively. We demonstrate this mechanism of competition using high-density DNA arrays, which allow for high-throughput measurement of protein binding levels to tens of thousands of DNA sequences containing DNA mismatches. Using the competition data, we train and evaluate machine learning models to predict the reduction in repair enzyme binding levels for unseen DNA sequences containing DNA mismatches, in the presence of transcription factors. We observe that transcription factors outcompete DNA repair enzymes for high affinity DNA mismatches depending on the sequence contexts and the position of the mismatch within or around the transcription factor binding site core. We found that this mechanism of competition can be modeled accurately, even when using only the independent transcription factor and DNA repair enzyme binding levels as input features. These findings implicate competition between transcription factors and DNA repair enzymes as a major determinant of the somatic hypermutation at transcription factor binding sites observed in cancer genomes.

Modeling the Competition Between Transcription Factors and DNA Repair Enzymes for Recognition of DNA Mismatches

Anthony Lau[1]*, Kyle Pinheiro[2], Wei Zhu[2], Harshit Sahay[2], Raluca Gordân[1]

[1]Department of Genomics and Computational Biology, University of Massachusetts Chan Medical School, Worcester, MA 01605, USA
[2]Center for Advanced Genomic Technologies, Duke University, Durham, NC 27708, USA

DNA mismatches occur frequently throughout the human genome due to either nucleotide misincorporation events that occur during DNA replication or spontaneous deamination of 5-methylcytosine. During DNA repair, the initial step in most DNA repair pathways – including DNA mismatch repair and base excision repair – is the precise recognition of the DNA lesion by the appropriate DNA repair enzyme. Here, we show that transcription factors can directly compete and interfere with the recognition of DNA mismatches by the DNA repair enzymes MutSα and TDG/MBD4, the primary initiators of the DNA mismatch repair and base excision repair pathways, respectively. We demonstrate this mechanism of competition using high-density DNA arrays, which allow for high-throughput measurement of protein binding levels to tens of thousands of DNA sequences containing DNA mismatches. Using the competition data, we train and evaluate machine learning models to predict the reduction in repair enzyme binding levels for unseen DNA sequences containing DNA mismatches, in the presence of transcription factors. We observe that transcription factors outcompete DNA repair enzymes for high affinity DNA mismatches depending on the sequence contexts and the position of the mismatch within or around the transcription factor binding site core. We found that this mechanism of competition can be modeled accurately, even when using only the independent transcription factor and DNA repair enzyme binding levels as input features. These findings implicate competition between transcription factors and DNA repair enzymes as a major determinant of the somatic hypermutation at transcription factor binding sites observed in cancer genomes.