

Researchers develop novel data representation for transcription factor-binding sequences

October 20 2023



Credit: Unsplash/CC0 Public Domain

The diverse characteristics of the human body's various cells are reflected in their gene expression patterns. The regulation of such gene expression is based on transcription factors that bind to specific sequences in the genome.

The elucidation of unique transcription factor-binding sequences for each cell type is pivotal in unraveling the regulatory mechanisms governing [gene expression](#) within these cell types. Nonetheless, a comprehensive understanding of transcription factor-binding sequences, encompassing commonalities and variations across transcription factor types and cell types, remains elusive.

Using data concerning the binding sites of numerous human transcription factors, researchers at the University of Tsukuba have developed "MOCCS profiles," a novel data representation for transcription factor-binding sequences and analyzed binding sequences across various transcription factors and cell types. The work is [published](#) in the journal *BMC Genomics*.

Their findings reveal that approximately half of the examined transcription factors possess distinct binding sequences for specific cell types. Moreover, utilizing the MOCCS profiles, the researchers developed an index to predict the effect of single nucleotide polymorphisms (SNPs) on DNA binding of transcription factors and showed that the effect of disease-related SNPs on transcription factor binding can be appropriately evaluated from the perspective of transcription factors and [cell types](#).

The MOCCS profiles hold vast potential for various applications such as combining with epigenomic data to decipher cell type-specific gene expression regulatory mechanisms and evaluating the impact of somatic mutations in cancer cells on the binding of [transcription factors](#).

More information: Saeko Tahara et al, Transcription factor-binding k-mer analysis clarifies the cell type dependency of binding specificities and cis-regulatory SNPs in humans, *BMC Genomics* (2023). [DOI: 10.1186/s12864-023-09692-9](https://doi.org/10.1186/s12864-023-09692-9)

Provided by University of Tsukuba

Citation: Researchers develop novel data representation for transcription factor-binding sequences (2023, October 20) retrieved 2 May 2024 from <https://phys.org/news/2023-10-representation-transcription-factor-binding-sequences.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.