

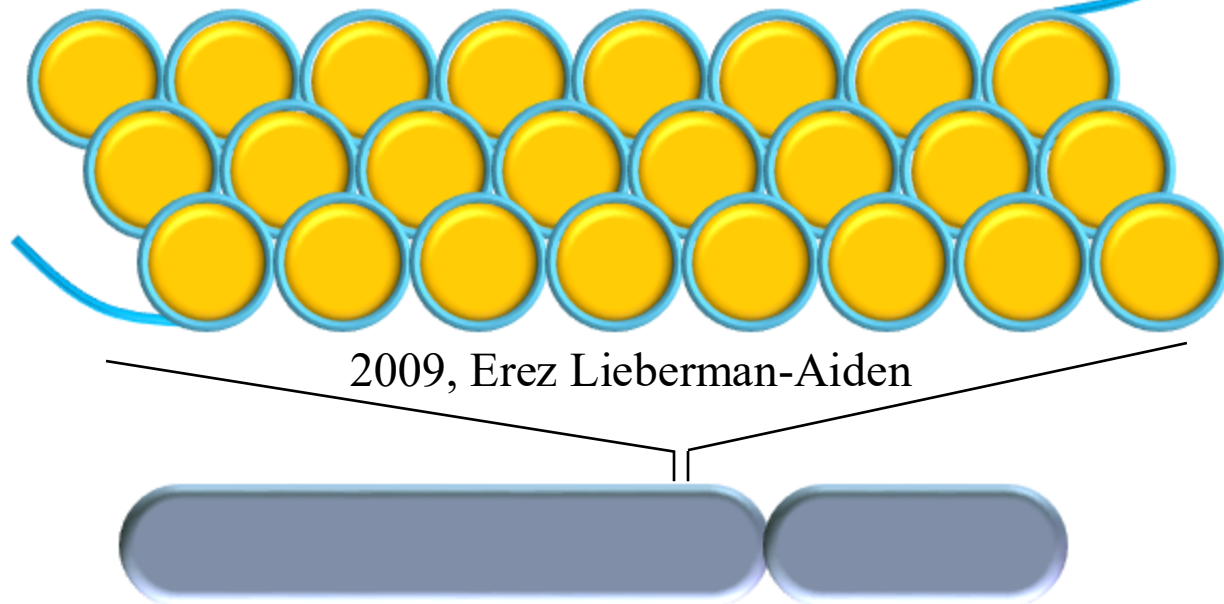
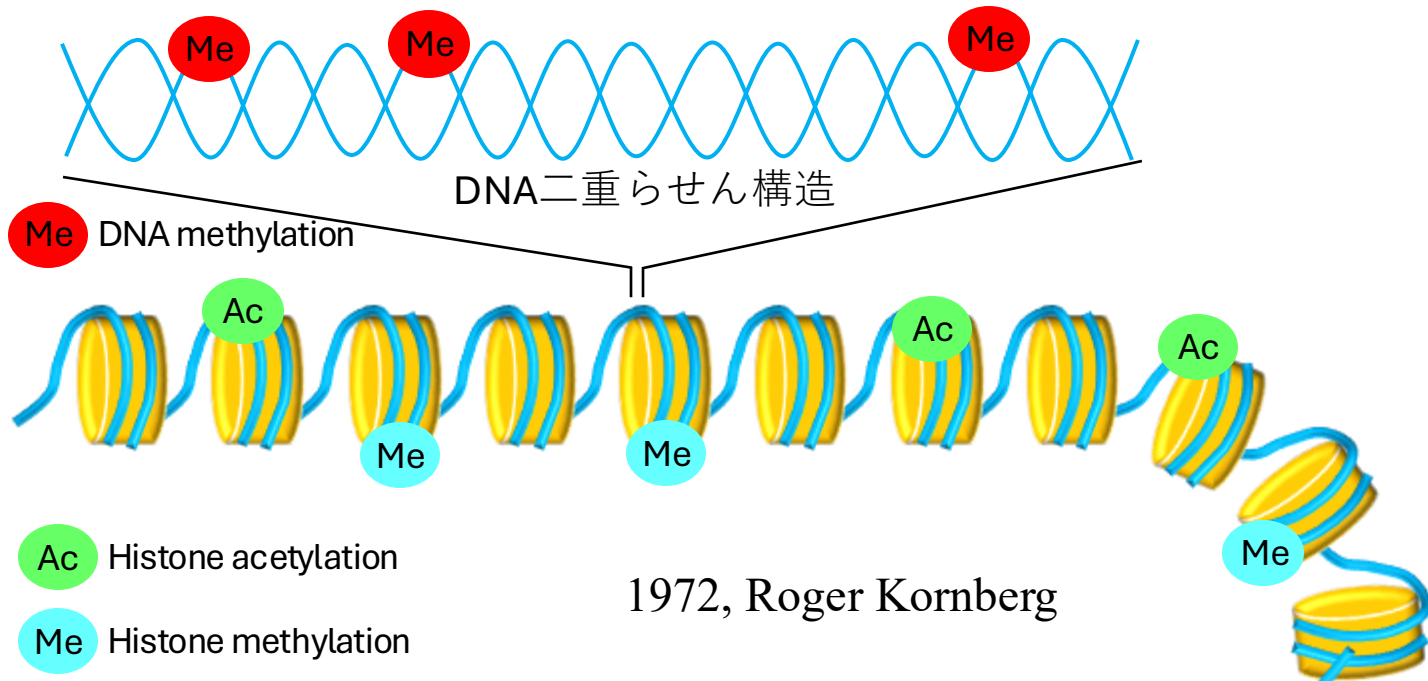
Genome Language Model

今後の方向性

森下 真一

東京大学

2025年7月30日 @ ROIS

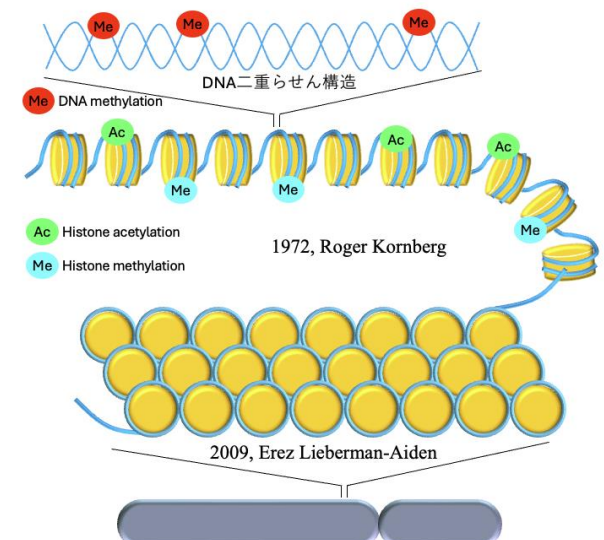
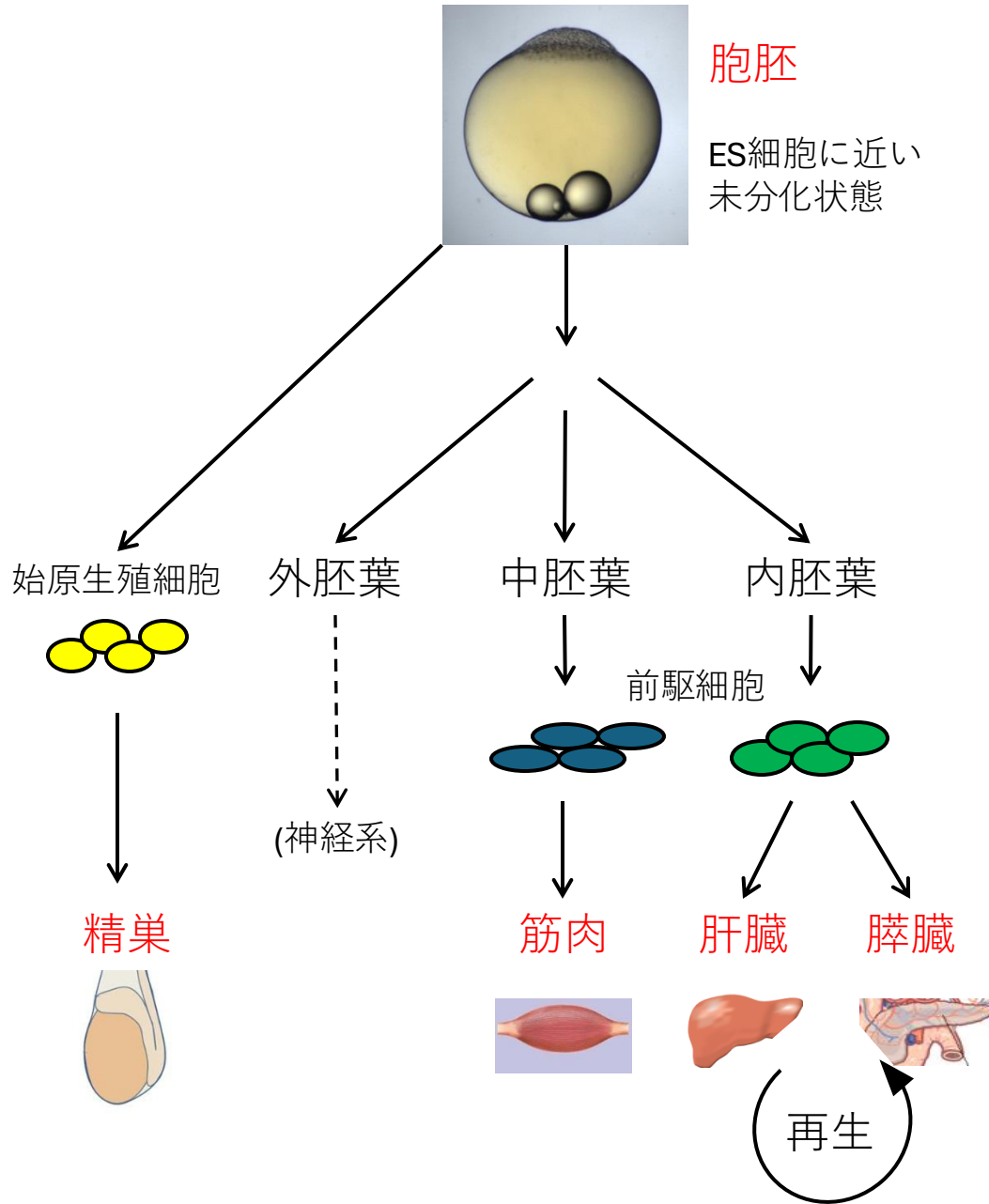


発生過程では多分化能が
獲得・維持される

疾患における変化

変化しやすい指標:

- 遺伝子発現量
- Epigenetic code
 - ✓ DNAメチル化
 - ✓ ヒストン修飾
 - ✓ ヌクレオソーム分布
- DNA 3次元折畳み構造



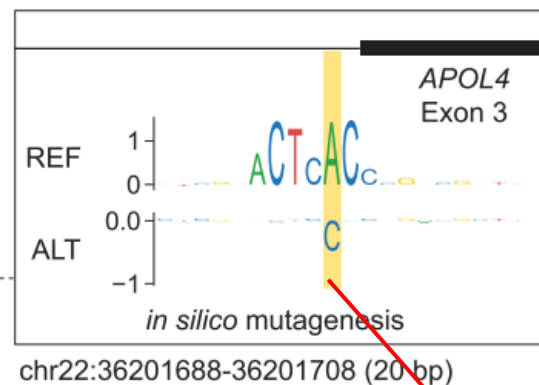
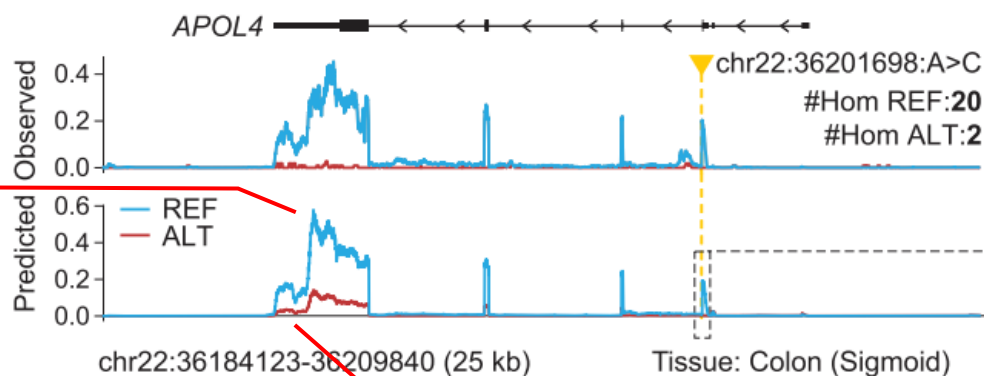
Variant (変異) は病気の診断に使われ重要

疾患 X に影響する variant を新たに探してください

- AlhphaGenome 論文の目玉も variant “advancing regulatory **variant** effect prediction with a unified DNA sequence model. preprint (2025/06/25).”

ヒントが不可欠 variant 探索用データ (UK BioBank, GWAS DB, 家系情報)
遺伝子発現量、ゲノム修飾

Predictions for a known eQTL



遺伝子発現
量の差

これほど差
が出るか？

過去の文献で説明可能な例があるか？

探索した
variant を
与える

疾患 X に影響する variant を新たに探してください

Genome Language Model への問合せ例 (ChatGPT風)

1. UK BioBank 50万人のデータから候補遺伝子上の variant を探索し、
2. 周辺の転写物発現量(total RNA-seq)が顕著に上昇する variant とその組織を AlphaGenome で推定し、
3. 転写量が有意に上昇する variant を見つけ、過去の文献の情報にあるか否かを確認してください

ChatGPT では望ましい回答は無し

GLM の目標設定案

- 疾患関連 variant の探索
- 高精度ゲノムの活用
- Precision Medicine
- 項目
 - 構造多型 (variant) の分析
 - 家系データ分析 孤発性疾患と *de novo* variant
 - AlphaGenome, Enformer, Nucleotide Transformer 等の高速実行
 - データ UK BioBank, HPRC, ENCODE, 4D Nucleosome 等の取込

関連研究 Zeng, W., Guo, H., Liu, Q. & Wong, W. H. Improving polygenic prediction from whole-genome sequencing data by leveraging predicted epigenomic features. Proc. Natl. Acad. Sci. 122, 2017 (2025).

高精度ゲノム
Human
Pangenome
Reference
Consortium
(HPRC)
代表
Benedict Paten,
Karen Miga @
UCSC

データ公開
2025/5～

Human Pangenome 関係の論文
(今月号の *Nature*)

Article

Structural variation in 1,019 diverse humans
based on long-read sequencing

<https://doi.org/10.1038/s41586-025-09290-7>
Received: 22 April 2024
Accepted: 16 June 2025
Published online: 23 July 2025
Open access
 Check for updates

Siegfried Schloissnig^{1,10}, Samarendra Panj^{2,3,10}, Jana Ebler^{2,3}, Carsten Hain⁴, Vasiliki Tsapalou⁴, Arda Söylev^{2,3}, Patrick Hühner^{1,5}, Hufsah Ashraf^{2,3}, Timofey Prodanov^{2,3}, Mila Asparuhova^{1,6}, Hugo Magalhães^{2,3}, Wolfram Höps⁷, Jesus Emiliano Sotelo-Fonseca^{8,9}, Tomas Fitzgerald¹⁰, Walter Santana-Garcia¹⁰, Ricardo Moreira-Pinhal^{11,12}, Sarah Hunt¹⁰, Francy J. Pérez-Llanos^{13,14}, Tassilo Erik Wollenweber¹³, Sugirthan Sivalingam¹⁵, Dagmar Wiczorek¹⁶, Mario Cáceres^{11,12,16}, Christian Gilissen⁷, Ewan Birney¹⁰, Zhihao Ding¹⁷, Jan Nygaard Jensen¹⁷, Nikhil Podduturi¹⁷, Jan Stutzki¹⁸, Bernardo Rodríguez-Martín^{4,8,9,10}, Tobias Rausch^{1,10}, Tobias Marschall^{1,3,10} & Jan O. Korbei^{4,10,10}

Article

Complex genetic variation in nearly
complete human genomes

<https://doi.org/10.1038/s41586-025-09140-6>
Received: 23 September 2024
Accepted: 12 May 2025
Published online: 23 July 2025
Open access
 Check for updates

Glennis A. Logsdon^{1,2,4,2}, Peter Ebert^{3,4,4,2}, Peter A. Audano^{5,4,2}, Mark Loftus^{6,7,41,42}, David Porubsky¹, Jana Ebler^{4,8}, Feyza Yilmaz⁵, Pille Hallast⁵, Timofey Prodanov^{4,8}, DongAhn Yoo¹, Carolyn A. Paisie⁵, William T. Harvey¹, Xuefang Zhao^{9,10,11}, Gianni V. Martino^{6,7,12}, Mir Henglin^{4,8}, Katherine M. Munson¹, Keon Rabbani¹³, Chen-Shan Chin¹⁴, Bida Gu¹³, Hufsah Ashraf^{4,8}, Stephan Scholz^{4,15}, Olanrewaju Austine-Orimoloye¹⁶, Parithi Balachandran¹, Marc Jan Bonder^{17,18,19}, Haoyu Cheng²⁰, Zechen Chong²¹, Jonathan Crabtree²², Mark Gerstein^{23,24}, Lisbeth A. Guethlein²⁵, Patrick Hasenfeld²⁶, Glenn Hickey²⁷, Kendra Hoekzema¹, Sarah E. Hunt¹⁶, Matthew Jensen^{23,24}, Yunzhe Jiang^{23,24}, Sergey Koren²⁸, Youngjun Kwon¹, Chong Li^{29,30}, Heng Li^{31,32}, Jiaqi Li^{23,24}, Paul J. Norman^{33,34}, Keisuke K. Oshima², Benedict Paten²⁷, Adam M. Phillippy²⁸, Nicholas R. Pollock³³, Tobias Rausch²⁶, Mikko Rautiainen³⁵, Yuwei Song²¹, Arda Söylev^{4,8}, Arvis Sulovari¹, Likhitha Surapaneni¹⁶, Vasiliki Tsapalou²⁶, Weichen Zhou³⁶, Ying Zhou³¹, Qihui Zhu^{3,37}, Michael C. Zody³⁸, Ryan E. Mills³⁹, Scott E. Devine²², Xinghua Shi^{29,30}, Michael E. Talkowski^{40,11}, Mark J. P. Chaisson¹³, Alexander T. Dilthey^{41,5}, Miriam K. Konkel^{42,7,5}, Jan O. Korbei^{26,5}, Charles Lee^{5,5}, Christine R. Beck^{1,5,39,5}, Evan E. Eichler^{1,40,5} & Tobias Marschall^{4,8,5}