preciseTAD: A machine learning framework for precise 3D domain boundary prediction at base-level resolution

References List

Al Bkhetan, Ziad, and Dariusz Plewczynski. 2018. “Three-Dimensional Epigenome Statistical Model: Genome-Wide Chromatin Looping Prediction.” *Sci Rep* 8 (1): 5217. <https://doi.org/10.1038/s41598-018-23276-8>.

Alipour, Elnaz, and John F Marko. 2012. “Self-Organization of Domain Structures by Dna-Loop-Extruding Enzymes.” *Nucleic Acids Research* 40 (22): 11202–12.

Ay, Ferhat, Timothy L Bailey, and William Stafford Noble. 2014. “Statistical Confidence Estimation for Hi-c Data Reveals Regulatory Chromatin Contacts.” *Genome Res* 24 (6): 999–1011. <https://doi.org/10.1101/gr.160374.113>.

Bonev, Boyan, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L Papadopoulos, Yaniv Lubling, Xiaole Xu, et al. 2017. “Multiscale 3D Genome Rewiring During Mouse Neural Development.” *Cell* 171 (3): 557–72.

Boulesteix, Anne-Laure, Silke Janitza, Jochen Kruppa, and Inke R König. 2012. “Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Bioinformatics.” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2 (6): 493–507.

Chang, Li-Hsin, Sourav Ghosh, and Daan Noordermeer. 2020. “TADs and Their Borders: Free Movement or Building a Wall?” *Journal of Molecular Biology* 432 (3): 643–52.

Chawla, Nitesh V, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. “SMOTE: Synthetic Minority over-Sampling Technique.” *Journal of Artificial Intelligence Research* 16: 321–57.

Cresswell, Kellen G, John C Stansfield, and Mikhail G Dozmorov. 2020. “SpectralTAD: An R Package for Defining a Hierarchy of Topologically Associated Domains Using Spectral Clustering.” *BMC Bioinformatics* 21 (1): 319. <https://doi.org/10.1186/s12859-020-03652-w>.

Davidson, Iain F, Benedikt Bauer, Daniela Goetz, Wen Tang, Gordana Wutz, and Jan-Michael Peters. 2019. “DNA Loop Extrusion by Human Cohesin.” *Science* 366 (6471): 1338–45. <https://doi.org/10.1126/science.aaz3418>.

Dixon, Jesse R, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. 2012. “Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions.” *Nature* 485 (7398): 376.

Durand, Neva C, James T Robinson, Muhammad S Shamim, Ido Machol, Jill P Mesirov, Eric S Lander, and Erez Lieberman Aiden. 2016. “Juicebox Provides a Visualization System for Hi-c Contact Maps with Unlimited Zoom.” *Cell Systems* 3 (1): 99–101.

Durand, Neva C, Muhammad S Shamim, Ido Machol, Suhas SP Rao, Miriam H Huntley, Eric S Lander, and Erez Lieberman Aiden. 2016. “Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-c Experiments.” *Cell Systems* 3 (1): 95–98.

Flyamer, Ilya M, Johanna Gassler, Maxim Imakaev, Hugo B Brandão, Sergey V Ulianov, Nezar Abdennur, Sergey V Razin, Leonid A Mirny, and Kikuë Tachibana-Konwalski. 2017. “Single-Nucleus Hi-c Reveals Unique Chromatin Reorganization at Oocyte-to-Zygote Transition.” *Nature* 544 (7648): 110–14. <https://doi.org/10.1038/nature21711>.

Forcato, Mattia, Chiara Nicoletti, Koustav Pal, Carmen Maria Livi, Francesco Ferrari, and Silvio Bicciato. 2017. “Comparison of Computational Methods for Hi-c Data Analysis.” *Nature Methods* 14 (7): 679.

Franke, Martin, Daniel M Ibrahim, Guillaume Andrey, Wibke Schwarzer, Verena Heinrich, Robert Schöpflin, Katerina Kraft, et al. 2016. “Formation of New Chromatin Domains Determines Pathogenicity of Genomic Duplications.” *Nature* 538 (7624): 265–69. <https://doi.org/10.1038/nature19800>.

Fraser, James, Carmelo Ferrai, Andrea M Chiariello, Markus Schueler, Tiago Rito, Giovanni Laudanno, Mariano Barbieri, et al. 2015. “Hierarchical Folding and Reorganization of Chromosomes Are Linked to Transcriptional Changes in Cellular Differentiation.” *Molecular Systems Biology* 11 (12).

Fudenberg, Geoffrey, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. 2016. “Formation of Chromosomal Domains by Loop Extrusion.” *Cell Reports* 15 (9): 2038–49.

Hansen, Anders S, Claudia Cattoglio, Xavier Darzacq, and Robert Tjian. 2018. “Recent Evidence That Tads and Chromatin Loops Are Dynamic Structures.” *Nucleus* 9 (1): 20–32.

Harrold, Caroline L, Matthew E Gosden, Lars L P Hanssen, Rosa J Stolper, Damien J Downes, Jelena M. Telenius, Daniel Biggs, et al. n.d. “A Functional Overlap Between Actively Transcribed Genes and Chromatin Boundary Elements.” *bioRxiv*, 2020.07.01.182089. <https://doi.org/10.1101/2020.07.01.182089>.

Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C Lin, Peter Laslo, Jason X Cheng, Cornelis Murre, Harinder Singh, and Christopher K Glass. 2010. “Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities.” *Mol Cell* 38 (4): 576–89. <https://doi.org/10.1016/j.molcel.2010.05.004>.

Hnisz, Denes, Abraham S Weintraub, Daniel S Day, Anne-Laure Valton, Rasmus O Bak, Charles H Li, Johanna Goldmann, et al. 2016. “Activation of Proto-Oncogenes by Disruption of Chromosome Neighborhoods.” *Science* 351 (6280): 1454–8.

Kai, Yan, Jaclyn Andricovich, Zhouhao Zeng, Jun Zhu, Alexandros Tzatsos, and Weiqun Peng. 2018. “Predicting Ctcf-Mediated Chromatin Interactions by Integrating Genomic and Epigenomic Features.” *Nature Communications* 9 (1): 1–14.

Krefting, Jan, Miguel A Andrade-Navarro, and Jonas Ibn-Salem. 2018. “Evolutionary Stability of Topologically Associating Domains Is Associated with Conserved Gene Regulation.” *BMC Biology* 16 (1): 1–12.

Krietenstein, Nils, Sameer Abraham, Sergey V Venev, Nezar Abdennur, Johan Gibcus, Tsung-Han S Hsieh, Krishna Mohan Parsi, et al. 2020. “Ultrastructural Details of Mammalian Chromosome Architecture.” *Mol Cell* 78 (3): 554–565.e7. <https://doi.org/10.1016/j.molcel.2020.03.003>.

Kuhn, Max. 2012. “The Caret Package.” *R Foundation for Statistical Computing, Vienna, Austria. URL Https://Cran. R-Project. Org/Package= Caret*.

Li, Xiao, Ziyang An, and Zhihua Zhang. 2019. “Comparison of Computational Methods for 3D Genome Analysis at Single-Cell Hi-c Level.” *Methods*.

Libbrecht, Maxwell W, Ferhat Ay, Michael M Hoffman, David M Gilbert, Jeffrey A Bilmes, and William Stafford Noble. 2015. “Joint Annotation of Chromatin State and Chromatin Conformation Reveals Relationships Among Domain Types and Identifies Domains of Cell-Type-Specific Expression.” *Genome Research* 25 (4): 544–57.

Lieberman-Aiden, Erez, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.” *Science* 326 (5950): 289–93.

Lupiáñez, Darı́o G, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, et al. 2015. “Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions.” *Cell* 161 (5): 1012–25. <https://doi.org/10.1016/j.cell.2015.04.004>.

Lupiáñez, Darı́o G, Malte Spielmann, and Stefan Mundlos. 2016. “Breaking Tads: How Alterations of Chromatin Domains Result in Disease.” *Trends in Genetics* 32 (4): 225–37.

Meaburn, Karen J, Erik Cabuy, Gisele Bonne, Nicolas Levy, Glenn E Morris, Giuseppe Novelli, Ian R Kill, and Joanna M Bridger. 2007. “Primary Laminopathy Fibroblasts Display Altered Genome Organization and Apoptosis.” *Aging Cell* 6 (2): 139–53. <https://doi.org/10.1111/j.1474-9726.2007.00270.x>.

Mendoza-Vargas, Alfredo, Leticia Olvera, Maricela Olvera, Ricardo Grande, Leticia Vega-Alvarado, Blanca Taboada, Verónica Jimenez-Jacinto, et al. 2009. “Genome-Wide Identification of Transcription Start Sites, Promoters and Transcription Factor Binding Sites in E. Coli.” *PLoS One* 4 (10): e7526.

Mirny, Leonid A, Maxim Imakaev, and Nezar Abdennur. 2019. “Two Major Mechanisms of Chromosome Organization.” *Curr Opin Cell Biol* 58 (June): 142–52. <https://doi.org/10.1016/j.ceb.2019.05.001>.

Moore, Benjamin L, Stuart Aitken, and Colin A Semple. 2015. “Integrative Modeling Reveals the Principles of Multi-Scale Chromatin Boundary Formation in Human Nuclear Organization.” *Genome Biol* 16 (May): 110. <https://doi.org/10.1186/s13059-015-0661-x>.

Mourad, Raphael, and Olivier Cuvier. 2016. “Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation.” *PLoS Computational Biology* 12 (5): e1004908.

Nora, Elphège P, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. “Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre.” *Nature* 485 (7398): 381.

Prati, Ronaldo C, GEAPA Batista, and Maria Carolina Monard. 2011. “A Survey on Graphical Methods for Classification Predictive Performance Evaluation.” *IEEE Transactions on Knowledge and Data Engineering* 23 (11): 1601–18.

Ramirez, Fidel, Devon P Ryan, Bjorn Gruning, Vivek Bhardwaj, Fabian Kilpert, Andreas S Richter, Steffen Heyne, Friederike Dundar, and Thomas Manke. 2016. “DeepTools2: A Next Generation Web Server for Deep-Sequencing Data Analysis.” *Nucleic Acids Research* 44 (W1): W160–W165.

Rao, Suhas SP, Miriam H Huntley, Neva C Durand, Elena K Stamenova, Ivan D Bochkov, James T Robinson, Adrian L Sanborn, et al. 2014. “A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.” *Cell* 159 (7): 1665–80.

Salameh, Tarik J, Xiaotao Wang, Fan Song, Bo Zhang, Sage M Wright, Chachrit Khunsriraksakul, and Feng Yue. 2019. “A Supervised Learning Framework for Chromatin Loop Detection in Genome-Wide Contact Maps.” *bioRxiv*, 739698.

Sanborn, Adrian L, Suhas SP Rao, Su-Chen Huang, Neva C Durand, Miriam H Huntley, Andrew I Jewett, Ivan D Bochkov, et al. 2015. “Chromatin Extrusion Explains Key Features of Loop and Domain Formation in Wild-Type and Engineered Genomes.” *Proceedings of the National Academy of Sciences* 112 (47): E6456–E6465.

Schmitt, Anthony D, Ming Hu, and Bing Ren. 2016. “Genome-Wide Mapping and Analysis of Chromosome Architecture.” *Nature Reviews Molecular Cell Biology* 17 (12): 743.

Schreiber, Jacob, Timothy Durham, Jeffrey Bilmes, and William Stafford Noble. 2020. “Avocado: A Multi-Scale Deep Tensor Factorization Method Learns a Latent Representation of the Human Epigenome.” *Genome Biol* 21 (1): 81. <https://doi.org/10.1186/s13059-020-01977-6>.

Schreiber, Jacob, Ritambhara Singh, Jeffrey Bilmes, and William Stafford Noble. 2019. “A Pitfall for Machine Learning Methods Aiming to Predict Across Cell Types.” *bioRxiv*, 512434.

Sexton, Tom, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. 2012. “Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome.” *Cell* 148 (3): 458–72.

Stunnenberg, Hendrik G, International Human Epigenome Consortium, and Martin Hirst. 2016. “The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery.” *Cell* 167 (5): 1145–9. <https://doi.org/10.1016/j.cell.2016.11.007>.

Sun, James H, Linda Zhou, Daniel J Emerson, Sai A Phyo, Katelyn R Titus, Wanfeng Gong, Thomas G Gilgenast, et al. 2018. “Disease-Associated Short Tandem Repeats Co-Localize with Chromatin Domain Boundaries.” *Cell* 175 (1): 224–238.e15. <https://doi.org/10.1016/j.cell.2018.08.005>.

Taberlay, Phillippa C, Joanna Achinger-Kawecka, Aaron TL Lun, Fabian A Buske, Kenneth Sabir, Cathryn M Gould, Elena Zotenko, et al. 2016. “Three-Dimensional Disorganization of the Cancer Genome Occurs Coincident with Long-Range Genetic and Epigenetic Alterations.” *Genome Research* 26 (6): 719–31.

Tyner, Cath, Galt P Barber, Jonathan Casper, Hiram Clawson, Mark Diekhans, Christopher Eisenhart, Clayton M Fischer, et al. 2017. “The Ucsc Genome Browser Database: 2017 Update.” *Nucleic Acids Res* 45 (D1): D626–D634. <https://doi.org/10.1093/nar/gkw1134>.

Van Bortle, Kevin, Michael H Nichols, Li Li, Chin-Tong Ong, Naomi Takenaka, Zhaohui S Qin, and Victor G Corces. 2014. “Insulator Function and Topological Domain Border Strength Scale with Architectural Protein Occupancy.” *Genome Biology* 15 (5): R82.

Wei, Qiong, and Roland L Dunbrack Jr. 2013. “The Role of Balanced Training and Testing Data Sets for Binary Classifiers in Bioinformatics.” *PloS One* 8 (7): e67863.

Weinreb, Caleb, and Benjamin J Raphael. 2016. “Identification of Hierarchical Chromatin Domains.” *Bioinformatics* 32 (11): 1601–9.

Zufferey, Marie, Daniele Tavernari, Elisa Oricchio, and Giovanni Ciriello. 2018. “Comparison of Computational Methods for the Identification of Topologically Associating Domains.” *Genome Biol* 19 (1): 217. <https://doi.org/10.1186/s13059-018-1596-9>.