

Probing the contribution of DNA low complexity regions in transcriptional regulations

Abstract 4.7

Océane Cassan, Christophe Vroland, Laurent Bréhélin, Charles Lecellier

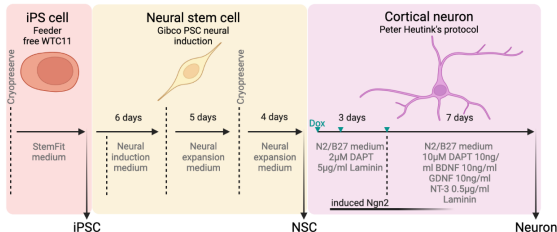


Montpellier Computational Regulatory Genomics group

October 3, 2023

Research question

Can we use DNA sequence to predict active **promoters and enhancers** during **neuron differentiation**? What is the role of **low complexity sequences** in this regulatory code?



Available data

Transcription Start Sites (TSS) delineated via ONT-CAGE (Table 4), tissue-specific expression, and candidate Cis-Regulatory Element signature (cCRE, SCREEN).

How can we extract functional sequence features from regulatory elements?

How can we extract functional sequence features from regulatory elements?

Unsupervised approaches such as dimension reduction:
SeqArchR [Sarvesh Nikumbh, 2022]

How can we extract functional sequence features from regulatory elements?

Unsupervised approaches such as dimension reduction:
SeqArchR [Sarvesh Nikumbh, 2022]

Supervised approaches linking sequence to expression:

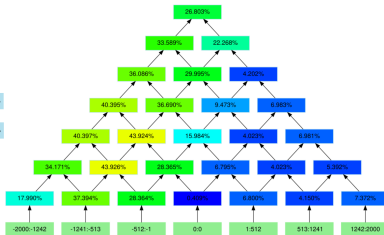
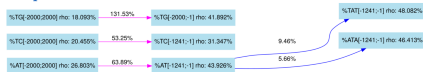
- Implicit features with deep learning → Interpretability?
- Explicit features modelling:
 - ① **TF binding motifs** (e.g JASPAR PWM)
 - ② **STR - microstallites** [Grapotte et al., 2021], Charles' talk
 - ③ **Biased kmer frequencies in informative regions**
 - Lattice search in binned sequence [Menichelli et al., 2021]
↳ **DExTER**
 - Hidden Markow Models delineating flexible regions
↳ **superHMM**

Extracting important low complexity regions : DExTER

Low complexity regions (LCR) delineated by supervised lattice search.

↳ **kmer frequency within region** as a predictor for TSS expression.

Step 1 - Feature extraction:



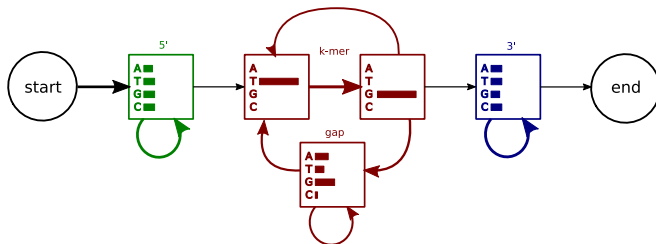
Step 2 - Feature selection and predictor learning:

	y	...	%GT [-1241,0]	%GG [-2000,-513]	%GC [-1241,2000]	%ATA [-1241,-1]	%TT [1,512]	...
gene 1	0.15		0.17	0.24	0.15	0.26	0.08	
gene 2	0.09		0.20	0.19	0.09	0.12	0.10	
gene 3	0.17		0.16	0.18	0.12	0.24	0.17	
...	...		...	...	...	...	...	

Extracting important low complexity regions : superHMM

Discriminative Hidden Markov Models (Laurent Bréhélin).

↳ **LCR size** as a predictor for TSS expression.

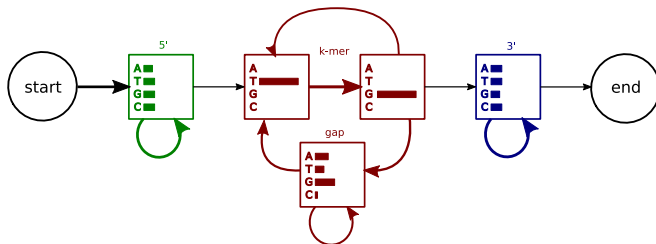


One sequence :

Extracting important low complexity regions : superHMM

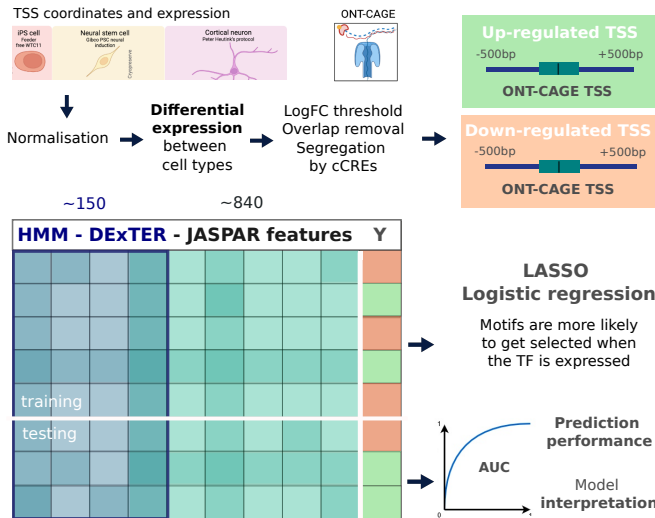
Discriminative Hidden Markov Models (Laurent Bréhélin).

↳ **LCR size** as a predictor for TSS expression.

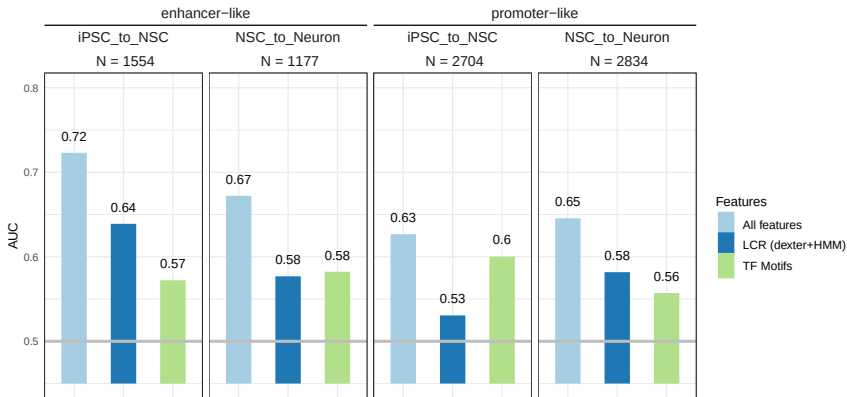


One sequence :

Can we predict the sequence features of TSS activity?

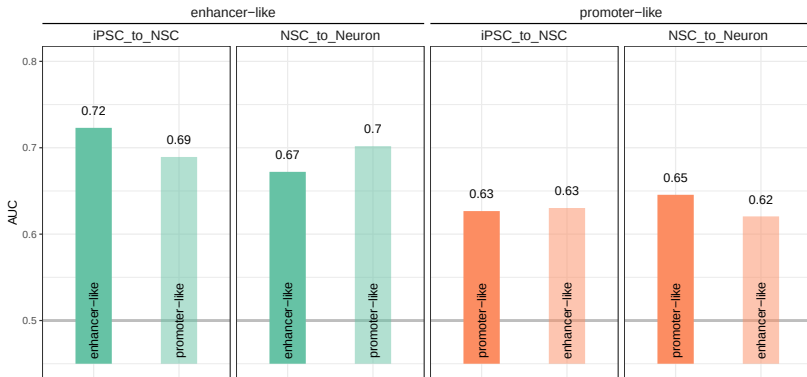


TSS change of activity during differentiation can be predicted



Both **LCR** and **TF motifs** contribute to model predictions.

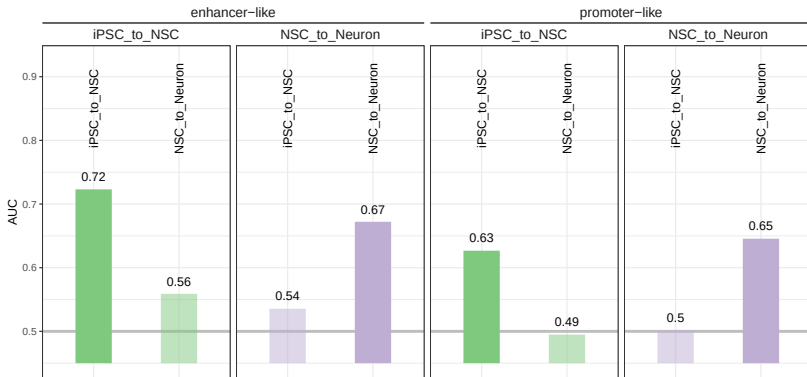
Similar rules govern the change in activity of enhancers and promoters



"A **unified architecture** of transcriptional regulatory elements"?

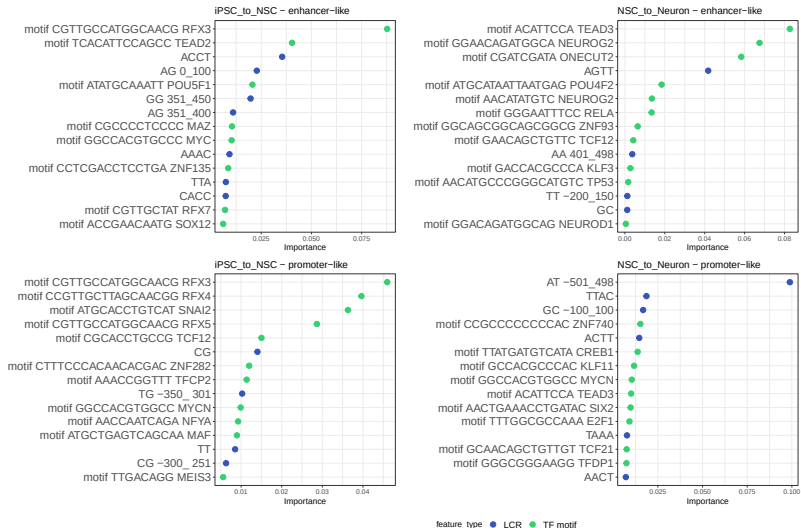
↳ Trends in Genetics [Andersson et al., 2015]

Distinct rules govern TSS changes in activity between cell transitions

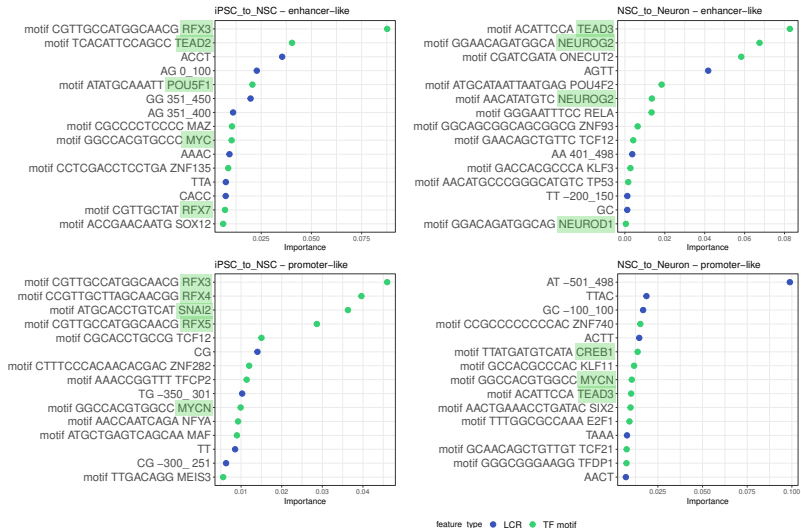


Classifiers are not interchangeable : they are **stage-specific**.

Important features are a mixture of LCR and TF motifs



Important features are a mixture of LCR and TF motifs



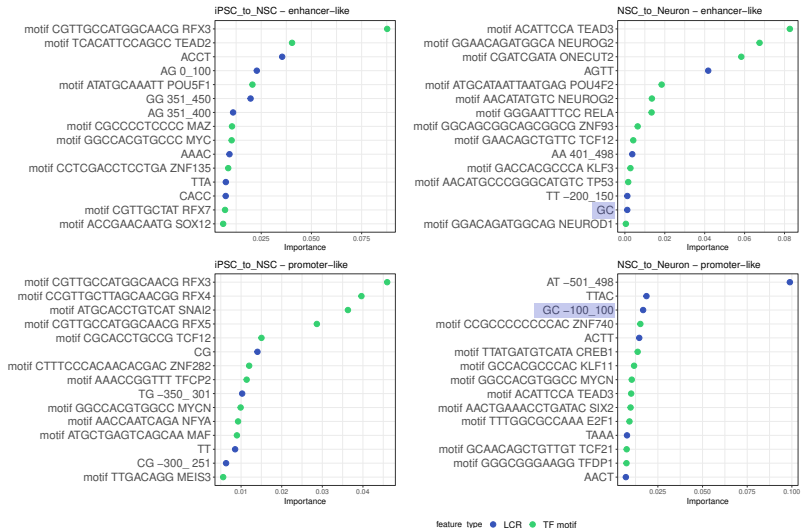
Important features are a mixture of LCR and TF motifs

Top TFs are documented for their role in neuron cell fate :

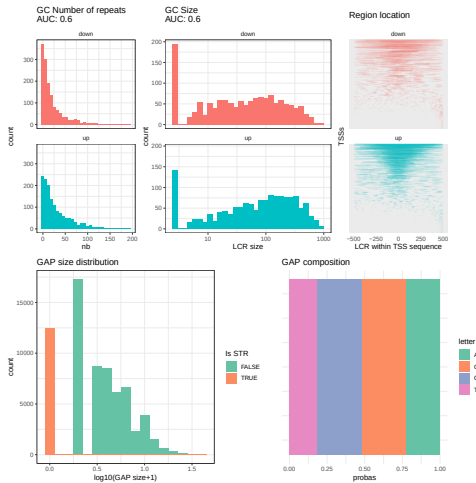
- **TEAD2** blocks Tbr1+ neuron formation, **TEAD1** and **TEAD3** promote this early fate
↳ [Mukhtar et al., 2020]
- **MYC proteins promote neuronal differentiation** by controlling the mode of progenitor cell division
↳ [Zinin et al., 2014]
- **RFX genes** are evolutionarily conserved TFs that act as master regulators of central nervous system development
↳ [Harris et al., 2021]

But also POU5 TFs, NEUGOG2, NEUROD1...

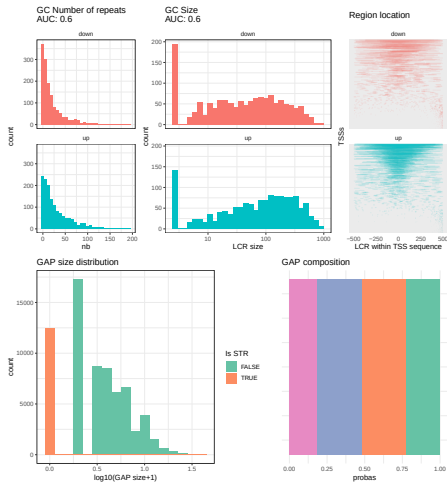
Important features are a mixture of LCR and TF motifs



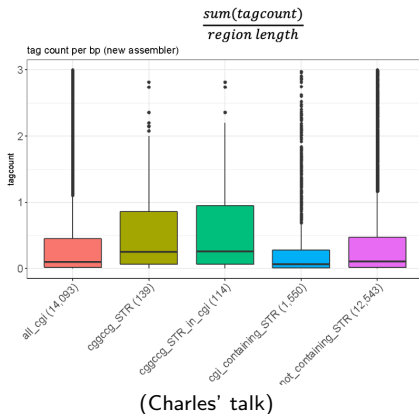
Important features are a mixture of LCR and TF motifs



Important features are a mixture of LCR and TF motifs



CG-rich regions identified by superHMM could help refine functional CpG islands



Conclusion

Main results

- A tangible part of stage specific **TSS activity can be predicted by sequence**, with LCR and TF motifs showing some **complementary**
- While enhancers and promoters share similar sequence rules, **distinct neuron differentiation stages have distinct sequence features**

Conclusion

Main results

- A tangible part of stage specific **TSS activity can be predicted by sequence**, with LCR and TF motifs showing some **complementary**
- While enhancers and promoters share similar sequence rules, **distinct neuron differentiation stages have distinct sequence features**

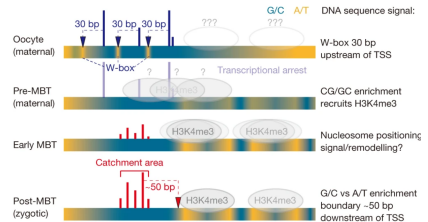
"Evidence that alternative transcriptional initiation is largely nonadaptive"? → PLOS Biology [Xu et al., 2019]

LETTER

doi:10.1038/nature12974

Two independent transcription initiation codes overlap on vertebrate core promoters

Vanja Haberle^{1,2*}, Nan Li^{3*}, Yavor Hadzhiev³, Charles Plessy^{4,5}, Christopher Previti^{6†}, Chitrag Nepal^{6†}, Jochen Gehrig^{2†}, Xianjun Dong^{7†}, Alina Akalin^{8*}, Asa Maria Suzuki^{1,5}, Wilfried F. J. van Ijcken⁹, Olivier Armant⁹, Marco Feng⁹, Uwe Strähle⁸, Piero Carninci^{1,5}, Ferenc Müller⁹ & Boris Lenhard^{2,9}



Conclusion

Perspectives

- Apply this analysis to **monocyte** activation
- Feature importance analyses will provide **valuable candidates** for experimental validation
- Include Hi-C and RADICL-Seq **interaction counts** as additional features
- Refine our **regulatory grammar** modelling
- **Charles Lecellier, Christophe Vroland** are collaborating with **Mathys Grapotte** and **Cedric Notredame** (CGR Barcelone) on a complementary approach

Acknowledgments

Computational Regulatory Genomics

Laurent Bréhélin	CR CNRS, LIRMM
Quentin Bouvier	PhD student, IGMM
Sophie Lèbre	MCF, IMAG & LIRMM
Charles Lecellier	DR CNRS, IGMM & LIRMM
Mathilde Robin	Engineer, LIRMM & IRCM
Océane Cassan	Post-doc, LIRMM
Christophe Vroland	Post-doc, IGMM
Kevin Yauy	MD, PhD, Univ. Montpellier



FANTOM6-Interactome Acknowledgements

Project co-ordination

Piero Carninci, Wallace Yip, Jay Shin, Hazuki Takahashi, Masaki Kato



Advanced Genomics Circuit

Kayoko Yasuzawa
Anika Prabhu
Callum Parr
Tsukasa Kouno
Yan-Jun Lan
Youtaro Shibayama
Fernando Lopez Redondo
Julio Jesus Leon Incio
Masayoshi Itoh

Large-scale Biomedical Data Technology

Takeya Kasukawa
Tomoe Nobusada
Imad Abugessaisa
Jessica Severin

Transcriptome Technology

Mitsuyoshi Murata
Satoshi Takizawa
Kentaro Kaji
Matthew Valentine
Lokesh Pati Tripathi
Xufeng Shu
Harsita Sharma
Hiromi Sueki

Genome Information Analysis

Chung-Chau Hon
Jonathan Moody

Comprehensive Genomic Analysis

Yasushi Okazaki
Kokoro Ozaki
Tsumugi Kawashima
Hiroko Kinoshita
Shohei Noma
Chitose Takahashi
Michihira Tagami

Cellular Epigenetics

Jen-Chien Chang

IMS sequencing Platform



Istituto Nazionale di Genetica Molecolare

Beatrice Bodega
Valeria Ranzani
Benedetto Polimeni



KTH Royal Institute of Technology

Pelin Sahlén
Artemy Zhigulev



Karolinska Institute

Andreas Lennartsson
Magda Bienko
Wenjing Kang
Wing Hin Yip



Istituto Italiano di Tecnologia

Camilla Ugolini



The University of Tokyo

Martin Frith



References I

- ▶ Andersson, R., Sandelin, A., and Danko, C. G. (2015).
A unified architecture of transcriptional regulatory elements.
Trends in Genetics, 31(8):426–433.
- ▶ Grapotte, M., Saraswat, M., Bessière, C., Menichelli, C., Ramilowski, J. A., Severin, J., Hayashizaki, Y., Itoh, M., Tagami, M., Murata, M., Kojima-Ishiyama, M., Noma, S., Noguchi, S., Kasukawa, T., Hasegawa, A., Suzuki, H., Nishiyori-Sueki, H., Frith, M. C., Abugessaisa, I., Aitken, S., Aken, B. L., Alam, I., Alam, T., Alasiri, R., Alhendi, A. M. N., Alinejad-Rokny, H., Alvarez, M. J., Andersson, R., Arakawa, T., Araki, M., Arbel, T., Archer, J., Archibald, A. L., Arner, E., Arner, P., Asai, K., Ashoor, H., Astrom, G., Babina, M., Baillie, J. K., Bajic, V. B., Bajpai, A., Baker, S., Baldarelli, R. M., Balic, A., Bansal, M., Batagov, A. O., Batzoglou, S., Beckhouse, A. G., Beltrami, A. P., Beltrami, C. A., Bertin, N., Bhattacharya, S., Bickel, P. J., Blake, J. A., Blanchette, M., Bodega, B., Bonetti, A., Bono, H., Bornholdt, J., Bttcher, M., Bougouffa, S., Boyd, M., Breda, J., Brombacher, F., Brown, J. B., Bult, C. J., Burroughs, A. M., Burt, D. W., Busch, A., Caglio, G., Califano, A., Cameron, C. J., Cannistraci, C. V., Carbone, A., Carlisle, A. J., Carninci, P., Carter, K. W., Cesselli, D., Chang, J.-C., Chen, J. C., Chen, Y., Chierici, M., Christodoulou, J., Ciani, Y., Clark, E. L., Coskun, M., Dalby, M., Dalla, E., Daub, C. O., Davis, C. A., de Hoon, M. J. L.,

References II

de Rie, D., Denisenko, E., Deplancke, B., Detmar, M., Deviatiiarov, R., Bernardo, D. D., Diehl, A. D., Dieterich, L. C., Dimont, E., Djebali, S., Dohi, T., Dostie, J., Drablos, F., Edge, A. S. B., Edinger, M., Ehrlund, A., Ekwall, K., Elofsson, A., Endoh, M., Enomoto, H., Enomoto, S., Faghihi, M., Fagiolini, M., Farach-Carson, M. C., Faulkner, G. J., Favorov, A., Fernandes, A. M., Ferrai, C., Forrest, A. R. R., Forrester, L. M., Forsberg, M., Fort, A., Francescato, M., Freeman, T. C., Frith, M., Fukuda, S., Funayama, M., Furlanello, C., Furuno, M., Furusawa, C., Gao, H., Gazova, I., Gebhard, C., Geier, F., Geijtenbeek, T. B. H., Ghosh, S., Ghosheh, Y., Gingeras, T. R., Gojobori, T., Goldberg, T., Goldowitz, D., Gough, J., Greco, D., Gruber, A. J., Guhl, S., Guigo, R., Guler, R., Gusev, O., Gustincich, S., Ha, T. J., Haberle, V., Hale, P., Hallstrom, B. M., Hamada, M., Handoko, L., Hara, M., Harbers, M., Harrow, J., Harshbarger, J., Hase, T., Hasegawa, A., Hashimoto, K., Hatano, T., Hattori, N., Hayashi, R., Hayashizaki, Y., Herlyn, M., Heutink, P., Hide, W., Hitchens, K. J., Sui, S. H., 't Hoen, P. A. C., Hon, C. C., Hori, F., Horie, M., Horimoto, K., Horton, P., Hou, R., Huang, E., Huang, Y., Hugues, R., Hume, D., Ienasescu, H., Iida, K., Ikawa, T., Ikemura, T., Ikeo, K., Inoue, N., Ishizu, Y., Ito, Y., Itoh, M., Ivshina, A. V., Jankovic, B. R., Jenjaroenpun, P., Johnson, R., Jorgensen, M., Jorjani, H., Joshi, A., Jurman, G., Kaczkowski, B., Kai, C., Kaida, K., Kajiyama, K., Kaliyaperumal, R., Kaminuma, E., Kanaya, T., Kaneda, H., Kapranov, P., Kasianov, A. S., Kasukawa, T., Katayama, T., Kato, S.,

References III

Kawaguchi, S., Kawai, J., Kawaji, H., Kawamoto, H., Kawamura, Y. I., Kawasaki, S., Kawashima, T., Kempfle, J. S., Kenna, T. J., Kere, J., Khachigian, L., Kiryu, H., Kishima, M., Kitajima, H., Kitamura, T., Kitano, H., Klaric, E., Klepper, K., Klinken, S. P., Kloppmann, E., Knox, A. J., Kodama, Y., Kogo, Y., Kojima, M., Kojima, S., Komatsu, N., Komiyama, H., Kono, T., Koseki, H., Koyasu, S., Kratz, A., Kukalev, A., Kulakovskiy, I., Kundaje, A., Kunikata, H., Kuo, R., Kuo, T., Kuraku, S., Kuznetsov, V. A., Kwon, T. J., Larouche, M., Lassmann, T., Law, A., Le-Cao, K.-A., Lecellier, C.-H., Lee, W., Lenhard, B., Lennartsson, A., Li, K., Li, R., Lilje, B., Lipovich, L., Lizio, M., Lopez, G., Magi, S., Mak, G. K., Makeev, V., Manabe, R., Mandai, M., Mar, J., Maruyama, K., Maruyama, T., Mason, E., Mathelier, A., Matsuda, H., Medvedeva, Y. A., Meehan, T. F., Mejhert, N., Meynert, A., Mikami, N., Minoda, A., Miura, H., Miyagi, Y., Miyawaki, A., Mizuno, Y., Morikawa, H., Morimoto, M., Morioka, M., Morishita, S., Moro, K., Motakis, E., Motohashi, H., Mukarram, A. K., Mummery, C. L., Mungall, C. J., Murakawa, Y., Muramatsu, M., Murata, M., Nagasaka, K., Nagase, T., Nakachi, Y., Nakahara, F., Nakai, K., Nakamura, K., Nakamura, Y., Nakamura, Y., Nakazawa, T., Nason, G. P., Nepal, C., Nguyen, Q. H., Nielsen, L. K., Nishida, K., Nishiguchi, K. M., Nishiyori, H., Nitta, K., Noguchi, S., Noma, S., Notredame, C., Ogishima, S., Ohkura, N., Ohno, H., Ohshima, M., Ohtsu, T., Okada, Y., Okada-Hatakeyama, M., Okazaki, Y., Oksvold, P., Orlando, V., Ow, G. S., Ozturk,

References IV

M., Pachkov, M., Paparountas, T., Parihar, S. P., Park, S.-J., Pascarella, G., Passier, R., Persson, H., Philippens, I. H., Piazza, S., Plessy, C., Pombo, A., Ponten, F., Poulain, S., Poulsen, T. M., Pradhan, S., Prezioso, C., Pridans, C., Qin, X.-Y., Quackenbush, J., Rackham, O., Ramilowski, J., Ravasi, T., Rehli, M., Rennie, S., Rito, T., Rizzu, P., Robert, C., Roos, M., Rost, B., Roudnicky, F., Roy, R., Rye, M. B., Sachenkova, O., Saetrom, P., Sai, H., Saiki, S., Saito, M., Saito, A., Sakaguchi, S., Sakai, M., Sakaue, S., Sakaue-Sawano, A., Sandelin, A., Sano, H., Sasamoto, Y., Sato, H., Saxena, A., Saya, H., Schafferhans, A., Schmeier, S., Schmidl, C., Schmocker, D., Schneider, C., Schueler, M., Schultes, E. A., Schulze-Tanzil, G., Semple, C. A., Seno, S., Seo, W., Sese, J., Severin, J., Sheng, G., Shi, J., Shimoni, Y., Shin, J. W., SimonSanchez, J., Sivertsson, A., Sjostedt, E., Soderhall, C., Laurent, G. S., Stoiber, M. H., Sugiyama, D., Summers, K. M., Suzuki, A. M., Suzuki, H., Suzuki, K., Suzuki, M., Suzuki, N., Suzuki, T., Swanson, D. J., Swoboda, R. K., Tagami, M., Taguchi, A., Takahashi, H., Takahashi, M., Takamochi, K., Takeda, S., Takenaka, Y., Tam, K. T., Tanaka, H., Tanaka, R., Tanaka, Y., Tang, D., Taniuchi, I., Tanzer, A., Tarui, H., Taylor, M. S., Terada, A., Terao, Y., Testa, A. C., Thomas, M., Thongjuea, S., Tomii, K., Triglia, E. T., Toyoda, H., Tsang, H. G., Tsujikawa, M., Uhlén, M., Valen, E., van de Wetering, M., van Nimwegen, E., Velmeshev, D., Verardo, R., Vitezic, M., Vitting-Seerup, K., von Feilitzen, K., Voolstra, C. R., Vorontsov, I. E., Wahlestedt,

References V

C., Wasserman, W. W., Watanabe, K., Watanabe, S., Wells, C. A., Winteringham, L. N., Wolvetang, E., Yabukami, H., Yagi, K., Yamada, T., Yamaguchi, Y., Yamamoto, M., Yamamoto, Y., Yamamoto, Y., Yamanaka, Y., Yano, K., Yasuzawa, K., Yatsuka, Y., Yo, M., Yokokura, S., Yoneda, M., Yoshida, E., Yoshida, Y., Yoshihara, M., Young, R., Young, R. S., Yu, N. Y., Yumoto, N., Zabierowski, S. E., Zhang, P. G., Zucchelli, S., Zwahlen, M., Chatelain, C., Carninci, P., de Hoon, M. J. L., Wasserman, W. W., Bréhélin, L., and and, C.-H. L. (2021).

Discovery of widespread transcription initiation at microsatellites predictable by sequence-based deep neural network.

Nature Communications, 12(1).

- ▶ Harris, H. K., Nakayama, T., Lai, J., Zhao, B., Argyrou, N., Gubbels, C. S., Soucy, A., Genetti, C. A., Suslovitch, V., Rodan, L. H., Tiller, G. E., Lesca, G., Gripp, K. W., Asadollahi, R., Hamosh, A., Applegate, C. D., Turnpenny, P. D., Simon, M. E., Volker-Touw, C. M., van Gassen, K. L., van Binsbergen, E., Pfundt, R., Gardeitchik, T., de Vries, B. B., Immken, L. L., Buchanan, C., Willing, M., Toler, T. L., Fassi, E., Baker, L., Vansenne, F., Wang, X., Ambrus, J. L., Fannemel, M., Posey, J. E., Agolini, E., Novelli, A., Rauch, A., Boonsawat, P., Fagerberg, C. R., Larsen, M. J., Kibaek, M., Labalme, A., Poisson, A., Payne, K. K., Walsh, L. E.,

References VI

Aldinger, K. A., Balciuniene, J., Skraban, C., Gray, C., Murrell, J., Bupp, C. P., Pascolini, G., Grammatico, P., Broly, M., Küry, S., Nizon, M., Rasool, I. G., Zahoor, M. Y., Kraus, C., Reis, A., Iqbal, M., Uguen, K., Audebert-Bellanger, S., Ferec, C., Redon, S., Baker, J., Wu, Y., Zampino, G., Syrbe, S., Brosse, I., Jamra, R. A., Dobyns, W. B., Cohen, L. L., Blomhoff, A., Mignot, C., Keren, B., Courtin, T., Agrawal, P. B., Beggs, A. H., and Yu, T. W. (2021).

Disruption of RFX family transcription factors causes autism, attention-deficit/hyperactivity disorder, intellectual disability, and dysregulated behavior.

Genetics in Medicine, 23(6):1028–1040.

- ▶ Menichelli, C., Guitard, V., Martins, R. M., Lèbre, S., Lopez-Rubio, J.-J., Lecellier, C.-H., and Bréhélin, L. (2021).

Identification of long regulatory elements in the genome of plasmodium falciparum and other eukaryotes.

PLOS Computational Biology, 17(4):e1008909.

- ▶ Mukhtar, T., Breda, J., Grison, A., Karimaddini, Z., Grobecker, P., Iber, D., Beisel, C., van Nimwegen, E., and Taylor, V. (2020).

Tead transcription factors differentially regulate cortical development.

Scientific Reports, 10(1).

References VII

- ▶ Sarvesh Nikumbh (2022).
seqarchr.
- ▶ Xu, C., Park, J.-K., and Zhang, J. (2019).
Evidence that alternative transcriptional initiation is largely nonadaptive.
PLOS Biology, 17(3):e3000197.
- ▶ Zinin, N., Adameyko, I., Wilhelm, M., Fritz, N., Uhlén, P., Ernfors, P., and Henriksson, M. A. (2014).
scpMYC/scp proteins promote neuronal differentiation by controlling the mode of progenitor cell division.
EMBO reports, 15(4):383–391.