

Supplementary Material - fourSynergy: Ensemble-based interaction
calling on 4C-seq data using gradient-free optimization

1 Supplementary Figures

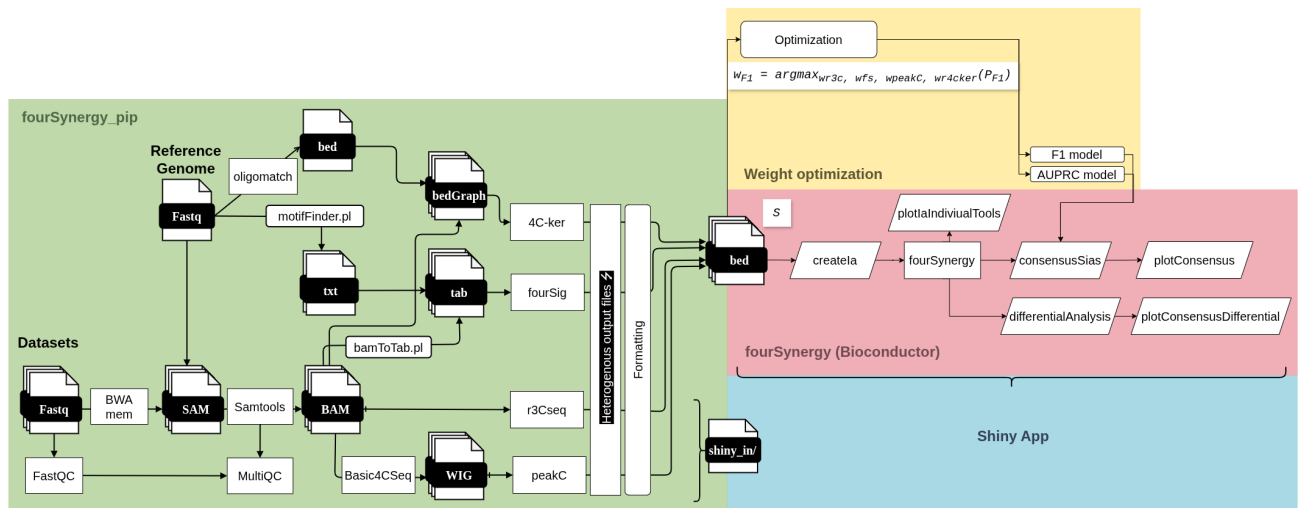


Figure S1: Schematic overview of fourSynergy framework. The framework features preprocessing pipeline, weight optimization, R Bioconductor package and R shiny application.

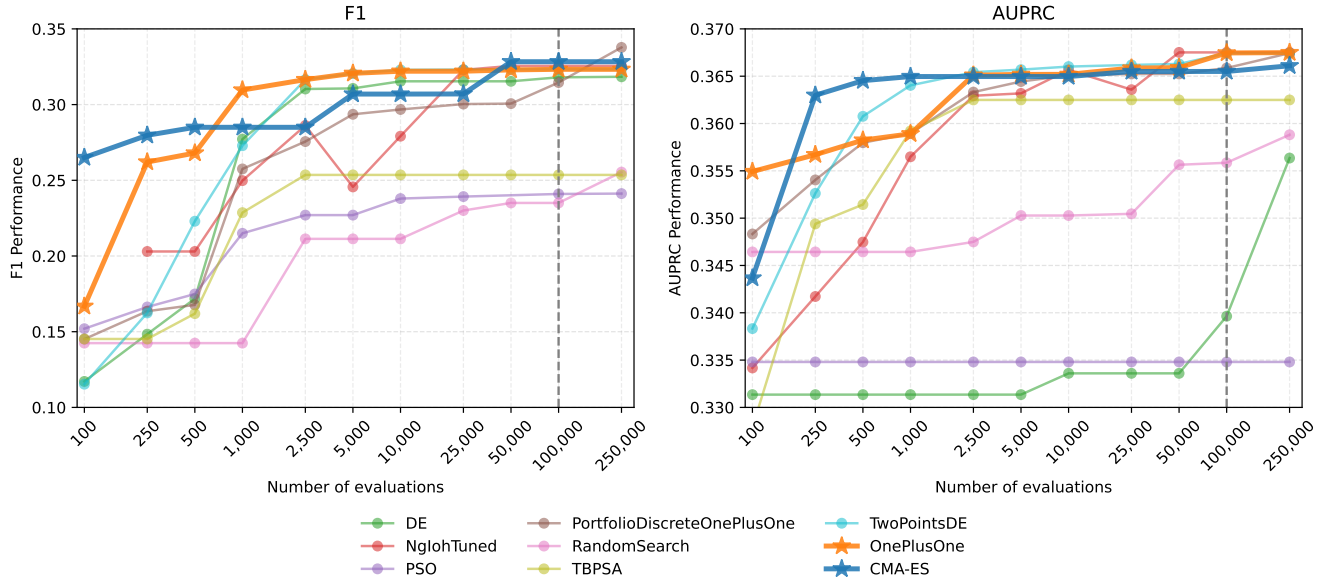


Figure S2: **Benchmarking Nevergrad optimizers.** Benchmarking Nevergrad optimizers across different iteration budgets using F1 score and area under precision recall curve (AUPRC).

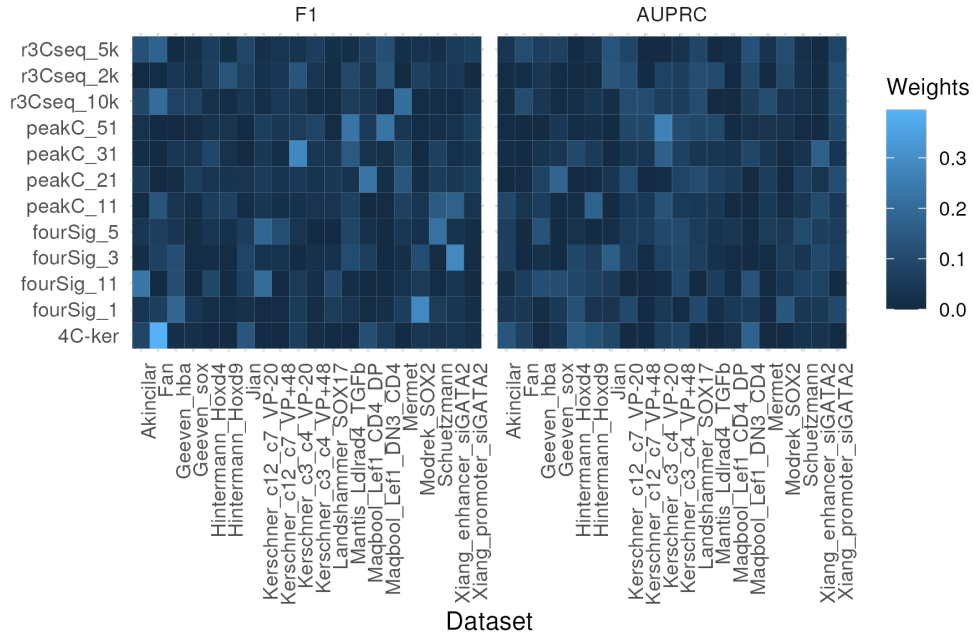


Figure S3: **Optimized weights for single dataset.** Heatmap showing the optimized weights for each tool (y-axis) across all datasets (x-axis), with color intensity indicating the weight.

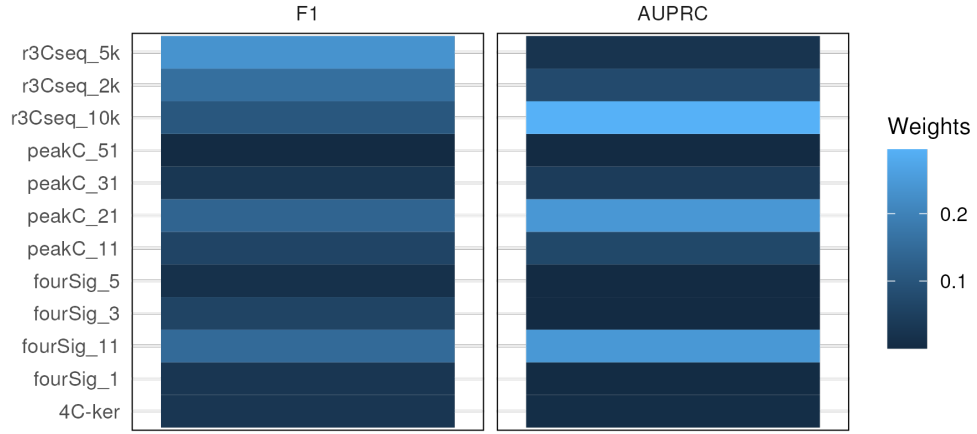


Figure S4: **Optimized weights over all datasets.** Heatmap showing the optimized weights for each tool (y-axis) over all datasets (x-axis), with color intensity indicating the weight.

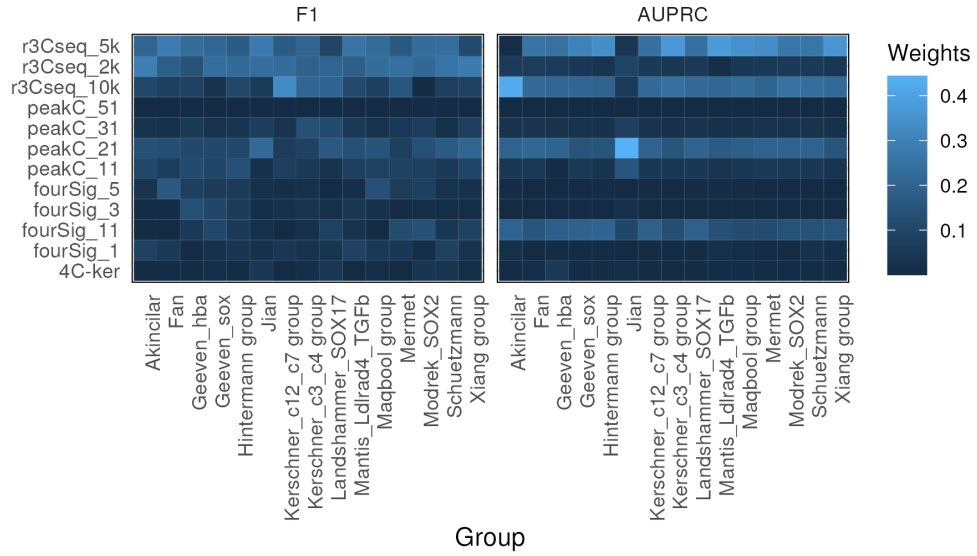


Figure S5: **Optimized weights for Leave-One-Group-Out cross validation.** Heatmap showing the optimized weights for each tool (y-axis) across all datasets groups (x-axis), with color intensity indicating the weight.

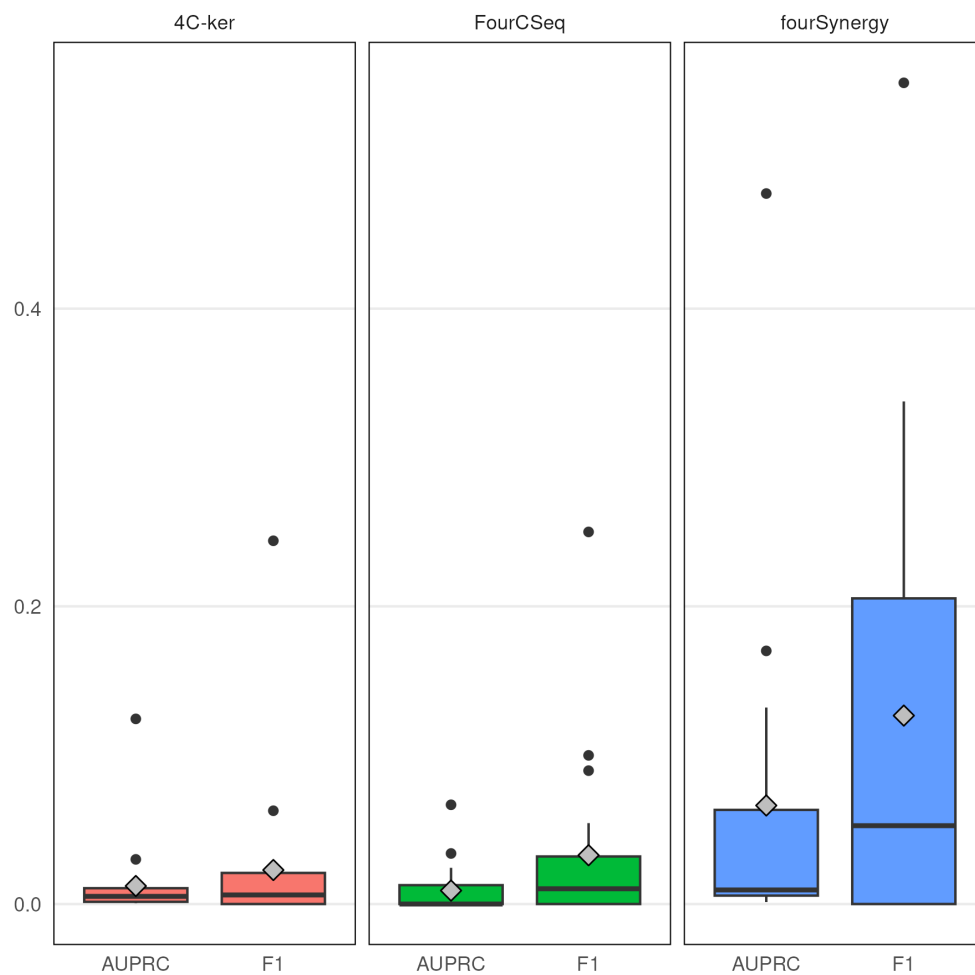


Figure S6: **Predictive performance of differential interaction calling using different 4C-seq algorithms.**

Boxplots illustrating the distribution of F1 scores and area under the precision-recall curve (AUPRC) values (y-axis) across different 4C-seq tools (x-axis), with diamonds indicating the median values.

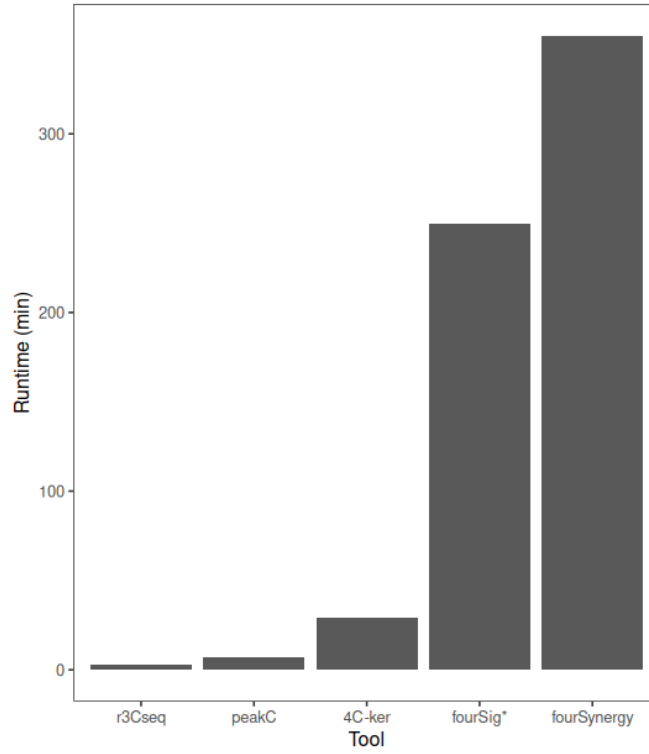


Figure S7: **Runtime Benchmarking of 4C-seq analysis tools.** Runtime of base tools (x-axis) in minutes (y-axis). Runtime includes preprocessing and tool execution with window sizes mentioned in Materials and Methods.

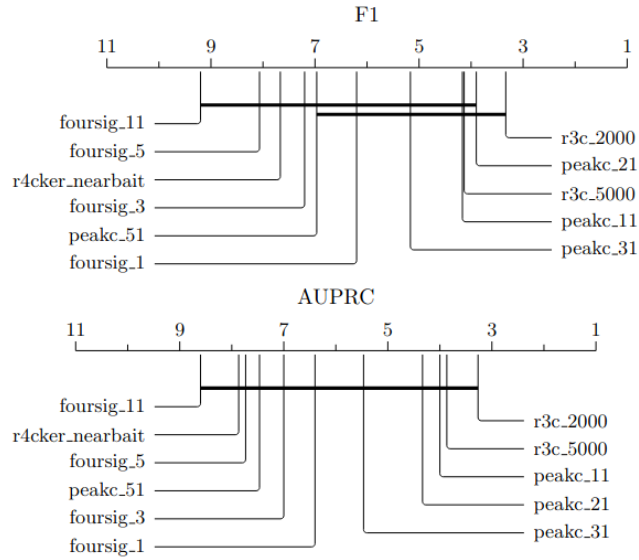


Figure S8: **Critical difference analysis of base tools.** Critical difference analysis of the predictive performance of the base tools across the curated collection of 4C-seq datasets (n=20).

2 Supplementary Formulas

$$S = w_{\text{r3c } 2\text{k}} \cdot x_{\text{r3c } 2\text{k}} + w_{\text{r3c } 5\text{k}} \cdot x_{\text{r3c } 5\text{k}} + w_{\text{r3c } 10\text{k}} \cdot x_{\text{r3c } 10\text{k}} + \quad (1)$$

$$w_{\text{foursig } 1} \cdot x_{\text{foursig } 1} + w_{\text{foursig } 3} \cdot x_{\text{foursig } 3} + w_{\text{foursig } 5} \cdot x_{\text{foursig } 5} + w_{\text{foursig } 11} \cdot x_{\text{foursig } 11} + \quad (2)$$

$$w_{\text{peakC } 11} \cdot x_{\text{peakC } 11} + w_{\text{peakC } 21} \cdot x_{\text{peakC } 21} + w_{\text{peakC } 31} \cdot x_{\text{peakC } 31} + w_{\text{peakC } 51} \cdot x_{\text{peakC } 51} + \quad (3)$$

$$w_{4\text{C-ker}} \cdot x_{4\text{C-ker}} \quad (4)$$

where

$$w_{\text{r3c } 2\text{k}}, w_{\text{r3c } 5\text{k}}, w_{\text{r3c } 10\text{k}},$$

$$w_{\text{foursig } 1}, w_{\text{foursig } 3}, w_{\text{foursig } 5}, w_{\text{foursig } 11},$$

$$w_{\text{peakC } 11}, w_{\text{peakC } 21}, w_{\text{peakC } 31}, w_{\text{peakC } 51},$$

$$w_{4\text{C-ker}} \in [0, 0.5]$$

$$w_{F_1}^* = \arg \max_{(w_{\text{r3c } 2\text{k}}, w_{\text{r3c } 5\text{k}}, w_{\text{r3c } 10\text{k}}, w_{\text{foursig } 1}, w_{\text{foursig } 3}, w_{\text{foursig } 5}, w_{\text{foursig } 11}, w_{\text{peakC } 11}, w_{\text{peakC } 21}, w_{\text{peakC } 31}, w_{\text{peakC } 51}, w_{4\text{C-ker}})} \left(\frac{1}{n} \sum_{D_0, \dots, D_n} P_{F_1} \right)$$

3 Supplementary Methods

3.1 Preprocessing

The entry point of the pipeline are FASTQ files downloaded via the Gene Expression Omnibus (GEO) [1]. Relevant information on the 4C-seq experiment is provided in a config file.

In the first step of preprocessing, initial quality control of the FASTQ files is performed using FASTQC [2]. The reads are aligned to the reference genome using BWA MEM (0.7.18-r1243-dirty) [3]. The resulting BAM files are subsequently indexed, sorted, and the alignment statistics are summarized using Samtools [4]. All results are then collected by MultiQC [5]. Basic4CSeq further performs a 4C-seq-specific quality control and basic analysis [6]. In this step, the reference genome is fragmented at the restriction sites to create a virtual fragment library (VFL), in silico mimicking the restriction enzyme digestion performed in the wet lab. This is particularly important as the virtual fragment library forms the basis of our analysis. In addition, we perform various steps to convert the input data into the appropriate formats for each 4C-seq tool. The different interaction callers have varying input requirements. r3Cseq and 4C-ker operate on BAM files, while 4C-ker additionally needs a reduced genome file to map reads on. peakC utilizes WIG files, while fourSig requires TAB files, which can be created using scripts from the fourSig software suite. Detailed information regarding preprocessing can be found in the workflow in the supplementary materials (Fig.S1).

3.2 Runtime benchmarking

The runtime was tested on a Linux system with an Intel(R) Xeon(R) CPU E5-2695 v4 @ 2.10GHz, running Debian 6.1.153-1, with 504 GB of total RAM, our pipeline utilized a maximum of 16 GB RAM during execution.

3.3 Optimizer benchmarking

For selecting the most suitable optimizer, we benchmarked several Nevergrad implementations across increasing evaluation budgets to assess their effectiveness for the average multi-optimization task. Therefore, this benchmark does not constitute the cross-validation procedure, as performance is evaluated directly on the entire set. For each optimizer, performance was tracked as a function of the number of objective-function evaluations. The tests are not performed only once per metric and optimizer, but restarted for every budget step. Results were summarized using F1 score and AUPRC. This analysis was intended to compare how quickly different optimizers reached strong solutions and how performance evolved with additional iterations. Together, these benchmarks provide an overview of both convergence behavior and final optimizer choice for both metrics.

Figure S2 summarizes the results from the described analysis. Comparing F1 score and AUPRC suggests that optimization with respect to F1 score is the more challenging problem, as most optimizers start below 0.2 and improve gradually over time. In contrast, AUPRC starts above 0.3 and reaches a similar maximum level at the end of the optimization. In addition, AUPRC remains relatively stable from a budget of 2500 onward. The analysis does not reveal a clear convergence point for either metric, as some optimizers still improve between 100,000 and 250,000 iterations. Nevertheless, for most of the best-performing optimizers, the improvement in this range is only marginal, which led us to select 100,000 iterations as a practical sweet spot for all subsequent analysis. Based on this benchmark, we choose the CMA-ES for all F1 score-related optimizations and OnePlusOne for the AUPRC-related optimizations, as these optimizers achieved the best performance at the budget of 100,000 iterations.

3.4 Ensemble potential and diversity

We calculated theoretical upper bounds on ensemble performance following [7] for an ensemble comprising the five top-ranked tools identified from a critical-difference analysis (Fig. S8). Let $\{M_1, \dots, M_5\}$ be our

five models and let $k = (L + 1)/2$ with L the number of models (i.e., 5), then the upper bound is given by

$$\max P_{maj} = \min\{1, \sum(k), \sum(k-1), \dots, \sum(1)\} \quad (5)$$

with $\sum m = 1/m \sum_{i=1}^{L-k+m} p_i$ with $m = 1, \dots, k$.

These tools were selected because they ranked consistency highly with respect to both F_1 and AUPRC. We compared this upper bound with the performance of the best individual tool according to the same critical-difference analysis (r3Cseq_2000). In 13 datasets, the upper bound of the ensemble exceeded the observed F_1 score of r3Cseq_2000. Averaged across datasets, the best-performing single tool (r3cseq_2000) achieved a mean F_1 score of 0.16, whereas the corresponding theoretical upper bound for the top-five ensemble was 0.25. Furthermore, we quantified the diversity among the top five tools by measuring entropy [7]. A fragment was classified as a "call" if the significance score was ≥ 1 .

$$entropy = \frac{1}{N} \frac{2}{L-1} \sum_{j=1}^N \min\{(\sum_{i=1}^L y_{ij}), (L - \sum_{i=1}^L y_{ij})\} \quad (6)$$

The mean entropy was 0.049, indicating limited disagreement among the top-ranked tools.

Overall, these findings suggest that, despite relatively low diversity, the top-ranked tools still contain complementary information that could be exploited using ensemble-based methods. In particular, the gap between the best individual tool and the theoretical ensemble upper bound indicates that performance gains are already achievable with simple aggregation strategies such as majority voting. More flexible approaches, such as our weighted voting procedure, further improve performance by assigning greater influence to consistently stronger tools. Further analyses demonstrated that increasing the number of integrated tools led to higher diversity but simultaneously to a lower theoretical ensemble upper bound. This suggests that ensemble performance depends on the balance between complementarity and the predictive quality of the included tools. In our weighted voting approach, this balance is incorporated directly into the optimization objective, allowing the method to determine both which tools should contribute to the final prediction and how strongly.

4 Supplementary Tables

Dataset	organism	VP	chromosome	condition	Rep	control	Rep	RE1	RE2	GEO	No. validated	interactions	Group	Reference
Akincilar	human	5	2	2	2	2	HindIII	DpnII	GSE77265	5	2		-	[8]
Fan	mouse	5	2	2	2	2	HindIII	DpnII	GSE87299	2	2		-	[9]
Schuetzmann	mouse	2	2	2	2	2	NlaIII	NlaIII	GSE110683	2	2		-	[10]
Geeven_hba	mouse	11	3	3	3	3	DpnII	Csp6I	GSE105177	2	2		-	[11]
Geeven_sox	mouse	3	3	3	3	3	DpnII	Csp6I	GSE105177	1	1		-	[11]
Mernnet	mouse	10	3	3	3	3	DpnII	NlaIII	GSE101423	1	1		-	[12]
Maqbool	mouse	3	2	2	2	2	DpnII	Csp6I	GSE144586	6	6		Maqbool group	[13]
Maqbool	mouse	3	2	2	2	2	DpnII	Csp6I	GSE144586	4	4		Maqbool group	[13]
Jian	human	11	3	3	3	3	BglII	NlaIII	GSE111863	1	1		-	[14]
Modrek_SOX2	mouse	3	2	2	2	2	DpnII	Csp6I	GSE94962	7	7		-	[15]
Mantis_Ldlrad4_TGFB	mouse	18	2	2	2	2	NlaIII	DpnII	GSE197013	8	8		-	[16]
Kerschner_c12.c7.VP-20	human	7	2	2	2	2	NlaIII	DpnII	GSE220855	1	1		Kerschner_c12.c7 group	[17]
Kerschner_c12.c7.VP+48	human	7	2	2	2	2	NlaIII	DpnII	GSE220855	1	1		Kerschner_c12.c7 group	[17]
Kerschner_c3.c4.VP-20	human	7	2	2	2	2	NlaIII	DpnII	GSE220855	1	1		Kerschner_c3.c4 group	[17]
Kerschner_c3.c4.VP+48	human	7	2	2	2	2	NlaIII	DpnII	GSE220855	1	1		Kerschner_c3.c4 group	[17]
Hintermann_Hoxd1	mouse	2	1	2	2	2	DpnII	NlaIII	GSE195591	2	2		Hintermann group	[18]
Hintermann_Hoxd4	mouse	2	2	2	2	2	DpnII	NlaIII	GSE195591	2	2		Hintermann group	[18]
Landshammer_SOX17	human	8	3	3	3	3	NlaIII	DpnII	GSE178990	1	1		-	[19]
Xiang-enhancer_siGATA2	human	X	2	2	2	2	DpnII	NlaIII	GSE275926	1	1		Xiang group	[20]
Xiang-promoter_siGATA2	human	X	2	2	2	2	DpnII	NlaIII	GSE275926	1	1		Xiang group	[20]

Table S1: Overview of datasets and validated interactions

Tool	Mean F1	Median F1	SD_F1	Mean AUPRC	Median AUPRC	SD AUPRC
foursig_1	0.04	0.02	0.04	0.02	0.01	0.01
foursig_11	0.02	0.01	0.03	0.01	0.00	0.02
foursig_3	0.03	0.01	0.04	0.02	0.01	0.02
foursig_5	0.03	0.01	0.04	0.02	0.00	0.02
opo_indiv_auprc	0.06	0.02	0.09	0.40	0.32	0.29
opo_indiv_f1	0.52	0.45	0.26	0.38	0.31	0.29
opo_multi_auprc	0.12	0.07	0.13	0.36	0.28	0.30
opo_multi_auprc_logo	0.07	0.03	0.09	0.34	0.23	0.31
opo_multi_f1	0.34	0.26	0.30	0.34	0.28	0.29
opo_multi_f1_logo	0.31	0.25	0.27	0.32	0.27	0.29
peakc_11	0.18	0.18	0.16	0.16	0.09	0.21
peakc_21	0.19	0.19	0.17	0.12	0.06	0.15
peakc_31	0.12	0.08	0.14	0.07	0.02	0.09
peakc_51	0.06	0.00	0.11	0.03	0.00	0.06
r3c_10000	0.12	0.07	0.13	0.07	0.04	0.08
r3c_2000	0.16	0.10	0.14	0.08	0.05	0.08
r3c_5000	0.15	0.08	0.15	0.08	0.05	0.10
r4cker_nearbait	0.03	0.02	0.04	0.02	0.01	0.02
single_tool_logo_auprc	0.18	0.18	0.16	0.16	0.09	0.21
single_tool_logo_f1	0.13	0.10	0.15	0.12	0.03	0.22

Table S2: Predictive performance statistics

Left-out dataset	F1		AUPRC	
	optimized	single tool	optimized	single tool
Fan	0.00	0.00	0.11	0.00
Geeven_hba	0.36	0.00	0.26	0.09
Geeven_sox	0.29	0.27	0.81	0.27
Mermet	0.00	0.00	0.04	0.00
Hintermann_Hoxd4	0.60	0.21	0.45	0.22
Hintermann_Hoxd9	0.24	0.18	0.07	0.16
Kerschner_c12_c7_VP-20	0.89	0.00	0.80	0.00
Kerschner_c12_c7_VP+48	0.48	0.39	0.86	0.86
Kerschner_c3_c4_VP+48	0.19	0.11	0.19	0.10
Kerschner_c3_c4_VP-20	0.00	0.00	0.01	0.00
Landshammer_SOX17	0.23	0.20	0.25	0.21
Mantis_Ldlrad4_TGFb	0.54	0.00	0.39	0.21
Xiang_enhancer_siGATA2	0.00	0.24	0.01	0.25
Xiang_promoter_siGATA2	0.67	0.35	1.00	0.50
Akincilar	0.00	0.00	0.26	0.00
Jian	0.69	0.45	0.21	0.16
Maqbool_Lef1_CD4_DP	0.13	0.00	0.22	0.01
Maqbool_Lef1_DN3_CD4	0.17	0.08	0.16	0.02
Modrek_SOX2	0.26	0.12	0.16	0.09
Schuetzmann	0.43	0.00	0.56	0.00

Table S3: Predictive performance per LOGO group

References

- [1] Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1):D991–D995, 11 2012.
- [2] FastQC, Jun 2015.
- [3] Heng Li and Richard Durbin. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14):1754–1760, 05 2009.
- [4] Heng Li, Bob Handsaker, Alec Wysoker, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 06 2009.
- [5] Philip Ewels, Måns Magnusson, Sverker Lundin, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048, October 2016.
- [6] Carolin Walter, Daniel Schuetzmann, Frank Rosenbauer, and Martin Dugas. Basic4cseq: an R/bioconductor package for analyzing 4c-seq data. *Bioinformatics*, 30(22):3268–3269, 07 2014.
- [7] Ludmila I Kuncheva. *Combining Pattern Classifiers*. John Wiley & Sons, Ltd, 2004.
- [8] Semih Can Akincilar, Ekta Khattar, Priscilla Li Shan Boon, et al. Long-Range Chromatin Interactions Drive Mutant TERT Promoter Activation. *Cancer Discovery*, 6(11):1276–1291, 11 2016.
- [9] Zenghua Fan, Meng Zhao, Parth D Joshi, et al. A class of circadian long non-coding RNAs mark enhancers modulating long-range circadian gene regulation. *Nucleic Acids Res.*, 45(10):5720–5738, June 2017.
- [10] Daniel Schuetzmann, Carolin Walter, Boet van Riel, et al. Temporal autoregulation during human PU.1 locus SubTAD formation. *Blood*, 132(25):2643–2655, 12 2018.
- [11] Geert Geeven, Hans Teunissen, Wouter de Laat, et al. peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Research*, 46(15):e91–e91, 05 2018.
- [12] Jérôme Mermet, Jake Yeung, Clémence Hurni, et al. Clock-dependent chromatin topology modulates circadian transcription and behavior. *Genes Dev.*, 32(5-6):347–358, March 2018.
- [13] Muhammad Ahmad Maqbool, Léo Pioger, and El Aabidine and others. Alternative enhancer usage and targeted polycomb marking hallmark promoter choice during T cell differentiation. *Cell Rep.*, 32(7):108048, August 2020.
- [14] Xing Jian and Gary Felsenfeld. *Insulin* promoter in human pancreatic β cells contacts diabetes susceptibility loci and regulates genes affecting insulin metabolism. *Proceedings of the National Academy of Sciences*, 115(20):E4633–E4641, 2018.
- [15] Aram S. Modrek, Danielle Golub, Themasap Khan, et al. Low-Grade Astrocytoma Mutations in IDH1, P53, and ATRX Cooperate to Block Differentiation of Human Neural Stem Cells via Repression of SOX2. *Cell Reports*, 21(5):1267–1280, 2017.
- [16] Marta Vicioso-Mantis, Raquel Fueyo, Claudia Navarro, et al. JMJD3 intrinsically disordered region links the 3D-genome structure to TGF β -dependent transcription activation. *Nature Communications*, 13(1):3263, 2022.
- [17] Jenny L Kerschner, Alekh Paranjapye, and Ann Harris. Cellular heterogeneity in the 16HBE14o-airway epithelial line impacts biological readouts. *Physiol. Rep.*, 11(11):e15700, June 2023.

- [18] Aurélie Hintermann, Isabel Guerreiro, Lucille Lopez-Delisle, et al. Developmental and evolutionary comparative analysis of a regulatory landscape in mouse and chicken. *Development*, 149(12):dev200594, 06 2022.
- [19] Alexandro Landshammer, Adriano Bolondi, Helene Kretzmer, et al. *T-REX17* is a transiently expressed non-coding RNA essential for human endoderm formation. *eLife*, 12:e83077, jan 2023.
- [20] Rachel R Xiang, Shin-Ai Lee, Caroline F Tyndall, et al. CRISPR screening identifies regulators of enhancer-mediated androgen receptor transcription in advanced prostate cancer. *Cell Rep.*, 44(2):115312, February 2025.