Table S1: The number of enhancers used to design EPIP.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cell line** | **Enhancers** | **Enhancers overlapped with ChromHMM** | **Enhancers overlapped with H3K27ac** | **Comments** |
| GM12878 | 32693 | 11619 | 7684 |  |
| HELA | 32693 | 7023 | 5100 |  |
| HMEC | 32693 | 13573 | 9038 |  |
| HUVEC | 32693 | 11936 | 9198 |  |
| IMR90 | 32693 | 7542 | 4888 |  |
| K562 | 32693 | 9914 | 5710 |  |
| KBM7 | 32693 | 32693 | 32693 | No ChromHMM and H3K27ac |
| NHEK | 32693 | 12498 | 9055 |  |

We overlapped the 32,693 Fantom enhancers with ChromHMM enhancers and then with H3K27ac signal for each of the 7 cell lines (except KBM7). We could not find available ChromHMM enhancers and H3K27ac data for KBM7.

Table S2: The number of promoters used to design EPIP.

|  |  |  |  |
| --- | --- | --- | --- |
| **Cell line** | **Promoters** | **Expressed Promoters** | **Comments** |
| GM12878 | 57820 | 15435 |  |
| HELA | 57820 | 14990 |  |
| HMEC | 57820 | 57820 | No expression data |
| HUVEC | 57820 | 14776 |  |
| IMR90 | 57820 | 14689 |  |
| K562 | 57820 | 14880 |  |
| KBM7 | 57820 | 57820 | No expression data |
| NHEK | 57820 | 15350 |  |

From 57,820 Gencode version 19 genes, we considered expressed genes in the corresponding cell lines. Then we defined promoters, considering 1kb upstream and 100 bp downstream region around the corresponding gene TSS. For HMEC and KBM7, we could not find the gene expression data, so for these two cell lines, we considered all the 57,820 genes to define promoters. To consider a gene expressed we consider at least 0.3 reads per kilobase of transcript per million mapped reads (RPKM) value and at most 0.1 irreproducible discovery rate (IDR) for that gene.

Table S3: The number of enhancer-promoter pairs for different read cutoffs and looplists.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Cell line** | **EP-pairs** | **Cell specific EPIs** | **EP-pairs within 2.5kb to 2Mb** |
| Cutoff 5 | GM12878 | 95159 | 72592 | 92771 |
| HMEC | 21387 | 2531 | 19500 |
| HUVEC | 13350 | 1413 | 12199 |
| IMR90 | 13050 | 4348 | 12778 |
| K562 | 22577 | 5391 | 21508 |
| KBM7 | 143379 | 101882 | 134144 |
| NHEK | 11675 | 1094 | 10659 |
| Cutoff 10 | GM12878 | 57891 | 48633 | 56634 |
| HMEC | 8687 | 962 | 7001 |
| HUVEC | 5116 | 451 | 4088 |
| IMR90 | 6467 | 2851 | 6279 |
| K562 | 10741 | 3103 | 9820 |
| KBM7 | 45668 | 30484 | 38454 |
| NHEK | 4897 | 456 | 3956 |
| Cutoff 20 | GM12878 | 33220 | 29458 | 32293 |
| HMEC | 4033 | 313 | 2514 |
| HUVEC | 2048 | 123 | 1142 |
| IMR90 | 2622 | 1248 | 2476 |
| K562 | 4808 | 1435 | 3935 |
| KBM7 | 16340 | 10066 | 10136 |
| NHEK | 2198 | 141 | 1312 |
| Cutoff 30 | GM12878 | 22566 | 20164 | 21728 |
| HMEC | 2742 | 189 | 1381 |
| HUVEC | 1279 | 43 | 480 |
| IMR90 | 1418 | 612 | 1285 |
| K562 | 2926 | 710 | 2080 |
| KBM7 | 12040 | 7469 | 6367 |
| NHEK | 1587 | 60 | 808 |
| Cutoff 50 | GM12878 | 13060 | 11677 | 12293 |
| HMEC | 1554 | 34 | 413 |
| HUVEC | 778 | 9 | 158 |
| IMR90 | 666 | 257 | 535 |
| K562 | 1714 | 361 | 947 |
| KBM7 | 7756 | 4853 | 2843 |
| NHEK | 947 | 30 | 286 |
| Cutoff 100 | GM12878 | 5434 | 4670 | 4706 |
| HMEC | 421 | 0 | 102 |
| HUVEC | 55 | 2 | 14 |
| IMR90 | 267 | 79 | 161 |
| K562 | 824 | 56 | 224 |
| KBM7 | 5718 | 3952 | 1202 |
| NHEK | 791 | 0 | 157 |
| Looplists | GM12878 | 217 | 206 | 152 |
| HELA | 3 | 3 | 2 |
| HMEC | 227 | 212 | 86 |
| HUVEC | 8 | 8 | 5 |
| IMR90 | 70 | 60 | 47 |
| K562 | 30 | 22 | 20 |
| KBM7 | 8 | 5 | 6 |
| NHEK | 0 | 0 | 0 |

To generate the positive EP-pairs for different normalized read cutoff, we considered enhancers and promoters that overlapped with the contact matrix regions that interact with more reads than the corresponding read cutoff.

Table S4: Availability of feature peak files for different cell lines is given in terms of GEO (Gene Expression Omnibus) accession number.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **GM12878** | **HELA** | **HMEC** | **HUVEC** | **IMR90** | **K562** | **KBM7** | **NHEK** |
| **CTCF** | GSM733752 | GSM733785 | GSM733724 | GSM733716 | GSM935404 | GSM733719 | - | GSM733636 |
| **Dnase-Seq** | GSM816665 | GSM816633 | GSM816669 | GSM816646 | GSM816665 | GSM816655 | - | GSM816635 |
| **H3K27ac** | GSM733771 | GSM733684 | GSM733660 | GSM733683 | GSM469966 | GSM733656 | - | GSM733674 |
| **H3K27me3** | GSM733758 | GSM733696 | GSM733722 | GSM733688 | GSM469968 | GSM733658 | - | GSM733701 |
| **H3K36me3** | GSM733679 | GSM733711 | GSM733707 | GSM733757 | GSM521890 | GSM733714 | - | GSM733726 |
| **H3K4me1** | GSM733769 | GSM798322 | GSM733654 | GSM733683 | GSM521895 | GSM733651 | - | GSM733686 |
| **H3K4me2** | GSM733769 | GSM733734 | GSM733654 | GSM733683 | GSM521899 | GSM733651 | - | GSM733686 |
| **H3K4Me3** | GSM733708 | GSM733682 | GSM733712 | GSM733673 | GSM469970 | GSM733680 | - | GSM733720 |
| **H3K79me2** | GSM733736 | GSM733669 | GSM1003557 | GSM1003555 | GSM521909 | H3K79me2 | - | GSM1003527 |
| **H3K9ac** | GSM733677 | GSM733756 | GSM733713 | GSM733735 | GSM469973 | GSM733778 | - | GSM733665 |
| **H4K20me1** | GSM733642 | GSM733689 | GSM733647 | GSM733640 | GSM521915 | GSM733675 | - | GSM733728 |
| **Pol2** | GSM803355 | GSM733759 | - | GSM733749 | GSM935513 | GSM733643 | - | GSM733671 |
| **Rad21** | GSM935332 | - | - | - | GSM935624 | GSM803447 | - | - |
| **SMC3** | GSM935376 | - | - | - | GSM2422871 | GSM935310 | - | - |

For each type of data, there will be two corresponding features. For instance, there will be 31 features in GM12878, with 28 features from the fourteen types of data listed in the table and 3 features from the three static features.

Table S5: 11 partitions used to design EPIP. The features in each of these 10 partitions are also shown. The “feature\_E” and “feature\_P” represent the corresponding feature in an enhancer and promoter regions respectively.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Partition 1** | **Partition 2** | **Partition 3** | **Partition 4** | **Partition 5** | **Partition 6** | **Partition 7** | **Partition 8** | **Partition 9** | **Partition 10** | **Partition 11** |
| distance | Distance | distance | distance | distance | distance | distance | distance | distance | distance | distance |
| corr | Corr | corr | Corr | corr | corr | corr | corr | corr | corr | corr |
| css | Css | css | Css | css | css | css | css | css | css | css |
|  | H3k4me1\_E | DNaseI\_E | DNaseI\_E | H3k4me1\_E | H3k4me1\_E | DNaseI\_E | H3k4me1\_E | H3k4me1\_E | H3k4me1\_E | H3k4me1\_E |
|  | H3k4me1\_P | DNaseI\_P | DNaseI\_P | H3k4me1\_P | H3k4me1\_P | DNaseI\_P | H3k4me1\_P | H3k4me1\_P | H3k4me1\_P | H3k4me1\_P |
|  |  |  | H3k27ac\_E | H3k27ac\_E | H3k27ac\_E | H3k27ac\_E | H3k4me2\_E | H3k4me2\_E | H3k4me2\_E | H3k4me2\_E |
|  |  |  | H3k27ac\_P | H3k27ac\_P | H3k27ac\_P | H3k27ac\_P | H3k4me2\_P | H3k4me2\_P | H3k4me2\_P | H3k4me2\_P |
|  |  |  |  |  | H3k4me3\_E | H3k4me3\_E | H3k4me3\_E | H3k4me3\_E | H3k4me3\_E | H3k4me3\_E |
|  |  |  |  |  | H3k4me3\_P | H3k4me3\_P | H3k4me3\_P | H3k4me3\_P | H3k4me3\_P | H3k4me3\_P |
|  |  |  |  |  |  |  | H3k27ac\_E | H3k9ac\_E | H3k9ac\_E | H3k9ac\_E |
|  |  |  |  |  |  |  | H3k27ac\_P | H3k9ac\_P | H3k9ac\_P | H3k9ac\_P |
|  |  |  |  |  |  |  | DNaseI\_E | H3k27ac\_E | H3k27ac\_E | H3k27ac\_E |
|  |  |  |  |  |  |  | DNaseI\_P | H3k27ac\_P | H3k27ac\_P | H3k27ac\_P |
|  |  |  |  |  |  |  |  | H3k27me3\_E | H3k27me3\_E | H3k27me3\_E |
|  |  |  |  |  |  |  |  | H3k27me3\_P | H3k27me3\_P | H3k27me3\_P |
|  |  |  |  |  |  |  |  | H3k36me3\_E | H3k36me3\_E | H3k36me3\_E |
|  |  |  |  |  |  |  |  | H3k36me3\_P | H3k36me3\_P | H3k36me3\_P |
|  |  |  |  |  |  |  |  | H3k79me2\_E | H3k79me2\_E | H3k79me2\_E |
|  |  |  |  |  |  |  |  | H3k79me2\_P | H3k79me2\_P | H3k79me2\_P |
|  |  |  |  |  |  |  |  | H4k20me1\_E | H4k20me1\_E | H4k20me1\_E |
|  |  |  |  |  |  |  |  | H4k20me1\_P | H4k20me1\_P | H4k20me1\_P |
|  |  |  |  |  |  |  |  | Ctcf\_E | Ctcf\_E | Ctcf\_E |
|  |  |  |  |  |  |  |  | Ctcf\_P | Ctcf\_P | Ctcf\_P |
|  |  |  |  |  |  |  |  | DNaseI\_E | DNaseI\_E | DNaseI\_E |
|  |  |  |  |  |  |  |  | DNaseI\_P | DNaseI\_P | DNaseI\_P |
|  |  |  |  |  |  |  |  |  | Pol2\_E | Pol2\_E |
|  |  |  |  |  |  |  |  |  | Pol2\_P | Pol2\_P |
|  |  |  |  |  |  |  |  |  |  | Rad21\_E |
|  |  |  |  |  |  |  |  |  |  | Rad21\_P |
|  |  |  |  |  |  |  |  |  |  | Smc3\_E |
|  |  |  |  |  |  |  |  |  |  | Smc3\_P |

Table S6: Performance of EPIP (trained on 30% of 30+5 cutoff balanced and unbalanced data) on the rest (70%) of the balanced and unbalanced data and all pairs within 2.5kb to 2Mb distance data labelled by 30+5 cutoff. The training data was removed from each of these datasets.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Cell line** | **# Pos** | **# Neg** | **AUROC** | **AUPR** | **F1** | **Accuracy** | **Precision** | **Sensitivity/**  **Recall** | **Specificity** | **# of condition specific EPIs** | **# of predicted condition specific EPIs** | **% of predicted condition specific EPIs** |
| Balanced (30+5) | GM12878 | 15902 | 15902 | 0.7661 | 0.7669 | 0.9967 | 0.9967 | 0.9967 | 0.9967 | 0.9967 | 14077 | 14026 | 0.9964 |
| HMEC | 2187 | 2187 | 0.9933 | 0.9931 | 0.9837 | 0.9835 | 0.9770 | 0.9904 | 0.9767 | 131 | 110 | 0.8397 |
| HUVEC | 1056 | 1056 | 0.9962 | 0.9957 | 0.9900 | 0.9901 | 0.9915 | 0.9886 | 0.9915 | 30 | 18 | 0.6000 |
| IMR90 | 1020 | 1020 | 0.9977 | 0.9976 | 0.9971 | 0.9971 | 0.9961 | 0.9980 | 0.9961 | 429 | 427 | 0.9953 |
| K562 | 2188 | 2188 | 0.9987 | 0.9987 | 0.9931 | 0.9931 | 0.9931 | 0.9931 | 0.9931 | 499 | 486 | 0.9739 |
| KBM7 | 9433 | 9433 | 0.9818 | 0.9802 | 0.9795 | 0.9795 | 0.9804 | 0.9787 | 0.9804 | 5869 | 5668 | 0.9658 |
| NHEK | 1245 | 1245 | 0.9959 | 0.9952 | 0.9892 | 0.9892 | 0.9880 | 0.9904 | 0.9880 | 45 | 33 | 0.7333 |
| **Overall** | **33031** | **33031** | **0.9600** | **0.9600** | **0.9902** | **0.9902** | **0.9900** | **0.9905** | **0.9900** | **21080** | **20768** | **0.9852** |
| Unbalanced (30+5) | GM12878 | 15902 | 159012 | 0.7657 | 0.5686 | 0.9827 | 0.9968 | 0.9691 | 0.9967 | 0.9968 | 14089 | 14038 | 0.9964 |
| HMEC | 2187 | 21868 | 0.9908 | 0.9670 | 0.9084 | 0.9819 | 0.8400 | 0.9890 | 0.9812 | 141 | 117 | 0.8298 |
| HUVEC | 1056 | 10556 | 0.9965 | 0.9793 | 0.9576 | 0.9920 | 0.9242 | 0.9934 | 0.9919 | 27 | 20 | 0.7407 |
| IMR90 | 1020 | 10199 | 0.9967 | 0.9854 | 0.9695 | 0.9943 | 0.9442 | 0.9961 | 0.9941 | 445 | 441 | 0.9910 |
| K562 | 2188 | 21875 | 0.9987 | 0.9959 | 0.9581 | 0.9921 | 0.9258 | 0.9927 | 0.9920 | 480 | 467 | 0.9729 |
| KBM7 | 9433 | 94325 | 0.9800 | 0.9344 | 0.8888 | 0.9778 | 0.8162 | 0.9756 | 0.9780 | 5834 | 5604 | 0.9606 |
| NHEK | 1245 | 12446 | 0.9960 | 0.9791 | 0.9314 | 0.9867 | 0.8772 | 0.9928 | 0.9861 | 39 | 30 | 0.7692 |
| **Overall** | **33031** | **330281** | **0.9600** | **0.9200** | **0.9447** | **0.9895** | **0.9037** | **0.9896** | **0.9895** | **21055** | **20717** | **0.9839** |
| All pairs within 2.5 kb to 2 Mb (30+5 cutoff) | GM12878 | 10652 | 33604 | 0.7322 | 0.5761 | 0.8993 | 0.9462 | 0.8180 | 0.9985 | 0.9296 | 9790 | 9774 | 0.9984 |
| HMEC | 662 | 10837 | 0.9768 | 0.6714 | 0.2914 | 0.7221 | 0.1707 | 0.9924 | 0.7055 | 77 | 72 | 0.9351 |
| HUVEC | 237 | 9130 | 0.9925 | 0.6575 | 0.4233 | 0.9314 | 0.2688 | 0.9958 | 0.9297 | 12 | 11 | 0.9167 |
| IMR90 | 612 | 6067 | 0.9875 | 0.9248 | 0.7416 | 0.9412 | 0.6205 | 0.9216 | 0.9431 | 301 | 269 | 0.8937 |
| K562 | 1021 | 14390 | 0.9974 | 0.9664 | 0.6412 | 0.9267 | 0.4746 | 0.9882 | 0.9224 | 303 | 295 | 0.9736 |
| KBM7 | 3162 | 72386 | 0.9722 | 0.6455 | 0.2155 | 0.6982 | 0.1209 | 0.9905 | 0.6854 | 2044 | 2014 | 0.9853 |
| NHEK | 389 | 8462 | 0.9851 | 0.6473 | 0.3880 | 0.8617 | 0.2408 | 0.9974 | 0.8555 | 21 | 20 | 0.9524 |
| **Overall** | **16735** | **154876** | **0.9500** | **0.7300** | **0.5114** | **0.8149** | **0.3443** | **0.9932** | **0.7956** | **12548** | **12455** | **0.9926** |

Here EPIP is trained with 30% of the 30 cutoff positives and 5 cutoff balanced and unbalanced negatives to create the balanced and unbalanced models. Then these models were used to predict the rest of the 70% balanced and unbalanced data and all EP-pairs having enhancers and promoters that are located within 2.5kb to 2Mb distance between them.

Table S7: Performance of EPIP (trained on 30+5 cutoff data) on 5kb resolution looplists data from Rao et al., 2014, IMR90 data from Jin et al., 2013 and ChIAPET data from Li et al., 2012.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cell line | # Pos | # Neg | AUROC | AUPR | F1 | Accuracy | Precision | Sensitivity/Recall | Specificity | # of cell specific EPIs | # of predicted cell specific EPIs | % of predicted cell specific EPIs |
| Looplists | GM12878 | 217 | 2170 | 0.8105 | 0.3732 | 0.9840 | 0.9971 | 0.9773 | 0.9908 | 0.9977 | 199 | 198 | 0.9950 |
| HeLa | 3 | 0 | - | - | 0.8000 | 0.6667 | 1.0000 | 0.6667 | - | 3 | 2 | 0.6667 |
| HMEC | 227 | 2270 | 0.8536 | 0.4196 | 0.9020 | 0.9808 | 0.8403 | 0.9736 | 0.9815 | 0 | - | - |
| HUVEC | 8 | 80 | 0.8531 | 0.5837 | 0.3636 | 0.9205 | 0.6667 | 0.2500 | 0.9875 | 0 | - | - |
| IMR90 | 70 | 700 | 0.9536 | 0.7400 | 0.7833 | 0.9662 | 0.9400 | 0.6714 | 0.9957 | 15 | 14 | 0.9333 |
| K562 | 30 | 300 | 0.9548 | 0.7638 | 0.9123 | 0.9848 | 0.9630 | 0.8667 | 0.9967 | 4 | 4 | 1.0000 |
| KBM7 | 8 | 80 | 0.8438 | 0.4318 | 0.7500 | 0.9545 | 0.7500 | 0.7500 | 0.9750 | 0 | - | - |
| NHEK | 0 | 0 | - | - | - | - | - | - | - | - | - | - |
| **Overall** | **563** | **5600** | **0.8782** | **0.5520** | **0.8606** | **0.9755** | **0.8962** | **0.8277** | **0.9904** | **221** | **218** | **0.9864** |
| Jin | IMR90 | 4601 | 147900 | 0.9497 | 0.5799 | 0.6709 | 0.9836 | 0.8492 | 0.5544 | 0.9969 | 195 | 191 | 0.9795 |
| **Overall** | **4601** | **147900** | **0.9497** | **0.5799** | **0.8508** | **0.9914** | **0.8920** | **0.8133** | **0.9969** | **195** | **191** | **0.9795** |
| ChIA-PET | K562 | 1433 | 25940 | 0.9233 | 0.6016 | 0.6978 | 0.9735 | 0.8665 | 0.5841 | 0.9950 | 21 | 21 | 1.0000 |
| MCF7 | 534 | 0 | - | - | 0.9457 | 0.8970 | 1.0000 | 0.8970 | - | 534 | 479 | 0.8970 |
| **Overall** | **1967** | **25940** | **0.9233** | **0.6016** | **0.9080** | **0.9874** | **0.9311** | **0.8861** | **0.9950** | **555** | **500** | **0.9009** |

Here, EPIP trained on 30+5 cutoff data was used to predict the strictly defined datasets by other studies including Rao et al., 2014 looplists (Rao et al., 2014, Jin et al., 2013 and Li et al., 2012). These strictly defined dataset did not provide negative data. So, we added randomly chosen 5 cutoff negative EP-pairs that are 10 times in size of the positive EP-pairs of the corresponding sample.

Table S8: Performance of EPIP (trained on 30+5 cutoff data) on all pairs within 2.5kb to 2Mb data labelled by other cutoffs (10+5 cutoff, 20+5 cutoff, 30+5 cutoff, 50+5 cutoff, 100+5 cutoff).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cutoff** | **Cell line** | **# Pos** | **# Neg** | **AUROC** | **AUPR** | **F1** | **Accuracy** | **Precision** | **Sensitivity/Recall** | **Specificity** | **# of condition specific EPIs** | **# of predicted condition specific EPIs** | **% of predicted condition specific EPIs** |
| 10 | GM12878 | 45558 | 33604 | 0.5515 | 0.6231 | 0.9477 | 0.9398 | 0.9480 | 0.9474 | 0.9296 | 9790 | 9774 | 0.9984 |
| HMEC | 6282 | 10837 | 0.9680 | 0.9317 | 0.7943 | 0.8112 | 0.6617 | 0.9935 | 0.7055 | 77 | 72 | 0.9351 |
| HUVEC | 3845 | 9130 | 0.9747 | 0.9403 | 0.8665 | 0.9185 | 0.8423 | 0.8921 | 0.9297 | 12 | 11 | 0.9167 |
| IMR90 | 5606 | 6067 | 0.9056 | 0.8856 | 0.6564 | 0.7392 | 0.8939 | 0.5186 | 0.9431 | 301 | 269 | 0.8937 |
| K562 | 8761 | 14390 | 0.9378 | 0.9195 | 0.7615 | 0.8356 | 0.8446 | 0.6932 | 0.9224 | 303 | 295 | 0.9736 |
| KBM7 | 35249 | 72386 | 0.9296 | 0.8783 | 0.7511 | 0.7852 | 0.6051 | 0.9900 | 0.6854 | 2044 | 2014 | 0.9853 |
| NHEK | 3537 | 8462 | 0.9560 | 0.9004 | 0.8134 | 0.8752 | 0.7274 | 0.9225 | 0.8555 | 21 | 20 | 0.9524 |
| **Overall** | **108838** | **154876** | **0.8900** | **0.8700** | **0.8315** | **0.8463** | **0.7595** | **0.9185** | **0.7956** | **12548** | **12455** | **0.9926** |
| 20 | GM12878 | 21217 | 33604 | 0.6155 | 0.5280 | 0.9430 | 0.9536 | 0.8989 | 0.9916 | 0.9296 | 9790 | 9774 | 0.9984 |
| HMEC | 1795 | 10837 | 0.9795 | 0.8452 | 0.5277 | 0.7468 | 0.3590 | 0.9955 | 0.7055 | 77 | 72 | 0.9351 |
| HUVEC | 899 | 9130 | 0.9854 | 0.8597 | 0.7238 | 0.9335 | 0.5765 | 0.9722 | 0.9297 | 12 | 11 | 0.9167 |
| IMR90 | 1803 | 6067 | 0.9650 | 0.9048 | 0.7792 | 0.9013 | 0.7990 | 0.7604 | 0.9431 | 301 | 269 | 0.8937 |
| K562 | 2876 | 14390 | 0.9897 | 0.9707 | 0.7856 | 0.9183 | 0.6981 | 0.8981 | 0.9224 | 303 | 295 | 0.9736 |
| KBM7 | 6931 | 72386 | 0.9616 | 0.7708 | 0.3741 | 0.7117 | 0.2308 | 0.9860 | 0.6854 | 2044 | 2014 | 0.9853 |
| NHEK | 893 | 8462 | 0.9820 | 0.8014 | 0.5907 | 0.8686 | 0.4204 | 0.9933 | 0.8555 | 21 | 20 | 0.9524 |
| **Overall** | **36414** | **154876** | **0.9300** | **0.8100** | **0.6840** | **0.8291** | **0.5278** | **0.9715** | **0.7956** | **12548** | **12455** | **0.9926** |
| 30 | GM12878 | 10652 | 33604 | 0.7322 | 0.5761 | 0.8993 | 0.9462 | 0.8180 | 0.9985 | 0.9296 | 9790 | 9774 | 0.9984 |
| HMEC | 662 | 10837 | 0.9768 | 0.6714 | 0.2914 | 0.7221 | 0.1707 | 0.9924 | 0.7055 | 77 | 72 | 0.9351 |
| HUVEC | 237 | 9130 | 0.9925 | 0.6575 | 0.4233 | 0.9314 | 0.2688 | 0.9958 | 0.9297 | 12 | 11 | 0.9167 |
| IMR90 | 612 | 6067 | 0.9875 | 0.9248 | 0.7416 | 0.9412 | 0.6205 | 0.9216 | 0.9431 | 301 | 269 | 0.8937 |
| K562 | 1021 | 14390 | 0.9974 | 0.9664 | 0.6412 | 0.9267 | 0.4746 | 0.9882 | 0.9224 | 303 | 295 | 0.9736 |
| KBM7 | 3162 | 72386 | 0.9722 | 0.6455 | 0.2155 | 0.6982 | 0.1209 | 0.9905 | 0.6854 | 2044 | 2014 | 0.9853 |
| NHEK | 389 | 8462 | 0.9851 | 0.6473 | 0.3880 | 0.8617 | 0.2408 | 0.9974 | 0.8555 | 21 | 20 | 0.9524 |
| **Overall** | **16735** | **154876** | **0.9500** | **0.7300** | **0.5114** | **0.8149** | **0.3443** | **0.9932** | **0.7956** | **12548** | **12455** | **0.9926** |
| 50 | GM12878 | 6005 | 33604 | 0.8599 | 0.7164 | 0.8350 | 0.9401 | 0.7172 | 0.9992 | 0.9296 | 5146 | 5141 | 0.9990 |
| HMEC | 192 | 10837 | 0.9781 | 0.3792 | 0.1074 | 0.7107 | 0.0568 | 1.0000 | 0.7055 | 6 | 6 | 1.0000 |
| HUVEC | 86 | 9130 | 0.9934 | 0.4175 | 0.2113 | 0.9303 | 0.1181 | 1.0000 | 0.9297 | 0 | - | - |
| IMR90 | 249 | 6067 | 0.9974 | 0.9120 | 0.5891 | 0.9452 | 0.4182 | 0.9960 | 0.9431 | 70 | 69 | 0.9857 |
| K562 | 477 | 14390 | 0.9989 | 0.9431 | 0.4606 | 0.9249 | 0.2992 | 1.0000 | 0.9224 | 64 | 64 | 1.0000 |
| KBM7 | 1372 | 72386 | 0.9760 | 0.4549 | 0.1064 | 0.6911 | 0.0562 | 0.9891 | 0.6854 | 841 | 826 | 0.9822 |
| NHEK | 150 | 8462 | 0.9843 | 0.4147 | 0.1970 | 0.8580 | 0.1092 | 1.0000 | 0.8555 | 2 | 2 | 1.0000 |
| **Overall** | **8531** | **154876** | **0.9700** | **0.6100** | **0.3495** | **0.8062** | **0.2119** | **0.9975** | **0.7956** | **6129** | **6108** | **0.9966** |
| 100 | GM12878 | 2273 | 33604 | 0.9774 | 0.8212 | 0.6577 | 0.9341 | 0.4900 | 1.0000 | 0.9296 | 1438 | 1438 | 1.0000 |
| HMEC | 44 | 10837 | 0.9861 | 0.1272 | 0.0268 | 0.7067 | 0.0136 | 1.0000 | 0.7055 | 0 | - | - |
| HUVEC | 11 | 9130 | 0.9934 | 0.0840 | 0.0331 | 0.9298 | 0.0168 | 1.0000 | 0.9297 | 0 | - | - |
| IMR90 | 67 | 6067 | 0.9981 | 0.7444 | 0.2797 | 0.9438 | 0.1626 | 1.0000 | 0.9431 | 3 | 3 | 1.0000 |
| K562 | 117 | 14390 | 0.9990 | 0.8069 | 0.1732 | 0.9230 | 0.0948 | 1.0000 | 0.9224 | 1 | 1 | 1.0000 |
| KBM7 | 578 | 72386 | 0.9877 | 0.2734 | 0.0482 | 0.6879 | 0.0247 | 0.9983 | 0.6854 | 331 | 330 | 0.9970 |
| NHEK | 74 | 8462 | 0.9877 | 0.2624 | 0.1080 | 0.8567 | 0.0571 | 1.0000 | 0.8555 | 0 | - | - |
| **Overall** | **3164** | **154876** | **0.9900** | **0.4500** | **0.1666** | **0.7997** | **0.0908** | **0.9997** | **0.7956** | **1773** | **1772** | **0.9994** |

Here, EPIP trained on 30+5 cutoff was tested on all pairs within 2.5kb to 2Mb where the positive EP pairs are defined by all the five cutoffs considered (10, 20, 30, 50 and 100) and the negative EP-pairs are labelled by 5 cutoff negative EP-pairs.

Table S9: Performance of the cell specific EPIP (1 test cell) models, on the 7th cell line for balanced, unbalanced and all pairs data.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test Cell line** | **# Pos** | **# Neg** | **AUROC** | **AUPR** | **F1** | **Accuracy** | **Precision** | **Sensitivity/**  **Recall** | **Specificity** | **# of condition specific EPIs** | **# of predicted condition specific EPIs** | **% of predicted condition specific EPIs** |
| Balanced data (30+5 cutoff) | GM12878 | 22716 | 22716 | 0.7572 | 0.7890 | 0.5612 | 0.6947 | 0.9975 | 0.3904 | 0.9990 | 20164 | 6407 | 0.3177 |
| HMEC | 3124 | 3124 | 0.9935 | 0.9948 | 0.9847 | 0.9846 | 0.9791 | 0.9904 | 0.9789 | 189 | 159 | 0.8413 |
| HUVEC | 1508 | 1508 | 0.9975 | 0.9981 | 0.9920 | 0.9920 | 0.9953 | 0.9887 | 0.9954 | 43 | 26 | 0.6047 |
| IMR90 | 1457 | 1457 | 0.9986 | 0.9989 | 0.9952 | 0.9952 | 0.9979 | 0.9925 | 0.9979 | 612 | 601 | 0.9820 |
| K562 | 3125 | 3125 | 0.9984 | 0.9987 | 0.9918 | 0.9918 | 0.9952 | 0.9885 | 0.9952 | 710 | 685 | 0.9648 |
| KBM7 | 13475 | 13475 | 0.9887 | 0.9889 | 0.9793 | 0.9794 | 0.9836 | 0.9751 | 0.9837 | 8326 | 7990 | 0.9596 |
| NHEK | 1778 | 1778 | 0.9977 | 0.9984 | 0.9932 | 0.9933 | 0.9949 | 0.9916 | 0.9949 | 60 | 45 | 0.7500 |
| **Overall** | **47183** | **47183** | **0.9600** | **0.9700** | **0.8180** | **0.8449** | **0.9897** | **0.6971** | **0.9928** | **30104** | **15913** | **0.5286** |
| Unbalanced data (30+5 cutoff) | GM12878 | 22716 | 227160 | 0.7569 | 0.5570 | 0.5584 | 0.9439 | 0.9803 | 0.3904 | 0.9992 | 20164 | 6407 | 0.3177 |
| HMEC | 3124 | 31240 | 0.9936 | 0.9838 | 0.9154 | 0.9834 | 0.8509 | 0.9904 | 0.9827 | 189 | 159 | 0.8413 |
| HUVEC | 1508 | 15080 | 0.9974 | 0.9875 | 0.9582 | 0.9922 | 0.9296 | 0.9887 | 0.9925 | 43 | 26 | 0.6047 |
| IMR90 | 1457 | 14570 | 0.9985 | 0.9938 | 0.9783 | 0.9960 | 0.9646 | 0.9925 | 0.9964 | 612 | 601 | 0.9820 |
| K562 | 3125 | 31250 | 0.9984 | 0.9949 | 0.9741 | 0.9952 | 0.9602 | 0.9885 | 0.9959 | 710 | 685 | 0.9648 |
| KBM7 | 13475 | 134750 | 0.9892 | 0.9507 | 0.9061 | 0.9816 | 0.8463 | 0.9751 | 0.9823 | 8326 | 7990 | 0.9596 |
| NHEK | 1778 | 17780 | 0.9977 | 0.9948 | 0.9626 | 0.9930 | 0.9353 | 0.9916 | 0.9931 | 60 | 45 | 0.7500 |
| **Overall** | **47183** | **471830** | **0.9600** | **0.9200** | **0.7869** | **0.9657** | **0.9033** | **0.6971** | **0.9925** | **30104** | **15913** | **0.5286** |
| All pairs within 2.5kb to 2Mb (100+5 cutoff) | **GM12878** | **4706** | **37168** | **0.9816** | **0.9657** | **0.9002** | **0.9788** | **0.9578** | **0.8491** | **0.9953** | **3045** | **2396** | **0.7869** |
| All pairs within 2.5kb to 2Mb (30+5 cutoff) | GM12878 | 21728 | 37168 | 0.7379 | 0.7015 | 0.5347 | 0.7638 | 0.9785 | 0.3678 | 0.9953 | 20004 | 6356 | 0.3177 |
| HMEC | 1381 | 11498 | 0.9870 | 0.9119 | 0.4762 | 0.7651 | 0.3129 | 0.9957 | 0.7374 | 147 | 141 | 0.9592 |
| HUVEC | 480 | 9524 | 0.9938 | 0.9174 | 0.5203 | 0.9124 | 0.3529 | 0.9896 | 0.9085 | 30 | 25 | 0.8333 |
| IMR90 | 1285 | 6364 | 0.9966 | 0.9880 | 0.8744 | 0.9520 | 0.7806 | 0.9938 | 0.9436 | 605 | 597 | 0.9868 |
| K562 | 2080 | 15145 | 0.9980 | 0.9934 | 0.8219 | 0.9482 | 0.7029 | 0.9894 | 0.9426 | 655 | 642 | 0.9802 |
| KBM7 | 6367 | 75067 | 0.9711 | 0.6777 | 0.3951 | 0.7642 | 0.2471 | 0.9849 | 0.7455 | 4152 | 4056 | 0.9769 |
| NHEK | 808 | 8978 | 0.9974 | 0.9812 | 0.7043 | 0.9310 | 0.5451 | 0.9950 | 0.9253 | 41 | 37 | 0.9024 |
| **Overall** | **38835** | **200912** | **0.9600** | **0.8900** | **0.5490** | **0.8338** | **0.4899** | **0.6244** | **0.8743** | **28679** | **14250** | **0.4969** |

Here, using balanced, unbalanced and all pairs within 2.5kb to 2Mb data, six different cell-specific EPIP models were trained with the corresponding data of six different cell lines leaving one cell line as test cell line. These six cell-specific models were then tested to predict the test cell. This was done separately on the balanced, unbalanced and all pairs within 2.5kb to 2Mb data.

Table S10: EPIP vs TargetFinder and Ripple on TargetFinder data using cross validation

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cell\_line | Pos | Neg | AUROC | AUPR | F-score | Accuracy | Precision | Sensitivity/  Recall | Specificity | # of condition specific EPIs | # of predicted condition specific EPIs | % of predicted condition specific EPIs |
| TargetFinder | GM12878 | 2113 | 42200 | 0.9282 | 0.6033 | 0.5128 | 0.9644 | 0.7367 | 0.3933 | 0.993 | 1911 | 690 | 0.3611 |
| HeLa-S3 | 1740 | 34800 | 0.9311 | 0.6218 | 0.5459 | 0.9651 | 0.7168 | 0.4408 | 0.9913 | 1417 | 503 | 0.355 |
| HUVEC | 1524 | 30400 | 0.8944 | 0.5022 | 0.4097 | 0.9596 | 0.6757 | 0.294 | 0.9929 | 1268 | 404 | 0.3186 |
| IMR90 | 1254 | 25000 | 0.9097 | 0.5242 | 0.4283 | 0.9607 | 0.6998 | 0.3086 | 0.9934 | 1101 | 342 | 0.3106 |
| K562 | 1977 | 39500 | 0.9325 | 0.6455 | 0.5386 | 0.9662 | 0.7729 | 0.4133 | 0.9939 | 1749 | 800 | 0.4574 |
| NHEK | 1291 | 25600 | 0.948 | 0.6216 | 0.5344 | 0.9638 | 0.6979 | 0.433 | 0.9905 | 1025 | 552 | 0.5385 |
| Overall | 9899 | 197500 | 0.924 | 0.5864 | 0.5021 | 0.9636 | 0.7225 | 0.3848 | 0.9926 | 8471 | 3291 | 0.3885 |
| EPIP | GM12878 | 2113 | 42200 | 0.9496 | 0.8352 | 0.6866 | 0.9768 | 0.9632 | 0.5334 | 0.999 | 1911 | 1020 | 0.5338 |
| HELA | 1740 | 34800 | 0.9577 | 0.8779 | 0.6854 | 0.977 | 0.9839 | 0.5259 | 0.9996 | 1417 | 853 | 0.602 |
| HUVEC | 1524 | 30400 | 0.9302 | 0.7877 | 0.4824 | 0.9673 | 0.9838 | 0.3196 | 0.9997 | 1268 | 404 | 0.3186 |
| IMR90 | 1254 | 25000 | 0.9392 | 0.7859 | 0.5969 | 0.9721 | 0.9644 | 0.4322 | 0.9992 | 1101 | 550 | 0.4995 |
| K562 | 1977 | 39500 | 0.9484 | 0.841 | 0.6557 | 0.9753 | 0.9799 | 0.4927 | 0.9995 | 1749 | 890 | 0.5089 |
| NHEK | 1291 | 25600 | 0.9752 | 0.9038 | 0.6941 | 0.9774 | 0.9871 | 0.5352 | 0.9996 | 1025 | 634 | 0.6185 |
| Overall | 9899 | 197500 | 0.95 | 0.8386 | 0.6422 | 0.9746 | 0.9763 | 0.4784 | 0.9994 | 8471 | 4351 | 0.5136 |
| Ripple (on 3 common TargetFinder cell lines) | GM12878 | 2113 | 42200 | 0.772 | 0.203 | 0.01 | 0.952 | 0.55 | 0.005 | 1 | 1911 | 7 | 0.0037 |
| HELA | 1740 | 34800 | 0.743 | 0.209 | 0.03 | 0.953 | 0.844 | 0.016 | 1 | 1417 | 15 | 0.0106 |
| K562 | 1977 | 39500 | 0.729 | 0.165 | 0.005 | 0.952 | 1 | 0.003 | 1 | 1749 | 5 | 0.0029 |
| Overall | 5830 | 116500 | 0.7478 | 0.1922 | 0.0146 | 0.9526 | 0.7544 | 0.0074 | 0.9999 | 5077 | 27 | 0.0053 |
| EPIP (on 3 common cell lines between TargetFinder and Ripple) | GM12878 | 2113 | 42200 | 0.9496 | 0.8352 | 0.6866 | 0.9768 | 0.9632 | 0.5334 | 0.999 | 1911 | 1020 | 0.5338 |
| HELA | 1740 | 34800 | 0.9577 | 0.8779 | 0.6854 | 0.977 | 0.9839 | 0.5259 | 0.9996 | 1417 | 853 | 0.602 |
| K562 | 1977 | 39500 | 0.9484 | 0.841 | 0.6557 | 0.9753 | 0.9799 | 0.4927 | 0.9995 | 1749 | 890 | 0.5089 |
| Overall | 5830 | 116500 | 0.9519 | 0.8514 | 0.6759 | 0.9764 | 0.9748 | 0.5173 | 0.9993 | 5077 | 2763 | 0.5442 |

Table S11: EPIP vs TargetFinder and Ripple on EPIP data (all pairs within 2.5kb to 2Mb) using cross validation.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cell\_line | Pos | Neg | AUROC | AUPR | F-score | Accuracy | Precision | Sensitivity/  Recall | Specificity | # of condition specific EPIs | # of predicted condition specific EPIs | % of predicted condition specific EPIs |
| TargetFinder | GM12878 | 21536 | 36832 | 0.9587 | 0.9513 | 0.8778 | 0.9154 | 0.9399 | 0.8234 | 0.9692 | 19850 | 16655 | 0.839 |
| HUVEC | 385 | 7807 | 0.9295 | 0.6754 | 0.4876 | 0.9672 | 0.9143 | 0.3325 | 0.9985 | 22 | 12 | 0.5455 |
| IMR90 | 1207 | 5961 | 0.9618 | 0.9062 | 0.7992 | 0.942 | 0.9572 | 0.686 | 0.9938 | 540 | 450 | 0.8333 |
| K562 | 1986 | 14398 | 0.9772 | 0.9405 | 0.8405 | 0.9662 | 0.9812 | 0.7351 | 0.9981 | 602 | 520 | 0.8638 |
| NHEK | 751 | 8465 | 0.968 | 0.8743 | 0.6473 | 0.9566 | 0.9582 | 0.4887 | 0.9981 | 30 | 20 | 0.6667 |
| Overall | 25865 | 73463 | 0.959 | 0.8695 | 0.8618 | 0.9338 | 0.9436 | 0.7932 | 0.9833 | 21044 | 17657 | 0.8391 |
| EPIP (on 5 common TargetFinder cell lines) | GM12878 | 21728 | 37168 | 1 | 1 | 0.9965 | 0.9974 | 0.9939 | 0.9991 | 0.9964 | 20004 | 20001 | 0.9999 |
| HUVEC | 480 | 9524 | 1 | 0.97 | 0.9697 | 0.997 | 0.9412 | 1 | 0.9969 | 30 | 30 | 1 |
| IMR90 | 1285 | 6364 | 1 | 0.98 | 0.9705 | 0.9898 | 0.9434 | 0.9992 | 0.9879 | 605 | 605 | 1 |
| K562 | 2080 | 15145 | 1 | 0.99 | 0.9855 | 0.9965 | 0.9719 | 0.9995 | 0.996 | 655 | 655 | 1 |
| NHEK | 808 | 8978 | 1 | 0.97 | 0.9854 | 0.9975 | 0.9712 | 1 | 0.9973 | 41 | 41 | 1 |
| Overall | 26381 | 77179 | 1 | 0.982 | 0.9935 | 0.9967 | 0.9879 | 0.9992 | 0.9958 | 21335 | 21332 | 0.9999 |
| Ripple | GM12878 | 21728 | 37168 | 0.6645 | 0.5456 | 0.3792 | 0.6668 | 0.6063 | 0.2759 | 0.8953 | 20004 | 5587 | 0.2793 |
| K562 | 2080 | 15145 | 0.6628 | 0.2392 | 0.0143 | 0.8799 | 0.7895 | 0.0072 | 0.9997 | 655 | 5 | 0.0076 |
| Overall | 23808 | 52313 | 0.6637 | 0.3924 | 0.3565 | 0.715 | 0.6066 | 0.2524 | 0.9255 | 20659 | 5592 | 0.2707 |
| EPIP (on 2 common Ripple cell lines) | GM12878 | 21728 | 37168 | 1 | 1 | 0.9965 | 0.9974 | 0.9939 | 0.9991 | 0.9964 | 20004 | 20001 | 0.9999 |
| K562 | 2080 | 15145 | 1 | 0.99 | 0.9855 | 0.9965 | 0.9719 | 0.9995 | 0.996 | 655 | 655 | 1 |
| Overall | 23808 | 52313 | 1 | 0.995 | 0.9955 | 0.9972 | 0.992 | 0.9992 | 0.9963 | 20659 | 20656 | 0.9999 |

Table S12. Comparison of EPIP (30+5) model performance with the performances of EPIP (100+5), EPIP (100+30+5) and EPIP (looplists+5) models on balanced test data, unbalanced test data, and all pairs within 2.5kb to 2Mb test data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Cell line | AUROC | AUPR | F1 | Precision | Sensitivity/Recall | Specificity | % of predicted cell specific EPIs |
| EPIP (trained on 30+5) | GM12878 | 0.73 (0.77,0.77) | 0.58 (0.77,0.57) | 0.9 (1,0.98) | 0.82 (1,0.97) | 1 (1,1) | 0.93 (1,1) | 1 (1,1) |
| HMEC | 0.98 (0.99,0.99) | 0.67 (0.99,0.97) | 0.29 (0.98,0.91) | 0.17 (0.98,0.84) | 0.99 (0.99,0.99) | 0.71 (0.98,0.98) | 0.94 (0.84,0.84) |
| HUVEC | 0.99 (1,1) | 0.66 (1,0.98) | 0.42 (0.99,0.96) | 0.27 (0.99,0.92) | 1 (0.99,0.99) | 0.93 (0.99,0.99) | 0.92 (0.6,0.56) |
| IMR90 | 0.99 (1,1) | 0.92 (1,0.99) | 0.74 (1,0.97) | 0.62 (1,0.94) | 0.92 (1,1) | 0.94 (1,0.99) | 0.89 (1,0.99) |
| K562 | 1 (1,1) | 0.97 (1,1) | 0.64 (0.99,0.96) | 0.47 (0.99,0.93) | 0.99 (0.99,0.99) | 0.92 (0.99,0.99) | 0.97 (0.97,0.98) |
| KBM7 | 0.97 (0.98,0.98) | 0.65 (0.98,0.93) | 0.22 (0.98,0.89) | 0.12 (0.98,0.82) | 0.99 (0.98,0.98) | 0.69 (0.98,0.98) | 0.99 (0.97,0.96) |
| NHEK | 0.99 (1,1) | 0.65 (1,0.98) | 0.39 (0.99,0.93) | 0.24 (0.99,0.88) | 1 (0.99,0.99) | 0.86 (0.99,0.99) | 0.95 (0.73,0.81) |
| **Overall** | **0.95 (0.96,0.96)** | **0.73 (0.96,0.92)** | **0.51 (0.99,0.94)** | **0.34 (0.99,0.9)** | **0.99 (0.99,0.99)** | **0.8 (0.99,0.99)** | **0.99 (0.99,0.98)** |
| EPIP (trained on 100+5) | GM12878 | 0.97 (0.97,0.97) | 0.93 (0.97,0.94) | 0.68 (1,1) | 0.51 (1,1) | 1 (1,1) | 0.98 (1,1) | 1 (1,1) |
| HMEC | 0.99 (1,1) | 0.11 (1,1) | 0.01 (1,0.95) | 0.01 (0.99,0.9) | 1 (1,1) | 0.74 (0.99,0.99) | - |
| HUVEC | 1 (0.99,0.99) | 0.04 (0.99,0.98) | 0 (0.97,0.94) | 0 (0.97,0.9) | 1 (0.97,0.97) | 0.76 (0.97,0.99) | - |
| IMR90 | 1 (1,1) | 0.95 (1,1) | 0.03 (1,0.97) | 0.01 (1,0.94) | 1 (1,1) | 0.82 (1,0.99) | 1 (1,1) |
| K562 | 1 (1,1) | 0.8 (1,1) | 0.01 (0.99,0.95) | 0.01 (0.98,0.9) | 1 (1,1) | 0.81 (0.98,0.99) | 1 (1,1) |
| KBM7 | 0.99 (1,1) | 0.65 (1,1) | 0.05 (1,0.96) | 0.03 (0.99,0.93) | 1 (1,1) | 0.8 (0.99,0.99) | 1 (1,1) |
| NHEK | 1 (1,1) | 0.24 (1,1) | 0.02 (0.99,0.94) | 0.01 (0.99,0.88) | 1 (1,1) | 0.75 (0.99,0.99) | - |
| **Overall** | **0.99 (0.99,0.99)** | **0.53 (0.99,0.99)** | **0.09 (1,0.97)** | **0.05 (0.99,0.95)** | **1 (1,1)** | **0.83 (0.99,0.99)** | **1 (1,1)** |
| EPIP (trained on 100+30+5) | GM12878 | 0.98 (0.98,0.98) | 0.42 (0.96,0.7) | 0.68 (1,1) | 0.51 (1,1) | 1 (1,1) | 0.98 (1,1) | 1 (1,1) |
| HMEC | 0.97 (0.97,0.97) | 0.3 (0.95,0.7) | 0.16 (0.99,0.93) | 0.09 (0.98,0.88) | 0.99 (0.99,0.99) | 0.79 (0.98,0.99) | 0.91 (0.85,0.85) |
| HUVEC | 0.99 (0.99,0.99) | 0.44 (1,0.99) | 0.13 (0.99,0.97) | 0.07 (1,0.96) | 0.99 (0.99,0.99) | 0.92 (1,1) | 0.83 (0.61,0.61) |
| IMR90 | 0.99 (1,1) | 0.54 (1,0.98) | 0.43 (0.99,0.98) | 0.28 (1,0.98) | 0.99 (0.99,0.99) | 0.94 (1,1) | 0.98 (0.97,0.97) |
| K562 | 1 (1,1) | 0.77 (1,0.99) | 0.26 (0.99,0.98) | 0.15 (1,0.96) | 1 (0.99,0.99) | 0.92 (1,1) | 0.99 (0.97,0.97) |
| KBM7 | 0.92 (0.95,0.95) | 0.17 (0.93,0.62) | 0.12 (0.94,0.62) | 0.06 (0.9,0.45) | 0.98 (0.98,0.98) | 0.55 (0.89,0.88) | 0.97 (0.97,0.97) |
| NHEK | 0.99 (1,1) | 0.6 (1,0.98) | 0.26 (0.99,0.97) | 0.15 (0.99,0.95) | 1 (0.99,0.99) | 0.9 (0.99,0.99) | 0.9 (0.79,0.79) |
| **Overall** | **0.98 (0.98,0.98)** | **0.46 (0.98,0.85)** | **0.2 (0.97,0.78)** | **0.11 (0.95,0.64)** | **0.99 (0.99,0.99)** | **0.83 (0.95,0.95)** | **0.99 (0.98,0.98)** |
| EPIP (trained on looplists+5) | GM12878 | 0.53 (0.96,0.97) | 0 (0.93,0.62) | 0 (0.98,0.86) | 0 (0.96,0.76) | 1 (1,1) | 0.17 (0.95,0.97) | 1 (1,1) |
| HMEC | 0.55 (0.96,0.97) | 0 (0.92,0.65) | 0 (0.97,0.85) | 0 (0.94,0.74) | 1 (1,1) | 0.23 (0.94,0.96) | 1 (1,1) |
| HUVEC | 0.57 (1,0.98) | 0 (1,0.75) | 0 (1,0.86) | 0 (1,0.75) | 1 (1,1) | 0.36 (1,0.96) | 1 (1,1) |
| IMR90 | 0.57 (0.97,0.97) | 0 (0.94,0.64) | 0 (0.97,0.91) | 0 (0.96,0.84) | 0.98 (0.98,0.98) | 0.37 (0.96,0.98) | 0.97 (0.98,0.98) |
| K562 | 0.58 (1,0.97) | 0 (1,0.66) | 0 (1,0.82) | 0 (1,0.7) | 1 (1,1) | 0.3 (1,0.96) | 1 (1,1) |
| KBM7 | 0.5 (0.92,0.95) | 0 (0.86,0.5) | 0 (0.91,0.77) | 0 (1,0.71) | 0.83 (0.83,0.83) | 0.18 (1,0.96) | 0.75 (0.75,0.75) |
| NHEK | - | - | - | - | - | - | - |
| **Overall** | **0.55 (0.97,0.97)** | **0 (0.94,0.63)** | **0 (0.97,0.86)** | **0 (0.95,0.75)** | **0.99 (0.99,0.99)** | **0.24 (0.95,0.97)** | **0.99 (0.99,0.99)** |

In each entry with three numbers, the three numbers in order are for all pairs within 2.5 kb to 2 Mb test data, balanced test data, and unbalanced test data, respectively.

Table S13: Comparison of EPIP (30+5) model performance with the performances of EPIP (100+5) and EPIP (100+30+5) models on 5kb resolution looplists data from Rao et al., 2014, IMR90 data from Jin et al., 2013 and ChIAPET data from Li et al., 2012.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | AUROC | AUPR | F1 | Precision | Sensitivity/Recall | Specificity | % of predicted cell specific EPIs |
| EPIP (trained on 30+5) | Looplist | 0.8800 | 0.5500 | 0.9137 | 0.9058 | 0.9218 | 0.9904 | 0.9864 |
| Jin | 0.9500 | 0.5800 | 0.6709 | 0.8492 | 0.5544 | 0.9969 | 0.9795 |
| ChIA-PET | 0.9200 | 0.6000 | 0.7714 | 0.9107 | 0.6690 | 0.9950 | 0.9009 |
| EPIP (trained on 100+5) | Looplist | 0.5900 | 0.2600 | 0.9296 | 0.9215 | 0.9378 | 0.9920 | 0.9774 |
| Jin | 0.6300 | 0.2800 | 0.8412 | 0.8170 | 0.8668 | 0.9940 | 1.0000 |
| ChIA-PET | 0.6400 | 0.3200 | 0.9017 | 0.8808 | 0.9237 | 0.9905 | 0.8829 |
| EPIP (trained on 100+30+5) | Looplist | 0.8800 | 0.5500 | 0.8388 | 0.8936 | 0.7904 | 0.9905 | 0.9683 |
| Jin | 0.9500 | 0.5800 | 0.6427 | 0.8802 | 0.5062 | 0.9979 | 0.9744 |
| ChIA-PET | 0.9200 | 0.6000 | 0.7323 | 0.9227 | 0.6070 | 0.9961 | 0.7459 |

Figure S1: The comparison between the performance of the model with different number of estimators and maximum depths. EPIP model with 200 estimators and 10 maximum depth showed the best AUROC and AUPR scores.

A screenshot of a cell phone

Description generated with very high confidence