An Empirical Bayes Test for Allelic-Imbalance Detection in ChIP-seq: Supplementary Materials

Qi Zhang
Department of Statistics, University of Nebraska Lincoln
and
Sündüz Keleş
Department of Biostatistics and Medical Informatics
and Department of Statistics , University of Wisconsin Madison

April 10, 2017

Supplementary Notes

Implementation details of Non-Parametric Binomial (NPBin) test

NPBin includes two tuning parameters. Specifically, estimation of the density g for the latent allelic probability requires choices for K, the order, and $T + K - 3$, the number of basis functions. For ALI detection, we chose $K = 4$ and $T = 10$. For the baseball example, we chose $K = 4$ and $T = 20$. Selecting the number of (equally spaced) knots (and therefore the number of basis) is an acknowledged challenging problem (Belitser *and others*, 2014; Ruppert, 2002; Spiriti *and others*, 2013), and could be very computationally expensive in

our ALI detection setting. Our empirical studies with the ALI detection data indicated that the number of basis functions does not need to be very large to get good data fits. Hence, we did not need additional penalty on smoothness. One may increase the number of basis functions and penalize on the smoothness, which will then require selection of the penalty term. We found this to be unnecessary in our setting; however, in principle, it could be useful if the density of the latent variable is very skewed or unbounded.

The initialization of the EM algorithm is also a potentially critical point. We initialize $(a_{-K+2}, \cdots, a_{T-2})$ by splitting $[0, 1]$ into $T + K - 3$ equally sized intervals, and counting the proportion of $\hat{p}_j$'s, where $\hat{p}_j = x_j/m_j$, in these intervals.

When estimating g_0, we select a dense grid in $[0, 1]$ where the functional in (2.5) is evaluated. We use $10^{-4} \times \{1, 2, \ldots, 10^4 - 1\}$ as this dense grid.

The user can specify other values of the above parameters and different initialization strategies as they wish. Our general suggestion is to decide based on the shape and smoothness of the histogram of $\hat{p}_j$'s, e.g., more knots and basis functions are needed to characterize spikier signal pattern.

Details of Empirical Bayes Oracle (EBO) test

EBO first directly estimates g using splines. Similar to our proposed method, the number of knots are fixed and estimation of g does not involve a smoothness penalty. Then the coefficients, a_ℓ's, in (2.2) can be estimated by maximizing the marginal likelihood g with an EM algorithm. Next, EBO estimates g_0 in Beta family by maximum likelihood using the data in the "bulk" region and also accounting for truncation. In the context of ALI detection, the bulk region should be an interval around 0.5. Similar to Efron *and others* (2001), EBO estimates its empirical null density using the following likelihood-based approach. Let $b \in (0, 0.5)$, and $q_b(p)$ be the b'th quantile of the distribution of p. We define the inclusion indicator d_j as

$$d_j = 1\left\{q_{0.5-b} \leq p_j \leq q_{0.5+b}\right\},$$

so the "bulk" region in our case is the interval $[q_{0.5-b}, q_{0.5+b}]$. Let $g_0(p) = Beta(p; \alpha_0, \beta_0)$. We then estimate $(\alpha_0, \beta_0, \pi_0)$ by maximizing the following likelihood

$$L(\alpha_0, \beta_0, \pi_0) = \sum_{j=1}^{n} d_j \log\left(\pi_0 Beta(p_j; \alpha_0, \beta_0)\right) + (1 - d_j)\log(1 - \pi_0 F_0),$$

where

$$F_0 \equiv P(d_j = 1 \text{ under the null}) = \int_{q_{0.5-b}}^{q_{0.5+b}} Beta(p; \alpha_0, \beta_0)dp.$$

After $g(p)$ and $g_0(p)$ are estimated, we then define local FDR and FDR using the corresponding f and f_0. In practice, we set $b = 0.45$ in real data analysis and in simulations with $\pi_0 \geq 0.9$, and $b = 0.4$ in simulations with $\pi_0 < 0.9$.

External validation criteria of ALI detection results in actual ChIP-seq data using the allelic difference in the strength of the transcription factor binding motif

We designed evaluation criteria for the actual ChIP-seq ALI detection analysis we performed. Specifically, we used DNA motif information as a benchmark. This is a natural evaluation criterion for ChIP-seq assays targeting sequence-specific transcription factors that bind to a specific short DNA sequence patterns. For example, CTCF factor interacts with sequences similar to one or multiple copies of "CCCTC" (or "GAGGG" in the reverse complement strand; Mathelier *and others* 2013). Such DNA sequence pattern is called a motif for the TF and is commonly characterized by a position weight matrix (PWM). PWMs for many TFs have been learned from previous studies, and can be viewed as known information, e.g., the JASPAR database (Mathelier *and others*, 2013) harbors PWMs of 205 known TFs in vertebrates. The information content of the PWM can be visualized as a sequence logo. The top panel of Supplementary Figure 4 depicts the sequence logo of the CTCF factor. Sequence logos also depict DNA sequences that are more likely to be bound the corresponding TF.

If a SNP resides within a TF binding site, the DNA sequences from the paternal and the maternal alleles at the same genomic location differ by one base, which in turn may cause a difference in the motif strength, i.e., the likelihood that the TF will interact with the sequence, between the two alleles. Because the motif strength difference is defined using the sequence information and the known results from previous studies, which are independent of the ChIP-seq data under consideration, it can serve as an external validation criterion. We expect the allele with the stronger motif to be the *winning allele* of ALI detection from ChIP-seq. In Supplementary Figure 4, the sequences from the maternal allele and

the paternal allele differ by only one letter. However, nucleotide G in the paternal allele is more consistent with the sequence logo compared to the nucleotide C in the maternal allele. Thus, CTCF is more likely to bind with the paternal allele than the maternal allele. As a result, for the CTCF ChIP-seq data, we would expect that the majority of reads overlapping this SNP are from the paternal allele and the paternal allele is the winning allele in ALI detection. We used `R` package `atSNP` (Zuo *and others*, 2015) to calculate p-values for the strengths of the binding motif on both alleles, and their allelic strength difference based on sequence information. SNPs that satisfy the following two conditional were labeled as having a significant allelic motif strength difference: (1) at least one allele had a p-value less than 10^{-4} in the test for the strength of the motif; and (2) the p-value in the test for allelic motif strength difference was less than 0.1. The expected winning allele of such a SNP was the allele leading to stronger motif (smaller p-value in the test of for the strength of the motif on this allele).

Individual DNase-seq and ATAC-seq signals do not associate with specific DNA sequences; hence, the above criterion of using a specific PWM to evaluate allelic motif strength difference does not work for validation purposes. Both DNase-seq and ATAC-seq aim to quantify regions of the genome accessible by transcription factors. To a first approximation, DNase-seq and ATAC-seq peaks are highly correlated with the union of the binding sites of all TFs (Pique-Regi *and others*, 2011). Therefore, the direction of ALI in DNase-seq and ATAC-seq at each SNP should be consistent with the direction of the allelic motif strength difference of some TF that may bind at the same SNP. We considered the significance of the allelic difference in motif strength for all the 205 known vertebrate PWMs in the JASPAR database, and used the direction and atSNP p-value of the most significant change to label the SNPs in evaluation the DNase-seq and ATAC-seq ALI applications. We labeled a SNP as having a significant allelic motif strength difference if

p-value<0.1 for at least one PWM at this SNP, and used the winning allele of the PWM with the smallest p-value as the expected winning allele. Such external validation criterion is only available for 5-10% of SNPs, and the other SNPs simply showed no significant allelic difference in motif strength. Reasons for this include other unknown factors affecting TF binding and chromatin accessibility and the incompleteness of the JASPAR database, i.e., PWMs for a large number of human TFs still remain unknown.

For each of the 5-10% SNPs that exhibit significant allelic difference in motif strength, the expected winning allele in ALI detection is implied by the PWM calculations. They were compared with the actual winning alleles in ALI detection from ChIP-seq data. Note that all ALI detection methods will report the same winning allele, e.g., if $x_j > m_j - x_j$, indicating more maternal reads than paternal reads at SNP j, all methods will report the maternal allele as the winning allele. However, their estimated significance could be dramatically different. We label a SNP as a potential True Positive (TP) if the expected winning allele based on sequence information and the actual winning allele in ALI detection from ChIP-seq data are the same and a potential False Positive (FP) otherwise. When such a potential TP (or FP) SNP has enough statistical significance to be chosen as an ALI SNP from ChIP-seq data, its winning allele will agree (or contradict) with the motif-based benchmark. Thus, it can be viewed as a true positive (or false positive) in the conventional setting. However, potential true negatives or false negatives cannot be defined in the same fashion due to the incompleteness of the databases of known TF motifs, and the limited knowledge on the other factors that may affect ChIP-seq signal.

References

Belitser, Eduard, Serra, Paulo *and others*. (2014). Adaptive priors based on splines with random knots. *Bayesian Analysis* **9**(4), 859–882.

Buenrostro, Jason D, Giresi, Paul G, Zaba, Lisa C, Chang, Howard Y and Greenleaf, William J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**(12), 1213–1218.

Consortium, ENCODE Project *and others*. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**(7414), 57–74.

Efron, Bradley, Tibshirani, Robert, Storey, John D and Tusher, Virginia. (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American statistical association* **96**(456), 1151–1160.

Mathelier, Anthony, Zhao, Xiaobei, Zhang, Allen W, Parcy, François, Worsley-Hunt, Rebecca, Arenillas, David J, Buchman, Sorana, Chen, Chih-yu, Chou, Alice, Ienasescu, Hans *and others*. (2013). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic acids research*, gkt997.

Pique-Regi, Roger, Degner, Jacob F, Pai, Athma A, Gaffney, Daniel J, Gilad, Yoav and Pritchard, Jonathan K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome research* **21**(3), 447–455.

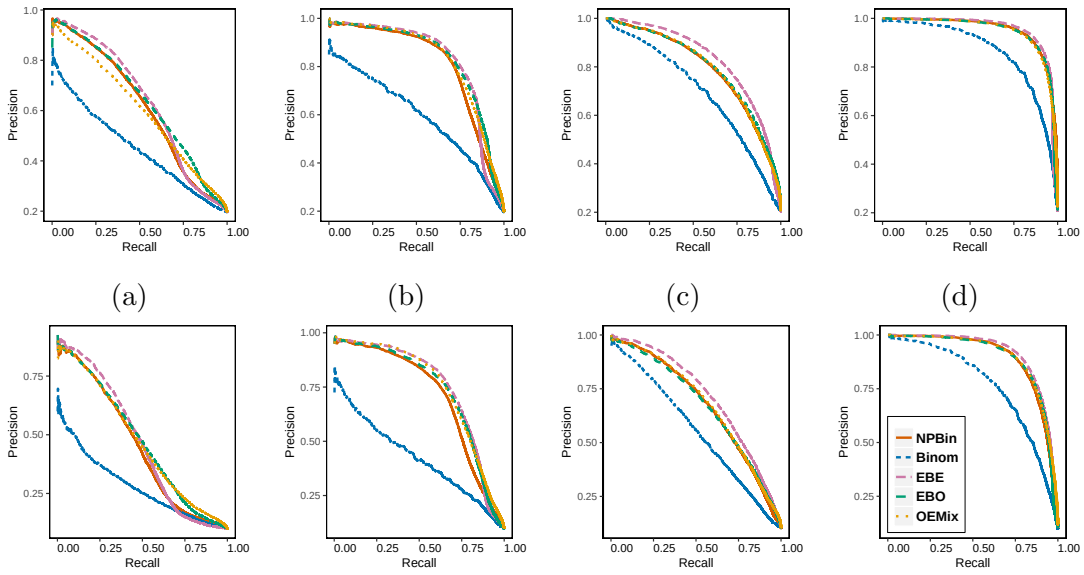
Ruppert, David. (2002). Selecting the Number of Knots for Penalized Splines. *Journal of Computational and Graphical Statistics* **11**(4), 735–757.

Spiriti, Steven, Eubank, Randall, Smith, Philip W and Young, Dennis. (2013). Knot selection for least-squares and penalized splines. *Journal of Statistical Computation and Simulation* **83**(6), 1020–1036.

Zhang, Qi, Zeng, Xin, Younkin, Sam, Kawli, Trupti, Snyder, Michael P and Keleş, Sündüz. (2016). Systematic evaluation of the impact of ChIP-seq read designs on genome coverage, peak identification, and allele-specific binding detection. *BMC bioinformatics* **17**(1), 1.

Zuo, Chandler, Shin, Sunyoung and Keleş, Sündüz. (2015). atSNP: transcription factor binding affinity testing for regulatory SNP detection. *Bioinformatics*, btv328.

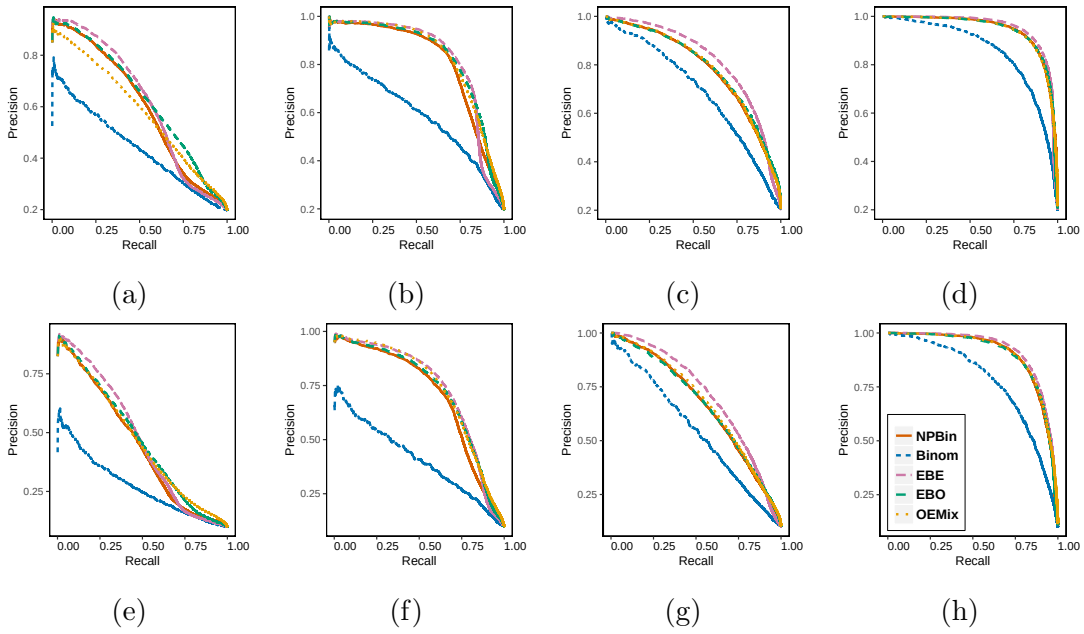
Supplementary Figures



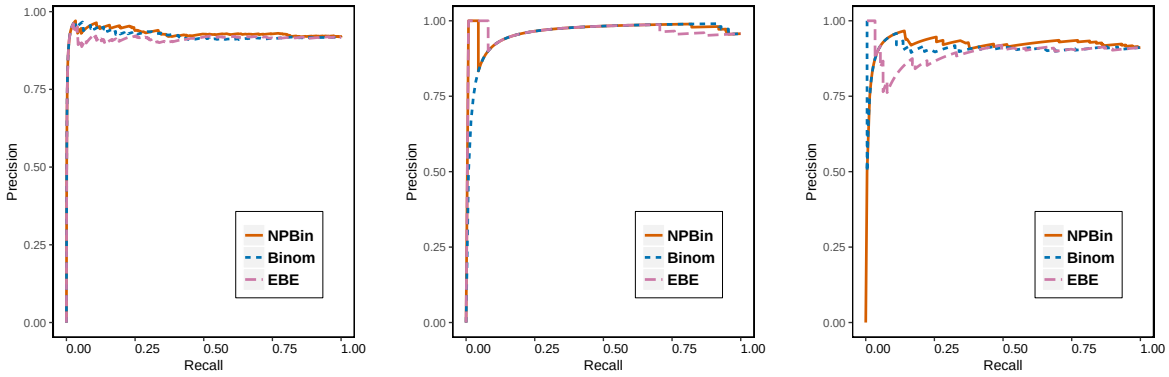
(a) (b) (c) (d)

(e) (f) (g) (h)

Supplementary Figure 1: The precision-recall plots for simulations with Beta null (π_0, α_0, d) = (a) (0.8,5,0.3); (b) (0.8,5,0.4); (c) (0.8,20,0.3); (d) (0.8,20,0.4); (e) (0.9,5,0.3); (f) (0.9,5,0.4); (g) (0.9,20,0.3); (h) (0.9,20,0.4).

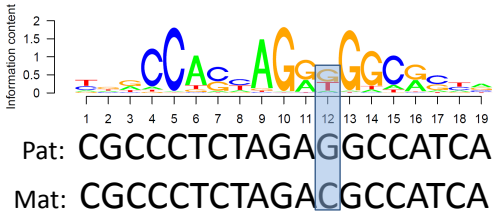


Supplementary Figure 2: The precision-recall plots for simulations with mis-specified null (not a beta distribution) (π_0, α_0, d) = (a) (0.8,5,0.3); (b) (0.8,5,0.4); (c) (0.8,20,0.3); (d) (0.8,20,0.4); (e) (0.9,5,0.3); (f) (0.9,5,0.4); (g) (0.9,20,0.3); (h) (0.9,20,0.4).



(a) (b) (c)

Supplementary Figure 3: Precision Recall curves for ChIP-seq ALI analysis. (a) ATAC-seq; (b) ChIP-seq of CTCF; (c) DNase-seq. Recall is defined as the proportion of potential TP that are selected, and precision is defined as the proportion of TP among the selected potential TP and FP.



Supplementary Figure 4: CTCF motif at SNP chr19:2,019,758. The paternal allele is more consistent with the sequence logo in the top panel.

Supplementary Tables

	Assay	Design	Read length
ATAC-seq (Buenrostro *and others*, 2013)		paired-end	50bps
CTCF ChIP-seq (Zhang *and others*, 2016)		paired-end	101bps
DNase-seq (Consortium *and others*, 2012)		single-end	36bps

Supplementary Table 1: List of ChIP-seq datasets analyzed in Section 3.

(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	2.37(0.57)	1.8(0.31)	2.14(0.27)	1.55(0.21)	2.81(0.86)	2.01(0.59)	1.92(0.27)	1.8(0.32)
OEMix	4.08(2.45)	0.85(0.14)	1.99(0.25)	1.15(0.19)	1.2(0.39)	0.94(0.22)	1.78(0.44)	1.33(0.35)
EBO	1(0.24)	1(0.13)	1(0.12)	1(0.05)	1(0.33)	1(0.25)	1(0.17)	1(0.08)
EBE	3.77(0.4)	2.26(0.18)	8.17(0.15)	6.2(0.12)	5.32(0.52)	3.66(0.22)	8.49(0.26)	7.9(0.26)

Supplementary Table 2: $L1$ loss in estimating g. Mean (standard error) over 20 simulation replications is reported. All reported numbers are the ratios of the original values and the original means of EBO in each column.

(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	-0.409(0.366)	0.635(0.347)	-0.7(1.208)	-2.049(0.668)	-0.159(0.355)	0.552(0.252)	-1.74(1.707)	-1.614(1.35)
OEMix	3.001(2.675)	-0.039(0.438)	-0.378(1.432)	-0.445(0.821)	-0.149(0.311)	-0.313(0.133)	-1.554(1.271)	-1.318(1.004)
EBO	-0.651(0.236)	-0.445(0.185)	-1.816(0.821)	-1.44(1.375)	-0.344(0.222)	-0.206(0.201)	0.04(0.779)	-0.374(1.289)
EBE	-1.795(0.184)	-1.735(0.114)	-12.846(0.226)	-12.469(0.196)	-1.575(0.21)	-1.489(0.14)	-11.822(0.387)	-11.747(0.382)
NPBin	0.067(0.019)	-0.023(0.027)	0.047(0.007)	0.032(0.007)	0.006(0.018)	-0.045(0.033)	0.035(0.013)	0.013(0.011)
OEMix	-0.048(0.065)	0.027(0.014)	0.037(0.008)	0.022(0.007)	0.027(0.018)	0.027(0.005)	0.032(0.007)	0.02(0.004)
EBO	0.097(0.012)	0.047(0.006)	0.021(0.005)	0.004(0.003)	0.048(0.015)	0.023(0.013)	0.005(0.012)	0.005(0.016)
EBE	0.141(0.012)	0.14(0.005)	0.079(0.004)	0.073(0.005)	0.069(0.019)	0.065(0.011)	0.027(0.01)	0.03(0.01)

Supplementary Table 3: Error in estimating the shape parameter α_0 under the null distribution g_0 (the upper panel) and π_0, the proportion of the null. Mean (standard error) of 20 simulation replications is reported.

FDR=0.05								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	44(24)	436(57)	224(32)	648(34)	10(8)	125(30)	61(11)	241(17)
OEMix	279(187)	357(51)	236(37)	673(34)	9(7)	76(11)	62(12)	244(15)
EBO	33(14)	309(27)	240(26)	696(21)	8(6)	82(14)	81(14)	273(21)
EBE	5(5)	86(17)	45(12)	400(23)	1(2)	15(6)	16(6)	136(19)
Binom	1163(46)	1426(46)	488(18)	850(31)	943(45)	1059(36)	204(32)	402(30)
FDR=0.10								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	117(44)	623(56)	360(37)	788(39)	33(18)	215(38)	103(18)	312(22)
OEMix	467(257)	533(60)	374(45)	816(38)	31(16)	147(18)	106(17)	316(18)
EBO	101(27)	507(34)	394(31)	846(22)	30(14)	168(22)	139(18)	350(23)
EBE	38(19)	265(25)	108(20)	571(23)	10(8)	72(18)	39(8)	213(20)
Binom	1501(42)	1756(62)	699(24)	1016(36)	1273(45)	1382(27)	319(41)	516(24)
FDR=0.20								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	262(75)	876(60)	584(42)	974(44)	84(36)	349(49)	176(31)	406(31)
OEMix	786(366)	759(80)	597(51)	1011(43)	80(34)	250(22)	179(27)	412(24)
EBO	246(46)	748(37)	645(33)	1039(25)	81(25)	291(28)	243(29)	455(29)
EBE	137(36)	536(30)	239(30)	779(29)	45(21)	191(26)	89(13)	315(24)
Binom	1975(61)	2169(41)	986(30)	1302(44)	1730(45)	1848(41)	554(61)	740(41)

Supplementary Table 4: Number of selected SNPs at varying nominal FDR levels. Mean (standard error) is reported across 20 simulation replications.

FDR=0.05								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	0.044(0.032)	0.063(0.013)	0.043(0.013)	0.02(0.005)	0.124(0.147)	0.07(0.027)	0.041(0.03)	0.02(0.009)
OEMix	0.192(0.117)	0.042(0.018)	0.045(0.014)	0.026(0.005)	0.158(0.251)	0.043(0.026)	0.045(0.033)	0.022(0.009)
EBO	0.055(0.037)	0.041(0.01)	0.048(0.013)	0.03(0.006)	0.123(0.162)	0.049(0.025)	0.069(0.036)	0.031(0.007)
EBE	0.053(0.123)	0.018(0.013)	0.001(0.004)	0.003(0.003)	0.043(0.109)	0.033(0.039)	0.011(0.027)	0.003(0.005)
Binom	0.571(0.017)	0.515(0.012)	0.181(0.015)	0.15(0.012)	0.742(0.012)	0.687(0.015)	0.257(0.034)	0.226(0.016)
FDR=0.10								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	0.081(0.039)	0.098(0.014)	0.076(0.012)	0.042(0.008)	0.144(0.057)	0.115(0.027)	0.068(0.035)	0.039(0.012)
OEMix	0.255(0.141)	0.071(0.027)	0.076(0.014)	0.051(0.008)	0.156(0.078)	0.057(0.023)	0.07(0.031)	0.04(0.009)
EBO	0.067(0.031)	0.068(0.013)	0.083(0.011)	0.057(0.008)	0.13(0.076)	0.081(0.023)	0.114(0.028)	0.063(0.014)
EBE	0.046(0.04)	0.028(0.008)	0.006(0.006)	0.008(0.004)	0.113(0.158)	0.048(0.025)	0.02(0.023)	0.007(0.007)
Binom	0.614(0.015)	0.565(0.011)	0.258(0.016)	0.221(0.013)	0.776(0.011)	0.733(0.011)	0.359(0.031)	0.317(0.023)
FDR=0.20								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	0.135(0.038)	0.2(0.024)	0.139(0.019)	0.1(0.009)	0.183(0.056)	0.21(0.044)	0.129(0.041)	0.1(0.021)
OEMix	0.35(0.158)	0.142(0.045)	0.14(0.017)	0.121(0.008)	0.175(0.056)	0.106(0.03)	0.129(0.033)	0.101(0.009)
EBO	0.129(0.033)	0.132(0.017)	0.156(0.012)	0.128(0.012)	0.187(0.047)	0.153(0.032)	0.191(0.03)	0.139(0.022)
EBE	0.069(0.035)	0.061(0.01)	0.026(0.01)	0.03(0.006)	0.126(0.065)	0.078(0.025)	0.034(0.029)	0.027(0.008)
Binom	0.656(0.013)	0.617(0.011)	0.354(0.016)	0.325(0.014)	0.809(0.01)	0.776(0.01)	0.495(0.023)	0.455(0.022)

Supplementary Table 5: Empirical FDR at varying nominal FDR levels. Mean (standard error) is reported across 20 simulation replications.

FDR=0.05								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	42(22)	434(46)	228(37)	621(35)	7(7)	119(22)	54(15)	252(21)
OEMix	373(200)	327(50)	238(45)	638(38)	7(5)	83(19)	54(14)	256(22)
EBO	32(10)	297(29)	245(25)	674(31)	7(6)	91(20)	76(14)	288(22)
EBE	4(4)	75(25)	44(16)	372(31)	1(1)	22(14)	13(6)	145(21)
Binom	1168(49)	1425(43)	487(25)	838(32)	955(35)	1053(40)	213(21)	418(34)
FDR=0.10								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	113(42)	620(48)	363(45)	765(37)	28(17)	208(28)	96(22)	322(22)
OEMix	590(279)	499(62)	376(55)	783(41)	28(11)	155(23)	96(19)	328(23)
EBO	101(19)	495(32)	397(30)	822(33)	26(12)	176(24)	136(19)	368(23)
EBE	30(13)	246(35)	107(27)	540(31)	9(6)	87(24)	33(11)	225(23)
Binom	1498(48)	1768(60)	690(39)	1001(49)	1277(37)	1374(48)	336(29)	532(31)
FDR=0.20								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	263(68)	873(59)	579(60)	950(40)	78(35)	338(37)	172(33)	420(27)
OEMix	951(394)	717(78)	591(72)	972(44)	83(27)	261(31)	172(27)	429(25)
EBO	252(29)	736(35)	641(38)	1013(33)	80(26)	300(27)	241(25)	480(27)
EBE	124(29)	520(33)	236(42)	750(34)	43(16)	207(31)	76(18)	331(25)
Binom	1989(42)	2192(44)	980(39)	1284(54)	1734(45)	1832(50)	566(38)	755(54)

Supplementary Table 6: Number of selected SNPs at varying nominal FDR levels of the simulation study with mis-specified null model. Mean (standard error) is reported across 20 simulation replications.

FDR=0.05								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	0.078(0.047)	0.061(0.013)	0.046(0.015)	0.023(0.008)	0.116(0.231)	0.085(0.022)	0.035(0.026)	0.018(0.009)
OEMix	0.251(0.104)	0.033(0.018)	0.047(0.016)	0.026(0.007)	0.096(0.145)	0.04(0.024)	0.034(0.025)	0.019(0.009)
EBO	0.072(0.052)	0.037(0.014)	0.051(0.011)	0.032(0.009)	0.114(0.223)	0.066(0.026)	0.061(0.032)	0.035(0.011)
EBE	0.083(0.169)	0.019(0.019)	0.008(0.012)	0.003(0.002)	0.112(0.245)	0.024(0.049)	0(0)	0.003(0.004)
Binom	0.571(0.013)	0.516(0.012)	0.188(0.017)	0.158(0.014)	0.747(0.018)	0.69(0.012)	0.279(0.025)	0.222(0.016)
FDR=0.10								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	0.094(0.035)	0.1(0.015)	0.075(0.019)	0.048(0.01)	0.109(0.084)	0.122(0.028)	0.064(0.033)	0.04(0.014)
OEMix	0.312(0.135)	0.056(0.022)	0.079(0.018)	0.053(0.008)	0.125(0.087)	0.072(0.023)	0.06(0.031)	0.044(0.011)
EBO	0.089(0.025)	0.061(0.014)	0.089(0.013)	0.063(0.008)	0.11(0.077)	0.097(0.027)	0.108(0.025)	0.068(0.011)
EBE	0.063(0.047)	0.024(0.012)	0.01(0.009)	0.008(0.004)	0.06(0.075)	0.051(0.025)	0.007(0.012)	0.009(0.006)
Binom	0.615(0.012)	0.568(0.008)	0.263(0.016)	0.233(0.016)	0.778(0.012)	0.734(0.011)	0.384(0.025)	0.314(0.016)
FDR=0.20								
(π_0, α_0, d)	(0.8,5,0.3)	(0.8,5,0.4)	(0.8,20,0.3)	(0.8,20,0.4)	(0.9,5,0.3)	(0.9,5,0.4)	(0.9,20,0.3)	(0.9,20,0.4)
NPBin	0.149(0.05)	0.208(0.029)	0.143(0.019)	0.109(0.016)	0.196(0.059)	0.209(0.034)	0.126(0.034)	0.1(0.026)
OEMix	0.41(0.146)	0.125(0.035)	0.144(0.021)	0.12(0.011)	0.2(0.073)	0.128(0.037)	0.12(0.029)	0.107(0.018)
EBO	0.141(0.023)	0.13(0.017)	0.166(0.013)	0.135(0.012)	0.189(0.035)	0.17(0.032)	0.196(0.026)	0.15(0.021)
EBE	0.074(0.027)	0.054(0.015)	0.027(0.015)	0.033(0.007)	0.127(0.058)	0.095(0.025)	0.027(0.023)	0.031(0.009)
Binom	0.658(0.011)	0.623(0.009)	0.359(0.017)	0.333(0.014)	0.809(0.01)	0.776(0.008)	0.5(0.02)	0.449(0.02)

Supplementary Table 7: Empirical FDR at varying nominal FDR levels of the simulation study with mis-specified null model. Mean (standard error) is reported across 20 simulation replications.

ATAC M=39659 (987, 89)			
Proportion of selected SNPs	NPBin	EBE	Binomial
0.05	112/6	89/8	115/6
0.1	176/9	149/16	192/14
0.15	235/13	205/20	247/18
0.2	286/21	258/23	291/20
CTCF M=19782 (112, 5)			
Proportion of selected SNPs	NPBin	EBE	Binomial
0.05	69/1	66/1	73/1
0.1	77/1	74/1	85/1
0.15	83/1	80/3	93/1
0.2	85/1	83/3	100/1
DNase M=28942 (206, 20)			
Proportion of selected SNPs	NPBin	EBE	Binomial
0.05	31/2	28/5	35/4
0.1	41/3	45/7	48/5
0.15	48/3	54/8	64/6
0.2	58/5	61/8	73/8

Supplementary Table 8: Number of TP and FP for fixed proportions of selected SNPs. The title for each panel indicates the total number of candidate SNPs (M) and the total numbers of potential TP and FP SNPs (when all M candidate SNPs are selected as ALI SNPs) in parentheses for each experiment.

	CTCF	DNase	ATAC(1,2)	ATAC(1,3)	ATAC(1,4)	ATAC(2,3)	ATAC(2,4)	ATAC(3,4)
NPBin	0.33	0.24	0.27	0.22	0.24	0. 22	0.22	0.23
EBE	0.20	0.12	0.12	0.16	0.17	0.10	0.10	0.19
Binom	0.36	0.29	0.28	0.23	0.25	0.23	0.22	0.23

Supplementary Table 9: Spearman's correlation of the rank measure w_j's calculated for each pair of replicates. For ATAC-seq, the numbers of in parenthesis are replicate ID's for ATAC-seq as there are four replicates.

Supplementary Data

The following are the data files and code used in this paper.

- npbin.r: R code of NPBin method
- demo_npbin.r: demo using a simulation example
- data_atac.txt, data_ctcf.txt, data_dnase.txt: the pre-processed ChIP-seq data analyzed in this paper. The following are the column names:
  - chr: chromosome
  - location: genomic location based on hg19
  - m: total number of reads covering the SNP
  - xm: total number of reads at the SNP from the maternal allele
  - winning.chip: the allele with more ChIP-seq reads. "P" if $xm < x/2$ and "M" otherwise.
  - motif: the ID and the transcription factor name of the motif in JASPAR database (Mathelier *and others*, 2013).
  - pval.mat.atSNP: the p-value of the motif on the maternal allele from R package `atSNP` (Zuo *and others*, 2015).
  - pval.pat.atSNP: the p-value of the motif on the paternal allele from `atSNP`.
  - pval.rank.atSNP: the p-value of the rank test of the allelic motif strength difference from `atSNP`.
  - winnig.motif: the allele with stronger motif, e.g., it is "M" if pval.mat.atSNP < pval.pat.atSNP.

- potential_TP: whether it is a potential TP based on our criteria described in the Supplementary notes. The users can define it differently using different thresholds on the various p-values from atSNP.
- potential_FP: whether it is a potential FP based on our criteria described in the Supplementary notes. The users can define it differently using different thresholds on the various p-values from atSNP.

If the motif related items of a SNP are all "NA", it means that there is no known motif at this SNP with either pval.mat.atSNP or pval.pat.atSNP less than 0.01.