Supplementary information

Using Beta-Binomial to call ASB events

To estimate the overdispersion parameter of our dataset, we need to determine an initial null distribution. Some tools (1-3) used all investigated regions to represent the null distribution. However, the investigated regions are a mixture of ASB events and non-ASB events. Including potential allele-specific binding events in the null distribution could overestimate the dispersion parameter. In the following process, we used the properties of non-ASB events with FDR > 0.05 (binomial test) to represent the null distribution.

Given the null distribution, the probability of a non-ASB event for a site with \( k \) reads on the favored allele and \( n \) reads in total can be calculated with the following equation:

\[
P(k|n, \pi) = P_{BetaBinomial}(k|n, \pi^2, (1 - \pi)/\gamma^2)
\]  

(1)

where \( \pi \) is mapping bias towards the favored allele, and \( \gamma \) is the dispersion parameter. The Beta-Binomial distribution is as following:

\[
P_{BetaBinomial}(k|n, a, b) = \binom{n}{k} \frac{B(a+k, n-k)}{B(a, b)}
\]  

(2)

Where \( a \) and \( b \) are called shape parameter, and \( B \) is the Beta function.

In equation (1), the two shape parameters were set to detect mapping bias from the read simulation. We used the maximum likelihood approach to estimate the overdispersion parameter \( \gamma \) across the null distribution set for each TF ChIP-Seq experiment. The probability of ASB event was calculated as the complementary event of non-ASB given by the beta-binomial test.

The overdispersion of allelic imbalance is significantly less in null distribution than in non-ASB events

The binomial test has been widely used to call ASB events (4-6). However, recent studies report the variance of allelic imbalance to be larger than expected in a binomial distribution across all heterozygous sites (1-3, 7). This is what we refer to as the “overdispersion” problem. The beta-binomial distribution has been introduced to correct for overdispersion in ASB calling, under the assumption that most allelic sites are balanced or follow the null distribution (2, 8). However, the potential for biological contributions to the overdispersion has not been fully investigated.

We attempted to assess to what extent overdispersion might be due to biological reasons, such as small TF binding alterations caused by motif alteration. We divided the non-ASB events with FDR > 0.05 (binomial test) into two classes: (1) TFBS alteration group, for which variants were found within the best predicted TFBS of the peak and exhibited motif score changes of at least 0.02; and (2) the remaining non-ASB events. We fit the distributions using beta-binomial, estimating the dispersion parameters of the two classes for each investigated TF. We found that the dispersion parameters were significantly higher in the TFBS alteration group than in the rest non-ASB events (Wilcoxon test, \( p \)-value=1.43e-06, Figure S8). Our results suggest that the observed overdispersion is at least in part due to mild TF binding alterations.

Replicate normalization method produces highly similar sets of ASB calls
In this manuscript, we used a direct sum approach in which we summed the read counts of each allele across the replicates, and then applied the binomial test on the derived sum of each allele. We also implemented a normalized approach regarding multiple replicates and compared the ASB calling between the two (direct sum and normalized). In the normalization approach, the read coverage at heterozygous positions is normalized between replicates following the scale factor-based procedures used in DEseq (9). The normalized count of each site is the original count divided by the scale factor. The normalized count values are thereafter processed using the same procedure as in direct sum approach.

The normalized approach resulted in 10,121 called ASB events, while direct-sum approach called 10,711. Overall, 9,511 ASB events were called by both approaches, and on average, 92% of the called ASB events using the normalized approach overlapped with those called with the direct-sum approach across investigated TFs. While a few datasets showed greater difference, as shown in Figure S8, most samples were clustered in the lower right corner reflecting high similarity between the results.

As one would expect, we observed that the overlap ratio was anti-correlated with the scale factor of the larger replicate (Spearman correlation coefficient = -0.64; Figure S8), showing a large replicate library difference in depth (large scale factor) correlated with greater divergence in ASB calling between the two approaches. However, the impact was modest, as there was still 85-95% overlap for the larger scale factor cases.

We explored the differences between methods, which confirmed our expectation that the normalized approach penalized those cases in which one replicate was strikingly lower than the other in terms of counts. This can be observed in Figure S9, in which we show the read coverage for the method-specific ASB calls for a TF experiment with high scale factor.

**Direct sum approach is used considering the characteristics of the data**

It is useful to recognize two key aspects of the ENCODE data prior to reviewing the findings. For the vast majority of TFs there are only two replicates (n=41), with only a few having three replicates (n=4). The second is a difference between standard RNA-Seq and ASB identification in ChIP-Seq. In standard RNA-Seq (as most published methods address), each sample are prepared and processed separately. For ASB detection in ChIP-Seq, two alleles are naturally controlled within the same single sample. These two aspects inform our decision about the selection of the ASB calling method.

Based on our perspective, with only two replicates for the vast majority of cases, we prefer to use the direct-sum approach. This reflects our view that the high coverage positions in a single ChIP-Seq replicate are well controlled (two alleles coming from the same nuclei). We believe that replicate normalization will be an important issue and should be deeply considered in future ASB analysis (particularly when greater replicate numbers are available).

**The sequence based classifier produces consistent predictions for lymphoblastoid cells across multiple individuals**

We tested the consistency of the random forest classifier in different individuals from the same cell type (that is lymphoblastoid cell line). Briefly, we collected CTCF ChIP-Seq data from multiple ENCODE samples. We trained a sequence based classifier with N-1 samples (N is the number of collected samples), and tested each model on the remaining sample. Results showed similar performance
between cross validation and testing (for instance, the mean AUPRC difference is equal to 0.02 and the standard deviation is 0.05, Figure S5). These results suggested that our sequence based model could be applied across individuals using a single training data set.

Figure S1. Comparing motif score of two alleles in heterozygous site binding events for all the investigated heterozygous site binding events. Each dot represented the relative motif score for favored allele (allele with higher ChIP-Seq read count) and unfavored allele in predicted TFBSs. ASB and non-ASB events were plotted separately. The black diagonal line indicated the cases with equal scores of two alleles. Heterozygous site binding data of all the TFs with known motif were presented together. This figure presented the entire data set partially depicted in Figure 1.
Figure S2. The positional impact and information content at each position of TF motifs.

For each TF, we plotted the positional impact of each motif position derived from ASB events (red bar) and its corresponding information content (blue line). This figure presented a TF specific perspective of Figure 2A.
The ChIP-Seq experiments were divided into TFs with comotif only (left panel; TFs with no known PWM), with both PWM and comotifs (middle panel), and with known PWMs only (right panel). ASB-SNVs were significantly enriched in the predicted TFBS of comotifs (p-value = 7.1e-41), combination of comotif and primary motif (p-value = 4.8e-42) and primary motif (p-value = 1.8e-128).
Figure S4. Allelic coordination between TFs and chromatin properties in GM12878 cell line.

The heatmap represented the -log(p-value) of correlation Pearson between allele imbalance of TF ChIP-Seq reads at heterozygous site binding events and chromatin properties (DHS and histone modifications).
Figure S5. Testing the performance of sequence models for CTCF in 7 individuals.

The figure showed the accuracy of cross validation within any 6 samples (red bar) and testing accuracy of the remaining individual (blue bar).
Figure S6. Compare the performance of the Seq model and BayesPI-Bar. Only 27 TFs experiments with available BayesPI-Bar models are presented in the comparison.
Figure S7. The most frequently selected key features in the Full models for the TFs without known motif. The suffix ‘favor’ (respectively ‘unfavor’) referred to the allele with higher (respectively lower) read counts at heterozygous sites. Details of each feature can be found in the Methods section and Supplementary Table S5.
Figure S8. Scale factor and overlap ratio between the direct sum and the normalized approach. Each point represents one investigated TF ChIP-Seq dataset.
Figure S9. Read coverage distribution of approach-specific ASB events in the smaller replicate. The data of CHD2 in HeLa-S3 are shown as it has a low overlap ratio (82.6%) and high scale factor 2.1. In this TF experiment, direct sum approach calls 53 approach-specific ASB events (red) and normalized approach calls 16 (blue). Two approaches share 76 ASB events in common.
| Data Type                          | Institution | URL                                                                 |
|-----------------------------------|-------------|----------------------------------------------------------------------|
| TF ChIP-Seq raw reads             | Haib        | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibTfbs/ |
| TF ChIP-Seq raw reads             | Sydh        | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhTfbs/ |
| TF ChIP-Seq raw reads             | Uta         | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeOpenChromChip/ |
| TF ChIP-Seq raw reads             | Uw          | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwTfbs/ |
| TF narrowPeak regions              | Awg         | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/ |
| DNase-Seq raw reads               | Uw          | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDnase/ |
| DNase-Seq raw reads               | Duke        | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeOpenChromDnase/ |
| Histone ChIP-Seq raw reads        | Broad       | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/ |
| Genotype data of Lymphoblastoid cell lines | Complete genomics | ftp://ftp2.completegenomics.com/vcf_files/Build37_2.0.0/ |
| Copy number variant region        | Haib        | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibGenotype/ |

*(Available for GM12878, HeLa-S3, and GM19238)*

**Supplementary Table 1.** Sources of the data used in ASB analysis.

Our analysis integrated multiple types of data, including ChIP-Seq, DNase-Seq, and genotype calling data. The categories, source institution, and URL of these data were listed in the table.
| Cell     | TF    | Peak Count | Heterozygous binding sites events | ASB |
|----------|-------|------------|-----------------------------------|-----|
| GM12878  | BATF  | 32427      | 1314                              | 201 |
| GM12878  | BCL11A| 17876      | 678                               | 60  |
| GM12878  | BCL3  | 15455      | 653                               | 60  |
| GM12878  | BHLHE40| 13986    | 661                               | 66  |
| GM12878  | CTCF  | 55551      | 2614                              | 396 |
| GM12878  | EBF1  | 33410      | 1445                              | 327 |
| GM12878  | EGR1  | 16331      | 718                               | 69  |
| GM12878  | ELF1  | 23008      | 1047                              | 58  |
| GM12878  | MEF2A | 17605      | 574                               | 55  |
| GM12878  | NFYB  | 13295      | 355                               | 77  |
| GM12878  | PAX5  | 19740      | 657                               | 43  |
| GM12878  | RAD21 | 40019      | 1963                              | 281 |
| GM12878  | RUNX3 | 67965      | 3304                              | 471 |
| GM12878  | SRF   | 8544       | 164                               | 31  |
| GM12878  | TCF12 | 20437      | 770                               | 82  |
| GM12878  | USF1  | 9778       | 305                               | 37  |
| HEla-S3  | TFAP2C| 25452      | 561                               | 84  |
| HEla-S3  | BRCA1 | 8114       | 274                               | 47  |
| HEla-S3  | CEBPB | 61004      | 2491                              | 922 |
| HEla-S3  | CHD2  | 20500      | 696                               | 129 |
| HEla-S3  | CTCF  | 58806      | 2506                              | 1076|
| HEla-S3  | GABPA | 6761       | 224                               | 72  |
| HEla-S3  | JUND  | 31633      | 641                               | 118 |
| HEla-S3  | MAFK  | 14185      | 431                               | 84  |
| HEla-S3  | MAX   | 29647      | 1111                              | 132 |
| HEla-S3  | NFYB  | 7156       | 149                               | 47  |
| HEla-S3  | NRSF  | 10247      | 372                               | 179 |
| HEla-S3  | P300  | 25854      | 661                               | 160 |
| HEla-S3  | POL2  | 25332      | 1284                              | 570 |
| HEla-S3  | PRDM1 | 4577       | 134                               | 50  |
| HEla-S3  | RAD21 | 43420      | 1883                              | 708 |
| HEla-S3  | RFX5  | 19284      | 664                               | 80  |
| HEla-S3  | SMC3  | 39567      | 1860                              | 565 |
| HEla-S3  | STAT3 | 13834      | 325                               | 56  |
| HEla-S3  | TAF1  | 16100      | 514                               | 83  |
| HEla-S3  | TBP   | 18489      | 466                               | 87  |
| HEla-S3  | TCF7L2| 19242      | 600                               | 124 |
| HEla-S3  | USF2  | 12306      | 373                               | 108 |
| HEla-S3  | ZNF143| 7048       | 261                               | 52  |
| GM12872  | CTCF  | 47151      | 2496                              | 488 |
| GM12873  | CTCF  | 51005      | 2575                              | 552 |
| GM19238  | CTCF  | 49938      | 2909                              | 500 |
| GM19239  | CTCF  | 41085      | 2473                              | 282 |
| GM19240  | CTCF  | 46036      | 2972                              | 573 |
| GM12864  | CTCF  | 46798      | 2390                              | 523 |
Supplementary Table 2. Processed heterozygous site binding data.

For each TF ChIP-Seq experiment, we listed the number of ChIP-Seq peaks (Peak count), heterozygous site binding events, and called ASB events.

|   |   | 1205998 | 51518 | 10765 |
|---|---|---------|-------|-------|
| Cell      | TF      | Comotif                                      |
|-----------|---------|----------------------------------------------|
| GM12878   | BATF    | IRF1, IRF(10), BZIP, IRF(10)                 |
| GM12878   | BHLHE40 | RUNX1, RUNT                                  |
| GM12878   | RAD21   | CTCF.ZF(11)                                 |
| GM12878   | RUNX3   | BATF, BZIP, RUNX1, RUNT, ETS1, ETS, FLI1, ETS|
| HeLa-S3   | CEBPB   | BATF, BZIP                                  |
| HeLa-S3   | P300    | NF-E2, BZIP, CEBP, BZIP(11), ATF3, BZIP(12,13)|
| HeLa-S3   | RAD21   | CTCF.ZF(11)                                 |
| HeLa-S3   | SMC3    | CTCF.ZF(11)                                 |
| HeLa-S3   | TCF7L2  | ATF3, BZIP                                  |
| Total     | 9       | 15                                           |

**Supplementary Table 3.** Discovered comotifs from heterozygous site binding events.

HOMER motifs were considered as comotifs when their motif change correlated with TF allelic binding imbalance in heterozygous site binding events (see Materials and Methods). The cell line, ChIP’ed TF, and correlated comotifs were provided respectively. TF-comotif pairs supported by external literature were given the corresponding references.
| Cell    | ASB-TF | TF       | Overlap ratio | -log(P-value) | Odd ratio |
|---------|--------|----------|---------------|---------------|-----------|
| GM12878 | BATF   | TBP      | 0.11          | 4.50          | 0.25      |
| GM12878 | CTCF   | ZNF143   | 0.34          | 10.45         | 0.43      |
| GM12878 | CTCF   | YY1      | 0.28          | 7.42          | 0.48      |
| GM12878 | CTCF   | POU2F2   | 0.10          | 4.83          | 0.37      |
| GM12878 | RAD21  | ZNF143   | 0.42          | 9.04          | 0.42      |
| GM12878 | RAD21  | RUNX3    | 0.28          | 5.61          | 0.47      |
| GM12878 | RAD21  | PAX5     | 0.14          | 4.69          | 0.37      |
| GM12878 | RUNX3  | POL2     | 0.21          | 12.31         | 0.34      |
| GM12878 | RUNX3  | ZNF143   | 0.13          | 9.73          | 0.29      |
| GM12878 | RUNX3  | POL24H8  | 0.17          | 9.29          | 0.37      |
| GM12878 | RUNX3  | ELF1     | 0.19          | 8.16          | 0.42      |
| GM12878 | RUNX3  | SMC3     | 0.14          | 7.86          | 0.37      |
| GM12878 | RUNX3  | YY1      | 0.24          | 7.71          | 0.48      |
| GM12878 | RUNX3  | CTCF     | 0.13          | 7.55          | 0.36      |
| GM12878 | RUNX3  | MAZ      | 0.19          | 6.71          | 0.46      |
| GM12878 | RUNX3  | PML      | 0.19          | 6.08          | 0.49      |
| GM12878 | RUNX3  | ELK1     | 0.07          | 5.65          | 0.28      |
| GM12878 | RUNX3  | EGR1     | 0.11          | 5.34          | 0.42      |
| GM12878 | RUNX3  | GR       | 0.11          | 5.34          | 0.42      |
| GM12878 | RUNX3  | RAD21    | 0.15          | 5.12          | 0.49      |
| GM12878 | RUNX3  | SIN3A    | 0.10          | 5.12          | 0.39      |
| GM12878 | RUNX3  | CMYC     | 0.04          | 5.01          | 0.17      |
| GM12878 | RUNX3  | ZEB1     | 0.04          | 5.01          | 0.17      |
| GM12878 | RUNX3  | MAX      | 0.13          | 4.82          | 0.47      |
| GM12878 | RUNX3  | GABP     | 0.06          | 4.61          | 0.31      |
| GM12878 | RUNX3  | TAF1     | 0.13          | 4.47          | 0.49      |
| GM12878 | RUNX3  | CHD2     | 0.18          | 4.45          | 0.54      |
| HeLa-S3 | CEBPB  | P300     | 0.37          | 16.61         | 2.07      |
| HeLa-S3 | CEBPB  | STAT3    | 0.21          | 6.52          | 1.68      |
| HeLa-S3 | CEBPB  | CTCF     | 0.08          | 4.56          | 0.50      |
| HeLa-S3 | CTCF   | SMC3     | 0.63          | 7.44          | 1.59      |
| HeLa-S3 | CTCF   | RAD21    | 0.68          | 6.59          | 1.58      |
| HeLa-S3 | MAX    | CMYC     | 0.13          | 6.77          | 3.31      |
| HeLa-S3 | MAX    | USF2     | 0.27          | 4.59          | 2.27      |
| HeLa-S3 | POL2   | TAF1     | 0.55          | 10.46         | 2.14      |
| HeLa-S3 | POL2   | HGFC1    | 0.53          | 8.65          | 1.98      |
| HeLa-S3 | POL2   | TBP      | 0.43          | 6.50          | 1.80      |
| HeLa-S3 | POL2   | E2F4     | 0.10          | 5.36          | 2.43      |
| HeLa-S3 | POL2   | GCN5     | 0.06          | 4.76          | 2.95      |
| HeLa-S3 | RAD21  | CTCF     | 0.72          | 16.29         | 2.58      |
| HeLa-S3 | RAD21  | CJUN     | 0.13          | 10.37         | 0.35      |
| HeLa-S3 | RAD21  | P300     | 0.15          | 9.96          | 0.38      |
| HeLa-S3 | RAD21  | CFOS     | 0.08          | 9.38          | 0.28      |
| HeLa-S3 | RAD21  | COREST   | 0.18          | 7.49          | 0.48      |
| HeLa-S3 | RAD21  | REST     | 0.18          | 7.49          | 0.48      |
| HeLa-S3 | RAD21  | CHD2     | 0.16          | 7.42          | 0.46      |
| HeLa-S3 | RAD21  | TCF7L2   | 0.14          | 7.15          | 0.44      |
| HeLa-S3 | RAD21  | SMC3     | 0.82          | 6.33          | 1.94      |
|                  | HeLa-S3 | SMC3  | TBP   | p-value | OR     |
|------------------|---------|-------|-------|---------|--------|
| RAD21            | 0.12    | 0.84  | 0.47  | 5.60    |
| TBP              | 0.12    | 0.73  | 0.49  | 7.77    |
| RAD21            | 0.09    | 0.15  | 0.44  | 4.11    |
| TAF1             | 0.14    | 0.16  | 0.48  | 4.97    |
| TBP              | 0.07    | 0.37  | 0.37  | 4.82    |
| RPC155           | 0.07    | 0.15  | 0.44  | 6.00    |
| HCFC1            | 0.08    | 0.16  | 0.48  | 5.68    |
| RAD21            | 0.12    | 0.14  | 0.51  | 4.71    |
| ZNF143           | 0.15    | 0.07  | 0.37  | 4.32    |

**Supplementary Table 4.** Presence of ASB events were associated with cobound TFs.

For each ASB dataset, we tested the association between ASB events and binding peaks of its cobound TFs (Fisher test, FDR < 0.005). The p-value and odds ratio of the significant pairs were listed.
| Category                          | Feature Name                                           | Description                                                                                     | Used in model(s)          |
|----------------------------------|--------------------------------------------------------|-----------------------------------------------------------------------------------------------|---------------------------|
| Motif of the ChIP'ed TF          | motif_favor,                                           | Motif score of two alleles                                                                     | Full, Seq+DHS, Seq        |
|                                  | motif_unfavor                                         |                                                                                               |                           |
|                                  | Motif_pvalue_ratio                                    | The log ratio of two alleles' binding potential (the P-value of the PWM score against random     |                           |
|                                  |                                                        | genome background)                                                                            |                           |
|                                  | Peak_motif_favor,                                      | Best motif scores within the peak regions for two alleles. The number of predicted TFBS in the  | Full, Seq+DHS, Seq        |
|                                  | Peak_motif_unfavor,                                   | peak region                                                                                   |                           |
|                                  | Peak_TFBS_num                                         |                                                                                               |                           |
| Position information             | Peak_dis                                              | SNV distance to the peak max position                                                         | Full, Seq+DHS, Seq        |
|                                  | PWM_position                                          | SNV position in the motif                                                                      |                           |
| Enriched motifs                  | comotif1_pavlue_ratio,                                | For each of the five enriched motifs, the log ratio of two alleles' binding potential (the P-   | Full, Seq+DHS, Seq        |
|                                  | comotif2_pavlue_ratio,                                | value of the enriched motif score against random genome background). We ranked the enriched    |                           |
|                                  |                                                        | motifs based on their enrichment within the peak regions.                                     |                           |
| DHS and 11 histone modifications | DHS_favor,                                             | The read count at each allele for each feature.                                                | Full, Seq+DHS (DHS only)  |
|                                  | DHS_unfavor,                                           |                                                                                               |                           |
|                                  | H3K27ac_favor,                                         |                                                                                               |                           |
|                                  |                                                        |                                                                                               |                           |
| Cobound TFs                      | Cobound1,                                              | Whether the SNV overlap with other TF binding peak regions. We ranked the cobound TFs based on | Full                      |
|                                  | Cobound2,                                              | the overlap ratio with heterozygous site binding events.                                       |                           |
|                                  |                                                        |                                                                                               |                           |

**Supplementary Table 5.** Input features used in three classification models (Seq, Seq+DHS, and Full).

The features were summarized into 5 categories based on the source or the nature of the data. “Feature names” referred to the features used in Figure 5(B) and supplementary Figure S6. Features were explained in the ‘Description’ column. The last column indicated the models which included corresponding features for training.

**References**

1. Chen, J., Rozowsky, J., Galeev, T.R., Harmanci, A., Kitchen, R., Bedford, J., Abyzov, A., Kong, Y., Regan, L. and Gerstein, M. (2016) A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nature communications*, 7, 11101.
2. van de Geijn, B., McVicker, G., Gilad, Y. and Pritchard, J.K. (2015) WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nature methods*, 12, 1061-1063.
3. Harvey, C.T., Moyerbrailean, G.A., Davis, G.O., Wen, X., Luca, F. and Pique-Regi, R. (2014) QuASAR: Quantitative Allele Specific Analysis of Reads. *Bioinformatics*. 
4. Reddy, T.E., Gertz, J., Pauli, F., Kucera, K.S., Varley, K.E., Newberry, K.M., Marinov, G.K., Mortazavi, A., Williams, B.A., Song, L. et al. (2012) Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome research, 22*, 860-869.

5. Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., Leng, J., Bjornson, R., Kong, Y., Kitabayashi, N. et al. (2011) AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Molecular systems biology, 7*, 522.

6. Younesy, H., Möller, T., Heravi-Moussavi, A., Cheng, J.B., Costello, J.F., Lorincz, M.C., Karimi, M.M. and Jones, S.J.M. (2014) ALEA: a toolbox for allele-specific epigenomics analysis. *Bioinformatics, 30*, 1172-1174.

7. Mayba, O., Gilbert, H.N., Liu, J., Haverty, P.M., Jhunjhunwala, S., Jiang, Z., Watanabe, C. and Zhang, Z. (2014) MBASED: allele-specific expression detection in cancer tissues and cell lines. *Genome biology, 15*, 405.

8. Skelly, D.A., Johansson, M., Madeoy, J., Wakefield, J. and Akey, J.M. (2011) A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome research, 21*, 1728-1737.

9. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome biology, 11*, R106.

10. Roy, S., Guler, R., Parihar, S.P., Schmeier, S., Kaczkowski, B., Nishimura, H., Shin, J.W., Negishi, Y., Ozturk, M., Hurdayal, R. et al. (2015) Batf2/Irf1 induces inflammatory responses in classically activated macrophages, lipopolysaccharides, and mycobacterial infection. *Journal of immunology, 194*, 6035-6044.

11. Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic acids research, 34*, D535-539.

12. Cherasse, Y., Maurin, A.C., Chaveroux, C., Jousse, C., Carraro, V., Parry, L., Deval, C., Chambon, C., Fafournoux, P. and Bruhat, A. (2007) The p300/CBP-associated factor (PCAF) is a cofactor of ATF4 for amino acid-regulated transcription of CHOP. *Nucleic acids research, 35*, 5954-5965.

13. Kim, W.H., Jang, M.K., Kim, C.H., Shin, H.K. and Jung, M.H. (2011) ATF3 inhibits PDX-1-stimulated transactivation. *Biochemical and biophysical research communications, 414*, 681-687.