

Optimized detection of transcription factor-binding sites in ChIP-seq experiments

Laura L. Elo^{1,2,*}, Aleksi Kallio³, Teemu D. Laajala^{1,2}, R. David Hawkins^{2,4,5},
Eija Korpelainen³ and Tero Aittokallio^{1,2,6,*}

¹Department of Mathematics, University of Turku, FI-20014 Turku, ²Turku Centre for Biotechnology, FI-20520 Turku, ³CSC - IT Center for Science Ltd., FI-02101 Espoo, Finland, ⁴Department of Medicine, Division of Medical Genetics, ⁵Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA and ⁶Institute for Molecular Medicine Finland (FIMM), University of Helsinki, FI-00014 Helsinki, Finland

Received June 14, 2011; Revised September 2, 2011; Accepted September 20, 2011

ABSTRACT

We developed a computational procedure for optimizing the binding site detections in a given ChIP-seq experiment by maximizing their reproducibility under bootstrap sampling. We demonstrate how the procedure can improve the detection accuracies beyond those obtained with the default settings of popular peak calling software, or inform the user whether the peak detection results are compromised, circumventing the need for arbitrary re-iterative peak calling under varying parameter settings. The generic, open-source implementation is easily extendable to accommodate additional features and to promote its widespread application in future ChIP-seq studies. The peakROTS R-package and user guide are freely available at <http://www.nic.funet.fi/pub/sci/molbio/peakROTS>.

INTRODUCTION

Chromatin immunoprecipitation coupled with deep sequencing (ChIP-seq) has offered a powerful means for genome-wide mapping of transcription factor-binding sites (1–4). Owing to the recent advances in the next-generation sequencing technology, the current ChIP-seq experiments are generating increasing amounts of data, the analysis of which is a computational challenge (4–6).

Despite the availability of a number of advanced software packages (4), the users are still facing the crucial challenge of deciding which package, along with its adjustable parameters, is most suitable for their specific needs so that they can extract full information from the data under

analysis. We have recently demonstrated that the choice of the software package may considerably affect the biological conclusions made from the ChIP-seq data (7), calling into question the validity of the binding site detections unless they are carefully confirmed in independent qPCR experiments. Another practical challenge is to decide whether the data is similar enough to those on which a specific peak calling algorithm was tuned to, in order to justify the use of its default parameters (6). However, even among the same type of data, variability in data quality may necessitate using various parameter settings (8). Accordingly, with the fixed default parameter settings, the choice of the best package is strongly dependent on the ChIP-seq data under analysis, making the selection between the different packages and optimization of their performance for a given data a challenging task (7,9,10).

To this end, we introduce here an adaptive procedure, which provides the user with an informed means to optimally adjust the parameters of a given software package to the intrinsic properties of each ChIP-seq data set separately. The procedure is based on the concept of maximizing the reproducibility of the binding site detections under random bootstrap sampling of the original data, while preserving the given ChIP and control sample labels. We have successfully used a similar concept in the context of other high-throughput profiling platforms, such as those based on gene-expression microarray or quantitative mass-spectrometry (MS) technologies (11). From an end-user perspective, rather than introducing new variants of the existing algorithmic solutions, some of which have been developed on—and perhaps also tuned to—particular data sets, it is more important to make the most of the currently used data analysis packages in a wide variety of application use cases. Here, using five human and one

*To whom correspondence should be addressed. Tel: +358 2 333 6027; Fax: +358 2 333 6595; Email: laliel@utu.fi
Correspondence may also be addressed to Tero Aittokallio. Tel: +358 2 333 6027; Fax: +358 2 333 6595; Email: tero.aittokallio@fimm.fi

mouse ChIP-seq data sets (Table 1), we demonstrate how our generic reproducibility-optimized test statistic ROTS (12) can provide significant practical benefits to two popular ChIP-seq peak detection packages MACS (13) and PeakSeq (14).

MATERIALS AND METHODS

The ROTS procedure for ChIP-seq studies

The generic data-adaptive procedure, based on reproducibility-optimized test statistic (ROTS), uses the maximal reproducibility across bootstrap samples as a systematic means to learn those parameters that are best adjusted to the high-throughput data under analysis (11,12). Here, the ROTS procedure was modified to deal with the genome-wide ChIP-seq data sets. Instead of making the bootstrap data pairs by re-sampling the individual samples, such as in the previous applications (11,12), the ChIP-seq ROTS procedure makes bootstrap samples of the reads within a single data set. More specifically, an equal number of reads as in the original data is sampled with replacement for each bootstrap data. The peak detection is then performed on each bootstrap data and the average peak list reproducibility between the bootstrapped data pairs is calculated at various top list sizes. To deal with the ambiguity that a peak in one data set may overlap with multiple peaks in another data set, the reproducibility calculations were made using the efficient approach introduced earlier (15). It first merges the two peak lists under comparison into a union set of n detected regions and then determines the number m of these regions found in both of the original lists. The peak list reproducibility is finally defined as $R = m/n$, which obtains value one if all the regions are overlapping and value zero if none of the regions overlaps. Two regions were considered overlapping if they shared at least one base pair.

The ROTS-based parameter combination for peak detection is selected by maximizing the reproducibility

Z-score over increasing top list sizes (k) and with respect to various parameter combinations (α):

$$Z_{k,\alpha} = \frac{R_{k,\alpha} - R_{k,\alpha}^0}{s_{k,\alpha}}.$$

The reproducibility $R_{k,\alpha}$ is defined as the average peak list reproducibility of the k top peaks as defined above over pairs of bootstrapped data sets, $s_{k,\alpha}$ is the estimated standard deviation of the bootstrap distribution of the peak list reproducibility at top list size k , and $R_{k,\alpha}^0$ corresponds to the null reproducibility in randomized data sets. In the present results, 1000 bootstrap data pairs were considered for each parameter combination α under investigation. To obtain a random reference for the null reproducibility, we used here data sets containing both a ChIP and a control sample, which are generally preferred in ChIP-seq studies (5–7,17,21,22), and applied the same peak detection procedure after switching the ChIP and control sample. The ROTS output is the peak list obtained from the original data using the parameters selected with the ROTS procedure.

The peak calling software packages

To test the benefits of ROTS in the parameter selection for binding site detection, we considered two popular software packages, MACS (version 1.3.5, <http://liulab.dfc.harvard.edu/MACS>) (13) and PeakSeq (version 1.01, <http://archive.gersteinlab.org/proj/PeakSeq>) (14), which both make use of the control sample and also include a number of user-adjustable parameters.

Model-based Analysis of ChIP-Seq (MACS) package (13) was selected because of its popularity in many ChIP-seq studies. MACS uses tag shifting and windowing to scan chromosome regions and a dynamic Poisson distribution to model the background signal. Within the ROTS procedure, we considered various parameter combinations, involving *shift size* between the strands (model-based,1,50,100,150,200), *band width* for the peak detection (100,300,500), and *background model*

Table 1. The ChIP datasets used in the present study

Transcription factor data	Study organism	ChIP-seq test data			ChIP-qPCR validation data		
		Reference(s)	#ChIP tags	#Control tags	Reference(s)	#Positives	#Negatives
STAT1_1 ^a	human	(14)	2 430 958	4 513 107	(15)	120	160
STAT1_2 ^a	human	(14)	8 184 450	4 513 107	(15)	120	160
NRSF_1 ^b	human	(1)	1 697 991	2 319 153	(16)	83	30
NRSF_2 ^c	human	(17)	5 349 088	10 162 151	(16)	83	30
FoxA1 ^d	human	(13)	3 909 804	5 233 682	(18)	26	12
FoxA2 ^e	mouse	(19)	2 813 847	4 428 744	(19,20)	55	11

^aFrom the STAT1 study, two replicate datasets were downloaded from the Gene Expression Omnibus (GEO accession GSE12782). STAT1_1: ChIP (rep1 lane A), control (rep1 lane C); STAT1_2: ChIP (rep2 lane B), control (rep1 lane C)

^bThe NRSF_1 data was downloaded from the Illumina website: http://www.illumina.com/downloads/Illumina_ChIPSeq_Demo_Data_Johnson_Science_2007.zip.

^cThe NRSF_2 data corresponding to the monoclonal antibody was downloaded from the QuEST website: <http://mendel.stanford.edu/sidowlab/downloads/quest/>.

^dThe FoxA1 data was downloaded from the MACS website: <http://liulab.dfc.harvard.edu/MACS/>.

^eFrom the mouse FoxA2 study, the first replicate pair of the ChIP and control samples was used (kindly provided by Dr Geetu Tuteja).

(global or local Poisson). The default values were used for the remaining parameters. In the bootstrap runs, the MACS algorithm was modified to allow a maximum of 100 reads having exactly the same location in the genome. In the original ChIP-seq data sets, MACS was applied without any modifications. In each data set, peaks with significance threshold $P < 10^{-3}$ according to MACS were recorded and the reproducibility was optimized over a wide range of top list sizes, here specifically $k = 25, 50, 100, 200, \dots, 1900, 2000, 2500, 3000, 4000, 5000, 7000, 10000, 20000, \dots, 80000, 90000$, or as long as $>90\%$ of the bootstrap peak lists with each parameter setting were at least of length k , using regions 500 bp around the peak summits.

PeakSeq is a more recent software package (14), which uses extended tag aggregation to profile genome regions, taking into account also the variability in genomic mappability, and a conditional binomial model to find enriched regions, under the assumption that the reads should occur with equal likelihood from the ChIP and control sample. Within the ROTS procedure, we varied two peak detection parameters, namely, *max gap* (100, 200, 300), which defines how distant peaks can be aggregated, and *read length* (50, 100, 150, 200, 250, 300, 350, 400), which defines the length of the extended tags. When varying some of the other parameters from their default values, we encountered stability problems with the original C-implementation that was used as part of the ROTS procedure. Therefore, the present results were obtained using this rather limited parameter space only. In the human data sets, the default parameter value for the number of windows per chromosome *W_PER_C* was used, whereas in the mouse FoxA2 data set, its value was changed from 250 to 200, to reflect better the chromosome size with the fixed window size *W_SIZE* of 1 Mb. In each data set, peaks with significance threshold $q < 0.2$ according to the PeakSeq definition of false discovery rate (FDR) were recorded and the reproducibility was calculated at the same top list sizes k as in the MACS runs (see above), using regions 500-bp around the peak centers.

The ChIP-seq and ChIP-qPCR data sets

We applied the ROTS procedure to five human and one mouse ChIP-seq data sets (Table 1). The data sets were selected on the basis of their public availability and the availability of ChIP-qPCR validation data for evaluation purposes. All the ChIP-seq data sets were sequenced with the Illumina/Solexa sequencing technology. The data sets were aligned in the original studies and these pre-processed data were used here. The two STAT1 data sets downloaded from the Gene Expression Omnibus (GEO) were further pre-processed before bootstrapping by removing reads not U0, U1, U2 in the ELAND format. The other data sets were used as downloaded and transformed into ELAND format if not already in that format. In the FoxA2 data set, we used initially the first replicate pair only due to computational reasons. Later, we extended the ROTS analyses to the other replicate pairs as well to study their relative performance. The genome build of the ChIP-seq data sets was used. If needed,

the qPCR regions were converted into the same build using the UCSC Genome Browser liftOver tool (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>).

The evaluation procedure

The relative performance of the ROTS-based parameter values was compared to that of the default parameters in terms of the accuracy of binding site detections. The peak calling was based solely on the ChIP-seq data, whereas the respective ground truth set of positive and negative sites came from independent confirmation studies performed by qPCR (Table 1). As an evaluation metric, we used the popular *F*-score, which takes values between zero (none correct detections) and one (perfect accuracy). The *F*-score takes into account both the *precision* (P or positive predictive value) and the *recall* (R or sensitivity) of the detections by calculating their harmonic mean:

$$F = \frac{2PR}{P+R}.$$

The ChIP-seq peaks were compared to the qPCR regions using the function *regionOverlap* in the Bioconductor Ringo package (version 1.10.0), which counts the number qPCR regions covered by at least one ChIP-seq peak (23). The regions 500 bp around the peak summit or the area center as reported by MACS or PeakSeq, respectively, were considered. The *F*-traces depicting the accuracy of the binding site detections were smoothed to better display the underlying detection performance of the ROTS and default settings in the relatively sparse ChIP-qPCR validation data sets (Figure 1 and Supplementary Figure S1). Here, we used the standard *R* smoothing function *lowess* with a smoothing parameter value of 0.1. To summarize the detection accuracies across the six data sets and two software in one histogram (Figure 2), the top- k levels at which the increase in the accuracy stabilized were used as cut-offs (*stable F-score*, indicated by arrows in Figure 1 and Supplementary Figure S1). However, all the original trace graphs are provided as Supplementary Figure S1.

The stable *F*-scores were determined separately for the ROTS-defined and default parameter settings of the particular peak detection algorithm (Figure 1). In the ROTS runs, we also recorded the ideal and worst *F*-scores, i.e. the highest and lowest possible accuracies that can be obtained given the independent qPCR validations and the pre-defined parameter space at each level of top peaks separately, even if this information is not available during the peak detection. The minimum and maximum accuracies were used to calculate the *scaled F-score*:

$$F' = \frac{F - \min F}{\max F - \min F}.$$

The scaled *F*-values were used here when comparing the performance of ROTS to that originating from the default settings of the peak calling software across the various data sets (Figure 2), because it can effectively normalize the differences between the ChIP data sets due to their different size, coverage, quality, etc. By taking into account the possible range of detection accuracies obtained with

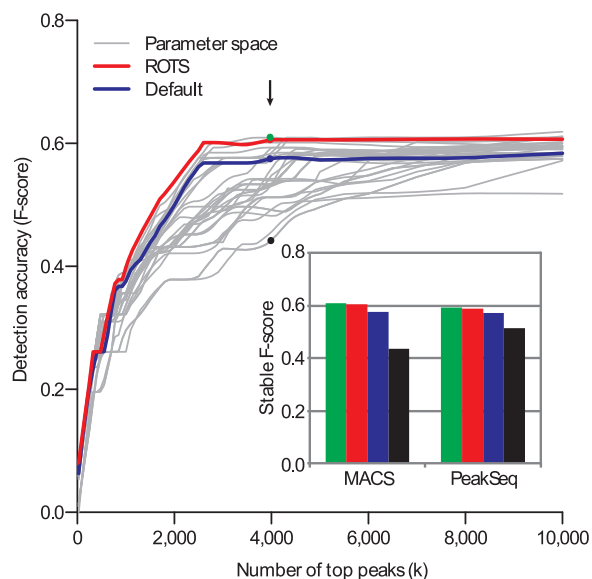


Figure 1. Accuracy of the binding site detections in the STAT1_1 data set as a function of top peaks identified by the *MACS* algorithm. The accuracy of the peak calling parameter combinations was evaluated with respect to independent qPCR validations using the *F*-score (see ‘Materials and Methods’ section). The grey traces show the variability in the accuracy when different parameter combinations were used. The red and blue traces, respectively, indicate the accuracy of the parameter values learned by the reproducibility optimization procedure (ROTS), compared to the default settings of the software package. The *insert* shows the *F*-levels at the cut-off point in which the increase in the accuracy stabilizes (the arrow). The green and black bars, respectively, indicate the highest and lowest *F*-scores among all the parameter combinations tested at the given cut-off point (the green and black points, respectively). The trace graphs were smoothed for displaying purposes. MACS detections in STAT1 were used here as an example; all the MACS and PeakSeq results are provided as Supplementary Figure S1.

different parameter settings (including the ROTS-defined and default settings), the scaled *F*-levels can quantify the relative difference between any given parameter combinations from a more practical point of view. The original *F*-scores (Default, ROTS, Min and Max) are available in Supplementary Figure S2.

The calculation of the *F* and *F'*-scores was based on the relative frequency of the true positives with respect to regions that were confirmed in independent qPCR experiments to be bound by the particular transcription factor (known positives), as well as regions that did not show binding in these experiments (known negatives). As has been noted before, assessing the relative frequency of false positives (or specificity) in the binding site detection is challenging, because of the question how to define reliably the true negative detections (9,22). Therefore, the specificity was not assessed in this study.

Implementation of the peakROTS package

We have made available an implementation of the ROTS procedure for ChIP-seq data (named peakROTS) as a stand-alone, open-source R package (Supplementary Tutorial, <http://www.nic.funet.fi/pub/sci/molbio/peakROTS>) The implementation is platform independent, requiring only an R environment (<http://www.r-project.org>). To facilitate

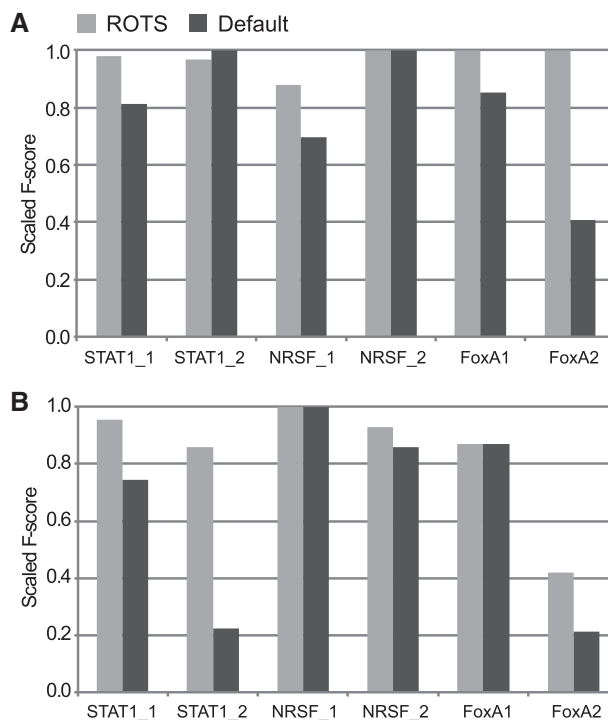


Figure 2. Accuracy of the binding site detections when using the ROTS or default parameter settings in MACS (A) and PeakSeq (B). The detection accuracy was evaluated using the scaled *F*-score (see ‘Materials and Methods’ section), which shows the practical difference between the two parameter combinations with respect to the highest and lowest possible accuracies that can be obtained, given the independent qPCR validations and the pre-defined parameter space. The scaled *F*-score was used here to compare the relative performance across the different data sets (FoxA2 data set is from a mouse system, while the others are human data sets); all the original *F*-scores (Default, ROTS, Min and Max) are available in Supplementary Figure S2. To summarize the detection accuracies across all the six data sets in one histogram, the stable *F*-scores are shown, which correspond to the top-*k* levels at which the increase in the accuracy stabilized (indicated by arrows in Figure 1 and Supplementary Figure S1). The overall difference between the ROTS and default parameters was statistically significant across the data sets (paired *t*-test, $P < 0.05$).

in-depth searching through large parameter spaces, we have modularized the implementation so that it can be efficiently distributed across multiple computing cores, allowing large computational resources to be utilized effectively. The infrastructure needed for the distributed computing is included in the *R* package. The current implementation supports both a local process-based distribution (single node, multiple cores), as well as an LSF batch processing system (multiple nodes). The distribution mechanism can be plugged-in to enable running different parts of any single analysis task even using different distribution mechanisms. The results presented here were computed on a HP CP4000 BL ProLiant cluster system (http://www.csc.fi/english/research/Computing_services/computing/servers/murska), using at maximum 512 computing cores via the LSF batch processing system.

When the analysis task is initialized, the peakROTS package generates a workflow graph, which describes the dependencies between the individual analysis steps (Supplementary Figure S3). In the actual computation,

the process reads in the workflow graph and becomes a master node for the current analysis task. New worker nodes are then spawned using the plugged-in distribution mechanism. Worker nodes report back to the master node, which takes care of the dependency tracking between the analysis steps. In the current implementation, the distribution mechanisms either start a new process or submit a new job to any LSF batch processing system. Plugging in new batch processing systems requires only providing the command that submits a job to the system. Only the submit command is specific to the batch processing system in question.

Some of the use cases reported here involve rather heavy computation. To perform such cases effectively, a key goal in the design of the distributed implementation has been error tolerance. The state of the distributed work is kept at disk all the time, meaning that the master node can crash or be shut down without losing any of the intermediate results. The master node synchronizes its state and continues to work once it has restarted. All the individual analysis jobs undergo three phases: pending, running and finished. The whole state of the system is represented in three text files, one for each of the phases. Each job corresponds to a single line in the text file and these are moved through the three files by the master node. When the master node is not running, the files can be edited manually, allowing, for instance, manual controlling of jobs for debugging and failure resolution.

In more general terms, the generic peakROTS implementation can be used not only for finding the optimal peak detection package and its parameter settings for each ChIP-seq data individually, but also for assessing the quality of each of the steps in the ChIP-seq data analysis pipeline. When optimizing such pipeline, it is essential not to be biased towards a specific data type or peak detection algorithm. Besides enabling the users to extract full information from their ChIP-seq data sets, the ROTS procedure can also be used by developers of new and improved peak detection algorithms as a benchmarking tool. The modular architecture of the peakROTS will accommodate additional new features, such as improved peak calling algorithms and large-scale cloud computing solutions, as dictated by future experimental and computational needs (Supplementary Documentation).

RESULTS

We first used data from the STAT1 study as an example to demonstrate the performance of ROTS with the two software packages in more detail (Figure 1). The peak calling parameters learned by the ROTS procedure provided systematic improvements in the precision and sensitivity of the default parameter settings, with an accuracy approaching the ideal case, corresponding to the situation where the qPCR validation information were already available in the peak calling phase. This information is typically not available in practice and it was not utilized by the ROTS procedure. The performance of the different parameter combinations presented with a considerably

large range of variation in terms of their qPCR-based detection accuracy, demonstrating that such a purely ChIP-seq data-driven adjustment of the parameter settings for the software packages is a highly non-trivial task. In particular, several profound differences in the peak detections between the ROTS and default parameter settings were observed (Supplementary Figure S4). It should be noted that the same STAT1 study was used also in the original PeakSeq work (14), further highlighting the relative improvements gained by ROTS over the default settings (Figure 1, *insert*).

To evaluate whether the benefits of the ROTS procedure generalize also to other studies, we repeated the same analyses in four human and in one mouse ChIP-seq data sets. The detection accuracies across all the six data sets, with notably different characteristics (Table 1), supported the idea that the ROTS procedure enables the user to adjust the peak calling parameters of the software packages for each data set individually, leading to significantly improved detection of binding sites, when compared to the default settings (Figure 2, paired *t*-test, $P < 0.05$). For comparison across diverse data sets, the scaled *F*-score was used, which quantifies the relative differences between the given parameter combinations (see 'Materials and Methods' section). As noted before (8), the software packages have been tested and trained on some of the older data sets, such as STAT1 and NRSF, which may explain why already the default parameters corresponded to the ideal performance in some cases (scaled $F = 1$). However, even if the NRSF and FoxA1 data sets were used in the original MACS work as testdata (13), the ROTS parameters could improve the binding site detection accuracies beyond its default settings (Figure 2A).

We further investigated whether the extent of sequencing attributed to the observed differences between the ROTS and default parameters by analyzing the accuracy of the binding site detections as a function of the number of sequenced tags in each data set (Supplementary Figure S5), as well as in subsamples of the STAT1 data set generated by randomly sampling 40–100% of the tags from the original data without replacement (Supplementary Figure S6). These results suggested that, in general, the number of tags cannot explain the differences in the detection accuracies between the ChIP-seq data sets. Intriguingly, however, the ROTS reproducibility levels can inform the user whether the peak calling was successful or not in a given data set. For instance, the relatively low detection accuracies in the mouse FoxA2 data set, especially with the PeakSeq algorithm, could be predicted from the poor reproducibility levels, as reported by the ROTS procedure (Supplementary Table S1). Further investigation of the replicates of the FoxA2 data set demonstrated how the reproducibility levels can indicate whether the peak calling results allow accurate detection of true binding sites (Supplementary Figure S7). Moreover, in case the user is willing to experiment with several peak detection software packages, the reproducibility levels may also be used to provide guidance on choosing the software solution for a given data set (Supplementary Figure S8).

DISCUSSION

Taken together, these proof-of-concept results demonstrate that the ROTS procedure provides the user with several advantages over the current practice when analyzing the massive data sets from the increasing number of ChIP-seq experiments. Not only does it make possible to avoid poor parameter settings for a given data set, but it can systematically improve the binding site detections, compared to those originating from the default settings. Beyond providing guidance on how to select the peak calling parameters, the procedure can also be used to inform whether the data quality and/or the software parameters were sufficient for reliable binding site detections with a selected software package, or even to choose the package which is optimal for a given data set. Therefore, the procedure should prove useful for optimizing a wide variety of existing and emerging ChIP-seq studies (a walk-through example use case is provided in Supplementary Tutorial).

While the potential of the ROTS procedure was demonstrated here using a relatively limited range of possible parameter combinations (36 in MACS and 24 in PeakSeq, Supplementary Table S1), it is likely that even higher improvements will be obtained after a more systematic and fine-scaled searching of those parts of the parameter space that are most potential for each data set and software separately. To enable its tailored application to future ChIP-seq experiments, we have made available an open-source and easily extendable implementation of the ROTS procedure, which can take advantage of local cluster or public cloud computing resources (Supplementary Documentation).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table S1, Supplementary Figures S1–S8, Supplementary Tutorial, Supplementary Documentation.

ACKNOWLEDGEMENTS

The authors thank Dr Geetu Tuteja for providing the separate FoxA2 replicates and Jukka Hiissa and Rolf Linden for technical assistance. The present results were made with the help of the computing resources of CSC - IT Center for Science Ltd.

FUNDING

Academy of Finland (grants 127575, 218591 to L.L.E., grants 120569, 133227, 140880 to T.A.). Funding for open access charge: The Academy of Finland.

Conflict of interest statement. None declared.

REFERENCES

- Johnson, D.S., Mortazavi, A., Myers, R.M. and Wold, B. (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science*, **316**, 1497–1502.

- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A. *et al.* (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods*, **4**, 651–657.
- Farnham, P.J. (2009) Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.*, **10**, 605–616.
- Park, P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, **10**, 669–680.
- Elo, L.L., Järvenpää, H., Tuomela, S., Raghav, S., Ahlfors, H., Laurila, K., Gupta, B., Lund, R.J., Tahvanainen, J., Hawkins, R.D. *et al.* (2010) Genome-wide profiling of interleukin-4 and STAT6 transcription factor regulation of human Th2 cell programming. *Immunity*, **32**, 852–862.
- Pepke, S., Wold, B. and Mortazavi, A. (2009) Computation for ChIP-seq and RNA-seq studies. *Nat. Methods*, **6**, S22–S32.
- Laajala, T.D., Raghav, S., Tuomela, S., Laheesmaa, R., Aittokallio, T. and Elo, L.L. (2009) A practical comparison of methods for detecting transcription factor-binding sites in ChIP-seq experiments. *BMC Genomics*, **10**, 618.
- Hawkins, R.D., Hon, G.C. and Ren, B. (2010) Next-generation genomics: an integrative approach. *Nat. Rev. Genet.*, **11**, 476–486.
- Wilbanks, E.G. and Facciotti, M.T. (2010) Evaluation of algorithm performance in ChIP-seq peak detection. *PLoS One*, **5**, e11471.
- Ho, J.W., Bishop, E., Karchenko, P.V., Nègre, N., White, K.P. and Park, P.J. (2011) ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis. *BMC Genomics*, **12**, 134.
- Elo, L.L., Hiissa, J., Tuimala, J., Kallio, A., Korpelainen, E. and Aittokallio, T. (2009) Optimized detection of differential expression in global profiling experiments: case studies in clinical transcriptomic and quantitative proteomic datasets. *Brief. Bioinform.*, **10**, 547–555.
- Elo, L.L., Filén, S., Laheesmaa, R. and Aittokallio, T. (2008) Reproducibility-optimized test statistic for ranking genes in microarray studies. *IEEE/ACM Trans Comput. Biol. Bioinform.*, **5**, 423–431.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
- Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carrier, N., Snyder, M. and Gerstein, M.B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol*, **27**, 66–75.
- Euskirchen, G.M., Rozowsky, J.S., Wei, C.L., Lee, W.H., Zhang, Z.D., Hartman, S., Emanuelsson, O., Stolc, V., Weissman, S., Gerstein, M.B. *et al.* (2007) Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. *Genome Res.*, **17**, 898–909.
- Mortazavi, A., Leeper Thompson, E.C., Garcia, S.T., Myers, R.M. and Wold, B. (2006) Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire. *Genome Res.*, **16**, 1208–1221.
- Valouev, A., Johnson, D.S., Sundquist, A., Medina, C., Anton, E., Batzoglu, S., Myers, R.M. and Sidow, A. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods*, **5**, 829–834.
- Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S. and Brown, M. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, **132**, 958–970.
- Tuteja, G., White, P., Schug, J. and Kaestner, K.H. (2009) Extracting transcription factor targets from ChIP-Seq data. *Nucleic Acids Res.*, **37**, e113.
- Wederell, E.D., Bilenky, M., Cullum, R., Thiessen, N., Dagginar, M., Delaney, A., Varhol, R., Zhao, Y., Zeng, T., Bernier, B. *et al.* (2008) Global analysis of in vivo Foxa2-binding sites in mouse adult

- liver using massively parallel sequencing. *Nucleic Acids Res.*, **36**, 4549–4564.
21. Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M. and Wong, W.H. (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat. Biotechnol.*, **26**, 1293–1300.
 22. Kharchenko, P.V., Tolstorukov, M.Y. and Park, P.J. (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.*, **26**, 1351–1359.
 23. Toedling, J., Skylar, O., Krueger, T., Fischer, J.J., Sperling, S. and Huber, W. (2007) Ringo—an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics*, **8**, 221.