

RESEARCH

Open Access



Reporter gene assays and chromatin-level assays define substantially non-overlapping sets of enhancer sequences

Daniel Lindhorst^{1,2} and Marc S. Halfon^{1,3,4,5,6*}

Abstract

Background Transcriptional enhancers are essential for gene regulation, but how these regulatory elements are best defined remains a significant unresolved question. Traditional definitions rely on activity-based criteria such as reporter gene assays, while more recently, biochemical assays based on chromatin-level phenomena such as chromatin accessibility, histone modifications, and localized RNA transcription have gained prominence.

Results We examine here whether these two types of definitions, activity-based and chromatin-based, effectively identify the same sets of sequences. We find that, concerning, the overlap between the two groups is strikingly limited. Few of the data sets we compared displayed statistically significant overlap, and even for those, the degree of overlap was typically small (below 40% of sequences). Moreover, a substantial batch effect was observed in which experiment set rather than experimental method was a primary driver of whether or not chromatin-defined enhancers showed a strong overlap with reporter gene-defined enhancers.

Conclusions Our results raise important questions as to the appropriateness of both old and new enhancer definitions, and suggest that new approaches are required to reconcile the poor agreement among existing methods for defining enhancers.

Keywords *Cis*-regulation, Enhancer, CRM, Gene regulation, Reporter gene, ATAC-seq, Histone modification

*Correspondence:

Marc S. Halfon
mshalfon@buffalo.edu

¹ Department of Biochemistry, University at Buffalo-State University of New York, 955 Main St. #5128, Buffalo, NY 14203, USA

² Present Address: Program in Biomedical Sciences, Columbia University, New York, NY 10032, USA

³ Department of Biomedical Informatics, University at Buffalo-State University of New York, Buffalo, NY 14203, USA

⁴ Department of Biological Sciences, University at Buffalo-State University of New York, Buffalo, NY 14260, USA

⁵ NY State Center of Excellence in Bioinformatics & Life Sciences, Buffalo, NY 14203, USA

⁶ Department of Molecular and Cellular Biology and Program in Cancer Genetics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA

Background

Transcriptional enhancers play an essential role in gene regulation and are primary mediators of development, homeostasis, disease, and evolution [1–3]. Among many unanswered questions about enhancer biology, one stands out as fundamental: how should an enhancer be defined? In the genomic era, the functional definition that for a quarter-century described enhancers as sequences with the ability to drive expression of a reporter gene from a minimal promoter [4, 5] has entered into an uneasy co-existence with transcription and chromatin-based definitions such as ability to bind specific sets of transcription factors or coactivators, presence of certain histone modifications, location in nucleosome-depleted regions, or transcription of enhancer RNA (eRNA) (e.g. [6–15]) (Fig. 1). It has become increasingly clear that



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

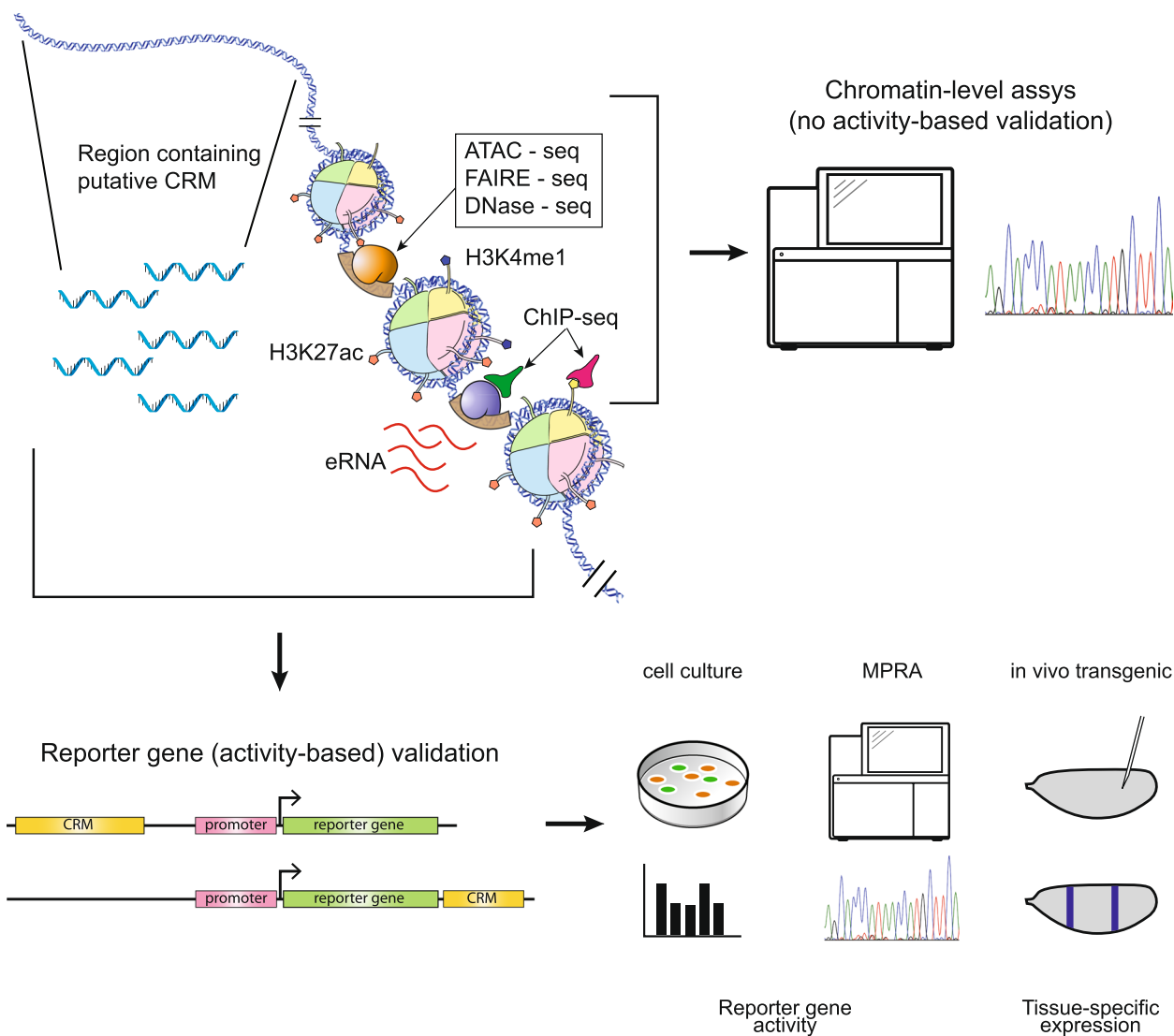


Fig. 1 Competing methods for defining enhancers. Classical definitions of enhancers, based on functional tests such as reporter gene assays (bottom), are in tension with enhancer definitions based on chromatin-level assays (top right). For either type of assay, sequences can be chosen based on a large number of specific or undirected criteria (top left)

these chromatin-level enhancer definitions identify sets of sequences with strikingly low levels of overlap, with concerning implications for regulatory genome annotation, ongoing studies of enhancer biology, and interpreting results from genome-wide association studies, expression quantitative trait locus studies, and the like [16–18]. However, it is unknown which of these various assays are in the best agreement with reporter gene data, which remains the “gold standard” for enhancer activity, and to what extent.

Here, we perform a comprehensive comparison to investigate whether one or a collection of chromatin-based assays are able to identify the majority of enhancers from

an extensive reporter-gene defined set. We show that not only do the chromatin-level assays show poor agreement among themselves, but also that they fail to discover a significant fraction of reporter-gene defined enhancers, often performing no better than random expectation. Our results raise questions as to whether any common current assays sufficiently interrogate the enhancer landscape, and about the accuracy of current regulatory genome annotations.

Results and discussion

In order to test for congruence between enhancers defined by reporter gene assays, and those defined by chromatin-based assays, we compared *Drosophila*

enhancers obtained from two sources: REDfly [19] and EnhancerAtlas2.0 [20]. The REDfly database contains over 38,000 empirically-defined enhancers manually curated from the literature, split roughly evenly between *in vivo* reporter gene assays and cell-culture based reporter gene assays. The EnhancerAtlas2.0 database, by contrast, is composed of 294,158 enhancers predicted by means of an unsupervised learning approach that combines data from ChIP-seq, ATAC-seq, FAIRE-seq, and other chromatin-level assays [20]. This comparison is uniquely possible for *Drosophila*: as the overwhelming majority of REDfly's enhancers are based on evidence developed without consideration of chromatin-level features, they are therefore identified fully independently from those in EnhancerAtlas, using different evidence.

For an initial comparison, we took all REDfly enhancers 2 kb or shorter (11,549 total) and determined how many of these sequences overlapped an EnhancerAtlas enhancer, without regard for the annotated tissue-specificity of the REDfly enhancers (Fig. 2A; see Methods). Of the 21 tissue-specific *Drosophila* EnhancerAtlas datasets, 17 had statistically significant overlap (adjusted $P < 0.01$, two-tailed z test), three had no significant overlap, and, surprisingly, one had significantly less overlap than expected by chance (Fig. 3A “STARR-seq included”, Table S1a). Examination of these data revealed that more than three-quarters of the REDfly enhancers we were using (8779/11549, 76%) were identified using a single method, STARR-seq [21], a cell-culture based, episomal reporter gene assay; these STARR-seq enhancers made up 80% of the REDfly-EnhancerAtlas overlapping enhancers (6027/7527). When we eliminated the STARR-seq enhancers from the REDfly test set and repeated the analysis, the results were strikingly different: only five of the EnhancerAtlas datasets (24%) now showed significant overlap (with one additional dataset just below our significance threshold), whereas ten datasets (48%) had significantly less overlap than expected by chance (Fig. 3A “STARR-seq not included”, Table S1b [Additional File 1]).

STARR-seq-defined enhancers thus have a profound effect on how well sequences from the two databases compare. Only four EnhancerAtlas datasets make use of STARR-seq data, indicating that a simple confounding of the source data cannot explain the results. In the absence

of STARR-seq enhancers, almost half of the REDfly datasets had less-than-expected overlap with EnhancerAtlas, suggesting that the majority of REDfly enhancers lack activity in most of the tissues covered by EnhancerAtlas. Conversely, these results suggest that STARR-seq, performed in a cultured cell line and using an episomal rather than an integrated reporter, may be identifying many sequences indiscriminately with respect to their tissue specific activity.

To examine more directly how REDfly enhancers active in specific tissues compare with the EnhancerAtlas datasets, we constructed tissue-specific REDfly enhancer sets using only *in-vivo* reporter gene tested enhancers < 1000 bp in length (Fig. 2B; Table S2 [Additional File 2]). Of the 21 EnhancerAtlas sets, 11 had one or more corresponding REDfly sets. We then determined how many enhancers from each paired set overlapped. Our expectation was that there should be significant overlap, with the EnhancerAtlas sets (based on undirected genome-wide assays) encompassing the great majority of REDfly enhancers (drawn from individual ad hoc experiments) (Fig. 2C). Surprisingly, only four (36%) of the eleven EnhancerAtlas sets showed significant overlap with their corresponding REDfly set ($P < 0.01$, Fig. 3A “all tissue-specific datasets”, Table S1c [Additional File 1]). Moreover, even for the sets with significant overlap, the number of in-common enhancers was strikingly limited (median 26%, range 5%-65%; Fig. 3B, Table S1c [Additional File 1]).

Since the EnhancerAtlas definitions integrate the data from multiple assays, we reasoned that the integration algorithm might be filtering out some of the true enhancers. To test this, we took the underlying data sets used by EnhancerAtlas (Table S1e [Additional File 1], referred to as EnhancerAtlas “subsets”; Fig. 2B) and compared them individually to the matched REDfly enhancer sets. Indeed, we saw a higher number of significantly overlapping enhancer sets (66%, Fig. 3A, Table S1d,e [Additional File 1]), but again, the number of in-common enhancers within each matched set was low (median 39%, Fig. 3B, Table S1d,e [Additional File 1]).

The low degree of in-common enhancers could represent a small number of REDfly enhancers that consistently match EnhancerAtlas enhancers, or different subsets of REDfly enhancers for each EnhancerAtlas

(See figure on next page.)

Fig. 2 Schematic of REDfly-EnhancerAtlas comparisons performed in this study. See text and Methods for details. **A** REDfly enhancers under 2000 bp in length were selected and tested for overlap against each of the 21 EnhancerAtlas *Drosophila* enhancer sets. Similar comparisons were made subsequently after filtering out all enhancers defined by STARR-seq assays. **B** REDfly enhancers were placed into tissue-specific sets after discarding all sequences longer than 1000 bp (600 bp for some sets). Eleven such sets were tested for overlap with eleven tissue-matched sets from EnhancerAtlas. Each EnhancerAtlas set was then split into multiple individual sets with the data from just a single component experiment (“subsets”), and each subset was tested for overlap against the corresponding REDfly tissue-matched set. **C** EnhancerAtlas data are based on genome-wide assays, while REDfly enhancers are drawn from individual reporter gene experiments. Therefore, we expect the majority of REDfly enhancers for a given tissue to comprise a subset of the EnhancerAtlas enhancers for the matched tissue

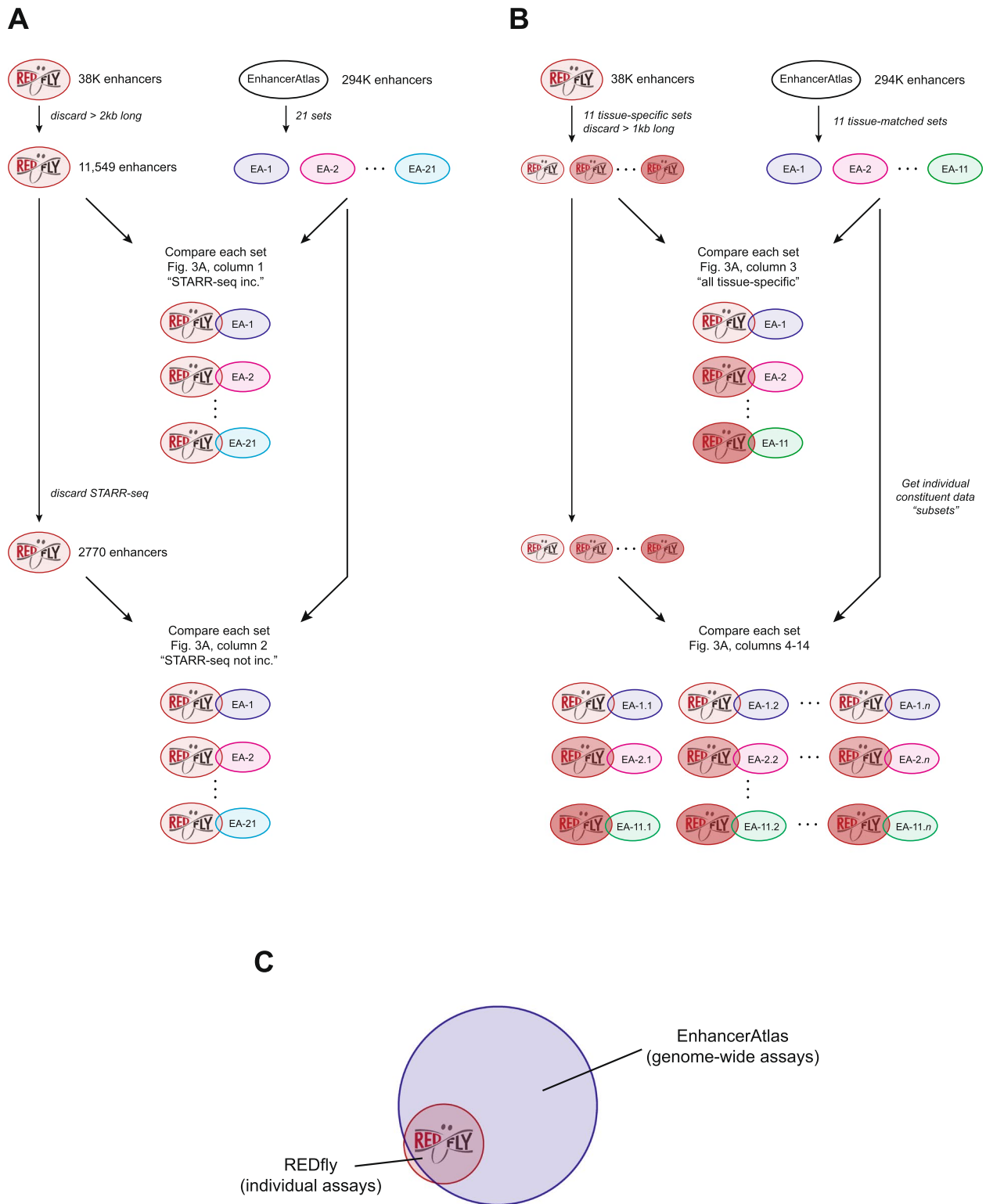


Fig. 2 (See legend on previous page.)

subset. To distinguish between these possibilities, we looked at the correlation between the sets of REDfly enhancers present in individual EnhancerAtlas subsets. The sets of enhancers found in different experiments were not overall well-correlated, suggesting that distinct REDfly enhancer sets are being identified (Fig. 4, Fig. S1 [Additional File 3]). However, a clear correlation structure was evident in which assays from a particular laboratory and method tended to cluster together. For example, the REDfly enhancers from one set of experiments (series GSE102839) are highly correlated (Fig. 4, white box; mean $r=0.74\pm 0.11$), and those from a different set of experiments (series GSE101827) are highly correlated (Fig. 4, yellow box; mean $r=0.67\pm 0.18$). However, even though both sets contain ATAC-seq experiments, they find a disjoint group of REDfly enhancers (mean $r=0.12\pm 0.08$). Although assay-specific correlation is occasionally observed (Fig. 4, yellow asterisks; $r=0.77$), in other cases similar assays are only weakly correlated (e.g., Fig. 4, white asterisks; $r=0.45$). EnhancerAtlas data based on ATAC-seq assays had a clear bias toward significant REDfly overlap (58/65 subsets significant; Table S1d [Additional File 1]), as did other open-chromatin assays, while CHIP-based methods tended to be more evenly distributed between significant and non-significant overlap (Table S1d [Additional File 1]). However, the fact that batch effects appear to dominate the correlation structure makes it difficult to draw conclusions as to the most effective enhancer identification methods. For instance, we note that over two-thirds of the ATAC-seq results are from a single experiment series, GSE101287, and the number of subsets with significant REDfly overlap for the ATAC-seq experiments in series GSE102441 and series GSE102839 was a more modest 60% and 50%, respectively.

These observations suggest two important conclusions: (1) identification of putative enhancers is highly dependent on not just the type of assay performed, but on the precise conditions under which it is performed; and (2) enhancer identification is reasonably robust given a particular set of assays and replicates. The fact that the sets of REDfly enhancers are stable within groups of replicate assays suggests that these enhancers are being specifically

(i.e., non-randomly) found, despite the fact that, as shown above, the number of REDfly enhancers identified through chromatin-level assays is frequently indistinguishable from random expectation. Thus, these assays do appear to be able to identify reporter-gene-defined enhancers, but with low efficiency and potentially high false-positive rates.

Recently, Gao et al. released “scEnhancer”, an EnhancerAtlas-like database based on single-cell ATAC-seq data drawn from different embryonic time points [22]. Unlike the results using EnhancerAtlas, the majority of scEnhancer datasets had a significant degree of overlap with their REDfly set (45/56, 80%; Fig. 1A, Table S1f [Additional File 1]), although similar to what we observed with EnhancerAtlas, the number of overlapping enhancers was low (median 27%, range 10%-65%; Table S1f [Additional File 1]). scATAC-seq may thus represent a more promising method for enhancer detection, although a proper assessment is difficult as all of the *Drosophila* data currently in scEnhancer are from a single source.

Conclusions

Reporter gene assays have long been considered the gold standard for defining enhancers. On these grounds, our results would seem to suggest that not only do chromatin-level assays frequently fail to identify common sets of enhancer sequences [16], but neither are they particularly effective at covering the majority of the enhancer landscape. However, we would caution against automatically accepting this conclusion, as there are well-known deficiencies that could lead to a substantial number of both false-positive and false-negative results from reporter gene assays. These include enhancer-promoter incompatibility, effects due to episomal expression or chromosomal integration, cell-type specificity, and ectopic expression due to missing repressor binding sites, as well as recent findings that enhancers can double as silencers or functionally overlap other regulatory features (see discussions in [23–27]). It is also possible that the “right” set of chromatin assays has yet to be applied. For instance, Koenecke et al. [28] suggest that a key parameter for enhancer identification is the relative,

(See figure on next page.)

Fig. 3 Comparisons between REDfly and EnhancerAtlas enhancer sets. REDfly and EnhancerAtlas datasets were compared and tested for significance as described in the text and Table S1 [Additional File 1]. Box plots show medians and the first and third quartiles. Data points shown in red are significant at a Bonferonni-adjusted P -value < 0.01 . Dataset names correspond to the names in Table S1 [Additional File 1]. For the “tissue-specific comparisons” (Table S1c [Additional File 1]), when multiple REDfly sets corresponded to the same EnhancerAtlas set, the one with the highest degree of overlap was selected for analysis. For the subset comparisons (Table S1d,e [Additional File 1]), the REDfly set with the largest number of significantly overlapping individual EnhancerAtlas component sets was used. **A** z-score distributions for each of the REDfly-EnhancerAtlas comparisons. **B** Percent of REDfly enhancers found that overlap EnhancerAtlas enhancers. For the “all tissue-specific datasets” plot, datasets with significant z-scores are indicated in red. For the subset comparisons, values represent the median percent overlap of the constituent data subsets, as shown in Table S1e [Additional File 1]

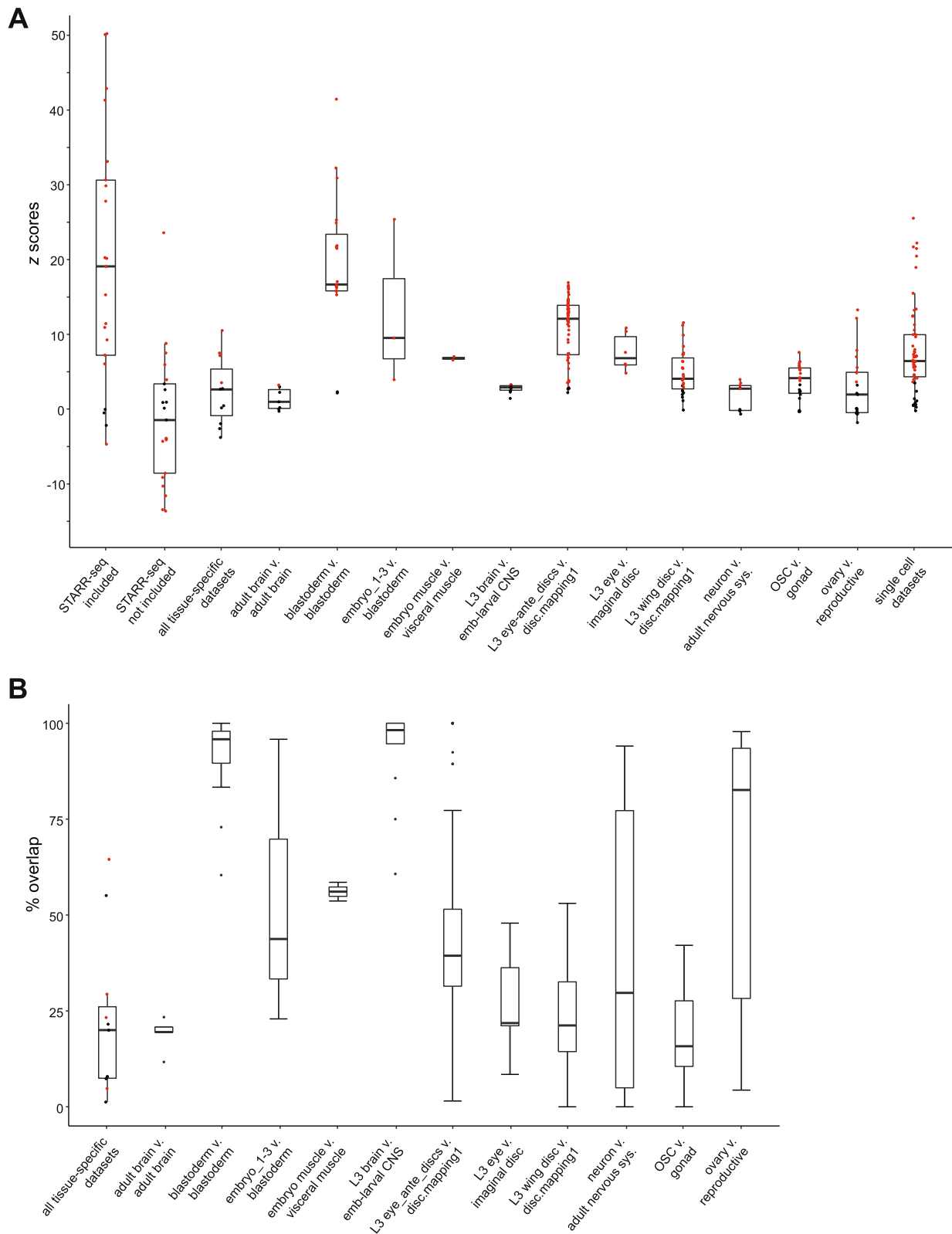


Fig. 3 (See legend on previous page.)

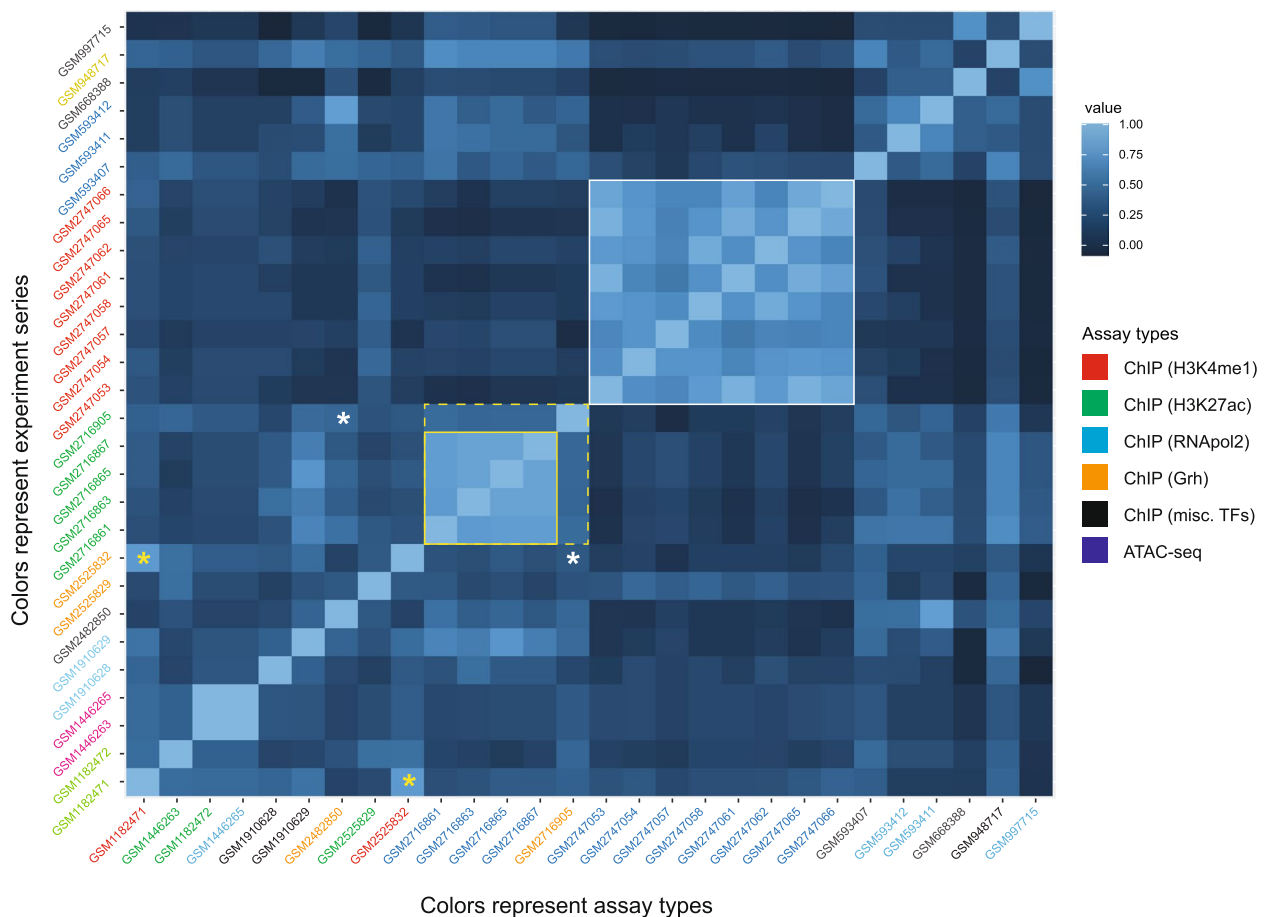


Fig. 4 Correlations between REDfly enhancers overlapping enhancers from each EnhancerAtlas subset used for the “L3_wing_disc” EnhancerAtlas set. The correlation structure demonstrates that experimental batch effects predominate over assay-type effects. High correlations are seen between sets from the same experiment group (see white box, yellow box), even when assay types differ (dotted yellow box). Although assay-specific correlations are sometimes present (yellow asterisks), identical assays can also be poorly correlated when from different experiment groups (white asterisks). Individual experiments on the y-axis (“GSM” identifiers) are colored according to common experimental series (i.e., performed by the same laboratory as a specific set of experiments) as provided in GEO (“GSE” identifiers, see Table S1d [Additional File 1]). Sets labeled in black are unrelated. Label colors on the x-axis indicate assay types as follows: red, H3K4me1 ChIP-seq; green, H3K27ac ChIP-seq; orange, Grh ChIP-seq; cyan, Pol2 ChIP-seq; dark blue, ATAC-seq; black, ChIP-seq against various transcription factors

rather than absolute, level of histone 3, lysine 27 acetylation (H3K27ac) flanking an enhancer, and many less-frequently assayed histone modifications appear to be present in various combinations at at least some subsets of enhancers (e.g., [29–31]).

Without true “known” enhancer data, therefore, it is impossible to say whether our results reflect significantly poor sensitivity in chromatin-level assays, or a much larger than heretofore recognized false positive rate in reporter gene assays. What is indisputable, however, is a clear need for approaches that can reconcile the poor agreement between and among the various reporter gene and chromatin-level enhancer identification methods. In this regard, the increasing tractability of genome-engineering approaches, e.g.,

via CRISPR/Cas9 sequence deletion and replacement, holds out an encouraging potential to interrogate the enhancer capability of sequences within their native genomic contexts.

Methods

Data sources

For comparisons using all REDfly data (Table S1a, S1b), sequences were obtained from REDfly v7.1.1 (Aug. 14 2020) by downloading all “CRM” entries in bed format, eliminating sequences >2000 bp, and then removing overlapping sequences using the script “SelectSmallestFeature.py” (Kazemian and Halfon 2019). All REDfly data used in this study are based on experimental reporter gene assays curated from the primary literature

by the professional REDfly biocuration team. The length cutoff of 2000 bp (and below, 600–1000 bp) was selected so as to reduce potential spurious results from the inclusion of non-regulatory sequences, or multiple independent enhancers, in our comparisons. Because REDfly data are based on reporter gene assays, the length of the tested sequences can be arbitrarily large, but data from numerous deletion studies suggests that typical enhancers are on the order of hundreds, not thousands, of base pairs.

For comparisons using specific REDfly subsets (Table S1c–S1f), “CRM” sequences were downloaded from REDfly v5.6.1 (Dec. 3 2019) in bed format with “cell-culture only” sequences excluded and a sequence length cutoff of either 1000 or 600 bp. Following removal of overlapping sequences using “SelectSmallestFeature.py” (Kazemian and Halfon 2019), the expression pattern annotations associated with each remaining sequence were used to place the sequences into one or more of 30 different tissue-specific groupings. Details of the composition of each set can be found in Supplementary Table S2.

For compatibility with EnhancerAtlas, genome coordinates were converted from dm6 to dm3 using LiftOver [32] with minMatch = 0.25.

EnhancerAtlas sequences were downloaded from EnhancerAtlas 2.0 (<http://enhanceratlas.org>, Nov. 19 2019). EnhancerAtlas sequences are compiled using the EnhancerAtlas unsupervised learning algorithm based on data from one or more of 12 types of chromatin-level assays [20]. In brief, data from each individual experiment (“track”) are normalized and merged based on Jaccard overlap across the genome. Importantly, a peak is only merged into the consensus profile if it is supported by at least 50% of the included tracks.

Identities of the component datasets for each of the 21 tissue-specific EnhancerAtlas sets were obtained from the metadata files available in the “data source” section. The provided GEO accession codes were then used to obtain the sequence-level data from NCBI. Data processing was performed as described in the EnhancerAtlas paper [20] to ensure consistent results. Wig files were converted to bigWig format through the script wig2BigWig, downloaded from the UCSC genome browser at genome.ucsc.edu/goldenpath/help/bigwig.html. BigWig files were converted to bedgraph format through the script bigWigtoBedGraph. Bedgraph files were converted to bed format through peak calling using MACS2 [33] with a cutoff enrichment of 2. Data sets where we were unable to replicate the exact EnhancerAtlas processing pipeline (primarily, raw sequencing data) were omitted from further analysis.

scEnhancer sequences were downloaded from scEnhancer (<http://enhanceratlas.net/scenhancer/>, Feb. 28. 2022).

Comparison of data sets

Bed files were compared using BEDTools *intersect* [34] and the *-wa* and *-u* flags. Note that with these parameters, even a single intersecting basepair will cause the two sequences to be scored as an intersection, making our tests highly sensitive to any degree of sequence overlap.

Significance of comparisons

Significance of dataset overlap was determined by permuting the coordinates of each REDfly dataset 500 times using BEDTools *shuffle* [34] and repeating the tests for intersection. The mean and standard deviation of the permuted results were then used to calculate a z-score. A Bonferroni-corrected *P* value equivalent to $P < 0.01$ was determined for each set of comparisons.

Correlation analysis

For each comparison, each potential REDfly enhancer was scored as 1 (found) or 0 (not found). Correlations between all pairs of result vectors were determined using the R *cor* function and visualized as heat maps using *ggplot*.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09123-9>.

Additional file 1.

Additional file 2.

Additional file 3.

Acknowledgements

We thank Hasiba Asma for assistance with figures and Yungki Park, Michael Buck, and members of the Halfon lab for comments on the manuscript. Nanzeeba Ahmad performed some preliminary analyses for this project. This work was performed in part using the resources of the University at Buffalo’s Center for Computational Research (<http://hdl.handle.net/10477/79221>).

Authors’ contributions

DL performed the analysis. MSH conceived of the study and provided guidance on the analysis. Both authors wrote the manuscript. Both authors read and approved the final manuscript.

Funding

This work was funded by grants from the National Science Foundation (NSF DBI-1758252) and the National Institutes of Health (NIH U24 GM142435) to MSH. The funders did not play any role in the design, analysis, or interpretation of the study or in writing the manuscript.

Availability of data and materials

REDfly data are accessible at <http://redfly.ccr.buffalo.edu>. EnhancerAtlas data can be accessed at <http://www.enhanceratlas.org/>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 11 October 2022 Accepted: 9 January 2023

Published online: 13 January 2023

References

- Field A, Adelman K. Evaluating Enhancer Function and Transcription. *Annu Rev Biochem.* 2020;89:213–34.
- Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell.* 2016;167(5):1170–87.
- Rickels R, Shilatifard A. Enhancer logic and mechanics in development and disease. *Trends Cell Biol.* 2018;28(8):608–30.
- Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell.* 1981;27(2 Pt 1):299–308.
- Moreau P, Hen R, Wasyluk B, Everett R, Gaub MP, Chambon P. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* 1981;9(22):6047–68.
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A.* 2010;107(50):21931–6.
- Dogan N, Wu W, Morrissey CS, Chen KB, Stonestrom A, Long M, et al. Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin.* 2015;8:16.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature.* 2009;459(7243):108–12.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007;39(3):311–8.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature.* 2011;470(7333):279–83.
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 2009;457(7231):854–8.
- Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell.* 2008;132(2):311–22.
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10(12):1213–8.
- Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 2007;17(6):877–85.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507(7493):455–61.
- Benton ML, Talipineni SC, Kostka D, Capra JA. Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function. *BMC Genomics.* 2019;20(1):511.
- Halfon MS. Studying transcriptional enhancers: the founder fallacy, validation creep, and other biases. *Trends Genet.* 2019;35(2):93–103.
- Gasperini M, Tome JM, Shendure J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet.* 2020;21(5):292–310.
- Rivera J, Keranen SVE, Gallo SM, Halfon MS. REDfly: the transcriptional regulatory element database for *Drosophila*. *Nucleic Acids Res.* 2018;47(D1):D828–34.
- Gao T, Qian J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res.* 2020;48(D1):D58–64.
- Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science.* 2013;339(6123):1074–7.
- Gao T, Zheng Z, Pan Y, Zhu C, Wei F, Yuan J, et al. scEnhancer: a single-cell enhancer resource with annotation across hundreds of tissue/cell types in three species. *Nucleic Acids Res.* 2022;50(D1):D371–9.
- Halfon MS. Silencers, Enhancers, and the Multifunctional Regulatory Genome. *Trends Genet.* 2020;36(3):149–51.
- Atkinson TJ, Halfon MS. Regulation of gene expression in the genomic context. *Comput Struct Biotechnol J.* 2014;9:e201401001.
- Zabidi MA, Stark A. Regulatory enhancer-core-promoter communication via transcription factors and cofactors. *Trends Genet.* 2016;32(12):801–14.
- Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res.* 2017;27(1):38–52.
- Small S, Arnosti DN. Transcriptional Enhancers in *Drosophila*. *Genetics.* 2020;216(1):1–26.
- Koenecke N, Johnston J, Gaertner B, Natarajan M, Zeitlinger J. Genome-wide identification of *Drosophila* dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biol.* 2016;17(1):196.
- Pradeepa MM, Grimes GR, Kumar Y, Olley G, Taylor GC, Schneider R, et al. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat Genet.* 2016;48(6):681–6.
- Taylor GC, Eskeland R, Hekimoglu-Balkan B, Pradeepa MM, Bickmore WA. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Res.* 2013;23(12):2053–65.
- Narita T, Higashijima Y, Kilic S, Liebner T, Walter J, Choudhary C. A unique H2B acetylation signature marks active enhancers and predicts their target genes. *bioRxiv.* 2022;07(18):500459. <https://doi.org/10.1101/2022.07.18.500459>.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002;12(6):996–1006.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.Learn more biomedcentral.com/submissions