

RESEARCH

Open Access



Comprehensive evaluation of diverse massively parallel reporter assays to functionally characterize human enhancers genome-wide

Junke Zhang^{1,2}, Alden King-Yung Leung², Yutong Zhu^{1,2}, Li Yao^{1,2}, Avery Willis², Xiuqi Pan^{2,3}, Abdullah Ozer³, Zhou Zhou³, Keith Siklenka⁵, Alejandro Barrera⁵, Jin Liang², Nathaniel D. Tippens^{1,2,3,4}, Timothy E. Reddy⁵, John T. Lis^{1,3,4*} and Haiyuan Yu^{1,2,4*}

*Correspondence:
johnlis@cornell.edu; haiyuan.yu@cornell.edu

¹ Department of Computational Biology, Cornell University, Ithaca, NY, USA

² Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY, USA

³ Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

⁴ Tri-Institutional Training Program in Computational Biology and Medicine, Cornell University, Ithaca, NY, USA

⁵ Department of Biostatistics and Bioinformatics, Duke University, Durham, USA

Abstract

Background: Massively parallel reporter assays (MPRAs) and self-transcribing active regulatory region sequencing (STARR-seq) have revolutionized enhancer characterization by enabling high-throughput functional assessment of regulatory sequences.

Results: Here, we systematically evaluate six MPRA and STARR-seq datasets generated in the human K562 cell line and find substantial inconsistencies in enhancer calls from different labs that are primarily due to technical variations in data processing and experimental workflows. To address these variations, we implement a uniform enhancer call pipeline, which significantly improve cross-assay agreement. While increasing sequence overlap thresholds enhanced concordance in STARR-seq assays, cross-assay consistency in LentiMPRA is strongly influenced by assay-specific factors. Functional validation using candidate cis-regulatory elements (cCREs) confirms that epigenomic features such as chromatin accessibility and histone modifications are strong predictors of enhancer activity. Importantly, our study validates transcription as a critical hallmark of active enhancers, demonstrating that highly transcribed regions exhibit significantly higher active rates across assays. Furthermore, we show that transcription enhances the predictive power of epigenomic features, enabling more accurate and refined enhancer annotation.

Conclusions: Our study provides a comprehensive framework for integrating different enhancer datasets and underscores the importance of accounting for assay-specific biases when interpreting enhancer activity. These findings refine enhancer identification using massively parallel reporter assays and improve the functional annotation of the human genome.



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

Background

Enhancers are key *cis*-regulatory DNA elements that drive transcriptional activity and play a pivotal role in gene regulation. Their influence extends beyond individual gene expression, shaping broader regulatory networks that control cell identity and function. Variants within enhancers have been strongly implicated in complex traits and diseases, emphasizing the importance of systematically identifying and characterizing enhancers to elucidate their contributions to gene expression and disease mechanisms [1–3].

Traditional reporter gene assays have long been used to characterize enhancer activity by positioning candidate sequences upstream or downstream of a minimal promoter linked to a reporter gene [4–6]. However, enhancers present a much greater challenge for functional characterization than the ~25,000 protein-coding genes in the human genome due to their vast numbers, sequence variability, and highly context-dependent activity [7, 8]. While these traditional reporter gene assays remain functional, the advent of high-throughput sequencing technologies has revolutionized enhancer studies, enabling massively parallel reporter assays (MPRAs) and self-transcribing active regulatory region sequencing (STARR-seq) to profile the regulatory activity of millions of sequences simultaneously [9–11]. These innovations have dramatically expanded our ability to interrogate enhancers on a genome-wide scale, addressing the limitations of conventional low-throughput approaches.

MPRAs, which utilize synthesized oligonucleotide libraries, position candidate sequences upstream of a minimal promoter and tag them with unique barcodes in the 3' or 5' UTR of the reporter gene. Regulatory activity is inferred by sequencing RNA transcripts associated with these barcodes [9, 10]. Despite their robustness, MPRAs face challenges in testing long DNA sequences and complex libraries due to synthesis and cost limitations [6, 12]. Additionally, placing candidate sequences upstream of a promoter may inadvertently capture promoter rather than enhancer activity, confounding the interpretation of regulatory function [13].

STARR-seq overcomes some of these constraints by placing candidate sequences within the 3' UTR of a reporter gene, allowing them to self-transcribe and directly quantify enhancer activity based on transcript abundance [11]. Unlike MPRAs, STARR-seq does not rely on synthesis but instead uses fragmented genomic DNA, typically obtained through sonication [14, 15] or nuclease digestion [16], enabling genome-wide enhancer screening without sequence-length restrictions [17]. However, STARR-seq also has inherent challenges. The placement of candidate sequences in the 3'UTR can affect mRNA stability, and thereby introduce orientation biases in enhancer quantification [18]. Furthermore, genome-wide STARR-seq requires highly complex libraries, necessitating deep sequencing and high transfection efficiency to achieve sufficient coverage [15]. Since random fragmentation rarely generates multiple identical copies of the same fragment, most fragments produce only a single readout. As a result, fragment-level analysis is not feasible, requiring the use of peak-calling algorithm to identify enhancer regions [19, 20]. In general, these approaches lack the necessary resolution to precisely delineate enhancer boundaries.

In recent years, several MPRA and STARR-seq variants have been developed to facilitate the genome-wide functional characterization of human enhancers and their sequence variants [14, 15, 21–25]. Among these efforts, the ENCODE Consortium

has played a pivotal role by implementing large-scale, high-throughput reporter assays within and across multiple cell lines to systematically map enhancer activity across the genome [26–28]. These efforts have generated extensive datasets that provide a valuable resource for dissecting the regulatory architecture of the genome.

However, several critical questions must be addressed to fully leverage these resources and refine the application of massively parallel reporter assays for deeper functional dissection of enhancer sequences. One key uncertainty is the extent to which the human genome has been functionally characterized. While STARR-seq has the theoretical capacity to screen enhancers genome-wide, practical limitations such as sequencing depth can significantly impact coverage. Additionally, the consistency of enhancer identification across different experimental platforms remains unclear. A recent study systematically compared nine different MPRA and STARR-seq assay designs using a fixed set of 2440 sequences, demonstrating how variations in experimental design influence enhancer activity measurements [29]. While this study provided valuable insights, it was conducted under controlled conditions rather than real-world applications, where assay-specific factors—such as library design, sequencing depth, and data-processing pipelines—may introduce systematic biases. Furthermore, the extent to which reporter assays yield consistent regulatory activity profiles and how functionally characterized enhancers align with annotations derived from epigenomic features—such as histone modifications, chromatin accessibility, and transcriptional activity—remain largely unexplored.

To fully integrate these existing large-scale reporter assay datasets for enhancer sequence and functional studies and optimize the future application of massively parallel reporter assays, a systematic evaluation of their genome-wide coverage, cross-assay consistency, and concordance with existing enhancer annotations is needed. Without such an assessment, leveraging these datasets for meaningful biological insights remains challenging, limiting our ability to accurately interpret regulatory landscapes and develop a unified framework for enhancer characterization.

In this study, we systematically evaluated a total of six STARR-seq and MPRA datasets representing four major MPRA and STARR-seq assay types obtained in the human K562 cell line. Initial comparisons of lab-reported enhancer calls revealed limited overlap, prompting a deeper investigation into the factors contributing to cross-assay inconsistencies. We reprocessed all datasets using a unified analytical framework, assessing dataset quality while implementing a standardized enhancer identification pipeline and improving cross-assay comparisons by recording both active and inactive regions. Using this harmonized approach, we found significantly improved enhancer call consistency across assays, especially in cases testing similar sequence composition. Furthermore, we assessed the functional relevance of enhancer candidates defined by enhancer RNA (eRNA) transcription start sites (TSSs) and defined by epigenomic profiles from the ENCODE registry of candidate cis-regulatory elements (cCREs), finding existing enhancer annotations are concordant with massively parallel reporter assay data. We also demonstrated transcription emerged as a critical mark of enhancer function, improving the predictive power of epigenomic features and enhancing the enhancer annotation. Our study provides the first comprehensive assessment of diverse massively parallel reporter assay datasets, offering a framework for integrating these datasets to

enhance biological insights and refine functional characterization strategies for future applications.

Results

Assessment of cross-assay consistency in enhancer identification

We analyzed six distinct STARR-seq and MPRA datasets produced by laboratories within the ENCODE Consortium's Functional Characterization Center, comprising three TilingMPRA datasets, a LentiMPRA dataset, an ATAC-STARR-seq, and a WHG-STARR-seq dataset [26–28]. Although all assays were performed in the human K562 cell line, they differed in experimental objectives, design strategies, and data processing methods. An overview of these experimental designs is illustrated in Fig. 1a, with detailed dataset descriptions provided in the Additional file 1.

To evaluate the consistency of enhancer identification, we compared enhancer calls reported in each dataset. Data were retrieved from either the ENCODE portal [26–28] or original publications and processed according to each laboratory's guidelines. The

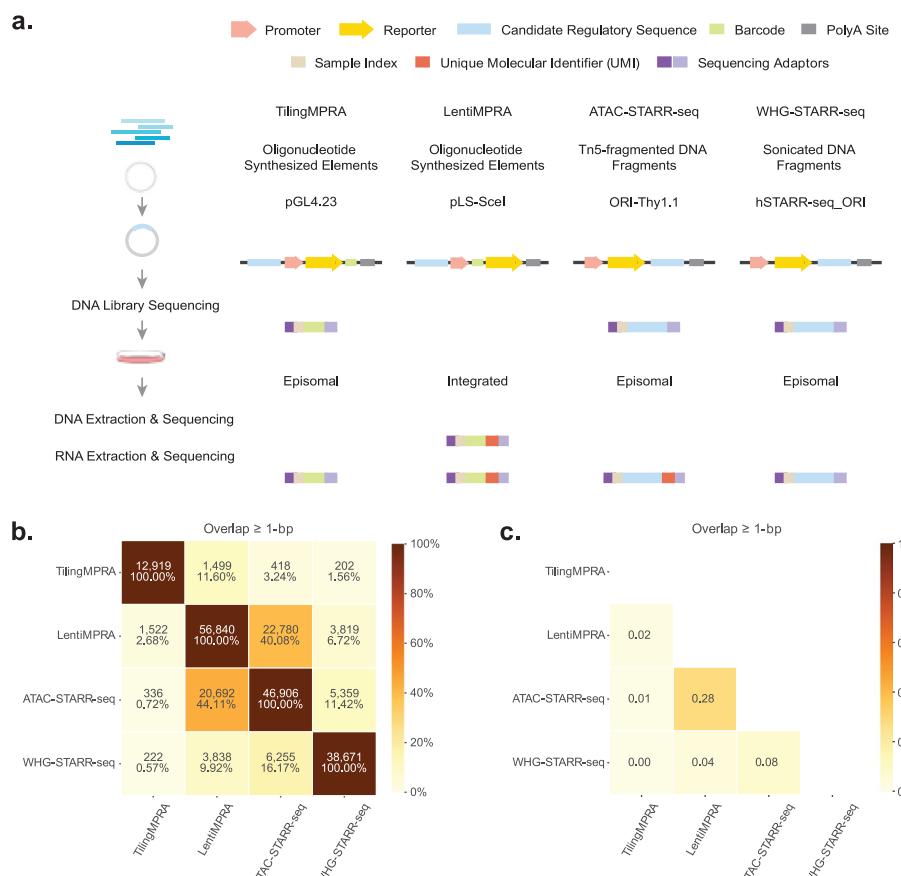


Fig. 1 Overview of experimental designs and assay consistencies across MPRA and STARR-seq assays. **a** Schematic representation of the experimental workflows for four types of MPRA and STARR-seq assays analyzed in this study. **b** Heatmap displaying the number of overlapping enhancer regions between assays and their percentage relative to the total number of enhancer regions identified in each assay, based on the $\geq 1\text{-bp}$ overlap criterion. **c** Heatmap presenting the Jaccard Index for pairwise comparisons between assays using $\geq 1\text{-bp}$ overlap criterion, quantifying overall similarity in enhancer identification

original number of lab-reported enhancer regions is summarized in Additional file 3: Table S1. To standardize comparisons, overlapping enhancer calls within each dataset were merged into unique regions, resulting in 12,919 enhancer regions across three TilingMPRA datasets, 56,840 regions from LentiMPRA, and 46,906 and 38,671 regions from ATAC-STARR-seq and WHG-STARR-seq, respectively.

We compared enhancer calls across assays by measuring the number of overlapping enhancer regions in each pairwise comparison, applying a minimal overlap threshold of 1 base pair (bp) to ensure inclusion of partially overlapping regions (Additional file 2: Fig. S1b). The highest overlap was observed between LentiMPRA and ATAC-STARR-seq, where approximately 40% (22,780 out of 56,840) of LentiMPRA regions overlapped with 44% (20,692 out of 46,906) of ATAC-STARR-seq regions. ATAC-STARR-seq and WHG-STARR-seq showed the second-highest overlap, with around 11% (5359 out of 46,906) of ATAC-STARR-seq regions overlapping with 16% (6255 out of 38,671) of WHG-STARR-seq regions. Comparisons involving LentiMPRA and WHG-STARR-seq, as well as TilingMPRA with other assays, exhibited lower overlap, reflecting differences in enhancer calls across these datasets (Fig. 1b).

To further quantify similarity across assays, we calculated the Jaccard Index (JI) for each pairwise comparison. Overall, enhancer identification exhibited low consistency, with most JI values approaching zero (Fig. 1c). The highest JI was observed between LentiMPRA and ATAC-STARR-seq (0.28), followed by ATAC-STARR-seq and WHG-STARR-seq (0.08). Applying stricter overlap criteria further reduced similarities (Additional file 2: Fig. S2a,b), highlighting the substantial variability in enhancer identification across different assays.

Unified processing of reporter assay datasets: initial data quality check

To better understand the limited consistency in enhancer identification observed across MPRA and STARR-seq datasets, we examined potential technical and biological contributors to assay variability, as detailed in the Additional file 1. While both biological and technical factors may contribute, a previous study [29] and our comparative analyses of assay design and data processing suggest that technical variation is the predominant driver of cross-assay discrepancies. These considerations underscore the importance of a standardized analytical framework that accounts for key technical differences to enable more accurate and equitable comparisons across assays. To this end, we implemented a unified processing strategy, beginning with a comprehensive quality assessment of all datasets. In addition to the factors discussed in the Additional file 1, the baseline quality of these datasets directly impacts assay consistency and reliability.

For TilingMPRA, LentiMPRA, count data were readily available through the ENCODE portal [26–28], or were re-processed with guidance from the original authors [30]. For ATAC-STARR-seq and WHG-STARR-seq datasets, we applied a unified genomic binning approach, creating 100-bp genomic bins with a 10-bp step size in both forward and reverse orientations (Additional file 2: Fig. S3a). Only fragments that fully covered each genomic bin were counted, allowing for orientation-independent enhancer identification and a more accurate assessment of genome-wide coverage of tested regions.

We assessed genome-wide coverage for the WHG-STARR-seq dataset and observed substantial library complexity, with over 96% of the human genome assayed after

processing using the genomic binning approach. However, a more detailed evaluation revealed that a notable subset of genomic bins exhibited low read depths in the DNA library (< 10; Methods) (Fig. 2a), raising concerns about limitations in sequencing depth and transfection efficiency. Because low-read-depth regions are typically excluded from downstream analyses, the reported genome-wide coverage likely overstates the extent of the genome that was effectively tested.

We also evaluated the coverage of accessible chromatin regions in the ATAC-STARR-seq and LentiMPRA datasets, where assayed fragments were either enriched in or selected from these regions. Both datasets demonstrated the ability to capture a substantial proportion of accessible regions with high read depths (Fig. 2b). Specifically, ATAC-STARR-seq achieved almost 100% of coverage of accessible regions characterized by ATAC-seq peaks, while LentiMPRA successfully covered 44% of DNase hypersensitive sites (DHSs) at higher read-depth threshold (≥ 10) in DNA libraries.

To assess reproducibility between replicates, we calculated Pearson correlations (ρ) for log-transformed counts per million (logCPM) of DNA and RNA counts, as well as $\log_2(\text{RNA}/\text{DNA})$ ratios, as these ratios represent the primary measurement

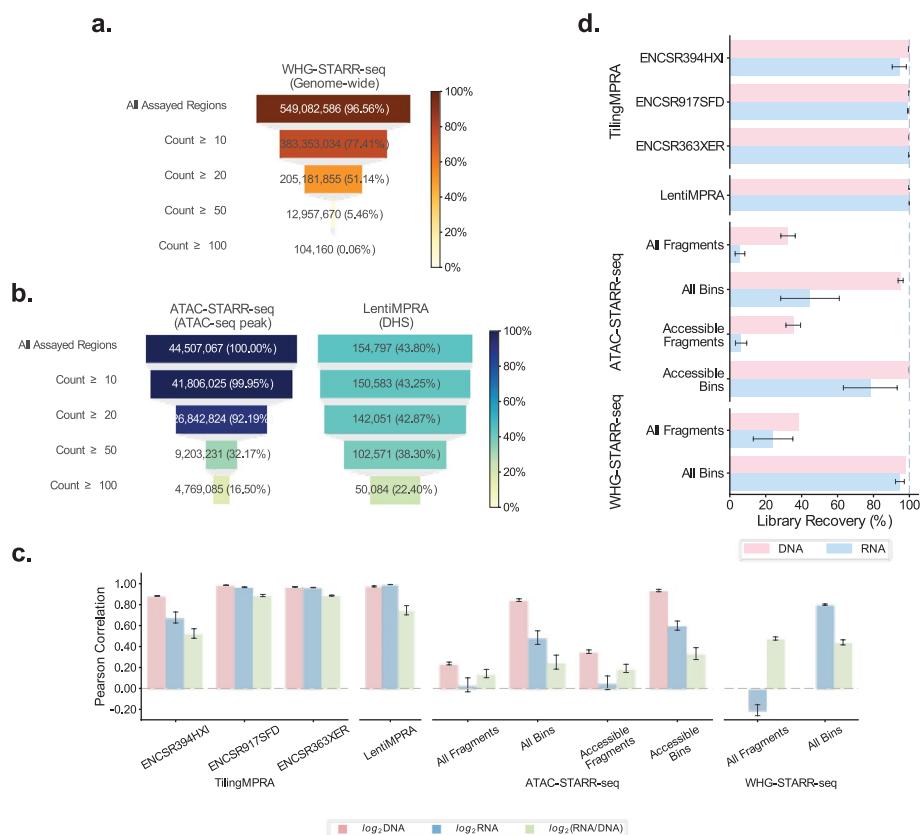


Fig. 2 Evaluation of data quality across MPRAs and STARR-seq assays. **a** Funnel plot showing the genome-wide coverage distribution of WHG-STARR-seq at varying read depths thresholds in DNA libraries. **b** Funnel plot illustrating the coverage distribution of accessible regions at different read depths thresholds in DNA libraries for ATAC-STARR-seq and LentiMPRA. Accessible regions are defined by ATAC-seq peaks from ATAC-STARR-seq DNA libraries and DNase-seq narrow peaks for LentiMPRA. **c** Bar plot presenting average Pearson correlation coefficients for \log_2 -transformed DNA CPM and RNA CPM, and $\log_2(\text{RNA}/\text{DNA})$ ratios across assays. **d** Bar plot depicting average library recovery rates in DNA and RNA libraries across assays.

of enhancer activity in downstream analyses. Overall, TilingMPRA and LentiMPRA demonstrated strong replicate correlations, indicating high reproducibility across libraries (Fig. 2c). Specifically, LentiMPRA showed robust correlations for both logCPM of DNA and RNA counts ($0.97 < \rho < 0.99$) and $\log_2(\text{RNA/DNA})$ ratios ($0.72 < \rho < 0.80$). Among the TilingMPRA datasets, ENCSR917SFD and ENCSR363XER displayed consistently high correlations ($0.96 < \rho < 0.99$ for logCPM, $0.87 < \rho < 0.90$ for $\log_2(\text{RNA/DNA})$), while ENCSR394HXI had moderately lower values ($0.62 < \rho < 0.89$ for logCPM, $0.47 < \rho < 0.58$ for $\log_2(\text{RNA/DNA})$), suggesting some variability within this dataset.

In contrast, ATAC-STARR-seq and WHG-STARR-seq demonstrated considerably lower fragment-level reproducibility (Fig. 2c). ATAC-STARR-seq showed weak agreement between replicates ($0.001 < \rho < 0.26$ for logCPM, $0.12 < \rho < 0.22$ for $\log_2(\text{RNA/DNA})$), while WHG-STARR-seq exhibited even greater variability, including negative RNA correlations (Fig. 2c). Aggregating fragments into genomic bins markedly improved replicate reproducibility for DNA and RNA counts and $\log_2(\text{RNA/DNA})$ ratios in the ATAC-STARR-seq dataset and RNA counts in WHG-STARR-seq dataset (Fig. 2c). Despite these improvements, the correlations for $\log_2(\text{RNA/DNA})$ ratios remained low in both datasets ($0.18 < \rho < 0.37$ for ATAC-STARR-seq, $0.42 < \rho < 0.47$ for WHG-STARR-seq). Further restricting analysis to accessible genomic bins in ATAC-STARR-seq provided marginal improvements but did not reach the high reproducibility observed in MPRA datasets, highlighting persistent variability in genome-wide STARR-seq measurements.

We also evaluated library recovery rates by calculating the proportion of fragments or genomic bins with at least one read in each library. TilingMPRA and LentiMPRA had high recovery rates (89–100%), whereas ATAC-STARR-seq exhibited an average library recovery rate below 40% in DNA libraries and even lower in RNA libraries (Fig. 2d). These findings suggest that many fragments were not consistently detected in ATAC-STARR-seq, possibly due to low sequencing depth or low transfection efficiency. Further analysis of fragments overlapping ATAC-seq peaks showed similar discrepancies in recovery rates between DNA and RNA libraries, pointing to limitations in data quality (Fig. 2d). WHG-STARR-seq also had low recovery rates at the fragment level (18–38%), but most genomic bins were represented in both DNA and RNA libraries (93–98%) (Fig. 2d), indicating that issues with sequencing depth and transfection efficiency were not as severe.

These results revealed substantial variability in data quality across different datasets. While MPRA assays exhibited consistently high data quality and reproducibility, genome-wide STARR-seq datasets were more susceptible to limitations such as insufficient sequencing depth and potential low transfection efficiency. These factors likely contributed to higher variability and reduced reliability in enhancer identification, and this issue can remain significant even when genomic binning is applied. Our findings highlight the necessity of applying stringent filtering criteria to exclude low-read-depth regions in the downstream analysis while also ensuring that the final reported tested regions accurately represent sequences with sufficient read depth, rather than using all assayed regions as a proxy for measuring tested region coverage.

Uniform processing of reporter assay datasets: enhancer call pipeline

While future studies should further address experimental challenges, to address the role of data processing in contributing to the observed inconsistencies, we implemented a unified enhancer call pipeline and applied it consistently across all datasets. The workflow is illustrated in Fig. 3a, with detailed methodology provided in the Methods section.

The pipeline begins with a raw count matrix as input and applies dataset-specific filters to remove fragments or genomic bins with low read depth. We then adapted the Trimmed Mean of M-values (TMM) normalization [31] and linear model approach from the Limma-Voom pipeline [32] to calculate $\log_2(\text{RNA/DNA})$ as a measure of regulatory activity for each fragment or genomic bin in each orientation. For targeted assays that included negative control sequences, we modified the original TMM normalization method to rely solely on negative controls for adjusting library size and composition bias. This approach provides greater accuracy in normalization, particularly for targeted assays where the assumption that most fragments lack regulatory effects may not hold.

After computing the $\log_2(\text{RNA/DNA})$ values, we assessed the regulatory activity of each fragment or genomic bin in both orientations by comparing it to the activity levels of negative controls through a Z-score analysis rather than relying on an arbitrary $\log_2(\text{RNA/DNA})$ cutoff. This comparison allowed for the identification of regions with significantly elevated activity relative to the basal transcription level defined by the negative controls in each orientation. To mitigate orientation bias, we incorporated regulatory activity in both orientations as a criterion for determining whether a fragment or genomic bin qualifies as a potential enhancer.

For genome-wide STARR-seq datasets that lacked negative controls in the original assays, we used genomic bins within exonic regions as surrogate negative controls, as enhancers are predominantly located in non-coding regions [26, 33, 34]. To ensure a clear distinction between potential enhancer regions and those likely to exhibit basal transcription, we excluded genomic bins overlapping the 300-bp flanking regions on either side of exons. This approach minimizes the risk of using genomic bins that may have counted fragments overlapping with enhancers in intronic regions, increasing the reliability of these surrogate negative controls.

Finally, our pipeline recorded both active and inactive regions identified in an orientation-independent manner, ensuring an accurate assessment of genome-wide coverage of tested regions. This comprehensive reporting approach also enables robust cross-assay comparisons. Detailed numbers of fragments or genomic bins tested in one or both orientations, the numbers of negative controls, and the numbers of enhancer regions identified are provided in the Additional file 3: Table S2.

Improved enhancer identification through unified enhancer call pipeline

We applied the uniform enhancer call pipeline to all datasets to standardize the identification of enhancer regions. In the ATAC-STARR-seq dataset, while all accessible regions characterized by ATAC-seq peaks were initially included in the assay, 91.20% were statistically tested for regulatory activity in at least one orientation (Fig. 3b). Furthermore, the effective coverage of regions tested in both orientations within accessible chromatin was reduced to 64.72% (Fig. 3b). Similarly, for the WHG-STARR-seq dataset, 96.61% of

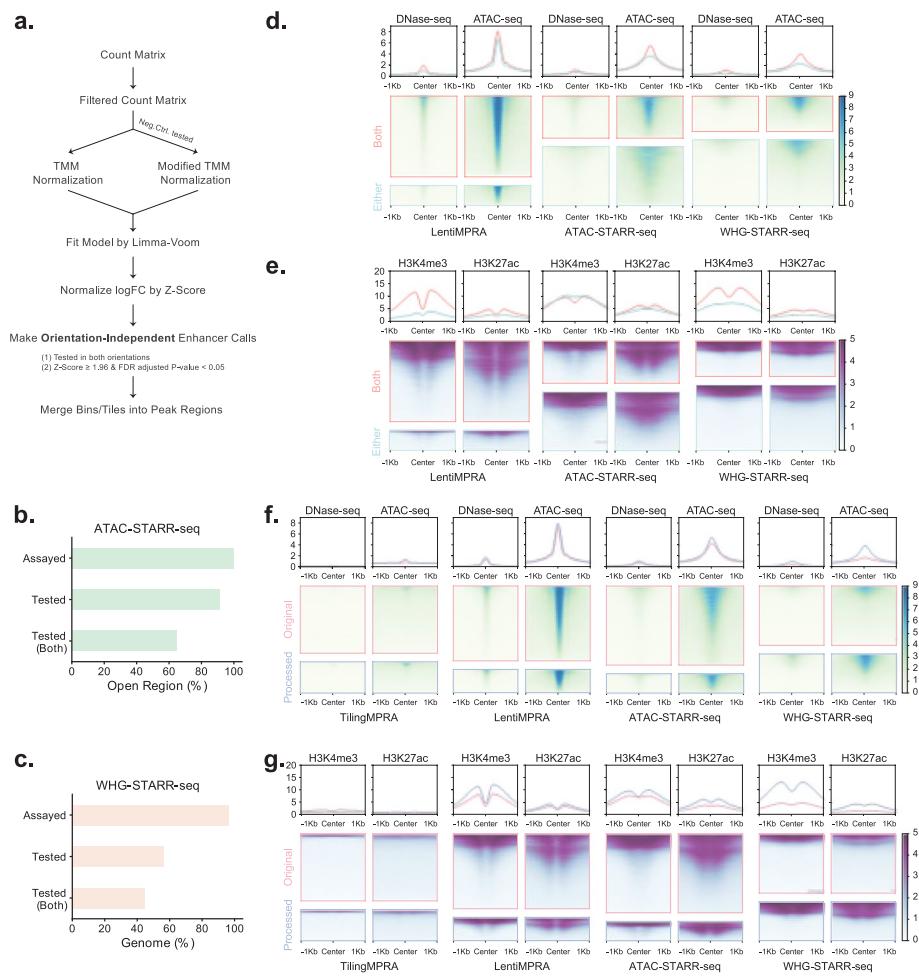


Fig. 3 Enhancer identification using a unified pipeline. **a** Schematic of the uniform enhancer call pipeline. The workflow begins with a raw count matrix as input, applies dataset-specific filters to exclude low-depth regions, and normalizes library size using TMM normalization. Regulatory activity is calculated as $\log_2(\text{RNA/DNA})$ values and Z-score analysis is performed to identify regions with significantly higher regulatory activity than negative control regions as enhancer regions in an orientation-independent manner. **b** Bar plot showing the assayed coverage, tested coverage in either orientation and tested coverage in both orientation for open chromatin regions characterized by ATAC-seq peaks derived from DNA libraries in ATAC-STARR-seq. **c** Bar plot summarizing genome-wide assayed coverage, tested coverage in either orientation and tested coverage in both orientations. **d** Meta-plots comparing the average DNase-seq and ATAC-seq signal profiles (± 1 kb from the center) for 2000 enhancer regions randomly sampled from those identified in both orientations versus 2000 regions tested in both orientations but active in only one orientation. **e** Meta-plots comparing the average H3K4me3 and H3K27ac histone modification profiles (± 1 kb from the center) for 2000 enhancer regions randomly sampled from those identified in both orientations versus 2000 regions tested in both orientations but active in only one orientation. **f** Meta-plots comparing the average DNase-seq and ATAC-seq signal profile (± 1 kb from the center) for 2000 randomly sampled enhancer regions from laboratory-reported enhancer calls versus those identified using the uniform enhancer call pipeline. **g** Meta-plots comparing the average H3K4me3 and H3K27ac histone profiles (± 1 kb from the center) for 2000 randomly sampled enhancer regions from laboratory-reported enhancer calls versus those identified using the uniform enhancer call pipeline.

the entire human genome was included in the assay; however, only 56.15% of regions were statistically assessed in at least one orientation, with just 44.59% tested in both orientations (Fig. 3c). These findings reveal that the effective coverage of genome-wide

STARR-seq datasets is significantly lower than expected, underscoring the importance of comprehensive reporting of tested regions to accurately evaluate assay performance and coverage.

Using our unified pipeline, we identified 57 enhancer regions in TilingMPRA (ENCSR394HXI), 26,874 in LentiMPRA, 11,507 in ATAC-STARR-seq, and 25,274 in WHG-STARR-seq. Notably, these enhancer regions exhibited significant regulatory activity in both orientations. For the two TilingMPRA datasets (ENCSR817SFD and ENCSR363XER), which tested elements exclusively in one orientation, we adapted our pipeline to perform orientation-dependent analysis, identifying 2117 enhancer regions in ENCSR817SFD and 3752 in ENCSR363XER.

To evaluate the significance of making orientation-independent enhancer calls, we investigated their epigenomic features by analyzing 2000-bp windows centered on these regions. Specifically, we compared the epigenomic features of orientation-independent enhancer regions to those of regions that were tested in both orientations but exhibited significant activity in only one, leveraging ENCODE datasets for DNase-seq, ATAC-seq, and ChIP-seq (H3K4me3 and H3K27ac) in the K562 cell line. Orientation-independent enhancers displayed higher chromatin accessibility, as indicated by stronger DNase-seq and ATAC-seq signal intensities compared to enhancers active in only one orientation across all datasets (Fig. 3d). Additionally, they exhibited greater enrichment of both promoter- and enhancer-associated histone modifications, with a more pronounced bimodal pattern around their centers (Fig. 3e). These findings suggest that orientation-independent enhancers are more robustly marked by epigenomic features characteristic of active regulatory elements and highlight the importance of making orientation-independent enhancer calls.

We also compared the enhancer regions identified through our unified processing pipeline with the original enhancer calls reported by each laboratory. Across all datasets, uniformly processed enhancer regions exhibited higher chromatin accessibility, as evidenced by stronger DNase-seq and ATAC-seq signals (Fig. 3f). Notably, while some enhancer calls from the unified pipeline were in inaccessible regions, they were still more enriched in accessible regions compared to original lab-reported peaks in the WHG-STARR-seq dataset (Fig. 3f). Additionally, histone modification profiles confirmed that orientation-independent enhancer regions identified by the unified pipeline were more strongly marked by H3K4me3 and H3K27ac compared to lab-reported enhancer regions (Fig. 3g). These results highlight the advantages of our unified pipeline in enhancing the confidence of enhancer identification and providing a more reliable foundation for comparative and functional studies.

Enhanced consistency across assay using uniform processed enhancer calls

With both active and inactive regions recorded through our uniform enhancer call pipeline, we reassessed assay consistency by evaluating how many enhancers identified in one assay were also identified as enhancers in others. To achieve this, we conducted pairwise comparisons by assessing the overlap between enhancer regions from one assay and all tested regions in another. Because our enhancer regions were defined

in an orientation-independent manner, inactive regions were also generated by merging elements or genomic bins tested in both orientations that lacked significant enhancer activity.

For each pairwise comparison, enhancer regions from assay A were evaluated against all tested regions in assay B, and vice versa, as overlaps were not necessarily symmetric. In cases where an enhancer region overlapped multiple tested regions in another assay, or multiple enhancer regions overlapped a single tested region, we assigned the best overlap based on the highest number of overlapping base pairs to minimize redundancy. We then calculated the JI and recorded both the number of enhancer regions that were also classified as enhancers in the other assay and the total number of enhancer regions tested. By restricting comparisons to commonly tested regions, this approach provided a more accurate and comprehensive assessment of cross-assay consistency.

Using the minimal overlap threshold (≥ 1 bp), we observed a statistically significant improvement in assay consistency, reflected by higher JI values (Fig. 4a; one-sided Wilcoxon paired test, $p=0.02$). Applying a stricter threshold of $\geq 50\%$ reciprocal overlap likewise yielded a significant increase in cross-assay consistency, with JI values substantially exceeding those based on lab-reported enhancer regions (one-sided Wilcoxon paired test, $p=0.001$). Together, these results demonstrate that adopting a uniform enhancer calling pipeline and employing refined comparison strategies enhances cross-assay consistency, underscoring the importance of standardized processing in functional characterization studies.

Assay-specific factors influence cross-assay consistency

While previous comparisons using lab-reported enhancer regions showed lower agreement across assays when a stricter overlap criterion ($\geq 50\%$ reciprocal overlap) was applied, comparisons using uniformly processed data demonstrated the opposite trend: most pairwise comparisons exhibited increased JI values under the stricter criterion compared to the ≥ 1 -bp threshold (Fig. 4a). For instance, when comparing LentiMPRA enhancers to tested regions in ATAC-STARR-seq and WHG-STARR-seq, the proportion of consistently active regions rose from 17 and 19% (using a ≥ 1 -bp threshold) to 74 and 78% (using $\geq 50\%$ reciprocal overlap), respectively. A similar pattern was observed in pairwise comparisons between ATAC-STARR-seq and WHG-STARR-seq (Fig. 4b,c), indicating that enhancer identification is more consistent when comparing sequences with greater overlap.

Despite the overall increase in consistency under more stringent overlap criteria, cross-assay agreement remained largely unchanged when comparing enhancer regions identified by ATAC-STARR-seq and WHG-STARR-seq to those tested in LentiMPRA, regardless of the overlap threshold (Fig. 4b,c). This suggests that assay-specific factors, rather than sequence overlap alone, dominate cross-assay agreement with LentiMPRA. Because LentiMPRA positions candidate sequences immediately upstream of a reporter gene, we suspected that its propensity to capture promoter activity, rather than enhancer activity, may explain this difference.

To test this, we stratified comparisons by TSS proximity, excluding TilingMPRA due to limited sample size. Tested regions were defined as proximal if $\geq 90\%$ of their sequence overlapped within 500 bp of a protein-coding TSS (GENCODE [35] annotation v45) and

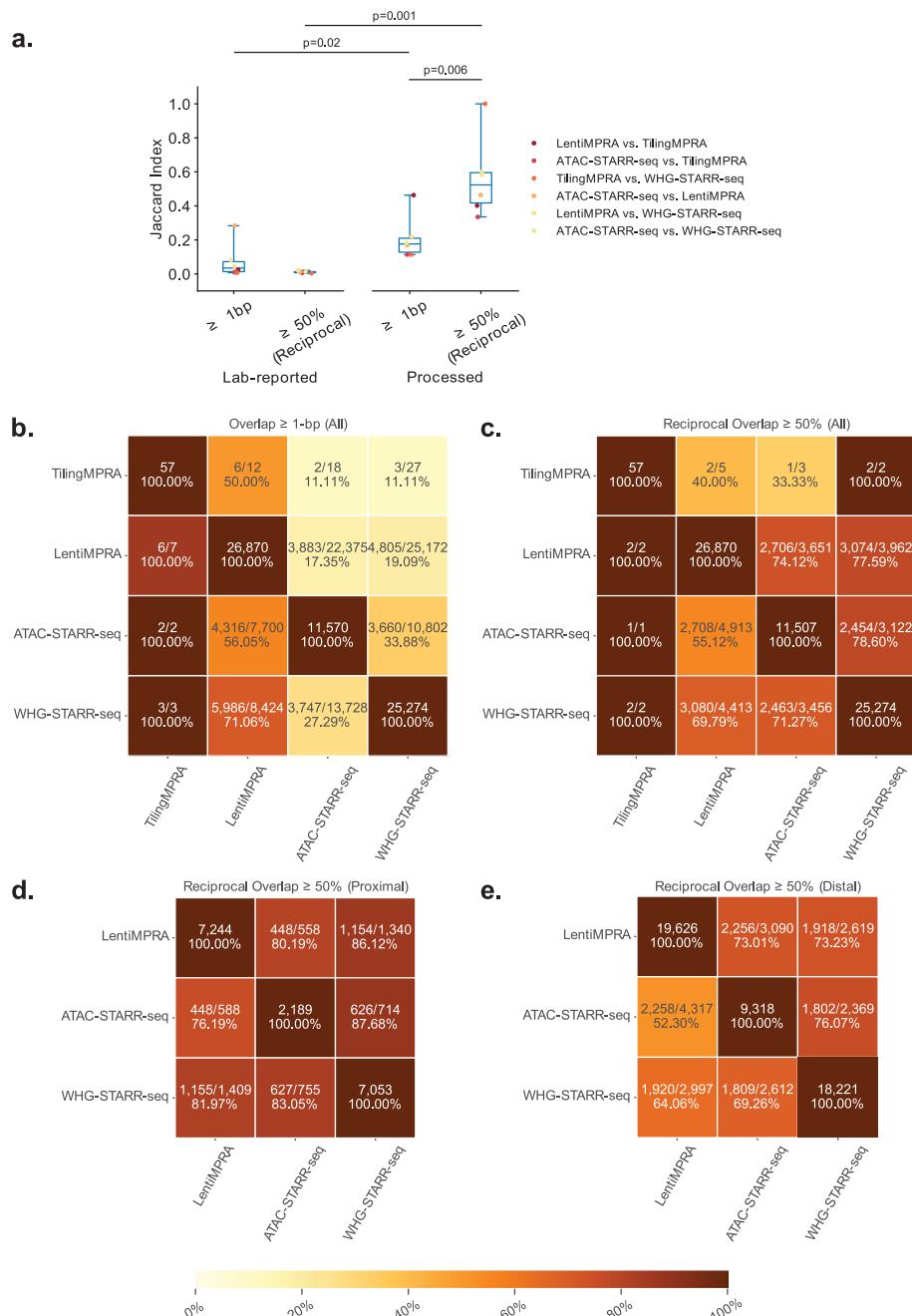


Fig. 4 Enhanced consistency in cross-assay comparisons using uniformly processed enhancer calls. **a** Box plot showing the Jaccard Index for pairwise comparisons between assays, calculated using the minimal overlap criterion of 1-bp and the stricter criterion of $\geq 50\%$ reciprocal overlap. Results are shown for both laboratory-reported and uniformly processed enhancer calls, illustrating the improved consistency achieved through uniform processing. **b,c** Heatmaps displaying the number of overlapping enhancer regions between assays under the $\geq 1\text{-bp}$ overlap criterion (**b**) and under the $\geq 50\%$ reciprocal overlap criterion (**c**). Each cell shows the ratio of the number of enhancer regions in the row dataset that overlap with enhancer regions in the column dataset to the number of enhancer regions in the row dataset overlapping with tested regions in the column dataset. Diagonal cells display the total number of enhancer regions identified in each dataset. **d,e** Heatmaps displaying the number of overlapping enhancer regions between assays under the $\geq 50\%$ reciprocal overlap criterion in proximal regions (**d**) and distal regions (**e**).

distal otherwise. Stratification revealed that ATAC-STARR-seq and WHG-STARR-seq showed significantly higher consistency with LentiMPRA in proximal regions than in distal regions. Specifically, using a $\geq 50\%$ reciprocal overlap threshold, $\sim 76\text{--}82\%$ of proximal enhancer regions identified by STARR-seq assays were also active in LentiMPRA, compared to only $\sim 52\text{--}64\%$ of distal regions (Fig. 4d,e). Similar trends were observed under the ≥ 1 bp overlap threshold (Additional file 2: Fig. S4a,b). These findings indicate that LentiMPRA aligns more closely with STARR-seq assays in detecting regulatory sequences in proximal regions. However, because MPRA-based assays may preferentially capture promoter rather than enhancer activity, further investigation is required to determine whether sequences active in both LentiMPRA and STARR-seq assays reflect enhancer-driven or promoter-driven regulation.

To test this, we assessed assay consistency separately in proximal and distal regions. TilingMPRA is excluded from this analysis due to limited sample size. Tested regions were classified as proximal if $\geq 90\%$ of their sequence overlapped within 500 bp of a protein-coding TSS (based on GENCODE [35] annotation v45) and distal otherwise. Stratifying comparisons by TSS proximity revealed that ATAC-STARR-seq and WHG-STARR-seq exhibited significantly higher consistency with LentiMPRA in proximal regions than in distal regions. Specifically, $\sim 62\text{--}73\%$ of proximal enhancer regions identified by STARR-seq assays were also active in LentiMPRA, whereas only $\sim 33\text{--}47\%$ of distal enhancer regions showed consistent activity (Fig. 4d,e and Additional file 2: Fig. S4b,c). Notably, these proportions differed only when comparing distal versus proximal regions but remained largely unchanged across different overlap thresholds (Fig. 4d,e). These findings suggest that LentiMPRA is more likely capturing promoter activity rather than enhancer activity as measured in genome-wide STARR-seq assays, emphasizing that assay-specific factors play a dominant role in determining cross-assay consistency when comparing to LentiMPRA.

Taken together, these results demonstrate that implementing a uniform enhancer calling pipeline effectively reduces technical variation, ensuring that cross-assay comparisons are not confounded by differences in data processing. The observed patterns of agreement or disagreement instead primarily reflect assay-specific biological factors intrinsic to each experimental design. This underscores the importance of standardized data processing for fair cross-assay evaluation, while also indicating that the remaining inconsistencies arise from biological properties of the assays rather than technical variation.

Evaluating functional support for enhancer-like and promoter like sequences in cCREs

Epigenomic features such as DNA accessibility and histone modifications have long been recognized as key indicators of active enhancers [5, 8, 34]. Leveraging these features, the ENCODE Consortium established a registry of cCREs [8]. To assess how well these elements are functionally validated by massively parallel reporter assays, we examined their coverage and activity in LentiMPRA, ATAC-STARR-seq, and WHG-STARR-seq datasets.

Since cCREs were not specifically designed as targeted sequences in these assays, we assessed their coverage by identifying overlaps between cCRE elements and tested regions. A cCRE was considered covered if it had at least a 1-bp overlap with

a tested region. To further characterize their representation across assays, we categorized covered cCREs into three mutually exclusive groups based on their overlap extent: high ($\geq 80\%$ reciprocal overlap), moderate (50–80% reciprocal overlap), and low (all other overlap). Detailed coverage statistics are provided in Additional file 2: Fig. S5a and Additional file 3: Table S3.

To evaluate the functional relevance of cCREs, we analyzed their active rates across LentiMPRA, ATAC-STARR-seq, and WHG-STARR-seq (Additional file 2: Fig. S5b). In both genome-wide STARR-seq datasets, cCREs associated with enhancer-like and promoter-like signatures—dELS, pELS, and PLS—demonstrated the highest active rates among all cCRE subtypes, whereas other cCRE categories exhibited lower active rates. Specifically, high-overlap dELS, pELS, and PLS each showed active rates ranging from 46 to 89% in ATAC-STARR-seq and WHG-STARR-seq (Additional file 2: Fig. S5b), highlighting their strong functional relevance in both genome-wide STARR-seq datasets. In contrast, the active rates of other cCRE subtypes declined sharply, with CA-H3K4me3 and CA-TF elements exhibiting moderate active rates (24–47%), followed by CA-CTCF and CA-only elements, which showed more limited active rates (5–9%). As expected, low-DNase elements, which are generally classified as inactive cCREs, displayed the lowest active rates (2–4%), only slightly higher than regions without any cCRE overlap (0.4–0.5%).

While the overall active-rate patterns were consistent across cCRE subtypes in genome-wide STARR-seq datasets, LentiMPRA displayed a distinct trend. PLS elements showed the highest active rate (57%), whereas dELS (24%) and pELS (22%) exhibited activity levels comparable to CA-H3K4me3 (30%), CA-TF (23%), and CA-only (31%). Moreover, low-DNase elements showed a 15% active rate in LentiMPRA, substantially higher than the 2–4% observed in ATAC-STARR-seq and WHG-STARR-seq, and regions without overlap with any cCREs also exhibited elevated activity (4%) compared to the minimal levels detected in STARR-seq assays (0.4–0.5%) (Additional file 2: Fig. S5b). Together, these observations highlight assay-specific biological factors shaping LentiMPRA activity profiles and suggest multiple, non-exclusive explanations: LentiMPRA may be more sensitive to promoter-associated activity, and its in-genome, integration-based readout likely introduces chromatin-context effects that differ from plasmid-based STARR-seq assays, potentially positioning certain sequences (e.g., low-DNase elements) into accessible chromatin environments and thereby inflating their apparent activity.

Collectively, these findings highlight the predictive power of cCREs in identifying active enhancers in reporter assays, particularly for dELS, pELS, and PLS, which exhibited significantly higher activity than other cCRE categories. The near absence of enhancer activity in regions lacking biochemical features underscores the essential role of chromatin accessibility and histone modifications in defining functional enhancers. At the same time, the distinct activity patterns observed in LentiMPRA, likely reflecting its preference for promoter-associated sequences and other assay-specific influences, emphasize the need to carefully consider assay-specific factors when interpreting results and integrating data from different massively parallel reporter assays. Further in-depth investigation will be required to dissect the

biological factors underlying the inconsistencies observed between LentiMPRA and STARR-seq assays.

Transcription as a critical mark of active enhancers

In addition to epigenomic features, enhancers are distinguished by their ability to generate eRNAs through divergent transcription [36, 37]. Tippens et al. demonstrated that divergent transcription serves as a more precise marker of active enhancers than histone modifications and identified a fundamental enhancer unit based on divergent transcription start sites (TSSs) [18]. Expanding on this, Yao et al. showed that GRO/PRO-cap is the most effective experimental approach to identify eRNAs and their divergent TSSs, and further compiled an enhancer compendium with a unified definition of enhancers based on divergent transcription [38].

Leveraging uniformly processed enhancer calls from large-scale reporter assays, we next examined these transcriptional characteristics of enhancers. Using the same analytical framework applied to cCREs, we assessed the coverage of GRO-cap enhancers [38] (divergent elements identified by PINTS from GRO-cap data) across the three assays. Detailed statistics are provided in Additional file 3: Table S4 and Additional file 2: Fig. S5d.

High-overlap GRO-cap enhancers exhibited strong enhancer activity, with 87 and 78% being active in ATAC-STARR-seq and WHG-STARR-seq, respectively (Additional file 2: Fig. S5e). Furthermore, GRO-cap enhancers consistently displayed significantly higher active rates compared to regions that neither overlapped with any GRO-cap elements nor exhibited GRO-cap signals (Additional file 2: Fig. S5e,f). Notably, while regions devoid of both transcriptional signals and overlap with GRO-cap elements exhibited the lowest active rates across all three assays (0.7–10%), regions that did not overlap with any annotated GRO-cap elements but still contained detectable GRO-cap signals showed slightly higher, albeit low, levels of activity (2–21%) (Additional file 2: Fig. S5f). These findings reinforce the strong functional relevance of GRO-cap enhancers in reporter assays, demonstrating that divergent transcription is a defining characteristic of active enhancers and supporting the enhancer architecture defined by previous studies [18, 38].

To further explore the functional relevance of transcriptional level, we categorized tested regions in LentiMPRA, ATAC-STARR-seq, and WHG-STARR-seq into four transcription-level classes (high, medium, low, and none) based on GRO-cap signals [39] (see Methods) and calculated the active rates within each category. Our analysis revealed a clear positive relationship: regions with higher transcription levels were significantly more likely to be classified as active across all three assays (Fig. 5a). Regions with no or low GRO-cap signals exhibited minimal active rates, particularly in ATAC-STARR-seq and WHG-STARR-seq, where values remained below 1%. Regions with medium transcription levels showed moderate active rates (6–24%), whereas highly transcribed regions reached the highest rates, with 31–50% of tested regions classified as active (Fig. 5a). These findings reinforce transcription level as a key predictor of enhancer activity across reporter assays.

Despite the low active rates observed in regions with little or no transcription, thousands of such regions were still identified as active enhancers across all three assays

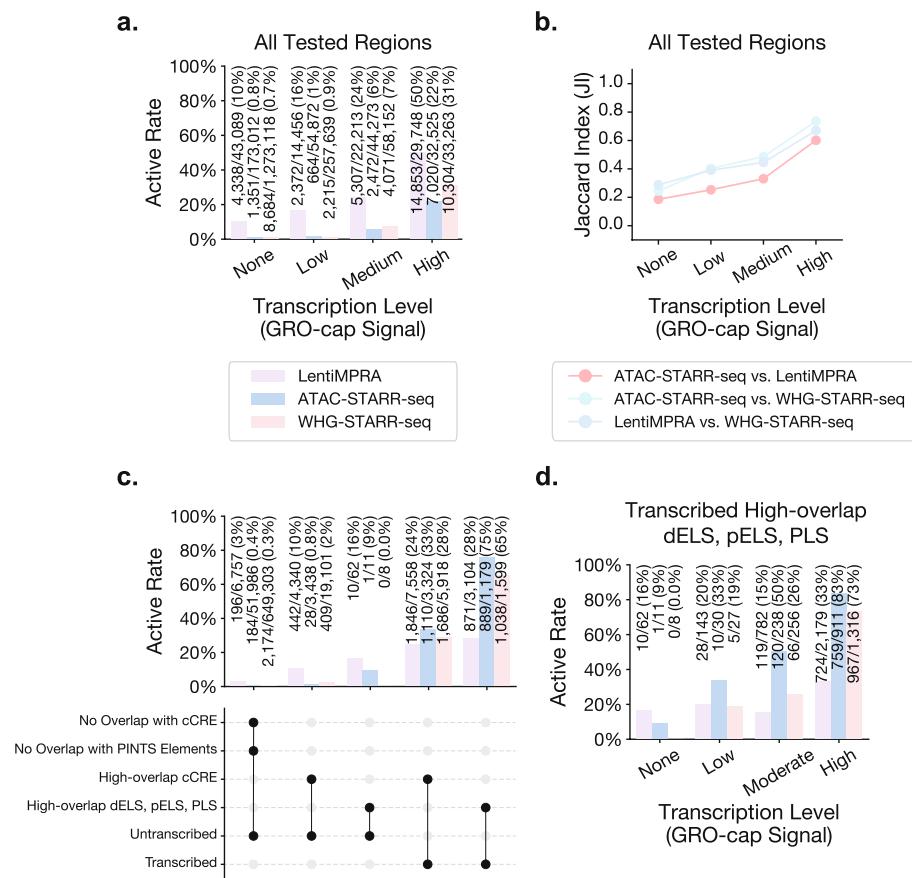


Fig. 5 Impact of transcription levels on active rates and assay consistencies. **a** Bar plot illustrating the active rates of all tested regions in LentiMPRA, ATAC-STARR-seq, and WHG-STARR-seq, categorized by transcription levels (none, low, medium, and high) determined by GRO-cap signals. **b** Line plot depicting Jaccard Index values for pairwise comparisons between LentiMPRA, ATAC-STARR-seq, and WHG-STARR-seq across all tested regions with varying transcription levels, calculated using the $\geq 50\%$ reciprocal overlap criterion. **c** Bar plot illustrating the active rate of transcribed and untranscribed regions with high-overlap with any types of cCREs or with high-overlap with dELS, pELS, and PLS or without any overlap with cCRE and PINTS elements. **d** Bar plot showing the active rate of high-overlap dELS, pELS, and PLS regions with different transcription levels (low, medium, high) determined by GRO-cap signals

(Fig. 5a). This raised concerns that a subset of these enhancer calls might represent false-positive hits. To explore this possibility, we examined assay consistency across transcription classes, hypothesizing that regions with lower transcription levels would exhibit reduced cross-assay agreement, suggesting a higher prevalence of false positives. Indeed, using $\geq 50\%$ reciprocal overlap as the comparison criterion, we observed a positive relationship between transcription levels and assay consistency (Fig. 5b, Additional file 2: Fig. S6). Regions lacking detectable transcription signals exhibited the lowest Jaccard Index values across all pairwise comparisons (Fig. 5b), indicating poor reproducibility across assays. Conversely, high-transcription regions exhibited the highest assay consistencies (Fig. 5b, Additional file 2: Fig. S6d). These results support the hypothesis that these reporter assays may yield a greater proportion of false positives in regions with lower transcription.

Transcription enhances the predictive power of biochemical features for enhancer activity

Next, we assessed whether transcription improves the ability of biochemical features to predict active enhancers. We analyzed tested cCREs with high overlap ($\geq 80\%$ reciprocal overlap) with reporter assay regions and classified them as either transcribed or untranscribed based on detectable GRO-cap signals. We then compared their active rates across assays.

Untranscribed cCREs exhibited low active rates in all three assays ($\sim 0.8\text{--}10\%$), with active rates only slightly higher than untranscribed regions that lacked cCRE or PINTS annotations ($\sim 0.3\text{--}3\%$) (Fig. 5c). Untranscribed dELS, pELS, and PLS showed slightly elevated active rates ($\sim 0\text{--}16\%$), though their sample sizes were limited.

In contrast, transcribed cCREs displayed significantly higher active rates across all assays ($\sim 24\text{--}33\%$) (Fig. 5c). This trend was particularly pronounced for transcribed dELS, pELS, and PLS, which exhibited much higher active rates ($\sim 28\text{--}75\%$) than their untranscribed counterparts (Fig. 5c). These results indicate that dELS, pELS, and PLS contain a higher proportion of functional enhancers than other cCRE categories and suggest that transcription serves as an additional predictive layer beyond traditional biochemical features such as chromatin accessibility and histone modifications (H3K4me3 and H3K27ac).

Further stratification of tested dELS, pELS, and PLS by transcription levels reinforced the strong relationship between transcription and enhancer activity across all assay types (Fig. 5d). Highly transcribed dELS, pELS, and PLS exhibited particularly high active rates, reaching 83% in ATAC-STARR-seq and 73% in WHG-STARR-seq (Fig. 5d). These findings emphasize transcription as a critical defining feature of active enhancers, complementing biochemical features and improving the precision of enhancer annotation.

Discussion

This study provides a comprehensive evaluation of a total of six reporter assay datasets generated by different laboratories, representing four major MPRA and STARR-seq assay types. Our analysis revealed substantial inconsistencies in enhancer identification across assays using original lab-reported enhancer calls, primarily driven by technical variations in experimental workflows and data processing methodologies. By applying a standardized analytical framework, we systematically assessed dataset quality, cross-assay consistency in enhancer identification, and the functional validation of enhancers based on epigenomic features and transcriptional features. Our findings highlight both the strengths and limitations of current high-throughput reporter assays in capturing enhancer activity and underscore the need for standardized experimental and analytical approaches in functional characterization studies.

Through re-processing and quality evaluation of all datasets, we identified insufficient fragment coverage, possibly stemming from inadequate sequencing depth and low transfection efficiency, as the critical limitation in genome-wide STARR-seq assays. These factors compromise not only the reproducibility of enhancer identification but also the effective coverage of tested genomic regions. Particularly in genome-wide assays, large proportions of the genome may remain untested or excluded due to low read depth, leading to an overestimation of genome-wide coverage. Addressing these technical

challenges is essential for improving the reliability and completeness of genome-wide enhancer screens.

To address technical discrepancies across assays, we developed and applied a uniform enhancer call pipeline designed to produce orientation-independent enhancer calls. This pipeline incorporated features such as normalization to negative controls, stringent statistical thresholds, and a requirement for enhancer activity in both orientations. Our results demonstrated that this unified approach successfully mitigated many sources of technical variation, yielding a more reliable and consistent set of enhancer regions across datasets. Significantly, our findings emphasized the critical role of orientation-independent analysis and the inclusion of negative controls in enhancing the reliability of enhancer identification. Testing fragments in both orientations and evaluating regulatory activities relative to negative controls proved essential for reducing technical biases.

We acknowledge that requiring significant regulatory activity in both orientations may reduce sensitivity by excluding regions with limited or missing data, as well as those showing enhancer potential in only one orientation. Some of these may represent potentially functional enhancers, rather than regions that are truly inactive. We applied this criterion intentionally to define the most confident set of enhancers, thereby minimizing technical variations and enabling fairer cross-assay comparisons that are more likely to reflect biological differences rather than technical variations. To complement this stringent set, we also reported enhancer regions identified by merging regions active in either orientation, which captures additional candidates, including those potentially overlooked by the strict orientation-independent approach, and provides a broader, more inclusive view of enhancer activity.

Moreover, the primary goal of our unified enhancer call pipeline was to address technical factors underlying inconsistencies in enhancer identification across assays, rather than to comprehensively optimize sensitivity and specificity for all applications. Future studies should aim to systematically evaluate the trade-offs between sensitivity and specificity in various enhancer call pipelines. Such efforts will be crucial for refining enhancer identification methodologies, particularly as functional characterization assays become increasingly diverse and complex.

Using uniformly processed enhancer calls, we conducted a comprehensive evaluation of cross-assay consistency and found improved agreement in enhancer identification across assays. Further analysis demonstrated that increasing sequence overlap thresholds substantially improved agreement, particularly in genome-wide STARR-seq datasets. However, LentiMPRA exhibited a distinct pattern, with its enhancer calls showing stronger agreement with STARR-seq assays in proximal regions, reinforcing its tendency to capture promoter-associated activity rather than distal enhancer activity. Additionally, LentiMPRA's random integration mechanism likely introduces variability by positioning sequences into different chromatin environments, which may either enhance or suppress activity depending on the local chromatin state. These findings emphasize the importance of considering assay-specific characteristics when integrating data from different reporter assays to ensure accurate interpretation of enhancer function.

By evaluating the functional relevance of candidate *cis*-regulatory elements (cCREs), we confirmed that epigenomic features, such as chromatin accessibility and histone modifications, serve as strong predictors of enhancer activity. cCREs associated with enhancer- and promoter-like signatures—dELS, pELS, and PLS—exhibited significantly higher active rates across genome-wide STARR-seq datasets compared to other cCRE subtypes, reinforcing their biological relevance. Conversely, elements lacking chromatin accessibility and histone modifications displayed minimal activity, underscoring the essential role of these epigenomic features in defining active enhancers.

LentiMPRA, however, displayed distinct activity patterns, with higher active rates for PLS and relatively lower activity for dELS and pELS compared to STARR-seq datasets. These differences suggest that LentiMPRA preferentially identifies promoter-driven regulatory elements rather than enhancers, further highlighting the need to consider assay-specific biases when interpreting MPRA data. Additionally, LentiMPRA showed unexpectedly higher active rates for low-DNase elements, possibly due to its random genomic integration placing these elements into more accessible chromatin regions, altering their apparent activity. These findings reinforce the need for careful interpretation of MPRA data.

Beyond epigenomic features, transcription emerged as a key determinant of enhancer function, with regions with higher transcription level displaying significantly higher activity across reporter assays. High transcription levels were strongly correlated with active rates of tested regions, whereas regions with low or no transcription exhibited greater cross-assay variability, suggesting a higher likelihood of false-positive enhancer calls. This highlights the importance of incorporating transcriptional markers to refine enhancer predictions and reduce misclassification. Furthermore, integrating transcriptional activity with epigenomic evidence improved enhancer annotation, as transcribed cCREs—particularly dELS, pELS, and PLS—showed significantly higher active rates than their untranscribed counterparts. These results suggest that transcription serves as an additional predictive layer beyond traditional chromatin features and should be considered when defining functional enhancers.

Conclusions

This study represents the first systematic evaluation of MPRA and STARR-seq datasets in real-world applications. By identifying critical technical factors and implementing a standardized analytical framework, we provide a foundation for improving experimental protocols and data processing methods in high-throughput reporter assays. Our uniform enhancer call pipeline offers a robust approach to enhancing data consistency and can serve as a benchmark for future studies.

The analytical framework established in this study can be extended to compare results across diverse functional characterization assays, such as CRISPR-based screens. Furthermore, the reliable sets of enhancer regions identified through this pipeline can be leveraged to investigate sequence features, enhancer-promoter interactions, and the structural basis of enhancer activity. Such analyses will deepen our understanding of enhancer biology and elucidate the mechanisms underlying assay-specific variability.

In summary, this study highlights the importance of standardization in enhancer characterization assays and demonstrates the value of integrating transcriptional and biochemical evidence for more accurate enhancer predictions. By addressing the technical and analytical challenges identified here, future studies can advance the functional characterization of human enhancers, ultimately improving our understanding of gene regulation and its implications for human health and disease.

Methods

Original reporter assay data acquisition and processing

Element quantification data for TilingMPRA datasets and peak regions from ATAC-STARR-seq and WHG-STARR-seq datasets were obtained from the ENCODE portal [26–28], with corresponding accession numbers listed in Additional file 3: Table S1. LentiMPRA quantification data and enhancer classifications were retrieved from its original publication [30] and are also accessible through the ENCODE portal [26–28].

To define enhancer regions in TilingMPRA datasets, we applied a threshold of \log_2 fold change ($\log_2\text{FC}$) ≥ 1 with an adjusted p -value < 0.01 . For the ATAC-STARR-seq dataset, regions with $\log_2\text{FC} > 0$ and an adjusted p -value < 0.01 were classified as enhancer regions. The total number of enhancer regions identified in each dataset, as well as the final numbers after merging overlapping regions, are summarized in Additional file 3: Table S1.

Cross-assay comparison of lab-reported enhancer regions

To assess the overlap between enhancer regions reported by different laboratories, we measured the fraction of enhancer regions in one assay that overlapped with enhancer regions identified in another. We performed pairwise comparisons across all datasets using two criteria: a minimal overlap threshold of 1 bp to maximize inclusion of partially overlapping regions (Additional file 2: Fig. S1b) and a $\geq 50\%$ reciprocal overlap threshold to provide a stricter assessment of enhancer reproducibility (Additional file 2: Fig. S1c). The number of overlapping enhancer regions was recorded for each pairwise comparison (Fig. 1b and Additional file 2: Fig. S2a).

Cross-assay comparison of uniformly processed enhancer regions

To systematically evaluate enhancer identification consistency across assays, we first distinguished enhancer regions from inactive regions in genome-wide STARR-seq datasets. Inactive regions were defined as genomic bins tested in both orientations that did not overlap with any orientation-independent enhancer regions, with overlapping bins merged to form continuous inactive regions.

For each pairwise comparison between assay A and assay B, we first identified orientation-independent enhancer regions in assay A that overlapped with tested regions in both orientations in assay B. The tested regions in assay B included both orientation-independent enhancer regions and inactive regions. We then quantified the proportion of enhancer regions in assay A that were not only tested but also identified as enhancers in assay B. This proportion was calculated as the number of enhancer

regions identified in both assays divided by the total number of enhancer regions in assay A that overlapped with tested regions in assay B.

We applied the same two overlap criteria to assess cross-assay consistency: ≥ 1 bp overlap for broad inclusion and $\geq 50\%$ reciprocal overlap for a more stringent evaluation. These comparisons were conducted across all datasets, and the results were reported in Heatmaps.

Jaccard index calculation

To quantitatively assess enhancer identification consistency across assays, we computed the Jaccard Index (JI) for each pairwise comparison. The Jaccard Index measures the similarity between two datasets, ranging from 0 to 1, with lower values indicating weaker agreement between assays.

The Jaccard Index for a given pair of assays, A and B, is defined as:

$$JI(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

For comparisons based on lab-reported enhancer regions, A and B represent the sets of enhancer regions identified in two different assays. Given that enhancer regions vary in size across datasets and that multiple enhancer regions in one dataset may overlap multiple regions in another, $|A \cap B|$ is defined as the maximum number of overlapping enhancer regions observed in either direction of comparison (A vs. B and B vs. A). $|A \cup B|$ represents the total number of unique enhancer regions across both assays.

For comparisons based on uniformly processed enhancer calls, A represents the set of orientation-independent enhancer regions in assay A that were also tested in both orientations in assay B, and B represents the corresponding set in assay B tested in both orientations in assay A. $|A \cap B|$ is also determined by the maximum number of overlapping enhancer regions across the two directional comparisons (A vs. B and B vs. A).

Reprocessing of genome-wide STARR-seq datasets

BAM files for ATAC-STARR-seq and WHG-STARR-seq datasets were retrieved from the ENCODE portal [26–28]. We adapted parts of the STARRPeaker [19] pipeline to process these BAM files and obtain original fragment counts for each library.

To obtain a refined set of original fragments and their corresponding raw counts, we applied a series of stringent filtering criteria. Unmapped, secondary, and chimeric alignments were discarded to retain only primary alignments. Reads with a mapping quality score below 10 were excluded to ensure high-confidence sequencing data. To mitigate potential biases from PCR amplification, reads with identical genomic coordinates were collapsed, a step applied to DNA replicates in ATAC-STARR-seq and across all WHG-STARR-seq libraries. For RNA libraries in ATAC-STARR-seq, PCR duplicates were removed using unique molecular identifiers (UMIs) to distinguish true biological duplicates from amplification artifacts.

Genomic bin count generation

To generate genomic bin counts, we used *pybedtools* [40, 41] to partition the human genome into 100-bp bins with a 10-bp step size. For each bin, we summed counts of fragments that fully covered the genomic bin (Additional file 2: Fig. S3a).

Quality assessment of reporter assay datasets: replicate reproducibility

To assess the reproducibility of replicates across datasets, we calculated Pearson correlation coefficients (ρ) for log-transformed counts per million (logCPM) of DNA and RNA counts, as well as log₂(RNA/DNA) ratios. Correlations were computed between biological replicates within DNA and RNA libraries for each dataset, and the results were averaged to provide an overall measure of replicate reproducibility.

For TilingMPRA and LentiMPRA datasets, replicate reproducibility was evaluated at the fragment level, where enhancer activity was quantified per tested sequence. In contrast, for genome-wide STARR-seq datasets, reproducibility was assessed at both the fragment level and the genomic bin level to account for the different resolution of data processing. Additionally, for the ATAC-STARR-seq dataset, we separately evaluated Pearson correlations in two conditions: across the entire genome and within accessible regions characterized by ATAC-seq peaks identified from DNA libraries in ATAC-STARR-seq.

Quality assessment of reporter assay datasets: library recovery rate

The library recovery rate was defined as the proportion of unique fragments detected in a given library relative to the total number of unique fragments identified across the entire dataset, encompassing all DNA and RNA libraries. A fragment was considered part of the dataset's total unique fragments if it was detected in at least one library, rather than requiring its presence in every library. This total serves as an estimate of the full set of input candidate fragments.

This metric provides insight into the reproducibility of fragment detection across replicates and carries slightly different implications for DNA and RNA libraries. In DNA libraries, higher recovery rates indicate greater consistency in library preparation and sufficient sequencing depth, whereas in RNA libraries, higher recovery rates reflect both efficient transfection and adequate sequencing depth. Conversely, lower DNA library recovery rates may suggest insufficient sequencing depth or stochastic loss of fragments during library preparation, while lower RNA library recovery rates could indicate transfection inefficiencies or suboptimal sequencing depth.

By evaluating library recovery rates for DNA and RNA libraries, we can better assess dataset quality, identifying potential technical limitations affecting dataset quality. The average library recovery rates were calculated for DNA and RNA libraries separately across assays and are presented in Fig. 2d.

Evaluation of assay coverage in the DNA libraries in genome-wide STARR-seq datasets

To assess potential limitations in sequencing depth within genome-wide STARR-seq datasets, we examined library complexity and genome-wide coverage by applying various read depth thresholds in DNA libraries. We imposed minimum raw count

thresholds of 10, 20, 50, and 100 across all DNA libraries to segment the datasets and evaluate library complexity and genome-wide coverage for each remaining subset of the datasets. These assessments were conducted after binning original fragments into genomic bins.

Since the DNA libraries of genome-wide STARR-seq assays were sequenced before transfection, they reflect the original fragment distribution across the genome. The representation of a region in the input DNA libraries plays a crucial role in determining its likelihood of being transfected and subsequently detected in the output RNA libraries. If a region was underrepresented in the input, it is less likely to have been thoroughly tested for enhancer activity in the RNA output.

We utilized *pybedtools* [40, 41] to quantify the genomic coverage by computing the number of base pairs covered at each threshold. The percentage of genome-wide coverage was determined by dividing the number of covered base pairs by the total number of base pairs in the hg38 human reference genome [42].

For ATAC-STARR-seq, which was specifically designed to enrich open chromatin regions rather than provide full genome coverage, we evaluated its coverage within open chromatin regions, as defined by ATAC-seq peaks from its DNA libraries. The open chromatin coverage of ATAC-STARR-seq was calculated by dividing the number of base pairs covered within ATAC-seq peaks by the total number of base pairs in these peaks.

To provide a comparison, we also assessed open chromatin coverage for LentiMPRA, as its candidate sequences were selected from DNase-seq peaks. DNase-seq peak regions were obtained from the ENCODE portal [26–28] (Accession: ENCFF185XRG). The open chromatin coverage for LentiMPRA was calculated using the same approach as in ATAC-STARR-seq, by determining the fraction of base pairs covered within DNase-seq peak regions.

Uniform enhancer call pipeline

To quantify enhancer activity based on $\log_2(\text{RNA}/\text{DNA})$ ratios, we adapted the Limma-Voom pipeline [43] with key modifications tailored to different datasets. While the linear model framework was retained, we implemented dataset-specific filtering strategies, a modified TMM normalization approach, and a Z-score-based classification method to identify enhancer regions in an orientation-independent manner.

Uniform enhancer call pipeline: dataset-specific filtering strategy

In general, filtering should be applied to remove bins or fragments with insufficient read depth, as these elements are unlikely to be reliably assayed. We recommend following the default implementation of the `filterByExpr` function in `edgeR` [44], which is also used in the well-established limma-voom pipeline. Filtering should primarily rely on DNA libraries, particularly when input DNA libraries are sequenced prior to transfection, because elements/bins with very low DNA abundance are less likely to be transfected into cells. As a practical guideline, a minimal raw count threshold of 10 can be used to calculate the corresponding logCPM cutoff across DNA libraries. We further recommend requiring fragments to pass the logCPM threshold in at least several libraries equal to the smaller of the DNA library count and the RNA library count. For MPRA datasets, which typically have lower library complexity and sufficient sequencing depth,

this threshold is generally sufficient to retain high-confidence elements. For genome-wide STARR-seq datasets, where library complexity is much higher and sequencing depth may be limiting, we recommend testing multiple thresholds, evaluating their effect on coverage, and ensuring that the mean–variance trend remains valid. In cases of limited sequencing depth, adaptive filtering strategies may be necessary to balance read depth and genome-wide coverage.

In this study, our filtering strategy was adapted from `filterByExpr()` in `edgeR` [44]. Raw counts were transformed into log-counts per million (logCPM) to normalize for library size differences, and filtering thresholds were determined by computing the logCPM equivalent of a predefined raw count cutoff. For TilingMPRA and LentiMPRA, the filtering threshold was determined using the average DNA library size, applying a raw count cutoff of 10 to establish the logCPM threshold, and we required fragments to pass this threshold in at least as many DNA libraries as the smaller of the DNA and RNA library counts. For ATAC-STARR-seq, to account for the larger library complexity while maintaining sufficient coverage, the threshold was calculated using the average DNA library size with a raw count cutoff of 20, and we required at least four DNA libraries to pass this logCPM threshold. For WHG-STARR-seq, only a single DNA library was available, and applying a strict DNA-based threshold would have severely reduced coverage. To avoid this, we applied a more lenient criterion, retaining fragments if they met the logCPM threshold (equivalent to a raw count of 10) in at least one library of any type (DNA or RNA). This dataset-specific adaptation ensured that filtering was stringent enough to remove low-depth elements while preserving sufficient coverage for reliable enhancer identification.

Uniform enhancer call pipeline: normalization strategy

To normalize for library size differences, we applied the Trimmed Mean of M-values (TMM) normalization from `edgeR` [31, 44], following the standard Limma-Voom framework. In ATAC-STARR-seq and WHG-STARR-seq, where most genomic bins were expected to exhibit no regulatory activity (i.e., showing no significant difference between RNA and DNA libraries), we applied the conventional TMM normalization method, assuming that the majority of regions had minimal transcriptional changes.

For LentiMPRA and TilingMPRA, we implemented a modified TMM normalization approach to address assay-specific biases. These assays included designated negative control sequences, and in the case of LentiMPRA, candidate sequences were particularly enriched for protein-coding promoters and potential enhancer elements [30]. This enrichment could result in a dataset disproportionately composed of regulatory-active fragments, making the standard assumption that most fragments were not differentially expressed less applicable. To address this, we modified TMM normalization to rely exclusively on negative control elements, allowing for a more accurate adjustment of library size and composition biases without being influenced by the overrepresentation of active regulatory elements. This refinement optimized normalization for the unique design of these assays, ensuring more reliable quantification of enhancer activity.

Uniform enhancer call pipeline: enhancer classification and statistical significance

Using the Limma-Voom procedure, $\log_2(\text{RNA/DNA})$ ratios were then calculated to quantify enhancer activity for each fragment or genomic bin. Rather than applying an arbitrary threshold, we employed a *Z*-score-based approach to identify regions with significantly elevated activity compared to background transcription levels. Background transcription levels were estimated using negative control elements, and the $\log_2(\text{RNA/DNA})$ threshold was set at the 95th percentile of negative control distributions.

For genome-wide STARR-seq datasets that lacked dedicated negative controls, genomic bins located within exonic regions were used as surrogate controls to determine the $\log_2(\text{RNA/DNA})$ threshold. This approach was based on previous studies indicating that enhancers are primarily located in non-coding regions [26, 33, 34]. To further refine enhancer identification and mitigate orientation bias, enhancer regions were required to exhibit significant activity in both forward and reverse orientations. First, fragments or genomic bins tested in both orientations were identified. Regions were then classified as enhancers if their $\log_2(\text{RNA/DNA})$ ratios exceeded the threshold and had an adjusted *p*-value < 0.05 . Overlapping fragments or genomic bins were merged to generate the final set of enhancer regions, ensuring a robust and unbiased identification process.

Uniform enhancer call pipeline: comprehensive reporting and dataset summary

The pipeline provided a comprehensive reporting of both active and inactive regions, ensuring accurate estimation of genome-wide coverage and facilitating robust cross-assay comparisons.

For genome-wide STARR-seq datasets, we reported multiple levels of coverage to reflect the extent of assay representation. *Assayed coverage* included all genomic bins before applying filters, representing the initial set of regions targeted in the experiment. *Tested coverage* encompassed genomic bins that remained after applying filtering criteria, reflecting regions with sufficient read depth for reliable enhancer activity quantification. Additionally, *tested coverage in both orientations* was defined as the subset of tested genomic bins that were assayed in both forward and reverse orientations, ensuring a stringent assessment of enhancer activity independent of strand bias.

The final dataset included $\log_2(\text{RNA/DNA})$ ratios, *Z*-scores, and statistical significance metrics for all tested genomic bins, as well as for bins tested in both orientations. Additionally, merged orientation-independent enhancer regions were reported to provide a set of enhancer calls across datasets. A summary of enhancer region counts, negative controls, and tested regions for each dataset is provided in Additional file 3: Table S2.

Evaluation of genomic context to compare regions

To examine the genomic context of enhancer regions, including DNA accessibility and histone modifications (H3K4me3 and H3K27ac), we compared lab-reported enhancer regions with uniformly processed enhancer regions and orientation-independent enhancers with regions tested in both orientations but identified as active in only one orientation. We utilized publicly available datasets from the ENCODE portal [26–28], specifically DNase-seq data (ENCFF972GVB), ATAC-seq data (ENCFF102ARJ), H3K4me3 ChIP-seq data (ENCFF911JVK), and H3K27ac ChIP-seq data (ENCFF381NDD).

For visualization, we used deepTools [42] to generate metaplots of signal intensities across enhancer regions. Signal intensities were plotted within a 2000-bp window centered at the region midpoint to capture local epigenomics features.

For the comparison between orientation-independent enhancer regions and those tested in both orientations but were active in only one orientation, we first selected genomic bins that were tested in both orientations. We then identified bins that were active in only one orientation, merged overlapping bins into contiguous regions, and excluded any regions that overlapped with orientation-independent enhancers.

Negative control regions in genome-wide STARR-seq datasets

Since the original ATAC-STARR-seq and WHG-STARR-seq assays did not include dedicated negative controls, we leveraged their genome-wide coverage to define genomic bins overlapping exonic regions as surrogate negative controls. This allowed for the implementation of a Z-score approach in these datasets to establish a background transcription level for enhancer classification.

To identify suitable genomic bins as negative controls, we extracted all exons of protein-coding genes from the GENCODE v42 annotation (hg38) [45]. Because the original fragments covering the 100-bp genomic bins could be substantially longer, we further excluded 300-bp flanking regions on both sides of each exon to prevent potential overlap with adjacent intronic regions, which could confound the regulatory activity measurements of exonic genomic bins. This filtering step ensured that only mid-exonic regions were retained as the final negative control reference regions.

We then identified all genomic bins that were fully contained within these negative control reference regions. These bins were used exclusively in the Z-score approach to characterize background transcription levels but were not used in the TMM normalization process.

Analysis of coverage and active rate of cCREs and GRO-cap enhancers

The comprehensive reporting of both active and inactive regions in genome-wide STARR-seq datasets enabled a systematic evaluation of the coverage and active rates of alternative enhancer annotations, such as cCREs and GRO-cap enhancers, across assays.

We retrieved cCRE annotations for K562 cells (Accession: ENCFF286VQG) from the ENCODE Portal [26–28]. GRO-cap enhancers were defined as divergent elements identified by PINTS [38].

To determine the extent to which these enhancer annotations were tested in LentiMPRA, ATAC-STARR-seq, and WHG-STARR-seq, we applied three mutually exclusive overlap categories: (1) high-overlap ($\geq 80\%$ reciprocal overlap), (2) moderate-overlap (50–80% reciprocal overlap), and (3) low-overlap (all other overlaps). The number of cCREs and GRO-cap enhancers tested in each assay is provided in Additional file 3: Table S3 and Additional file 3: Table S4.

To assess the functional relevance of these tested elements, we calculated their active rates within each overlap category. The active rate for each element type was defined as the proportion of tested elements that exhibited significant regulatory activity.

Annotation of transcription levels for tested regions using GRO-cap signals

To compare the active rates of tested regions with different transcription levels, we annotated each tested region with transcription levels based on GRO-cap signal data extracted from bigWig files [39].

To quantify the transcriptional activity within each tested region, we summed the GRO-cap signal from both orientations and normalized it by the region size. The normalized transcription level for each region was computed as the total GRO-cap signal divided by the length of the tested region.

Based on the normalized GRO-cap signal, we classified transcription levels into four categories: (1) None, for regions with no detectable GRO-cap signal; (2) Low, for regions with normalized GRO-cap signal ≤ 0.01 ; (3) Medium, for regions with normalized GRO-cap signal > 0.01 and ≤ 0.08 ; and (4) High, for regions with normalized GRO-cap signal > 0.08 .

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-025-03828-8>.

Additional file 1. Supplementary Information. Detailed descriptions of all MPRA and STARR-seq datasets utilized in this study, including experimental designs, analysis pipelines, and technical and biological factors contributing to the observed cross-assay inconsistencies based on the lab-reported enhancer calls.

Additional file 2. Supplementary figures S1–S6. Contains all supplementary figures referenced in the manuscript, including schematics of overlap criteria, cross-assay consistency heatmaps, binning strategies, distal vs. proximal region comparisons, functional validation of cCREs and GRO-cap elements, and transcription-stratified enhancer comparisons. Legends for all supplementary figures are provided within this file.

Additional file 3. Supplementary tables S1–S5. Compiled multi-tab Excel file containing all supplementary tables referenced in the manuscript. Tables include processed dataset summaries, enhancer call summaries, tested coverage statistics of cCREs and GRO-cap enhancers, and list of ENCODE datasets used in this study.

Acknowledgements

We thank members of the Yu and Lis laboratories for insightful discussions and feedback throughout this study. We are also grateful to the ENCODE Consortium for generating and sharing high-quality datasets that were critical to this work. We thank Tewhey, H. Dewey, V. Agarwal, A. Du, and D. Lee, for their valuable discussions.

Peer review information

Claudia Feng was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions

J.Z. conceptualized the study with guidance from H.Y., J.T.L., N.T., and J.L. J.Z. developed the methodologies, performed the formal analysis, and generated the figures. L.Y., A.L., Y.Z., A.W., X.P., K.S., A.B., and T.E.R. contributed to data curation and testing the pipeline code. J.Z. wrote the manuscript with guidance from H.Y. and J.T.L. A.O., Z.Z., L.Y., and K.S. reviewed and edited the manuscript. H.Y. and J.T.L. oversaw the project and acquired funding. All authors read and approved the final manuscript.

Funding

This work was funded by the National Institutes of Health grants UM1HG009393, R01 HG012970, and R01 AG077899 to J.T.L. and H.Y., as well as R01 DK127778 and R01 DK115398 to H.Y.

Data availability

This study made use of publicly available datasets from the ENCODE Consortium, which were accessed through the ENCODE portal [26–28]. The specific datasets are listed with accession identifiers in Additional file 3: Table S5. This study also used GRO-cap dataset from Core, L.J. et al. [39, 46], and GRO-cap enhancer calls from Yao, L. et al. [38]. All data generated or analyzed during this study are included in this published article and its supplementary information files. All processed data files can be accessed at Zenodo: <https://doi.org/10.5281/zenodo.1688276> [47] under the Creative Commons Attribution 4.0 International License. The source code used in this study is publicly available under the Apache License 2.0 (OSI Approved License) at Github: https://github.com/haiyuan-yu-lab/Analysis_Code_Comprehensive_Evaluation_MPRA_and_STARR_seq [48]. The code is also archived with a DOI at Zenodo: <https://doi.org/10.5281/zenodo.17254278> [49] under the Apache License version 2.0 (OSI Approved License).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 May 2025 Accepted: 8 October 2025

Published online: 03 November 2025

References

1. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190–5.
2. Kircher M, Xiong C, Martin B, Schubach M, Inoue F, Bell RJA, et al. Saturation mutagenesis of twenty disease-associated regulatory elements at single base-pair resolution. *Nat Commun*. 2019;10:3583.
3. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature*. 2021;593:238–43.
4. Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*. 1981;27:299–308.
5. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet*. 2020;21:71–87.
6. Inoue F, Ahituv N. Decoding enhancers using massively parallel reporter assays. *Genomics*. 2015;106:159–64.
7. 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;507:455–61.
8. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583:699–710.
9. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, et al. Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nat Biotechnol*. 2012;30:265–70.
10. Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol*. 2012;30:271–7.
11. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013;339:1074–7.
12. Santiago-Algarra D, Dao LTM, Pradel L, España A, Spicuglia S. Recent advances in high-throughput approaches to dissect enhancer function. *F1000Res*. 2017;6:939.
13. Nguyen TA, Jones RD, Snavely AR, Pfenning AR, Kirchner R, Hemberg M, et al. High-throughput functional comparison of promoter and enhancer activities. *Genome Res*. 2016;26:1023–33.
14. Vanhille L, Griffon A, Maqbool MA, Zácarias-Cabeza J, Dao LTM, Fernandez N, et al. High-throughput and quantitative assessment of enhancer activity in mammals by capstarr-seq. *Nat Commun*. 2015;6:6905.
15. Liu Y, Yu S, Dhiman VK, Brunetti T, Eckart H, White KP. Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol*. 2017;18:219.
16. Wang X, He L, Goggin SM, Saadat A, Wang L, Sinnott-Armstrong N, et al. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat Commun*. 2018;9:5380.
17. Muerdter F, Boryń ŁM, Arnold CD. STARR-seq - principles and applications. *Genomics*. 2015;106:145–50.
18. Tippens ND, Liang J, Leung AK-Y, Wierbowski SD, Ozer A, Booth JG, et al. Transcription imparts architecture, function and logic to enhancer units. *Nat Genet*. 2020;52:1067–75.
19. Lee D, Shi M, Moran J, Wall M, Zhang J, Liu J, et al. Starrpeak: uniform processing and accurate identification of STARR-seq active regions. *Genome Biol*. 2020;21:298.
20. Kim Y-S, Johnson GD, Seo J, Barrera A, Cowart TN, Majoros WH, et al. Correcting signal biases and detecting regulatory elements in STARR-seq data. *Genome Res*. 2021;31:877–89.
21. Kwasnieski JC, Fiore C, Chaudhari HG, Cohen BA. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res*. 2014;24:1595–602.
22. Ernst J, Melnikov A, Zhang X, Wang L, Rogov P, Mikkelsen TS, et al. Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat Biotechnol*. 2016;34:1180–90.
23. 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;27:38–52.
24. Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, et al. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*. 2018;172:1132–4.
25. Hansen TJ, Hodges E. ATAC-STARR-seq reveals transcription factor-bound activators and silencers within chromatin-accessible regions of the human genome. *Genome Res*. 2022;32:1529–41.
26. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
27. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res*. 2020;48:D882–9.
28. Hitz BC, Lee J-W, Jolanki O, Kagda MS, Graham K, Sud P, et al. The ENCODE uniform analysis pipelines. *Res Sq*. 2023; <https://doi.org/10.21203/rs.3.rs-3111932/v1>

29. Klein JC, Agarwal V, Inoue F, Keith A, Martin B, Kircher M, et al. A systematic evaluation of the design and context dependencies of massively parallel reporter assays. *Nat Methods*. 2020;17:1083–91.
30. Agarwal V, Inoue F, Schubach M, Penzar D, Martin BK, Dash PM, et al. Massively parallel characterization of transcriptional regulatory elements. *Nature*. 2025. <https://doi.org/10.1038/s41586-024-08430-9>.
31. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;11:R25.
32. Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15:R29.
33. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*. 2010;28(8):817–25.
34. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A*. 2014;111:6131–8.
35. Mudge JM, Carbonell-Sala S, Diekhans M, Martinez JG, Hunt T, Jungreis I, et al. GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res*. 2025;53:D966–75.
36. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 2010;465:182–7.
37. De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol*. 2010;8:e1000384.
38. Yao L, Liang J, Ozer A, Leung AK-Y, Lis JT, Yu H. A comparison of experimental assays and analytical methods for genome-wide identification of active enhancers. *Nat Biotechnol*. 2022;40:1056–65.
39. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet*. 2014;46:1311–20.
40. Quinlan AR, Hall IM. BEDtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
41. Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics*. 2011;27:3423–4.
42. Ramírez F, Ryan DP, Grünig B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44:W160–5.
43. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
44. Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res*. 2025;53. <https://doi.org/10.1093/nar/gkaf018>
45. Frankish A, Carbonell-Sala S, Diekhans M, Jungreis I, Loveland JE, Mudge JM, et al. GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res*. 2023;51:D942–9.
46. Leighton J, Core, André L. Martins, Charles G. Danko, Colin T. Waters, Adam Siepel, John T. Lis. Analysis of transcription start sites from nascent RNA identifies a unified architecture of initiation at mammalian promoters and enhancers. *Gene Expression Omnibus*; 2014. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60456>
47. Junke Z, Alden King-Yung L, Yutong Z, Li Y, Avery W, Xiuqi P, Abdullah O, Zhou Z, Keith S, Alejandro B, Jin L, Nathaniel D. Tippens, Timothy E. R, John T. Lis, Haiyuan Yu. Comprehensive evaluation of diverse massively parallel reporter assays to functionally characterize human enhancers genome-wide. Datasets. Zenodo; 2025. <http://dx.doi.org/10.5281/zenodo.1688727>
48. Junke Z, Alden King-Yung L, Yutong Z, Li Y, Avery W, Xiuqi P, Abdullah O, Zhou Z, Keith S, Alejandro B, Jin L, Nathaniel D. T, Timothy E. R, John T. Lis, Haiyuan Yu. Comprehensive evaluation of diverse massively parallel reporter assays to functionally characterize human enhancers genome-wide. Github; 2025. https://github.com/haiyuan-yu-lab/Analysis_Code_Comprehensive_Evaluation_MPRA_and_STARR_seq
49. Junke Z, Alden King-Yung L, Yutong Z, Li Y, Avery W, Xiuqi P, Abdullah O, Zhou Z, Keith S, Alejandro B, Jin L, Nathaniel D. T, Timothy E. R, John T. Lis, Haiyuan Yu. Comprehensive evaluation of diverse massively parallel reporter assays to functionally characterize human enhancers genome-wide. Zenodo; 2025. <http://dx.doi.org/10.5281/zenodo.1725427>

Publisher's Note

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