Summarizing internal dynamics boosts differential analysis and functional interpretation of super enhancers

Xiang Liu[1], Bo Zhao[2], Timothy Shaw[1], Brooke Fridley[1], Derek Duckett[3], Aik-Choon Tan[1], and Mingxiang Teng[1]*

[1]Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA

[2]Division of Infectious Disease, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[3]Department of Drug Discovery, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA

*To whom correspondence should be addressed. Email: mingxiang.teng@moffitt.org

ABSTRACT

Super enhancers (SEs) are broad enhancer domains usually containing multiple constitute enhancers with significantly elevated activities. The constitute enhancers work together through chromatin looping to build up distinct regulatory properties of SEs. Aberrant SE activities, which are critical to understand disease mechanisms, could be raised by the alterations of one or more of their constitute enhancers. However, the state-of-art binary strategy in calling differential SEs only relies on overall activity changes, neglecting the local differences of constitute enhancers within SEs. We propose a computational method to identify differential SEs by accounting for the combinatorial effects of constitute enhancers weighted with their activities and locations (internal dynamics). In addition to overall changes, our method finds four novel types of differential SEs pointing to the structural differences within SEs. When applied to public datasets for six cancer cells, we demonstrate that different types of differential SEs complement each other with distinct sets of gene targets and varied degrees of regulatory impacts. More importantly, we found that some cell-specific genes are linked to SE structural differences specifically, suggesting improved sensitivity by our methods in identifying and interpreting differential SEs. Such improvements further lead to increased discernment of cell identifies.

INTRODUCTION

Super enhancers (SEs) were proposed as broad regulatory domains on genome, usually spanning a minimum of thousands of base pairs and consisting of multiple constitute enhancers(1). The constitute enhancers work together as a unit, instead of separately, to facilitate high enhancer activity, observed as dense enrichment of cell master regulators, coactivators, Medicators and chromatin factors at SEs(2). These characteristics were further demonstrated by the fact that, distinct from regular enhancers, SE is specifically linked to gene regulation associated with cell identity and disease mechanisms(3,4).

Recent studies further showed that, beyond the elevated activity, the inside mechanics of SEs are also critical in defining their prominent roles in gene regulation, known as multi-promoter targeting and long-range interactions(5-8). Some SEs form a clear hierarchical structure where a hub constitute enhancer is responsible for the functional and structural organization of the whole SE(6,9). Other SEs, in contrary, receive relative balanced contribution from the constitute enhancers. In addition, constitute enhancers could establish an open chromatin interaction network within individual SEs(7), indicating the internal crosstalk across constitute enhancers in supporting SEs' unique functions.

The activity and relations of individual constitute enhancers were well appreciated during computational identification of SEs. Existing algorithms usually contain two processing steps(2,10). First, the activity and locations of genome-wide enhancers are inferred through peak detection using chromatin immunoprecipitation sequencing (ChIP-seq) data(11), particularly that measuring the binding of medicators, master regulators, or active histone mark H3K27Ac. Second, the inferred activity and locations are summarized linearly to prioritize broad enhancer regions(2,3), *i.e.* SEs, that contain densely enriched enhancers with high activities, *i.e.* the constitute enhancers.

However, the organization of constitute enhancers were not considered by current approaches in differential analysis of SEs, a key aspect of research interest when comparing across biological conditions(12-16). The alteration of SEs has been found to highly associate with disease dysregulation and be important for drug target identification(14-16). These approaches, in which SEs are treated as individual entities, usually identify differential SEs based on binary alteration comparison. In other words, they compare the presence and absence of SEs between biological conditions regardless the statuses of their constitute enhancers. As a results, differential SEs are generated largely relying on the algorithm parameters during SE detection(2,3). In addition, the sensitivity of these approaches is limited due to the impacts of local differences are downplayed within the broad genomic regions occupied by SEs.

We propose a computational method here to identify differential SEs by summarizing their internal differences that include the organization and differential statuses of constitute enhancers. In addition to overall differences, our method detects four extra differential categories specifically pointing to the internal structural alterations of SEs. This complements our understanding of SE dynamics across biological conditions. We demonstrate the unique characteristics of differential SE categories using public datasets by linking their activity to TF binding and gene expression with 3D chromatin interactions(17-19). The results indicate that overall differences and structural alterations are linked to distinct sets of gene targets with varied degrees of impacts. They jointly build up the dynamics of SE regulatory landscape across cells. When applied to public cancer cell datasets, our analysis shows that structural alterations linked to unique sets of cancer-specific genes and can maximize the discerning of cell identify for similar cells from the same cancer type. In summary, our method provides sensitive and biologically meaningful identification of differential

SEs. We also implement an R package (https://github.com/tenglab/DASE) to easy the use of our method, DASE (Differential Analysis of Super Enhancers).

RESULTS

Internal dynamics underlie genome-wide SE differences

SEs usually contain multiple constitute enhancers that locate in close genomic proximity along the genome(2,20). It is important to understand the roles of constitute enhancers in contributing to the highly cell-specific nature of SE profiles. To do that, we explored the SE profiles for six cancer cell types that were well-studied by ENCODE project(21) and had genomic sequencing data available for H3K27Ac histone modification, gene expression and 3D chromatin interactions (*Methods*). Exploratory analysis revealed substantial activity alterations within SEs across cancer cells. For instance, the constitute enhancers, located at two previous reported SE loci responsible for MYC regulation in multiple cancers(22-25), showed altered composite activities across cells as illustrated by H3K27Ac sequencing signals (Fig 1a). Different combination of active constitute enhancers, termed as the internal dynamics of SEs hereinafter, highlighted the varied SE activity and, in some cases, the completely absence of SEs across cells.

To summarize the dynamics across the whole genome, we compared the profiles of all constitute enhancers between pair-wise cancer cells using H3K27Ac signals (*Methods*). On average, over 40% of constitute enhancers showed significant differential activity (with conservative criteria: fold-change > 4 and q-value < 0.05) between cancer cells which accounts for over 80% of total SEs in these cells (Fig 1b). This indicates that SEs undergo frequent internal alterations across cancer cells. We further estimated how the alteration of constitute enhancers contributed to the overall differences of SEs. To simplify, we identified differential SEs between cancer cells using the presence/absence (binary) strategy mentioned previously (*Methods*). Only a small portion of

SEs (about 10%) with altered constitute enhancers didn't show overall differences, while the majority of them held the same differential statuses as their constitute enhancers (Fig 1b). This highlights the divergent influences of constitute enhancers on overall differential statuses of SEs.

We thus examined the characteristics of constitute enhancers that might affect their contribution to the overall SE differences. Not surprisingly, the spanning width and regulatory activity of constitute enhancers, indicated by H3K27Ac sequencing coverage, showed significant associations (Fig 1c-d). In brief, differential constitute enhancers with smaller width or lower activity presented less impacts on the overall statuses of their belonged SEs. In below, we built our model to summarize SE internal dynamics by accounting for these characteristics.

Modeling internal dynamics leads to distinct patterns of differential SEs

We reasoned that incorporating the internal dynamics, instead of relying only on the overall activity differences, should better accommodate true SE alterations. We implemented that by summarizing differential constitute enhancers into SE overall statuses using a weighted spline model (*Methods* and Fig 2). In brief, differential constitute enhancers were first evaluated using existing strategies on differential analysis ChIP-seq peaks(26). Then, a spline was fitted stratified by enhancer positions to smooth the differential statuses for consecutive constitute enhancers, while their activities and width of constitute enhancers were taken as fitting weights. Finally, the spline curves were evaluated with permutations to define reliable differential sub-regions within SEs, which were further summarized towards the overall differential statuses of SEs (*Methods*).

To demonstrate our findings, we applied our method to systematically compare SE profiles between two cancer cells, K562 and MCF7 (*Methods*). We chose the two cells because they have the most and high-quality annotation datasets in ENCODE data portal to support the downstream illustrations, *e.g.* metrics in transcription factor binding, chromatin interactions and gene

expression. Other cancer cells were also compared and demonstrated with selected metrics. In summary, our method detected the *overall-change* as well as four novel patterns of differential SEs denoting the structural differences within SEs (*i.e. shortened/lengthened*, *shifted*, *hollowed* and *other complex scenarios*) (Fig 2). *Overall-change* requires significant overall activity alterations (similar to binary approaches) as well as unified differential behavior from most constitute enhancers within a SE, *i.e.* most constitute enhancers are strengthened or weakened. (Fig 2a). *Shortened/lengthened* (in short as *shortened* hereinafter) represents SEs that have significant changes on their sizes by gaining or dismissing constitute enhancers between cells (Fig 2b). *Shifted* represents SEs have shifted genomic locations without significant size changes, *i.e.* constitute enhancers gained at one end of SEs and dismissed at the other end (Fig 2c) between compared cell conditions. *Hollowed* represents those with altered constitute enhancers in the middle of SEs while the two ends remain unchanged (Fig 2d). *Other complex scenarios* represent all other complicated or rare cases. Typical examples reveal that structural differences may (*e.g. shortened and hollowed*) or may not (*e.g. shifted*) accompany with overall activity differences quantified by H3K27Ac ChIP-seq signals (Fig 2a-d). Together, they provide novel insights to understand SE profiles associated to cell identifies.

In total, about 52% of the differential SEs were *overall-change* between K562 and MCF7, reflecting the distinct nature between cancer types. *Shortened* category dominated among all structural differences (65%), suggesting the widely spreading of SE size changes. Other types of structural differences, although not prevailing, presented consistently in comparisons between different cancer cells and uniquely associated with cell-specific gene regulation (see later sections). In addition, we note that over-claimed differential SEs by binary approaches due to ad hoc parameters during SE detection, could be properly corrected with the consideration of SE internal dynamics (Fig 2e).

Diverse differential SEs synergistically build up gene regulation

We further characterized the functional roles of different SE patterns in gene regulation. As previously mentioned, individual SEs may regulate multiple genes simultaneously through the assembly of various regulators and cofactors(8). Using an example SE, we highlight that transcription factor binding could highly correlate with the active status of constitute enhancers, and consequentially aid building up interactions with gene targets (Fig 3a). We thus examined the binding profiles for more transcription factors across genome-wide SEs. We scanned 78 transcription factors and histone modifications that have high quality ChIP-seq data in ENCODE portal for both K562 and MCF7. In general, transcription factor binding showed high correlation with the statuses of constitute enhancers, regardless the categories/patterns of differential SEs they belonged to. This indicates the SEs under distinct differential patterns share similar mechanisms to build gene regulation with transcription factors. For instance, a few transcription factors (27 out of 78), *e.g.* FOXM1, MBD2 etc., bind in both cell types only at their gained constitute enhancers and untie at their dismissed enhancers (Fig 3b), suggesting they might be critical in maintaining key functions in both cells. Some other transcription factors (31 out of 78), *e.g.* NFRKB, FOXA1 etc., in contrary, bind only at the active constitute enhancers in one cell type but the other, indicating their roles in cell-specific gene regulation.

Beyond transcription factor binding, we further examined the downstream effects of differential SEs on target gene regulation. We identified SE target genes in each cell type using 3D chromatin interactions inferred from POLR2A targeted ChIA-PET data (*Methods*). As expected, the gained constitute enhancers usually established new gene targets, while the dismissed enhancers removed existing targets (Fig 3c). As a result, SEs with gained constitute enhancers (strengthened and lengthened) in one cell type usually connected to more gene targets compare to their altered forms (weakened or shortened) in the other cell (Fig 3d). Such effects diverged in different SE patterns, with heavier effects observed in *overall-change, shortened* and *hollowed*

SEs while nearly no differences for *shifted* SEs. This is as expected since *shifted* provides no imputation on the directions of SE activity changes. To minimize the sequencing coverage impacts of ChIA-PET data on gene target counting, we justified the count differences by subtracting the median count (*i.e.* 1) in the control SE group (non-differential SEs). The low median value might correspond to the small coverage differences of ChIA-PET data in the compared cells as we chose intentionally (*Methods*).

Furthermore, differential expression analysis of the connected genes between the two cells (*Methods*) indicated that gained constitute enhancers were significantly associated with upregulated gene expression (Fig 3e). Such effects were also summarized at the SE level with strengthened and lengthened SEs presenting higher amplification on gene expression (Fig 3f). Again, *overall-change* and *shortened* SEs presented higher effects, while *shifted* hardly showed any difference. In addition, *hollowed* showed no clear impacts on gene expression, unlike their effects on linked gene numbers. Together, these suggests that *hollowed* and *shifted* SEs may play unique roles in complementing regulatory landscape across cell settings. As a control, we observed no significant expression differences of the linked genes by non-differential SEs.

When comparing the genes linked by different patterns of SEs, we found that some genes are regulated by multiple SEs across differential patterns (Fig 3g). However, most of them only link to one SE pattern, with comparable numbers of genes linked by *overall-change* SEs and SEs with structural differences. We further found that different pathways are uniquely enriched in genes linked by different patterns of SEs (Fig 3h). For instance, FoxO signaling pathway(27-29), cAMP signaling pathway(30), Amino sugar and nucleotide sugar metabolism(31), and BRCA1-dependent Ub-ligase activity(32) are enriched with genes linked to differential SEs as overall-change, shortened, shifted and hollowed, respectively. Together, different SE alteration patterns

synergistically build up gene regulation by playing distinct roles in targeting different genes and functions as well as complementing each other with varied degrees of regulatory impacts.

Accounting for internal dynamics improves identification and interpretation of differential SEs

Beside the structural characterization of SE alterations, our methods present overall improvement of differential SE identification over current binary approaches. We demonstrated the improvements by summarizing pair-wise comparisons across the six cancer cells. In total, we observed discrepancies accounting for over 18% of total SEs and covered all the SE categories defined by our methods (Figure 4a). As expected, the newly identified differential SEs by our methods were dominated with structural differences. We further examined the impacts of discrepant differential SEs on gene expression. We observed that the overall-change SEs only identified by our methods showed strongest effects on gene expression across all discrepant differential SEs (Fig 4b), suggesting those were potential false negative calls by the binary strategy. The other groups of discrepant differential SEs presented comparable but much lower effects on gene expression as compared to the common differential SEs by both binary strategy and our methods. This suggests their marginal effects on changing the expression of genes. However, such marginal effects could be reinforced if combined with SE structure differences, as we demonstrated above that structural differences were associated to increased or decreased number of gene targets (Fig 3d). Here, we didn't examine the impacts on the numbers of linked genes due to the dramatic differences of ChIA-PET sequencing coverage across the six cells. Large coverage differences could significantly bias the gene count metric and currently there is lack of efficient algorithms to normalize coverage differences on ChIA-PET data(33-35). It is noted that we didn't show the *shifted* and *other complex scenarios* here as their effects on genes are non-directional. It is also noted that, to avoid confounding effects due to genes might be targeted

by multiple SEs, we left out genes that linked to the common differential or non-differential SEs in the analysis.

To further illustrate gene regulation effects by accounting for SE internal dynamics, we zoomed into the uniquely linked genes by the newly identified differential SEs by our methods. In brief, we firstly filtered genes that were only linked by increased SEs (overall increased, lengthened or hollowed with increased constitute enhancers) in one cancer cell compared to the other five cells (*Methods*). We then compared this list to that obtained with binary strategy. Surprisingly, our methods not only recovered almost all the uniquely linked genes identified by binary strategy, but also identified a significant list of genes mainly linked to the SE structural differences (Fig 4c). This implies the elevated sensitivity of our methods in functional interpretation of SE regulation. We further performed pathway enrichment analysis based on the uniquely linked genes identified by our methods but the binary strategy, highlighting a number of cancer-specific pathways associated to SE structural differences (Fig 4d). For example, insulin signaling pathway(36), ErbB signaling pathway(37), thyroid hormone signaling pathway(38), TGF-beta signaling pathway(39), cytochrome P450(40), and Neurotrophin signaling pathway(41), were uniquely identified by our methods to enrich with SE-regulated genes in A549, HCT116, HepG2, K562, MCF7 and SK-N-SH, respectively. In summary, our methods showed improved sensitivity in linking SEs to cancer specific gene regulation, particularly through the consideration of internal dynamics.

Accounting for internal dynamics maximize the discerning of cell identity

We further reasoned that the elevated sensitivity by our methods could benefit to maximize our understanding on SE regulation between similar cells. We applied our methods to compare SE profiles in two similar cells, BC1 and BC3, that are B lymphocyte cells from Lymphoma under different viral infecting settings. We previously demonstrated that different viral infection led to distinct enhancer connectome on these cells(16). Here, with the elevated sensitivity on detecting

SE differences, we found that cell-specific gene expression linked to differential SEs, especially those with structural differences.

Overall, similar cells presented much higher similarity of SE profiles (Fig 5a). SEs with structural differences dominated among all identified differential SEs. We further linked the differential SEs to their target genes using chromatin interactions identified by H3K27Ac HiChIP datasets. Similar gene effect patterns were observed across differential SE categories, as we found previously (Fig 3c-f). Both counts and expression of the linked genes were enriched in the same direction as constitute enhancers/SEs altering between BC1 and BC3 (Fig 5b-e). Specifically, such effects are stronger by overall-change SEs, followed by shortened SEs, consistent with the findings based on ENCODE cells (Fig 3c-f). Finally, we extracted the linked genes and performed pathway enrichment analysis on the uniquely linked genes by different SE categories. We found unique pathways such as viral carcinogenesis particularly linked to the *shortened* categories of SEs. This suggests shortened SEs played key roles in gene regulation in response to the different viral infection between the two cells(42,43). This again highlights the differentiated roles of SE categories and the importance of featuring internal dynamics in SE differential analysis.

DISCUSSION

In this manuscript, we proposed a novel method to identify differential SEs by summarizing the internal dynamics. We categorized differential SEs into 5 major groups depending on their structural differences: *overall-change*, *shortened*, *hollowed*, *shifted* and *others*. By examining the associations of the five SE groups to transcription factor binding and gene expression, we found distinct features enriched in each group, such as linking with different numbers of genes and affecting gene expression at varied degrees. When comparing our methods with the widely adapted binary approach, we found more uniquely regulated genes linked to the newly identified structural differences of SEs. Such differences highlighted the elevated sensitivity of our methods.

It further suggests the increased power in identifying cell-specific gene regulation by SEs especially when applied on similar cells.

Specifically, our improved performance is powered by the consideration of SE internal dynamics. For instance, SEs might show frequent internal alterations yet with no overall activity changes, as typically presented in the *shifted* or *hollowed* groups by our methods. These differences, however, if not accounted for, could under-estimate the genome-wide variation of SE profiles and consequentially bias the evaluation of SE effects on gene expression. On the other side, significant activity changes of SEs are usually combined with structural alterations, either globally or partially, indicating modeling structural differences won't lose specificity in detecting true SE differences. However, we did notice that some SEs holding marginal activity changes which were weighted differently as discrepant calls between binary strategy and our methods. Nevertheless, these SEs usually showed marginal effects on gene expression. Thus, they either had minor effects on the downstream interpretation of gene regulation if missed by our methods or were deemed as structure differences by our methods with other effects such as linking to increased or decreased number of gene targets.

One limitation of our methods is that we cannot identify structural differences when a SE contains only one constitute enhancer. We proposed weighted spline model to justify the contribution of constitute enhancers by their width and activities. Thus, the model requires at least two constitute enhancers within a SE to generate confident estimation. In practice, we identified SEs with only one constitute enhancer as either non-differentials or overall-change if the activity alterations are significant. In addition, we identified differential SEs as other complexity scenarios if their internal patterns cannot be attributed to all other categories. We detailed this in Methods section. In practice, we found this category only accounts for a small portion of SEs (Fig 5a). We leave the closer interpretation of such complexity to future work.

SEs were conceptionally defined based on the intensity and enrichment of consecutive enhancers(2,3). As a result, significant changes of SEs may correspond to two scenarios: activity alterations between two SEs or status transitions between SEs and regular enhancers. These scenarios might further associate with different functional interpretations as regular enhancers tend to regulate less and closer genes comparing to SEs. Although we didn't provide approaches to discriminate the two scenarios as that goes beyond the scope of our proposed study, feasible strategies could be implemented in future work to improve the downstream interpretation. For instance, scanning the locations of constitute enhancers from gene promoters could help filter regular enhancers as they are usually close to their gene targets(44,45). Also, we can link SE differences to the status changes of local chromatin such as phase separation(46), to help determine if transitions occur between SE and regular enhancers. However, this requires integration of additional datasets to define chromatin statuses.

MATERIALS AND METHODS

Data acquisition

H3K27ac enrichment, gene expression and 3D interaction data were downloaded from ENCODE data portal(47) and GEO repositories(48). Particularly, quality-controlled alignment files of H3K27Ac ChIP-seq and RNA-seq, and chromatin contacts files of POLR2A ChIA-PET were downloaded from ENCODE for six selected cancer cell lines (A549, HCT116, HepG2, K562, MCF7 and SK-N-SH) (accession ID documented in Supplementary Table 1). Raw sequencing files of H3K27ac ChIP-seq and RNA-seq and H3K27ac HiChIP for two similar cancer cell lines (BC1 and BC3) were downloaded from GEO with accession IDs GSE136090(16) & GSE114791(49) (Supplementary Table 1).

H3K27Ac ChIP-seq data pre-processing

Raw ChIP-seq data downloaded from GEO was first aligned to human genome using Bowtie2(50). Then, both alignment files downloaded from ENCODE and generated in-house were processed for peak calling using MACS2(11), followed by SE detection using ROSE(2). All tools were applied with default parameters. ChIP-seq blacklist regions were excluded for downstream analysis(51).

RNA-seq data analysis

RNA-seq alignment files downloaded from ENCODE were quantified for gene expression count table using featureCount(52) suites based on GENCODE annotations. Raw FASTQ files from GEO were processed with semi-alignment and quantification tool Salmon(53) to generate gene expression count table based on GENCODE transcriptome. Then, differential analysis of gene expression was estimated using DESeq2(54) for all two-condition comparisons. The shrunk fold-changes were extracted to represent gene expression differences(55).

3D chromatin contacts analysis

The chromatin contacts generated by ENCODE project from POLR2A ChIA-PET data were directly adapted to link genes and super enhancers in ENCODE cancer cells. Basically, ENCODE project applied strict quality controls, and filtered confident chromatin contacts with at least 3 normalized interactions. HiChIP data of BC1 and BC3 cells are analyzed the same as we did before(16). In brief, reads were aligned to human genome using HiC-Pro(56). Sequencing replicates were merged to call chromatin contacts using hichipper(57) with confident interactions defined as at least 3 normalized interactions.

Differential analysis of constitute enhancers and binary SE differences

For each comparison between two cells that both have two ChIP-seq replicates, a uniform peak list was fist created by merging overlapped peaks across the compared samples. ChIP-seq reads

were then quantified using featureCount(52) to generate a read count table for the peak list. Differential peak analysis was performed by adapting DESeq2(54) (with parameter type='mean') to account for the varied dispersion between peaks with low and high read counts. The differential statuses of constitute enhancers (H3K27Ac peaks within SEs) were extracted based on their estimated log2 fold-changes and corresponding q-values. We also extracted the normalized coverage for constitute enhancers as the weight inputs for SE differential analysis below.

Binary SE differences were estimated based on the presence and absence of SEs between compared conditions. Basically, if a SE presents in both compared conditions at the same given location regardless size or activity changes, it will be identified as non-differential. In contrast, if a SE only presents in one condition at a given location, it will be identified as differential.

Modeling differential SEs with SE internal dynamics by DASE

We propose a computational method to identify differential SEs by accounting for the combinatorial effects of constitute enhancers weighted with their activities and locations (internal dynamics). In detail, the methods include the following steps (Supplementary Figure 1).

Input preparation. A uniform list of SEs is generated by merging overlapped SEs between compared conditions. The differential statuses (log2 fold-change) of all constitute enhancers located within SEs are extracted as well as their activities (ChIP-seq coverage) and locations (genomic coordinates), as calculated above. In practice, we select the maximum ChIP-seq coverage between compared conditions for each constitute enhancer to provide better weights in the spline model below.

Weighted spline model. For each SE, the log2 fold-change values of constitute enhancers stratified by their genomic locations are fitted using b-spline model, where the importance of

constitute enhancers are weighted by their relative activities. As a result, constitute enhancers show less impacts on the spline fitting if they have low activities and stay close to other constitute enhancers. We implement the spline model using R package *splines*. In addition, to ensure the robustness of b-splines in the case of too many low- or mild-activity constitute enhancers, we pre-estimate the degree of freedom for each fitting based on the number of top ranked constitute enhancers in each SEs. We choose top ranked constitute enhancers as the minimum number of enhancers that build up at least 95% of total SE activity. In detail, we set the degree of freedom of b-spline as 2, 3 and 4 if this number of top ranked constitute enhancers is less than 4, between 4-6, and larger than 6, respectively.

Significance estimation. We use permutations to define significant fitted values by b-spline. In brief, we randomly shuffle enhance activities in each compared sample, re-estimate the differential statuses of all constitute enhancers and re-fit splines for all SEs. As a result, we generate a null distribution of fitted b-spline values for all constitute enhancers. We repeat the processes 10 times for a stable null distribution. Significant fitted values are defined as those having greater or smaller values than the upper or lower inflection points (significant thresholds) of the null distribution (Supplementary Figure 1).

Status summarization for SE sub-regions. We divide each SE into multiple sub-regions using the intersects of b-spline curves and the significant thresholds (Figure 2). For instance, the curve located above the upper threshold indicates an up-altered partial region within the SE, while the curve located below the lower threshold indicates a down-altered partial region. The curves in between indicate non-altered SE sub-regions. To decrease potential noises in SE segmentation, we ignore sub-regions in which constitute enhancers account for less than 1% of the SE activities.

Overall differential status. We further summarize the overall differential statuses for SEs with heuristic approaches based on the statuses, locations, and activity occupancies (*i.e.* the percentage of activity over the total activity of SEs) of segmented sub-regions. Specifically, if only one region is resulted from segmentation of b-spline curve of a SE, the SE will be identified as either non-differential or differential depending on the status of that segment. If two segments are resulted (i.*e* one is significant and one is un-altered), we determine differential SEs based on the activity occupancy of the significant segment. In detail, two-segment SEs are identified as non-differential, *shortened* or *overall-change* if the significant segment occupies less than 10%, between 10%-90% and more than 90% of total SE activities. For a three-segment SE, we first check if it is *hollowed* based on whether the middle segment is significantly altered. If not, we check if it is *shifted* based on whether the three segments cover three different statues (i.e up-altered, down-altered and on-altered) separately. Otherwise, the remaining three-segment SEs fall into the following situation: the middle segment is non-significant while the left and right segments are both significant with the same altering direction. We then identify the overall SE statues as non-differential, *shortened* and *overall-change* based on the total activity occupancies of the left and right segments as below 10%, between 10%-50% and above 50%. It is noted that the *overall-change* are filtered with different criteria (break points at 90% vs 50%) in two-segment and three-segment SEs, to account for the total size impacts from the altered segments. For a SE with more than 3 segments, it is identified as *other complex scenarios* except that a four-segment SE holding all three statuses is defined as *hollowed*. Finally, we rank the significance of differential SEs using the activity occupancies of the significant segments separately for *overall-change*, *shortened*, *hollowed* and *shifted* SEs.

Translation factor enrichment analysis

ChIP-seq bam files for 78 documented transcription factors in both K562 and MCF7 cells were downloaded from ENCODE with accession ID provided in Supplementary Table 1. After calling

differential SEs between MCF7 and K562 with DASE, we extracted transcription factor occupancy from ChIP-seq data for all differential (fold-change >4 & q-value < 0.05) constitute enhancers that locate within differential SE categories: *overall-change*, *shortened*, *hollowed*, and *shifted*. The occupancy heatmap for selected transcription factors were then generated with Deeptools v3.5.1(58).

SE-gene targeting

We identify SE-gene targeting relationship using 3D chromatin contacts generated from POLR2A ChIA-PET or H3K27Ac HiChIP. Basically, a valid targeting is defined if one end of chromatin contacts is overlapped with SEs, while the other end is overlapped with gene promoters (selected as -3000bp ~ +1000bp from gene transcription start sites). Targeting relations are ignored if the SE-promoter distances are less than 20kb or greater than 500kb.

Pathway enrichment analysis

For pathway enrichment in genes linked by different SE categories, gene sets were first identified for each SE categories based on SE-gene targeting relations in both compared conditions. Then, only uniquely linked genes by each SE category were selected for pathway enrichment using DAVID Bioinformatics Resources v6.8(59) based on pathway databases BBID(60), BIOCARTA(61) and KEGG(62). For pathway enrichment in genes linked by cancer-specific SEs, genes were selected as those only identified by our methods compared to binary approaches.

FUNDING

This work was supported in part by the Biostatistics and Bioinformatics Shared Resource at the Moffitt Cancer Center [NCI P30 CA076292].

REFERENCES

1. Pott, S. and Lieb, J.D. (2015) What are super-enhancers? *Nat Genet*, **47**, 8-12.
2. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I. and Young, R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307-319.
3. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A. and Young, R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934-947.
4. Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Program, N.C.S. *et al.* (2013) Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A*, **110**, 17921-17926.
5. Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.M., Branco, M.R. *et al.* (2017) Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, **543**, 519-524.
6. Huang, J., Li, K., Cai, W., Liu, X., Zhang, Y., Orkin, S.H., Xu, J. and Yuan, G.C. (2018) Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat Commun*, **9**, 943.
7. Li, T., Jia, L., Cao, Y., Chen, Q. and Li, C. (2018) OCEAN-C: mapping hubs of open chromatin interactions across the genome reveals gene regulatory networks. *Genome Biol*, **19**, 54.
8. Novo, C.L., Javierre, B.M., Cairns, J., Segonds-Pichon, A., Wingett, S.W., Freire-Pritchett, P., Furlan-Magaril, M., Schoenfelder, S., Fraser, P. and Rugg-Gunn, P.J. (2018) Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. *Cell Rep*, **22**, 2615-2627.
9. Kai, Y., Li, B.E., Zhu, M., Li, G.Y., Chen, F., Han, Y., Cha, H.J., Orkin, S.H., Cai, W., Huang, J. *et al.* (2021) Mapping the evolving landscape of super-enhancers during cell differentiation. *Genome Biol*, **22**, 269.
10. Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I. and Young, R.A. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, **153**, 320-334.
11. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, **9**, R137.
12. Amaral, P.P. and Bannister, A.J. (2014) Re-place your BETs: the dynamics of super enhancers. *Mol Cell*, **56**, 187-189.
13. Allahyar, A., Vermeulen, C., Bouwman, B.A.M., Krijger, P.H.L., Verstegen, M., Geeven, G., van Kranenburg, M., Pieterse, M., Straver, R., Haarhuis, J.H.I. *et al.* (2018) Enhancer hubs and loop collisions identified from single-allele topologies. *Nat Genet*, **50**, 1151-1160.
14. Jia, Y., Chng, W.J. and Zhou, J. (2019) Super-enhancers: critical roles and therapeutic targets in hematologic malignancies. *J Hematol Oncol*, **12**, 77.

15. Yuan, J., Jiang, Y.Y., Mayakonda, A., Huang, M., Ding, L.W., Lin, H., Yu, F., Lu, Y., Loh, T.K.S., Chow, M. *et al.* (2017) Super-Enhancers Promote Transcriptional Dysregulation in Nasopharyngeal Carcinoma. *Cancer Res*, **77**, 6614-6626.
16. Wang, C., Zhang, L., Ke, L., Ding, W., Jiang, S., Li, D., Narita, Y., Hou, I., Liang, J., Li, S. *et al.* (2020) Primary effusion lymphoma enhancer connectome links super-enhancers to dependency factors. *Nat Commun*, **11**, 6318.
17. Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H. *et al.* (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, **462**, 58-64.
18. Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J. and Chang, H.Y. (2016) HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*, **13**, 919-922.
19. Fullwood, M.J., Wei, C.L., Liu, E.T. and Ruan, Y. (2009) Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res*, **19**, 521-532.
20. Shin, H.Y., Willi, M., HyunYoo, K., Zeng, X., Wang, C., Metser, G. and Hennighausen, L. (2016) Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat Genet*, **48**, 904-911.
21. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
22. Schuijers, J., Manteiga, J.C., Weintraub, A.S., Day, D.S., Zamudio, A.V., Hnisz, D., Lee, T.I. and Young, R.A. (2018) Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. *Cell Rep*, **23**, 349-360.
23. Chen, D., Zhao, Z., Huang, Z., Chen, D.C., Zhu, X.X., Wang, Y.Z., Yan, Y.W., Tang, S., Madhavan, S., Ni, W. *et al.* (2018) Super enhancer inhibitors suppress MYC driven transcriptional amplification and tumor progression in osteosarcoma. *Bone Res*, **6**, 11.
24. Lancho, O. and Herranz, D. (2018) The MYC Enhancer-ome: Long-Range Transcriptional Regulation of MYC in Cancer. *Trends Cancer*, **4**, 810-822.
25. Wang, C., Jiang, S., Zhang, L., Li, D., Liang, J., Narita, Y., Hou, I., Zhong, Q., Gewurz, B.E., Teng, M. *et al.* (2019) TAF Family Proteins and MEF2C Are Essential for Epstein-Barr Virus Super-Enhancer Activity. *J Virol*, **93**.
26. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R. *et al.* (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, **481**, 389-393.
27. Pellicano, F., Scott, M.T., Helgason, G.V., Hopcroft, L.E., Allan, E.K., Aspinall-O'Dea, M., Copland, M., Pierce, A., Huntly, B.J., Whetton, A.D. *et al.* (2014) The antiproliferative activity of kinase inhibitors in chronic myeloid leukemia cells is mediated by FOXO transcription factors. *Stem Cells*, **32**, 2324-2337.
28. Lin, C.H., Chang, C.Y., Lee, K.R., Lin, H.J., Chen, T.H. and Wan, L. (2015) Flavones inhibit breast cancer proliferation through the Akt/FOXO3a signaling pathway. *BMC Cancer*, **15**, 958.
29. Hornsveld, M., Smits, L.M.M., Meerlo, M., van Amersfoort, M., Groot Koerkamp, M.J.A., van Leenen, D., Kloet, D.E.A., Holstege, F.C.P., Derksen, P.W.B., Burgering, B.M.T. *et al.*

(2018) FOXO Transcription Factors Both Suppress and Support Breast Cancer Progression. *Cancer Res*, **78**, 2356-2369.
30. Dong, H., Claffey, K.P., Brocke, S. and Epstein, P.M. (2015) Inhibition of breast cancer cell migration by activation of cAMP signaling. *Breast Cancer Res Treat*, **152**, 17-28.
31. Gerardy-Schahn, R., Oelmann, S. and Bakker, H. (2001) Nucleotide sugar transporters: biological and functional aspects. *Biochimie*, **83**, 775-782.
32. Brzovic, P.S., Keeffe, J.R., Nishikawa, H., Miyamoto, K., Fox, D., 3rd, Fukuda, M., Ohta, T. and Klevit, R. (2003) Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc Natl Acad Sci U S A*, **100**, 5646-5651.
33. Capurso, D., Tang, Z. and Ruan, Y. (2020) Methods for comparative ChIA-PET and Hi-C data analysis. *Methods*, **170**, 69-74.
34. Lareau, C.A. and Aryee, M.J. (2018) diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics*, **34**, 672-674.
35. Lee, B., Wang, J., Cai, L., Kim, M., Namburi, S., Tjong, H., Feng, Y., Wang, P., Tang, Z., Abbas, A. *et al.* (2020) ChIA-PIPE: A fully automated pipeline for comprehensive ChIA-PET data analysis and visualization. *Sci Adv*, **6**, eaay2078.
36. Kodama, Y., Baxter, R.C. and Martin, J.L. (2002) Insulin-like growth factor-I inhibits cell growth in the a549 non-small lung cancer cell line. *Am J Respir Cell Mol Biol*, **27**, 336-344.
37. Lee, H., Lee, H., Chin, H., Kim, K. and Lee, D. (2014) ERBB3 knockdown induces cell cycle arrest and activation of Bak and Bax-dependent apoptosis in colon cancer cells. *Oncotarget*, **5**, 5138-5152.
38. Liao, C.H., Yeh, C.T., Huang, Y.H., Wu, S.M., Chi, H.C., Tsai, M.M., Tsai, C.Y., Liao, C.J., Tseng, Y.H., Lin, Y.H. *et al.* (2012) Dickkopf 4 positively regulated by the thyroid hormone receptor suppresses cell invasion in human hepatoma cells. *Hepatology*, **55**, 910-920.
39. Dong, M. and Blobe, G.C. (2006) Role of transforming growth factor-beta in hematologic malignancies. *Blood*, **107**, 4589-4596.
40. Coumoul, X., Diry, M., Robillot, C. and Barouki, R. (2001) Differential regulation of cytochrome P450 1A1 and 1B1 by a combination of dioxin and pesticides in the breast tumor cell line MCF-7. *Cancer Res*, **61**, 3942-3948.
41. Capasso, M. and Diskin, S.J. (2010) Genetics and genomics of neuroblastoma. *Cancer Treat Res*, **155**, 65-84.
42. Cesarman, E., Moore, P.S., Rao, P.H., Inghirami, G., Knowles, D.M. and Chang, Y. (1995) In vitro establishment and characterization of two acquired immunodeficiency syndrome-related lymphoma cell lines (BC-1 and BC-2) containing Kaposi's sarcoma-associated herpesvirus-like (KSHV) DNA sequences. *Blood*, **86**, 2708-2714.
43. Arvanitakis, L., Mesri, E.A., Nador, R.G., Said, J.W., Asch, A.S., Knowles, D.M. and Cesarman, E. (1996) Establishment and characterization of a primary effusion (body cavity-based) lymphoma cell line (BC-3) harboring kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) in the absence of Epstein-Barr virus. *Blood*, **88**, 2648-2654.
44. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A. *et al.* (2019) Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet*, **51**, 1664-1669.

45. Moore, J.E., Pratt, H.E., Purcaro, M.J. and Weng, Z. (2020) A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. *Genome Biol*, **21**, 17.
46. Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C. *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, **361**.
47. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K. *et al.* (2018) The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*, **46**, D794-D801.
48. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*, **30**, 207-210.
49. Wong, J.P., Stuhlmiller, T.J., Giffin, L.C., Lin, C., Bigi, R., Zhao, J., Zhang, W., Bravo Cruz, A.G., Park, S.I., Earp, H.S. *et al.* (2019) Kinome profiling of non-Hodgkin lymphoma identifies Tyro3 as a therapeutic target in primary effusion lymphoma. *Proc Natl Acad Sci U S A*, **116**, 16541-16550.
50. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.
51. Amemiya, H.M., Kundaje, A. and Boyle, A.P. (2019) The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep*, **9**, 9354.
52. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923-930.
53. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, **14**, 417-419.
54. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
55. Zhu, A., Ibrahim, J.G. and Love, M.I. (2019) Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*, **35**, 2084-2092.
56. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J. and Barillot, E. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*, **16**, 259.
57. Lareau, C.A. and Aryee, M.J. (2018) hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat Methods*, **15**, 155-156.
58. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*, **42**, W187-191.
59. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, **4**, 44-57.
60. Becker, K.G., White, S.L., Muller, J. and Engel, J. (2000) BBID: the biological biochemical image database. *Bioinformatics*, **16**, 745-746.
61. Nishimura, D. (2001) BioCarta. *Biotech Software & Internet Report*, **2**, 117-120.
62. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, **28**, 27-30.

FIGURES

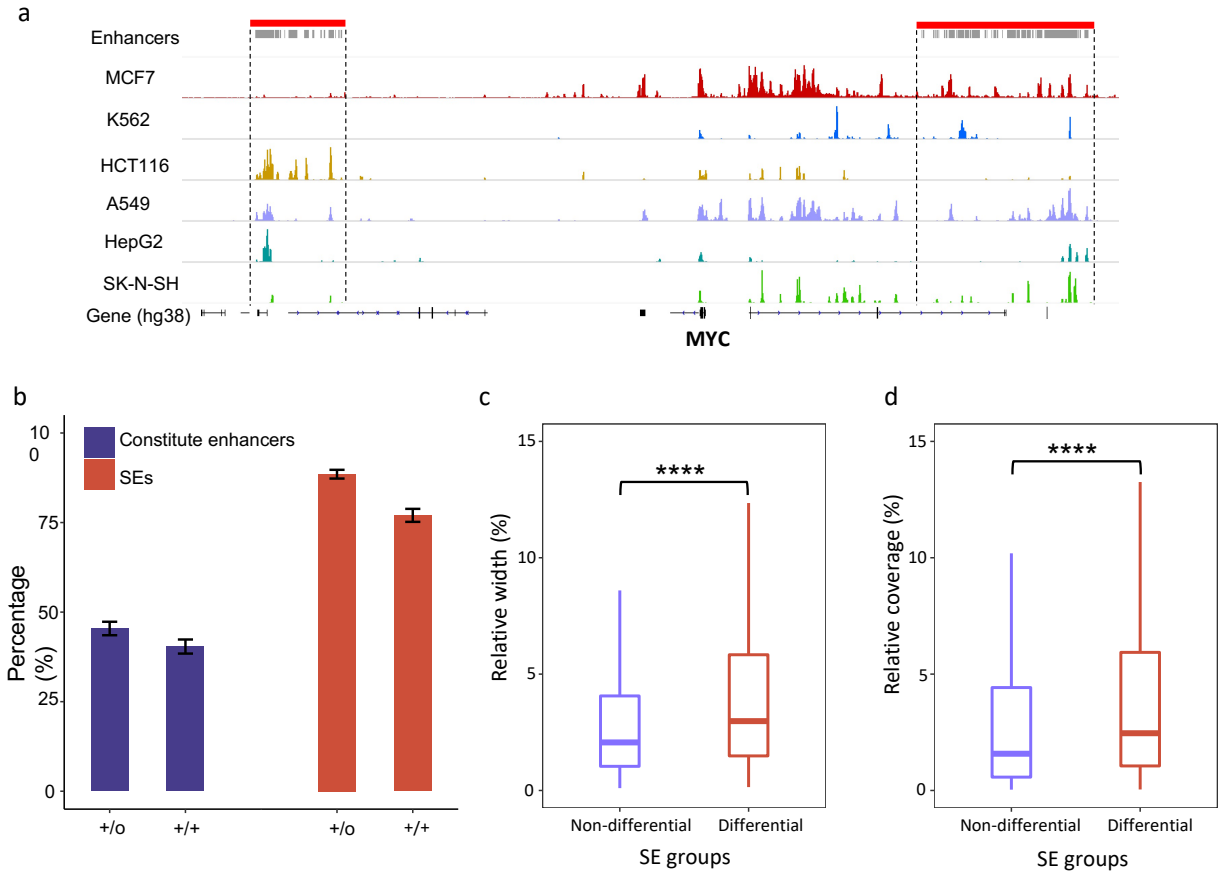

Fig.1. Internal dynamics of super enhancers.

a. Constitute enhancers show frequent alterations in selected cancer cells at two SE loci associated to MYC regulation. ChIP-seq tracks of H3K27Ac coverage are showed here. SEs and constitute enhancers are labeled above as red and grey bars. **b.** Frequencies of differential enhancers (Fold-change > 4 & q-value < 0.05) and differential SEs (based on binary strategy). Labels on the x-axis indicate features filtered by differential ("+") or no filtering ("o"), while left and right symbols correspond to filters on constitute enhancer and SEs, respectively. **c.** Relative width of differential constitute enhancers within non-differential and differential SEs. Relative width is defined as the percentage of constitute enhancer width over the summed width from all constitute enhancers within a SE. **d.** Relative coverage of differential constitute enhancers within non-differential and differential SEs. Relative coverage is defined as the percentage of constitute enhancer coverage over the summed coverage from all constitute enhancers within a SE.

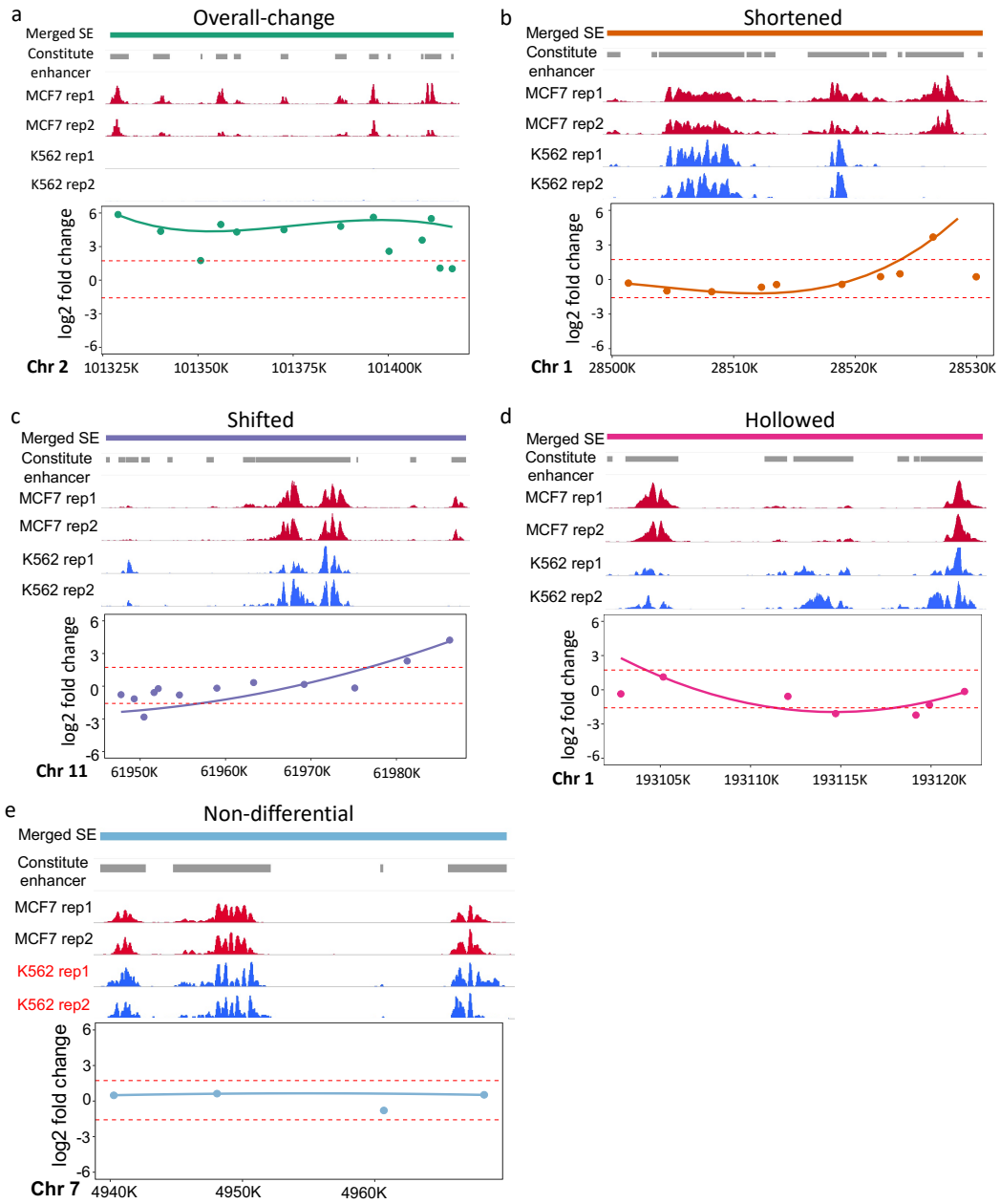

Fig.2. Differential SEs modeled with DASE.

SE examples are listed with overall-change (**a**), shortened (**b**), shifted (**c**), hollowed (**d**) and non-differential(**e**). In each sub-figure, the upper panel lists in order the SE regions, constitute enhancers, H3K27Ac ChIP-seq coverage in two cells with two replicates. The lower panel shows the fitted b-splines in addition to the original log2 fold-change values for constitute enhancers (points). Dashed lines indicate the estimated thresholds from permutation to define differential segments within SE regions. In **e**), red highlighted text indicates SE detected by binary strategy.

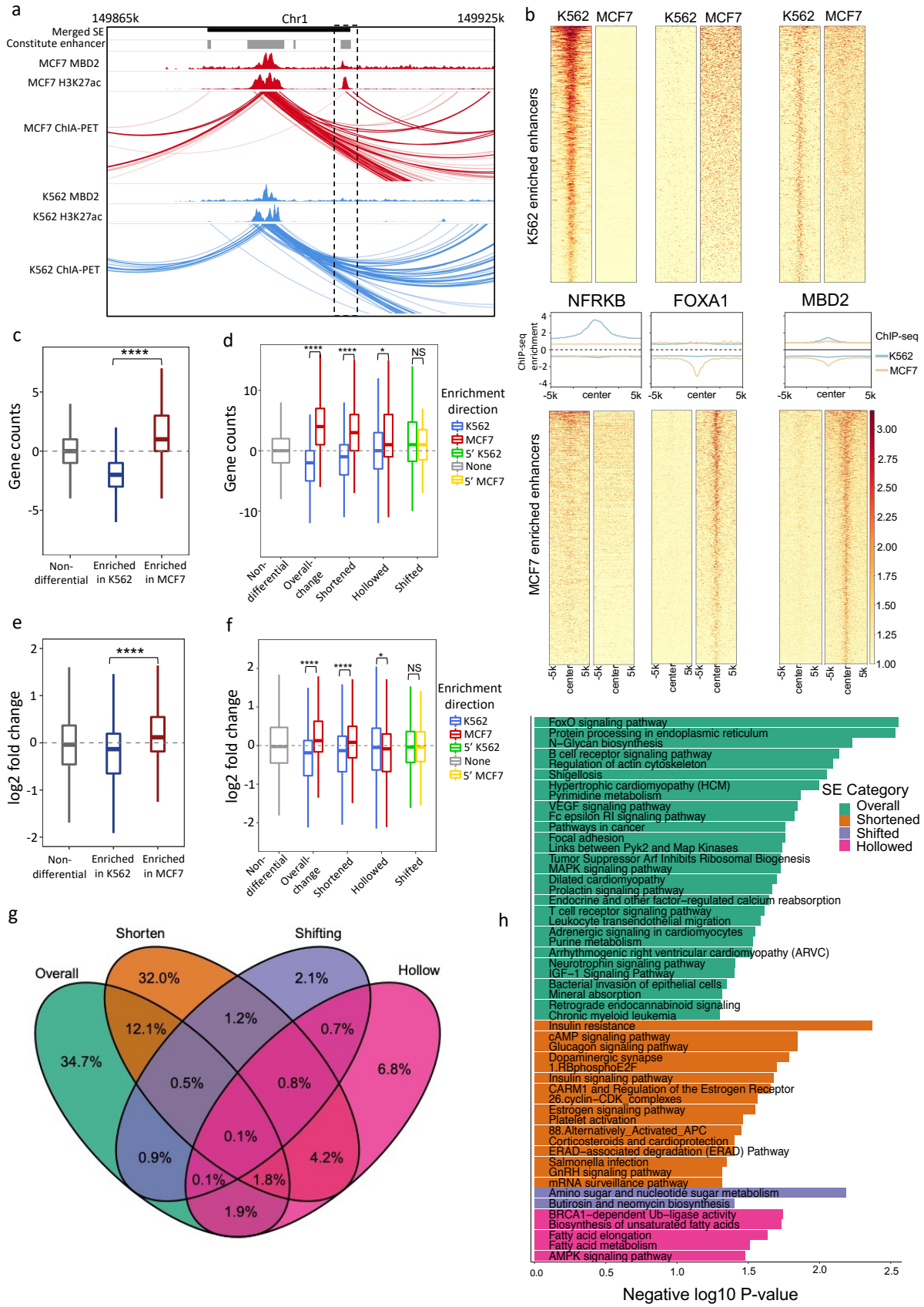
a
149865k
Chr1
149925k
Merged SE
Constitute enhancer
MCF7 MBD2
MCF7 H3K27ac
MCF7 ChIA-PET
K562 MBD2
K562 H3K27ac
K562 ChIA-PET
b
K562
MCF7
K562 enriched enhancers
NFRKB
FOXA1
MBD2
ChIP-seq enrichment
ChIP-seq
K562
MCF7
-5k
center
5k
MCF7 enriched enhancers
c
Gene counts
****
Non-differential
Enriched in K562
Enriched in MCF7
d
Gene counts
****
*
NS
Non-differential
Overall-change
Shortened
Hollowed
Shifted
Enrichment direction
K562
MCF7
5' K562
None
5' MCF7
e
log2 fold change
****
Non-differential
Enriched in K562
Enriched in MCF7
f
log2 fold change
****
*
NS
Non-differential
Overall-change
Shortened
Hollowed
Shifted
Enrichment direction
K562
MCF7
5' K562
None
5' MCF7
g
Overall
Shorten
Shifting
Hollow
32.0%
2.1%
12.1%
1.2%
0.7%
34.7%
0.5%
0.8%
6.8%
0.1%
0.9%
4.2%
0.1%
1.8%
1.9%
h
FoxO signaling pathway
Protein processing in endoplasmic reticulum
N-Glycan biosynthesis
B cell receptor signaling pathway
Regulation of actin cytoskeleton
Shigellosis
Hypertrophic cardiomyopathy (HCM)
Pyrimidine metabolism
VEGF signaling pathway
Fc epsilon RI signaling pathway
Pathways in cancer
Focal adhesion
Links between Pyk2 and Map Kinases
Tumor Suppressor Arf Inhibits Ribosomal Biogenesis
MAPK signaling pathway
Dilated cardiomyopathy
Prolactin signaling pathway
Endocrine and other factor-regulated calcium reabsorption
T cell receptor signaling pathway
Leukocyte transendothelial migration
Adrenergic signaling in cardiomyocytes
Purine metabolism
Arrhythmogenic right ventricular cardiomyopathy (ARVC)
Neurotrophin signaling pathway
IGF-1 Signaling Pathway
Bacterial invasion of epithelial cells
Mineral absorption
Retrograde endocannabinoid signaling
Chronic myeloid leukemia
Insulin resistance
cAMP signaling pathway
Glucagon signaling pathway
Dopaminergic synapse
1.RBphosphoE2F
Insulin signaling pathway
CARM1 and Regulation of the Estrogen Receptor
26.cyclin-CDK_complexes
Estrogen signaling pathway
Platelet activation
88.Alternatively_Activated_APC
Corticosteroids and cardioprotection
ERAD-associated degradation (ERAD) Pathway
Salmonella infection
GnRH signaling pathway
mRNA surveillance pathway
Amino sugar and nucleotide sugar metabolism
Butirosin and neomycin biosynthesis
BRCA1-dependent Ub-ligase activity
Biosynthesis of unsaturated fatty acids
Fatty acid elongation
Fatty acid metabolism
AMPK signaling pathway
SE Category
Overall
Shortened
Shifted
Hollowed
Negative log10 P-value

Fig.3. Differential SE categories linked to distinct regulatory features.

a. An example region highlights the correlation between constitute enhancer activity and transcription factor binding, which might be necessary to setup SE-gene interactions. **b**. Typical binding patterns of transcription factors at differential constitute enhancers, cell-specific (NKRKB & FOXA1) and shared (MBD2). Top panels are transcription factor signals at constitute enhancers enriched in K562. Bottom panels are transcription factor signals at constitute enhancers enriched in MCF7. The middle panels are the aggregated binding signals from the top and bottom panels with blue and orange lines indicating signals in K562 and MCF7 separately. The signs of y-axises in middle panels represent enrichment directions. **c**. Differences on the number of constitute-enhancer-linked genes between MCF7 and K562 for different constitute enhancers. Gene targets are identified based on chromatin contacts from ChIA-PET data. **d**. Differences on the number of SE-linked genes between MCF7 and K562 for different SE categories. In each differential SE category, SEs are separated into two groups based on their enrichment directions. **e**. log2 fold-change of constitute-enhancer-linked genes between MCF7 and K562. **f**. log2 fold-change of SE-linked genes between MCF7 and K562. In each differential SE category, genes are separated into two groups by the enrichment directions of their linked SEs. **g**. Overlaps of SE-linked genes across four differential SE categories in MCF7 and K562. **h**. Pathways uniquely enriched in each differential SE category.

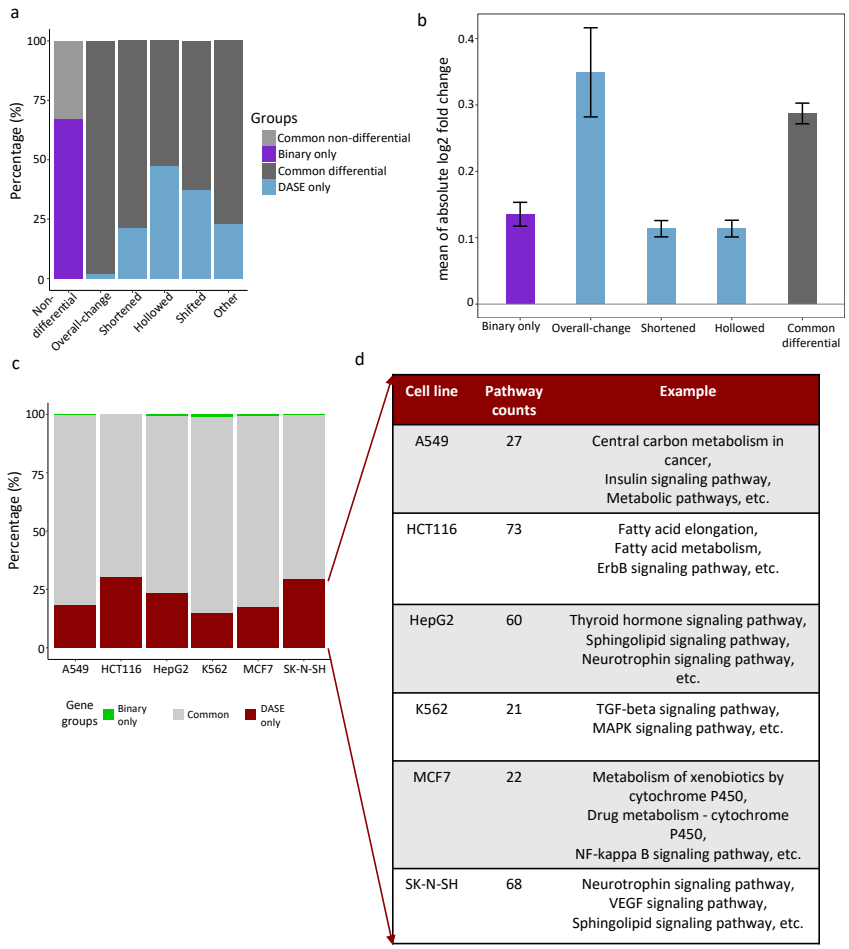

Fig.4. Comparisons between DASE and binary strategy based on six cancer cells.

a. Average discrepancies between DASE and binary strategy across 15 pair-wise comparisons. Light grey: common non-differentials; Dark grey: common differentials; purple: differential only by binary approach; blue: differential only by DASE. **b**. Expression impacts on SE-linked genes by differential SEs that are discrepantly called by DASE and binary strategy. Bars indicate the medians of log2 fold-changes in each comparison and error bars represent the standard error of medians. Results from two directions of enrichment are merged based on the absolute values of log2 fold-changes. **c.** Genes linked by activity increased SEs (overall increased, lengthened or hollowed with increased constitute enhancers) in each cancer cell compared to the other five cells. Green: genes uniquely found by binary strategy; red: genes uniquely found by DASE; grey: genes found by both methods. **d.** Pathways enriched in uniquely found genes by DASE.

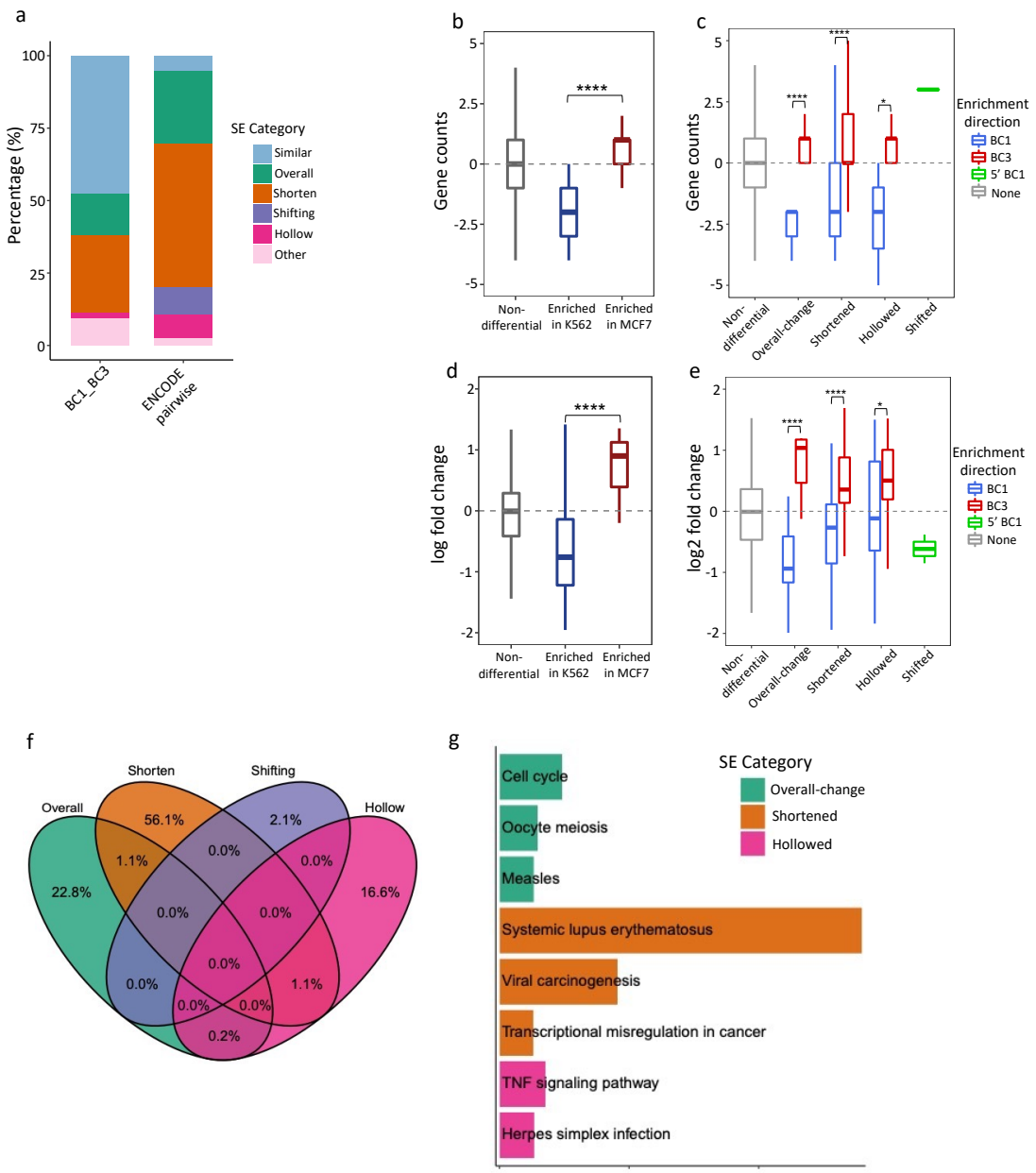

Fig.5. Identifying SE differences between similar cancer cells, BC1 and BC3.

a. Proportions of differential SE categories identified in same-cancer comparison and cross-cancer-type comparisons. Count differences (**b**,**c**) and log2 fold-changes (**d**,**e**) of linked genes by constitute enhancers (**b**,**d**) and SEs (**c**,**e**) between BC1 and BC3. Gene targets are identified based on chromatin contacts from HiChIP data. **f**. Overlaps of SE-linked genes across four differential SE categories in BC1 and BC3. **g**. Pathways that uniquely enriched in each differential SE category.