MicroC Database
Reference Manual

MicroC Database web app provides researchers with a visual interface of b-cell precursor acute lymphoblastic leukemia (BCP ALL) 3D chromatin landscape derived from the publication;

[INSERT PUBLICATION CITATION HERE]

Data Processing:

MicroC was performed on 35 BCP-ALL samples representing the following subtypes; High-hyperdiploid (HeH, n=14), *ETV6::RUNX1* (n=8), *TCF3::PBX1* (n=4), *BCR::ABL1* (n=3), DUX4-r (n=2), KMT2Ar (n=1), near haploid (n=1), iAMP21(n=1), B-other (n=1).

Merged loops from all samples, as well as loops merged by subtype, were generated at 1 kb and 10 kb resolutions respectively, as described in the Methods section of the publication.

Enhancer prediction was based on the presence of transcription start sites within the other end of a target loop anchor. For the 1kb loops, window of interest for transcription start site overlap was extended by 2kb on both directions to account for positional shifts and neighboring chromatin sites of interacting regions due to close proximity (see Methods).

Tracks:

Promoters: These represent -2000 bp and +500 bp bins of each transcription start site reported in ENSEMBL. They are represented as red bards under loops.

ENCODE tracks: H3K4me1, H3K4me3, H3K27ac, H3K27me3, DNase peaks downloaded from ENCODE Database using ChIP-seq experiments performed on GM12878 cell line.

Genes: Represents genomic location information of all protein coding genes in ENSEMBL. To ease visualization, HGNC nomenclature was used for labels.

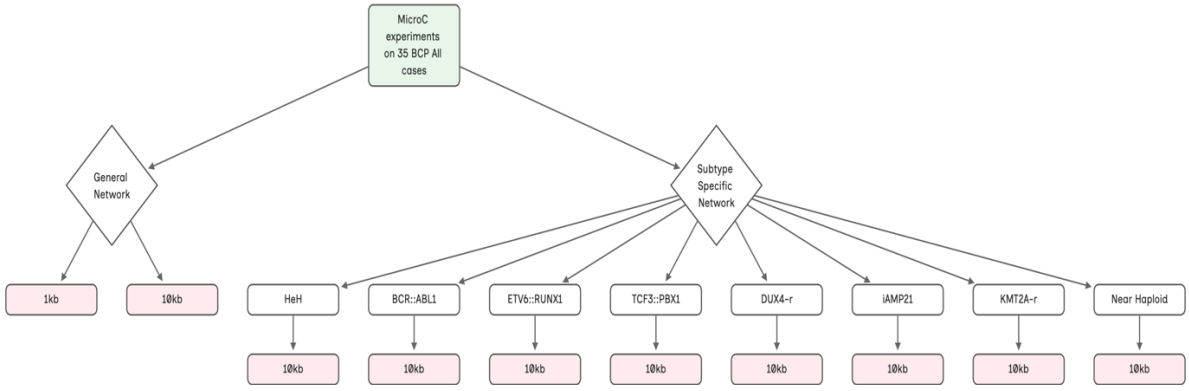
Loops: These are dynamic outputs of each query derived from loop data. Enhancer loops are depicted in red and others in blue.

Usage:

In this app, user can investigate chromatin interaction information processed from MicroC experiments ran on 35 primary BCP ALL samples (see methods in the paper) either from a general network or subtype specific network perspective.

General network represents loops extracted from the merged data of all cases. This module of choice is preferred if the user is more interested in leukemia focused and cis-regulatory element level investigations. Due to the high sequencing coverage, users can extract fine scale level interactions up to 1 kilobase resolution. Based on the research preference, user is also available to get this network information on 10kb resolutions. This resolution is still derived from the 1kb loops binned to 10kb. For example chr1:12000-13000 in 1kb transforms into chr1:10000-20000 in 10kb.

Our webapp also provides network data on 8 major subtypes of BCP ALL; High Hyperdiploidy, *BCR::ABL1*, *ETV6::RUNX1*, *TCF3::PBX1*, *DUX4*-rearrangement, iAMP21, KMT2A-r and near haploid cases. Using this module, users can investigate chromatin interactions and regulatory network in their subtype of interest or compare their findings against other subtypes. A general selection framework is depicted below. Due to the reduced amount of samples upon subtype specific selection, network is represented within a fixed resolution of 10 kb.



Submodules:

Gene or Location focused query:

This app allows users to examine 3d interaction network of BCP ALL from both gene-centric and location-centric points of view. In case the user is interested in a certain gene and its putative regulatory elements, gene focused module would represent all the loops that possess the promoter region of the given gene within an anchor.

If the user is interested in a genomic location and its behavior within a selected window, location focused module would visualize all the interactions found in the given region. If the anchors in the query region are hitting a promoter of a gene on their other ends, they are considered as enhancer loops and displayed with red arcs. To account for other chromatin interactions that are not representing an enhancer but still identified within the region, we also display them separately with blue arcs.

CRE prediction preference:

Briefly, this app predicts a loop to represent a cis-regulatory element (CRE) if one of its anchors overlap with a promoter region. However, promoter annotation is a complex subject. Although using only canonical promoters as an annotation constraint can be a safe approach since they are supported by stronger evidence on average, alternative promoter usage has been also demonstrated to be a feature in many cancers. Thereby, we decided to leave this preference to the user. Based on the research interest, the user can select which annotation preference to be used on CRE predictions.

If the user chooses to include all promoters for the analysis, every loop that hits the promoter region of any protein coding transcript reported on ENSEMBL would be predicted as a CRE loop. If canonical promoter only option is selected, then only the loops that hit the promoter of the canonical transcript would be reported as CRE while the others would be ignored even if they hit the transcription start site of an alternative transcript.

Example Usage:

1) What are the predicted cis-regulatory elements of ITPR1 gene in BCP ALL?

To address this question, we can do the following query; Analysis:General, Resolution:1k, Query interest:Gene, Predictions:All promoters. This query would generate a plot to represent all the loops found in the merged 1kb dataset for this gene region.

This query generates 7 enhancer predictions for this gene in the following regions;

chr3:4467000-4468000

chr3:4494000-4495000

chr3:4504000-4505000

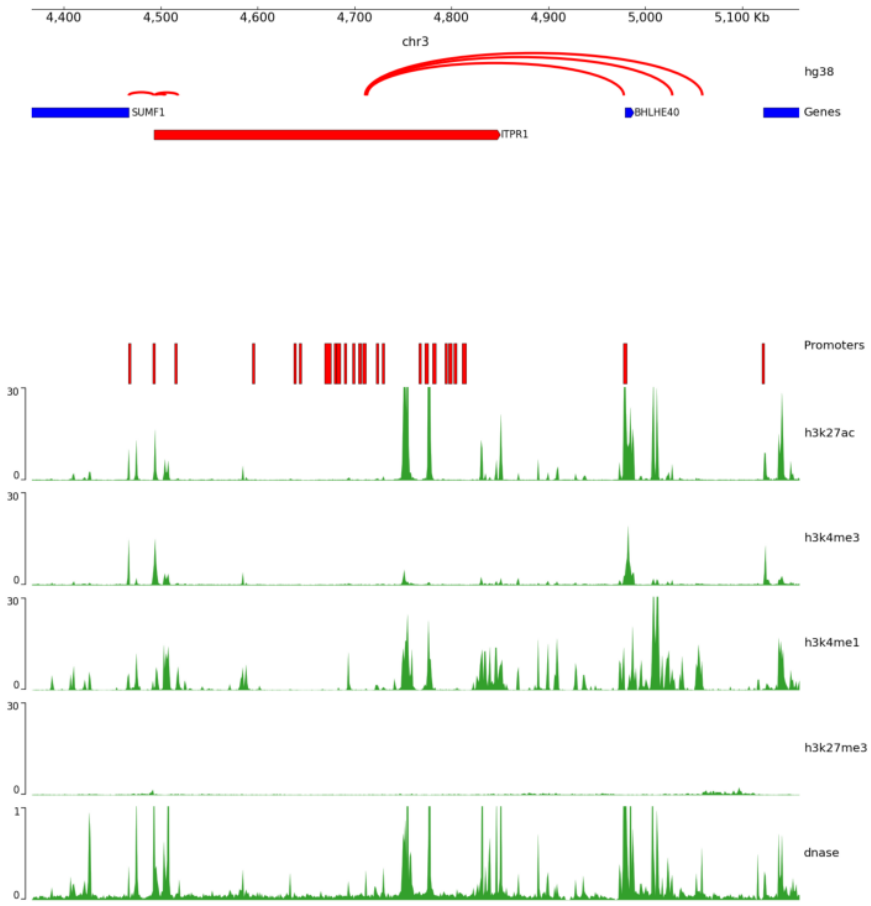
chr3:4517000-4518000

chr3:5058000-5059000

chr3:4977000-4978000

chr3:5027000-5028000

In the image below, we see that 7 promoter interacting loops found for *ITPR1* gene are represented by red arcs. Gene of interest (ITPR1) is highlighted in red
while the other genes in the region are blue. These predicted enhancers can be downloaded by the button above tracks.



2) Which one of these enhancers target the canonical promoter of *ITPR1*?

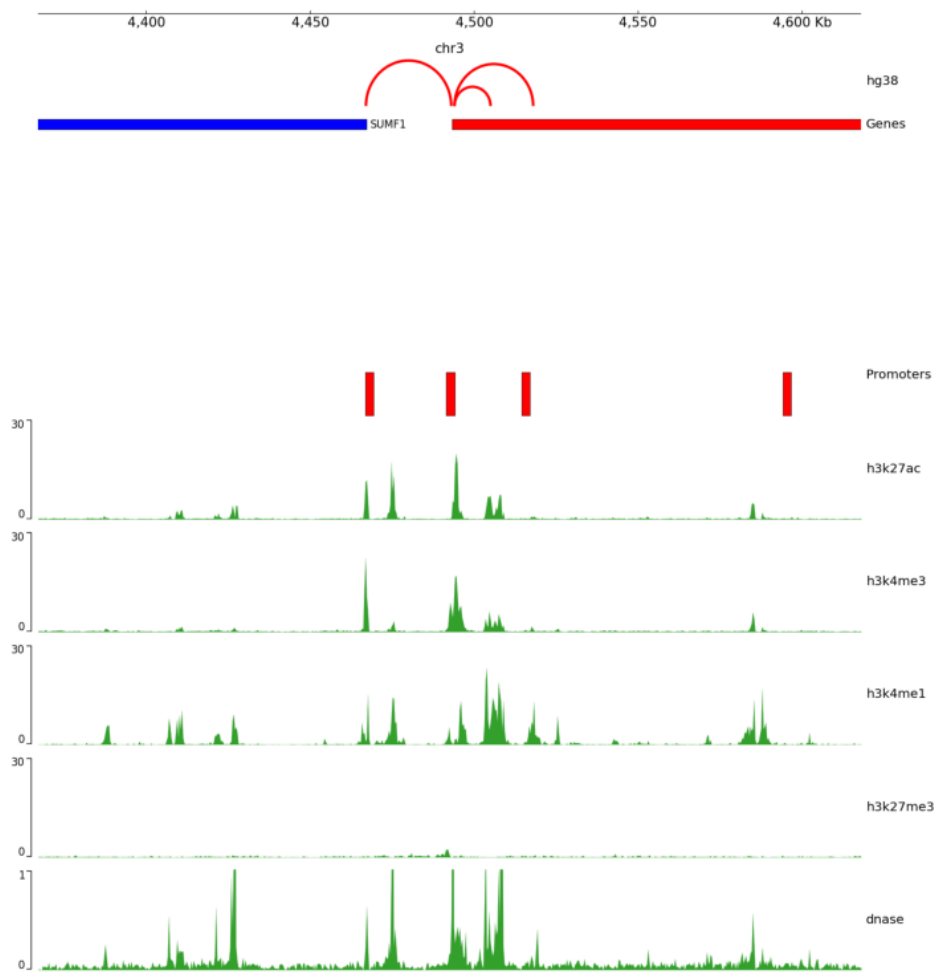
If you are interested only in the regulatory elements of the canonical transcripts for *ITPR1*, you can slightly modify the previous query by selecting canonical promoters on the "Predictions based on" button on the left panel. This modification would show that only 3 of the enhancers from the previous query are actually interacting with the promoter of the canonical transcript of *ITPR1*. These are found on;

chr3:4467000-4468000

chr3:4504000-4505000

chr3:4517000-4518000

A plot representing this query is given below.



3) What are the predicted cis-regulatory elements of *ITPR1* on high hyperdiploidy cases?

To address this question, we can do the following query; Analysis:Subtype specific, Subtype:High Hyperdiploidy, Resolution:10k (fixed), Query interest:Gene, Predictions:All promoters.

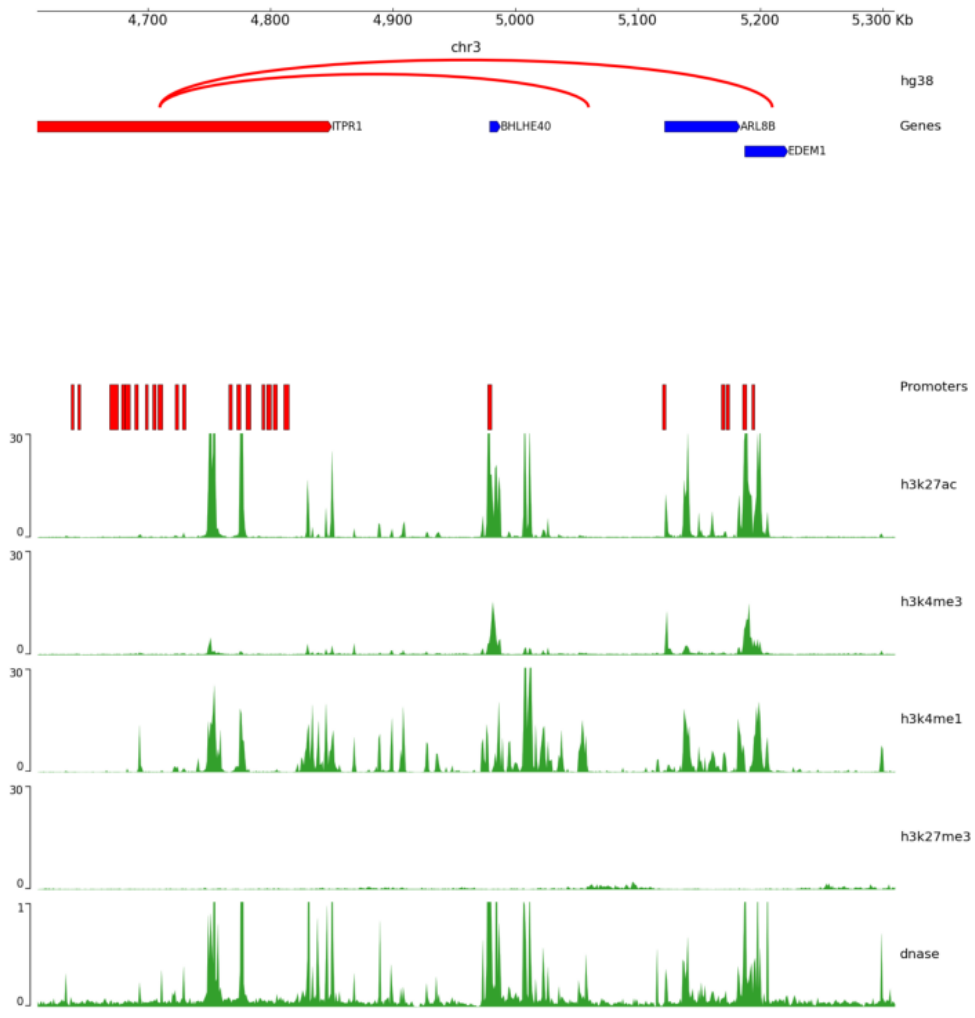
Here, you see that only two regulatory interactions were found for High hyperdiploidy cases compared to the 7 that was found in merged_1k data on example 1. These interacting anchors were found on;

chr3:5050000-5060000

chr3:5200000-5210000

There are two details that should be noticed here. First, you can see that the resolution is dropped to10k due to subtype specific analysis. As you can see, the enhancer found betweenchr3:5058000-5059000 on merged 1k samples also exist in lower resolution as chr3:5050000-5060000. However, the second enhancer found on 10k subtype specific result was not present on the merged_1k. This indicates an interaction intensity that

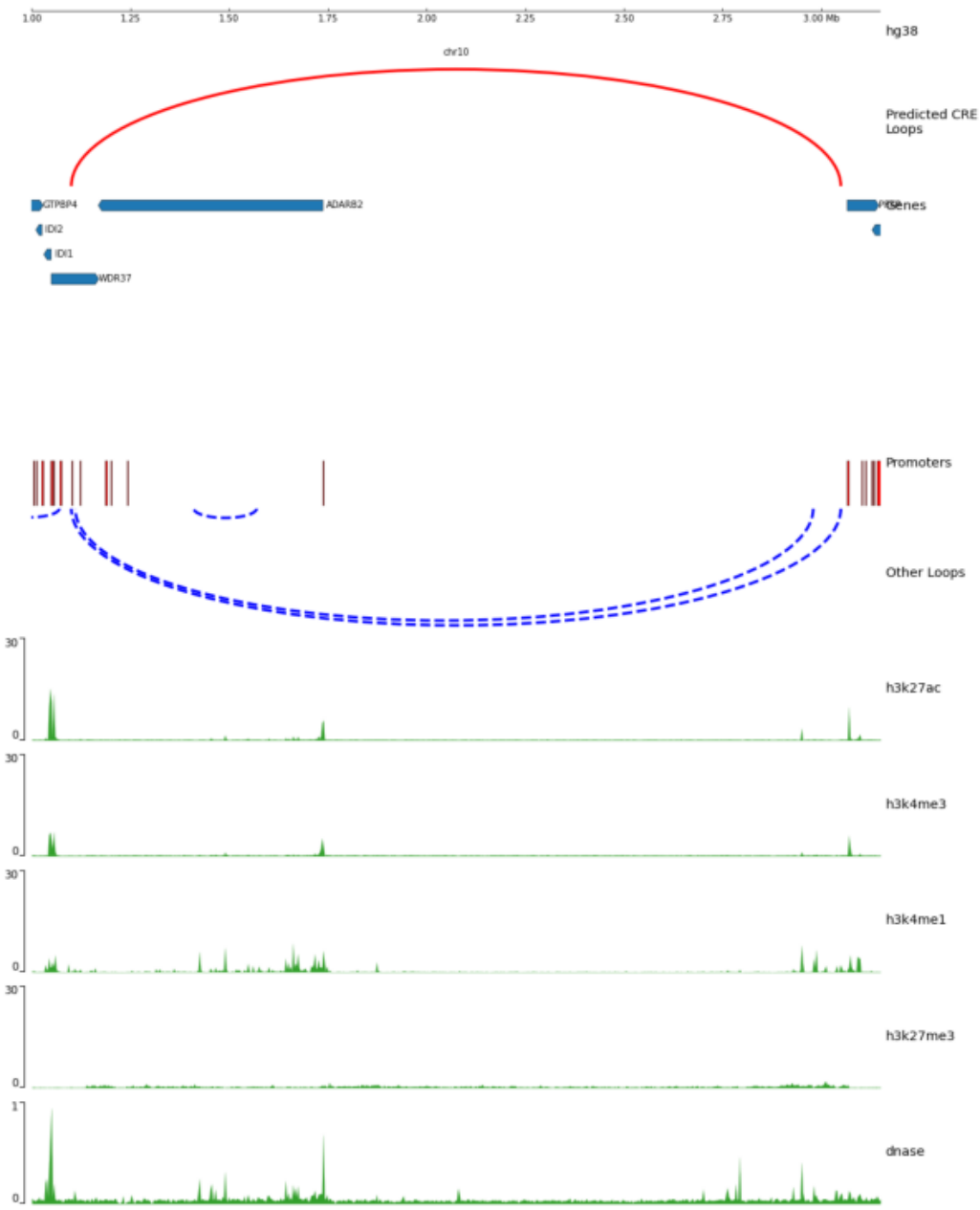
showed statistical significance when a subset of High hyperdiploidy samples were used for merging but disappeared when all of the samples are combined. Such cases may represent biologically significant differential looping but it can also be due to noise generated from different background correction rates in the given region. For these kinds of findings, further experimental analysis is recommended.



4) What are the loop interactions found in ETV6::RUNX1 cases within region chr10:1mb-2mb?

To address this question, we can do the following query; Analysis:Subtype specific, Subtype:ETV6::RUNX1, Resolution:10k (fixed), Query interest:Location, Predictions:All promoters. Then we put our coordinates of interest in the location panel. Location queries are limited to 1 mb for practical reasons.

This query predicts one enhancer in the region which putatively regulates *WDR37*.



As you examine the plot, you can see that the query spans a region approximately between 1mb-3.1mb. This is because there is only one enhancer loop predicted in the given query between 1mb to 2mb and it is located on chr10:3.04mb-3.05mb. The other end of this anchor which hits a promoter of *WDR37* which is located on chr10:1.1mb-1.11mb bin. For the ease of visualization we center our output based on the maximum range of CRE loops encountered in the given region with a slight addition on both sides. All enhancers found on this query (in this case only one) would be displayed as red arcs. On the track below the promoter bins, you see reverse orientation links in blue dashed lines. These also represent loops that are detected within the region but they do not overlap with a predicted enhancer.

Tracks below the loops are derived from ENCODE experiments on GM12878 to provide additional information regarding the regulatory structure of the region.