



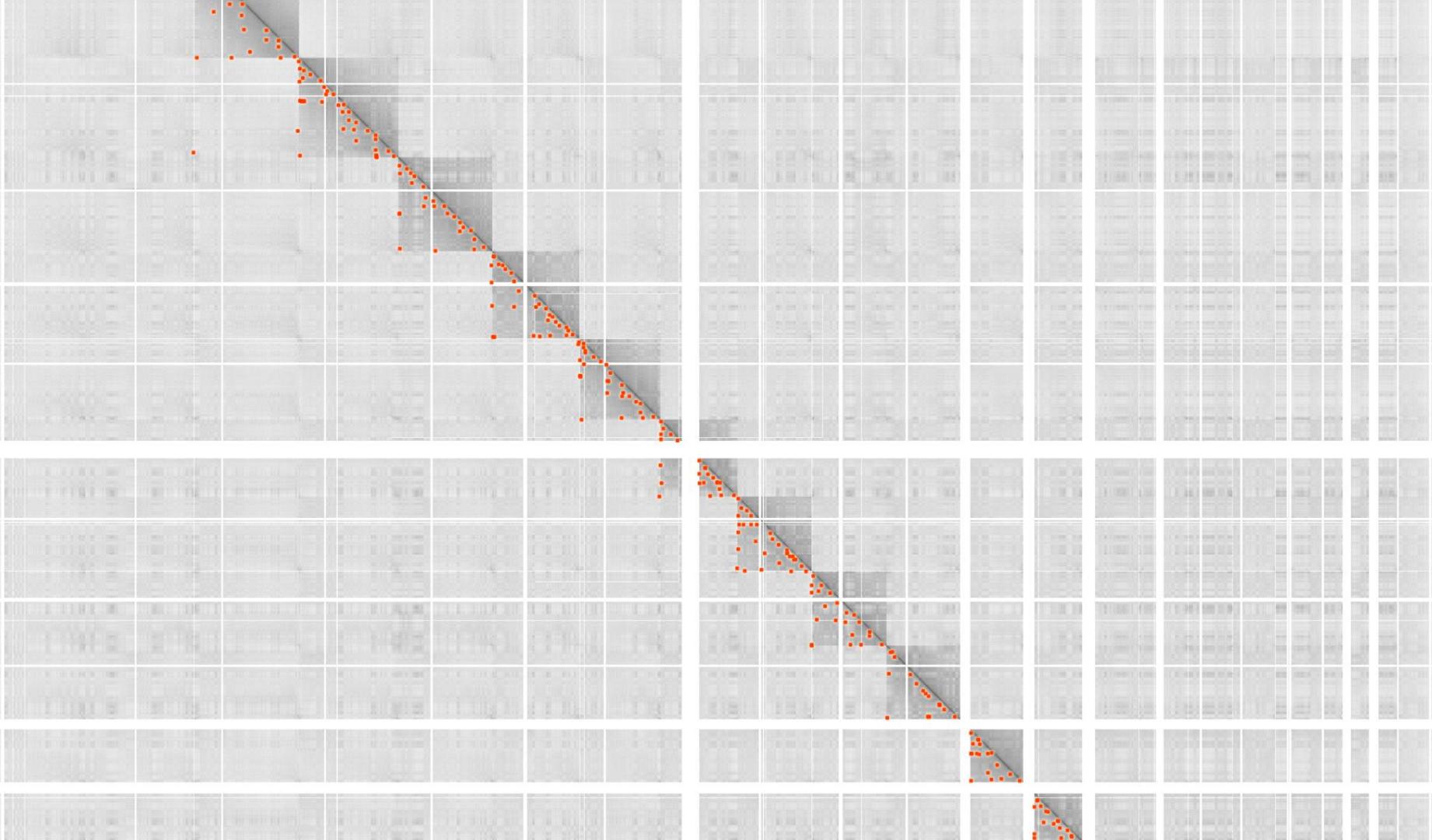
HiPiler

Exploring Many Hi-C Features Through Visual Decomposition

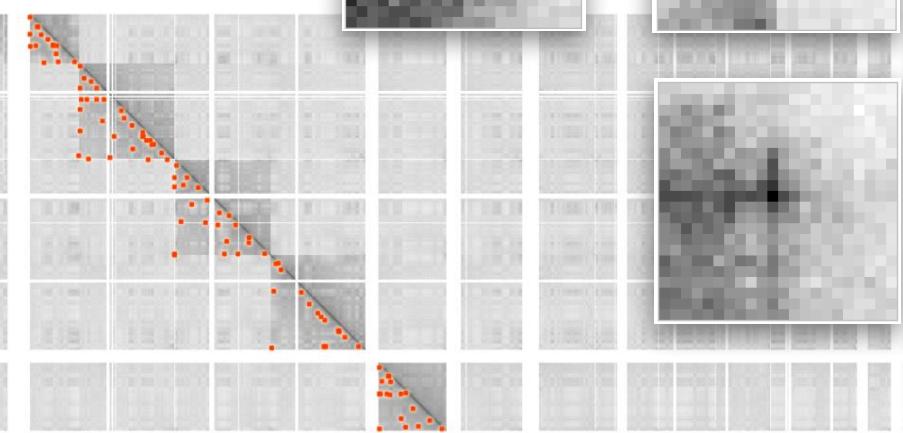
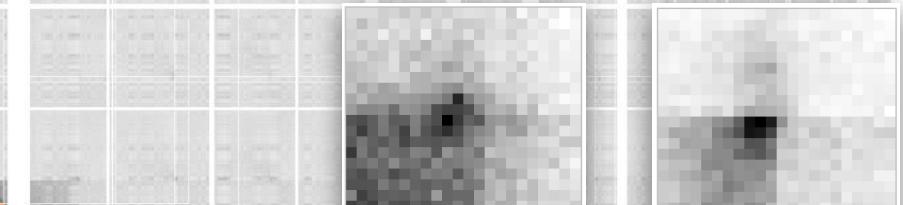
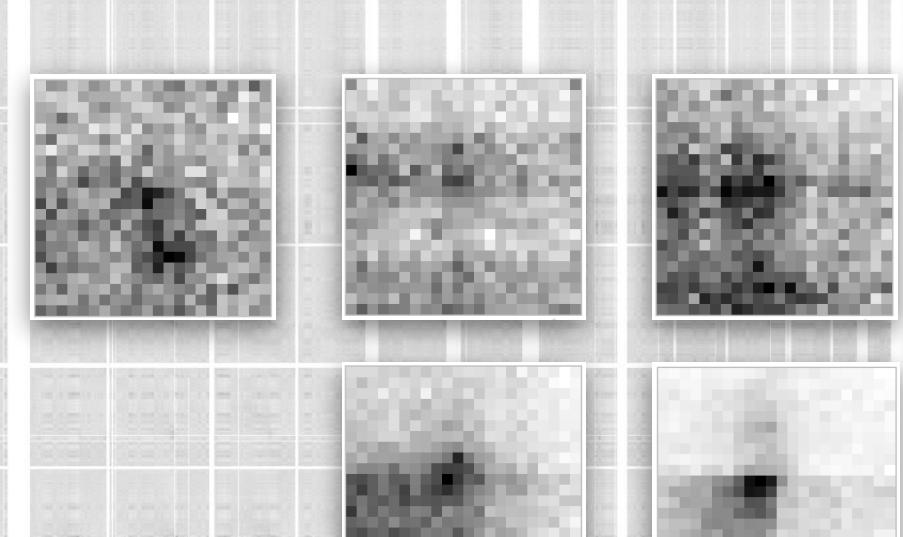
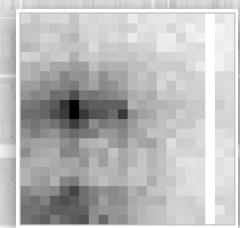
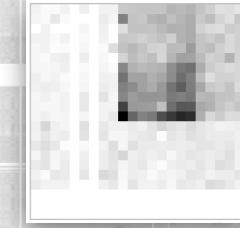
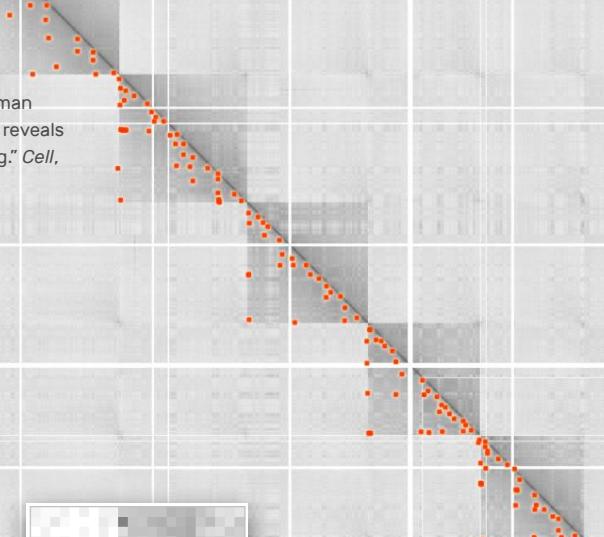
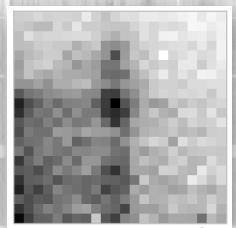
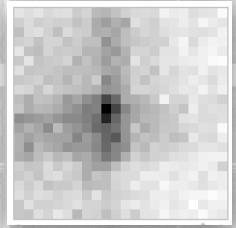
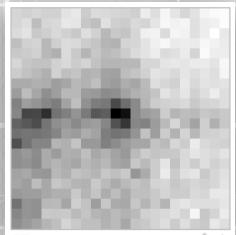
Fritz Lekschas, Benjamin Bach, Peter Kerpeljiev,
Nils Gehlenborg, and Hanspeter Pfister

... and special thanks to N. Abdennur, B. Alver, H. Belaghzal, A. van den Berg, J. Dekker, G. Fudenberg, J. Gibcus, A. Goloborodko, D. Gorkin, M. Imakaev, Y. Liu, L. Mirny, J. Nübler, P. Park, H. Strobelt, and S. Wang for their invaluable feedback during the development of HiPiler.





Rao et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." *Cell*, 159(7):1665–1680, 2014.





How does a specific or average Hi-C feature look?

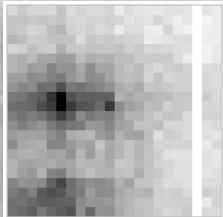
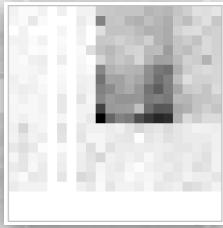
How variant and noisy are Hi-C features calls?

Are there subgroups among the extracted
Hi-C features?

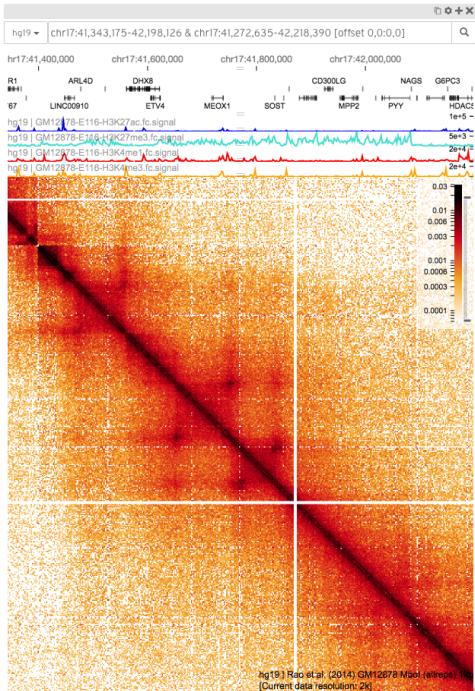
How do Hi-C features relate to other
derived attributes?

How do Hi-C features relate to
each other?

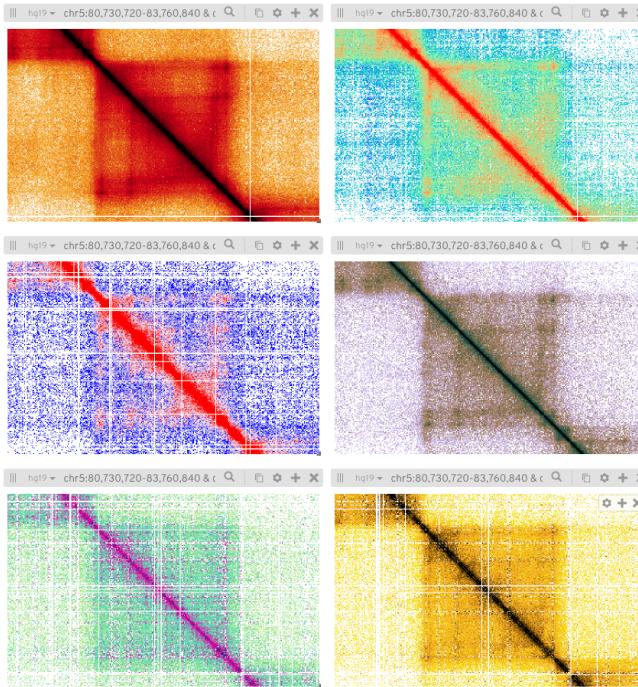
How to visually explore **many local** features?



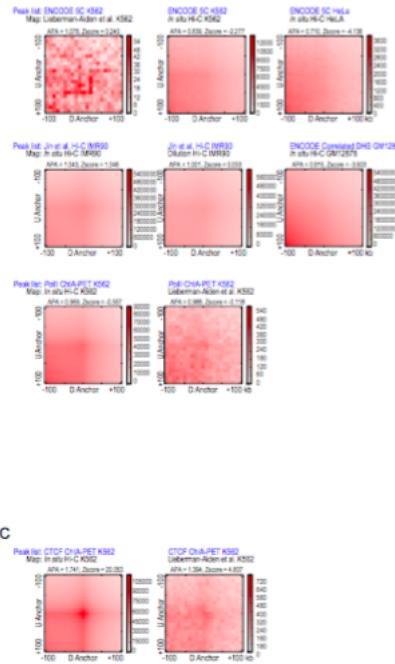
Single View



Multi View



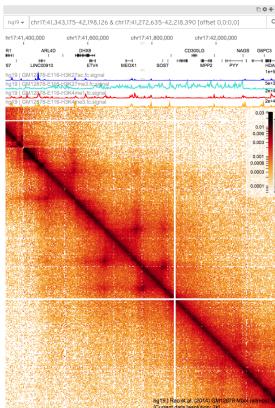
Custom View



Rao et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." *Cell*, 159(7):1665–1680, 2014.

Single View

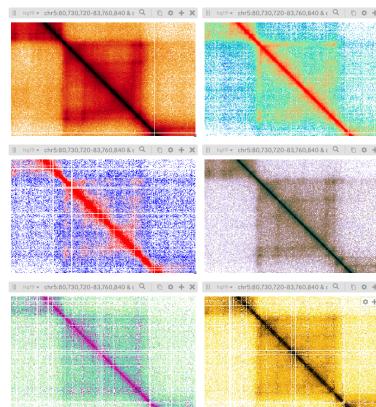
Simple to use



No comparisons

Multi View

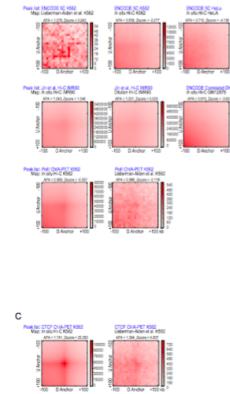
Comparison*



No aggregation

Custom View

Highly flexible



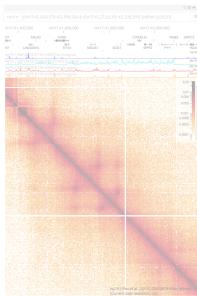
No interactions

***) Of up to handful of features**

Rao et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." *Cell*, 159(7):1665–1680, 2014.

Single View

Simple to use



No comparisons

???

Compare thousands
of features

Use metadata

Find subgroups

Inspect aggregates

Interactive

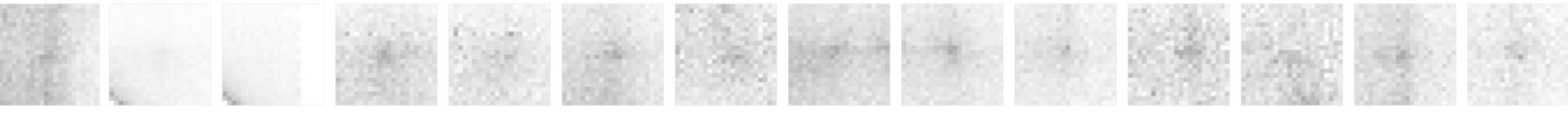
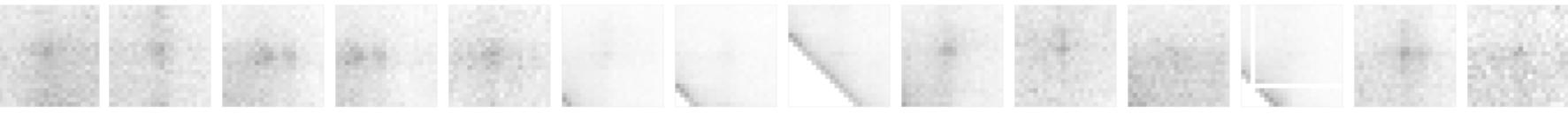
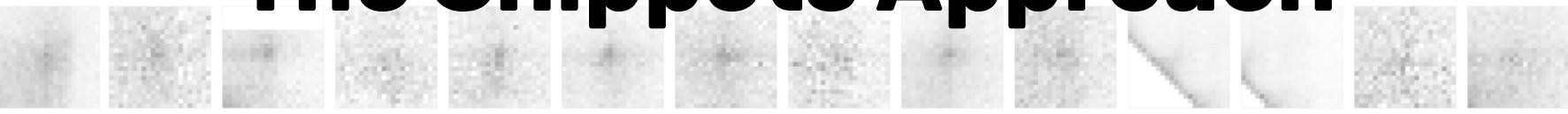
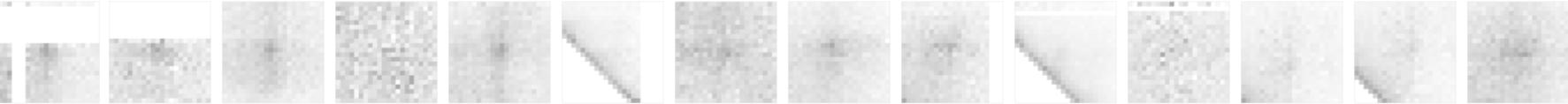
Custom View

Highly flexible



No interactions
Time consuming

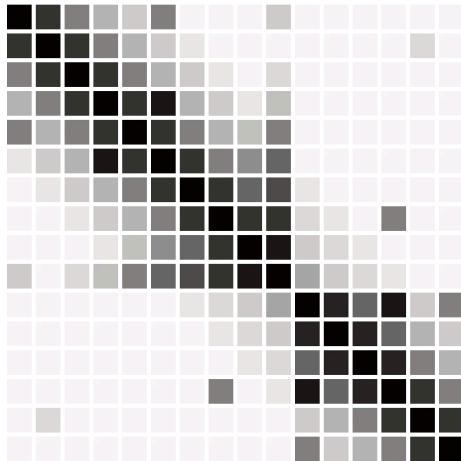
Rao et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." *Cell*, 159(7):1665–1680, 2014.



The Snippets Approach

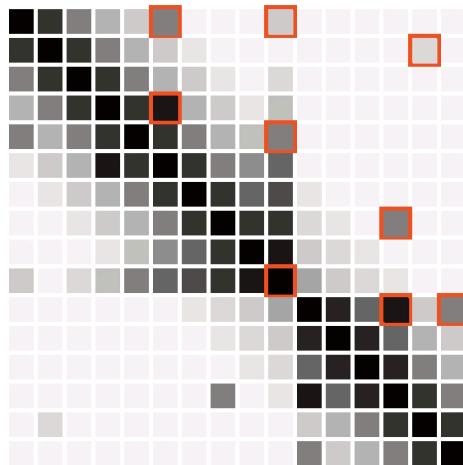
The Snippets Approach

Matrix



The Snippets Approach

Matrix

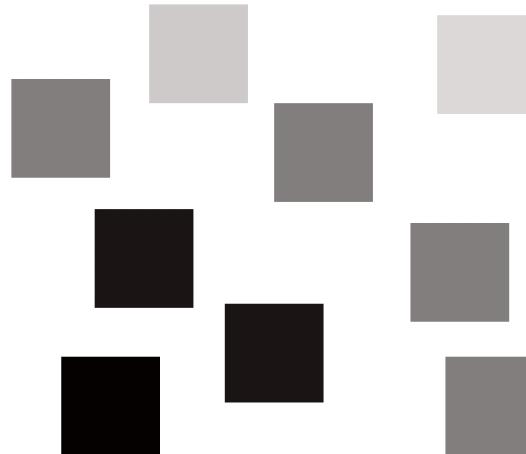


The Snippets Approach

Matrix



Snippets

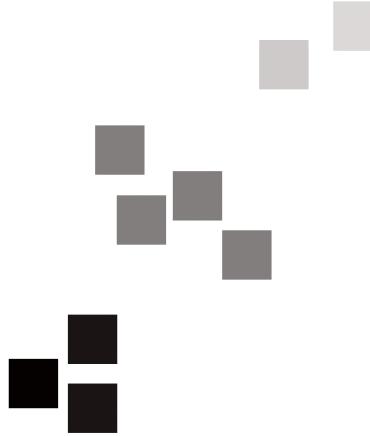


The Snippets Approach

Matrix



Snippets

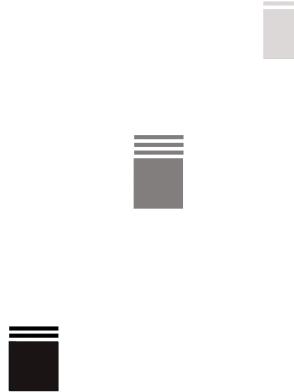


The Snippets Approach

Matrix



Snippets



Okay

but...



Okay

but...



hipiler.higlass.io

MATRIX

chr22:40,000,000



SNIPPETS

STA.



HiPiler

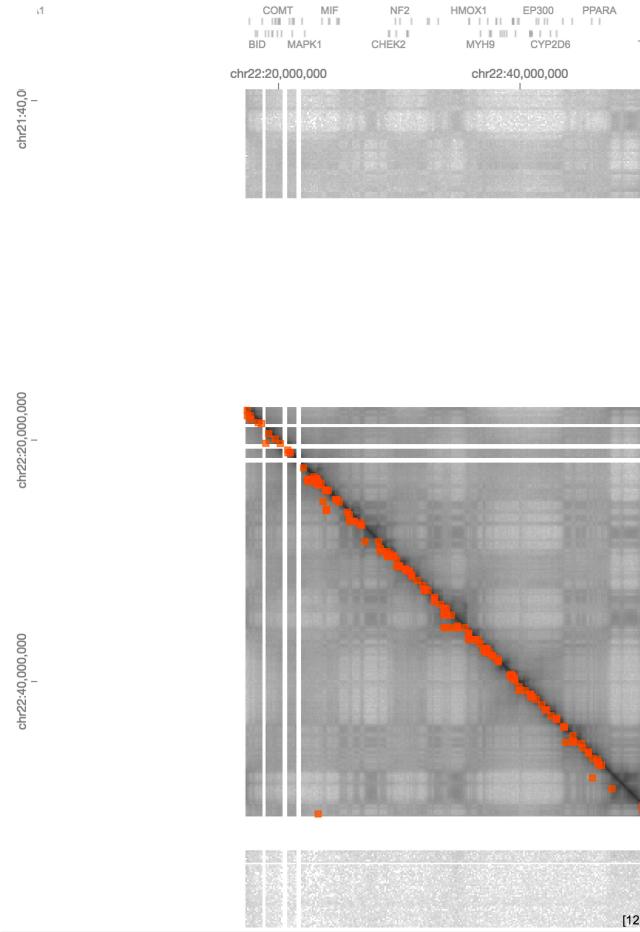
Interactive Exploration of Many Hi-C Features

chr22:40,000,000

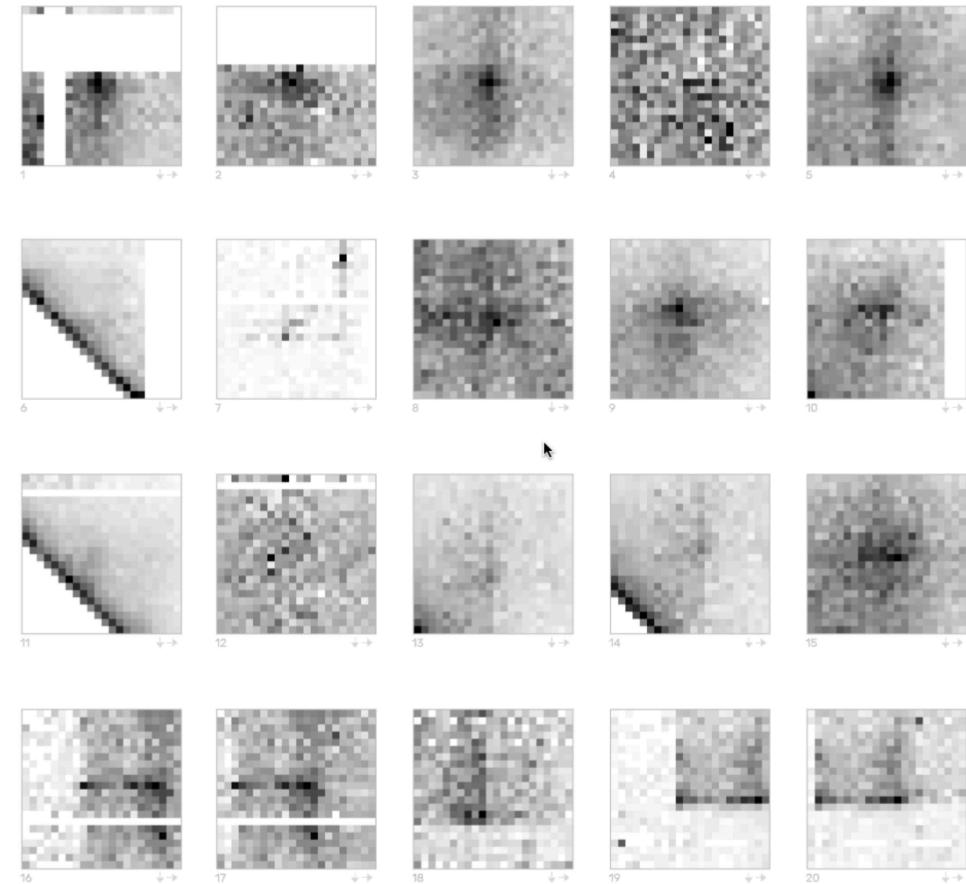


[128k]

MATRIX



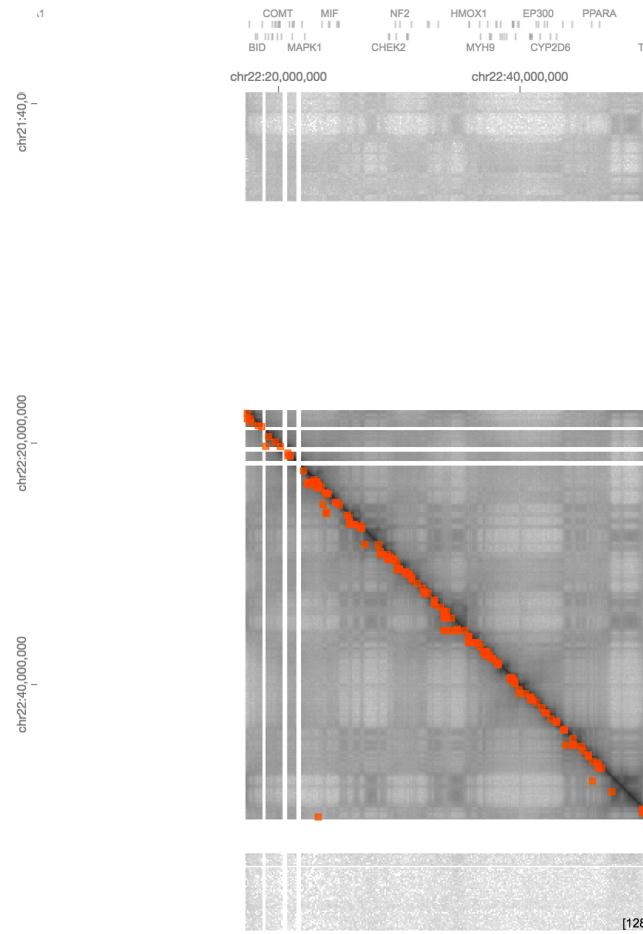
SNIPPETS



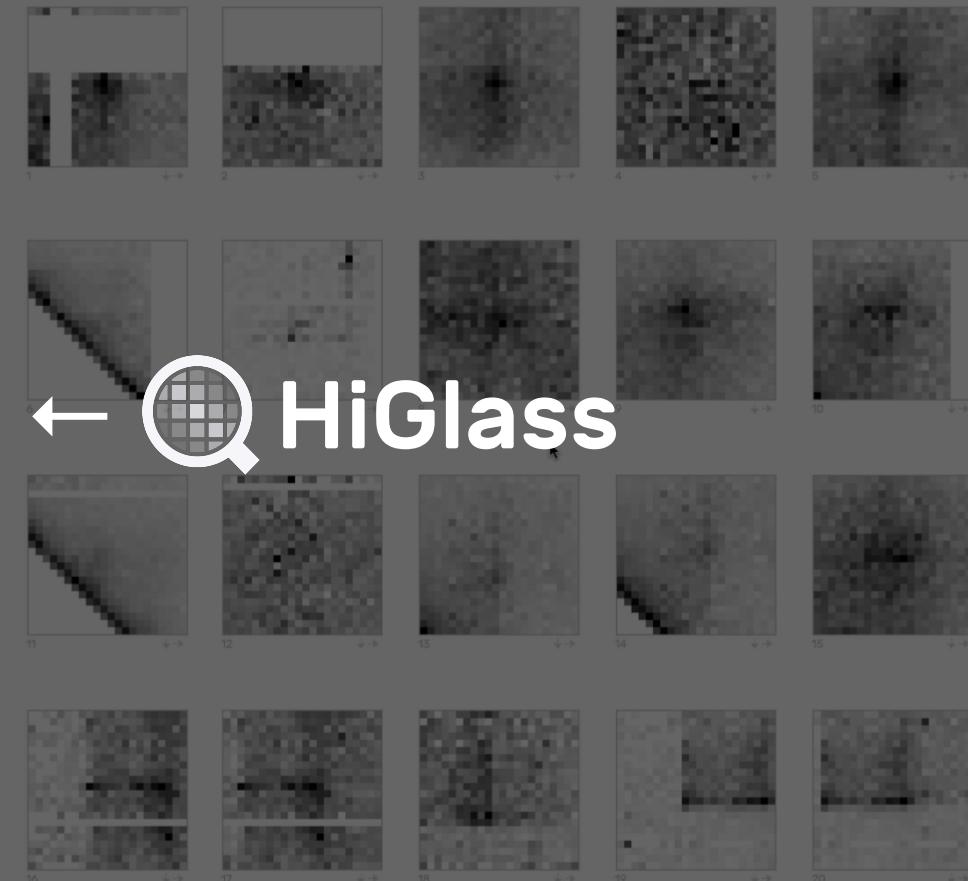
ARRANGE: Choose measures...

Cluster MORE

MATRIX



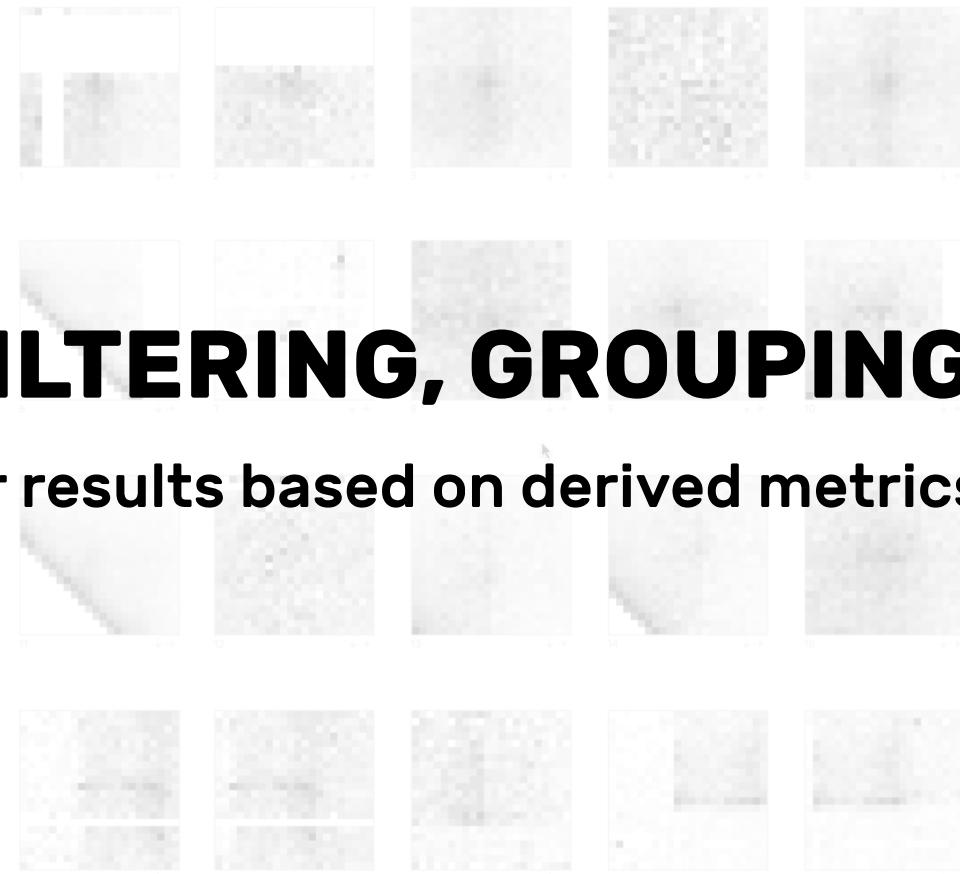
SNIPPETS



MATRIX



SNIPPETS

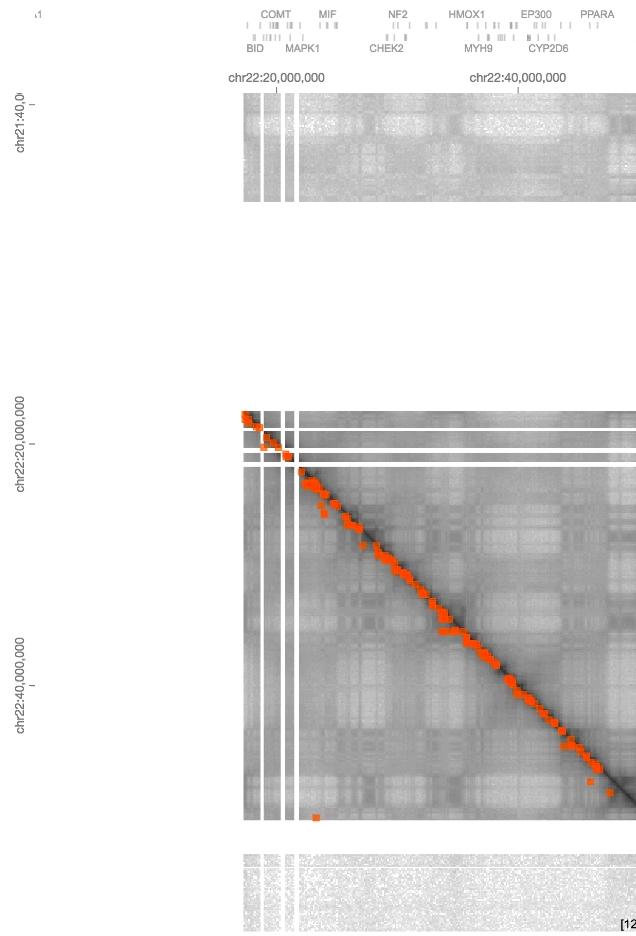


STA.

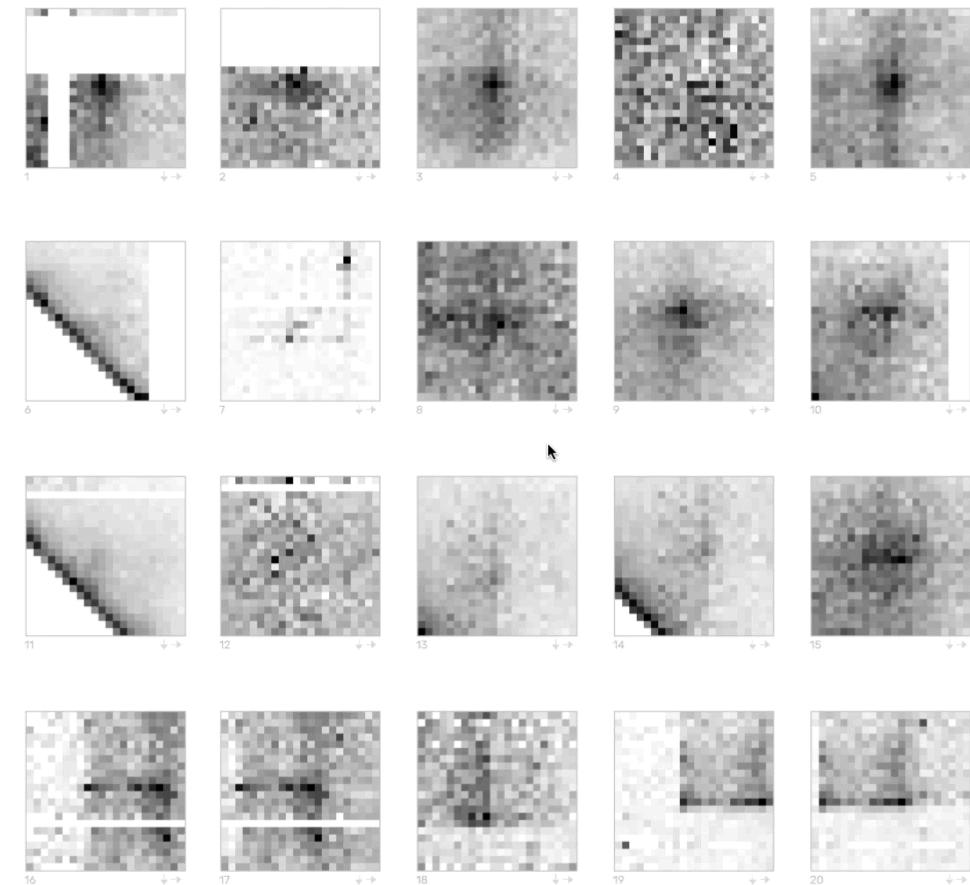
OVERVIEW, FILTERING, GROUPING

Understand and filter results based on derived metrics

MATRIX



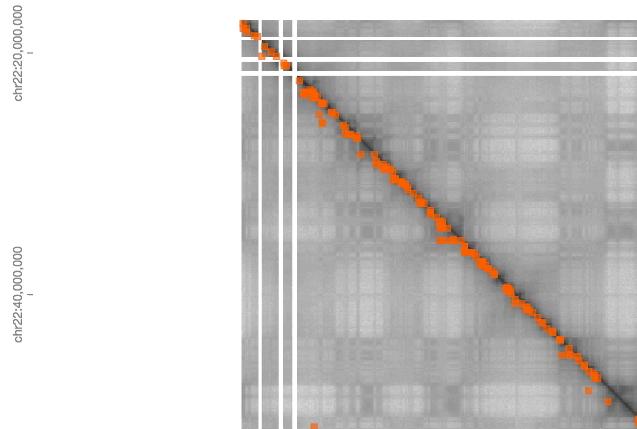
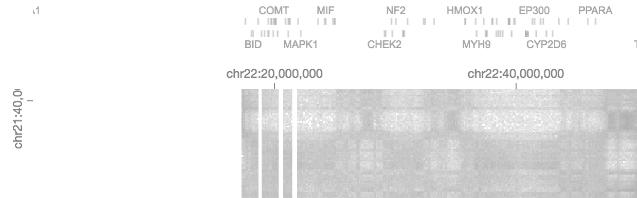
SNIPPETS



ARRANGE: Choose measures...

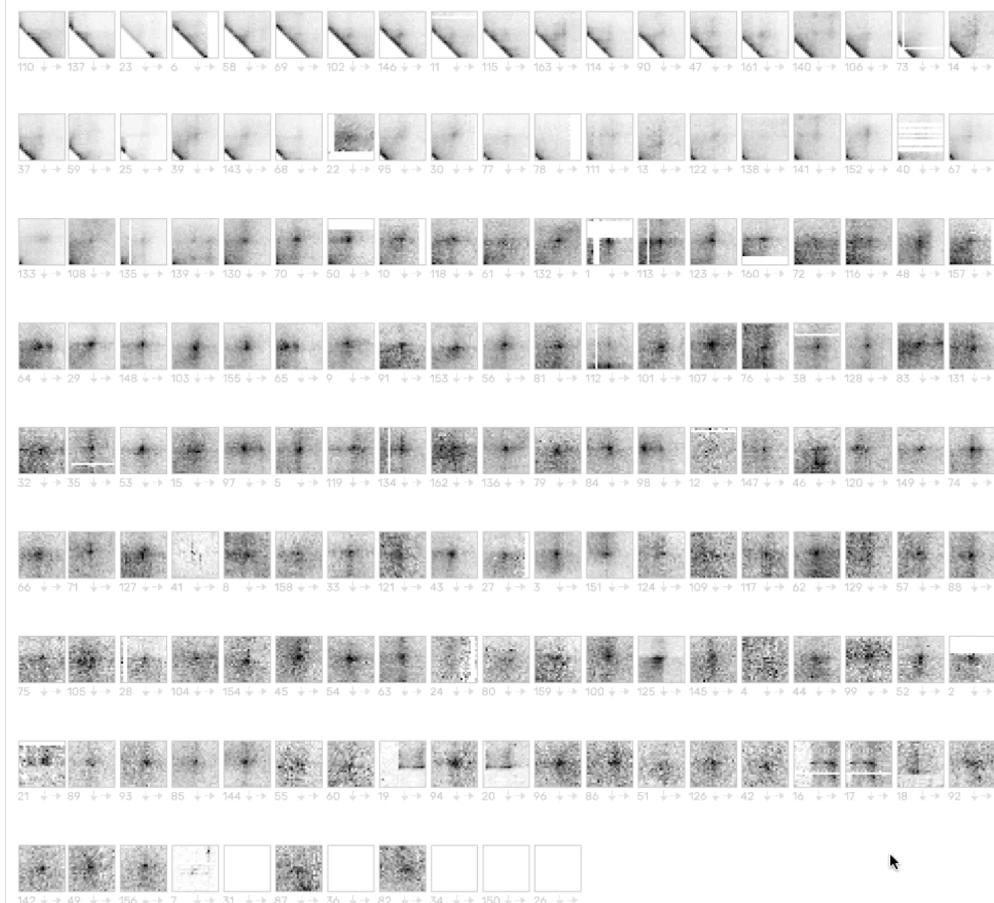
Cluster MORE

MATRIX



GRayscale HIGHL. SNIPP. DETAILS FADE OUT

SNIPPETS

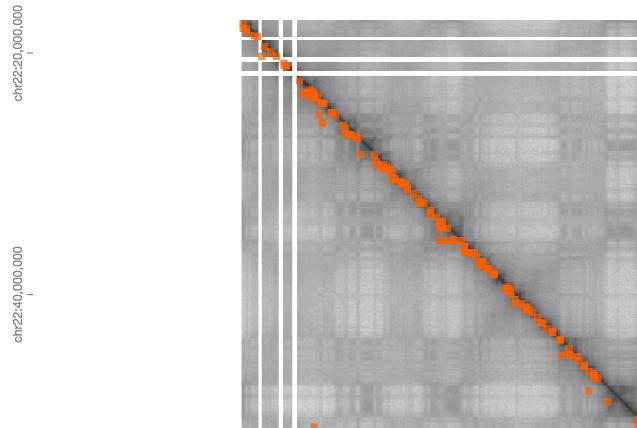
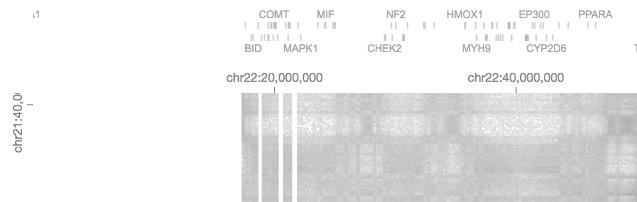


ARRANGE: Distance to diagonal ×

Cluster MORE



MATRIX



GRayscale HIGHL. SNIPP. DETAILS FADE OUT

SNIPPETS

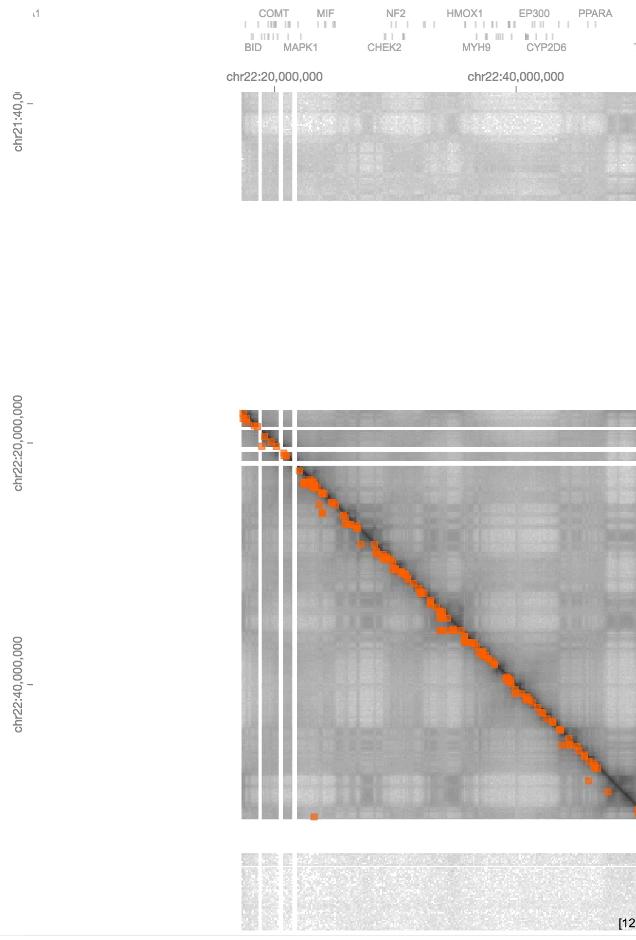


ARRANGE: Distance to diagonal ×

Cluster MORE

STA.

MATRIX



SNIPPETS

105000

0.222

Noise

0.094

Distance to diagonal

1820000

STA.

ARRANGE: Distance to diagonal × Noise ×

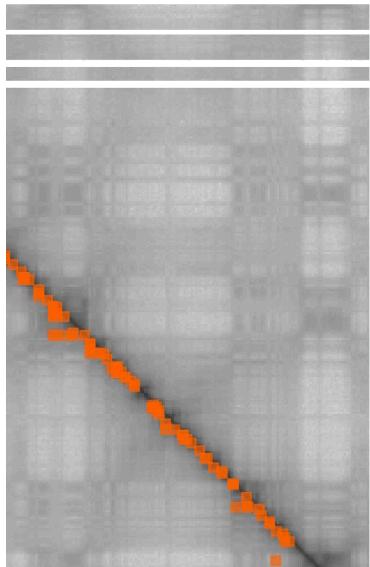
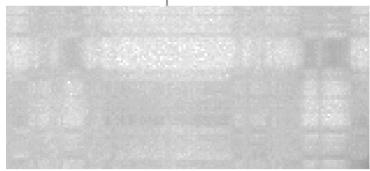
Cluster MORE

chr22:40,000,000

AGGREGATION

Assessing individual, average, and variance patterns

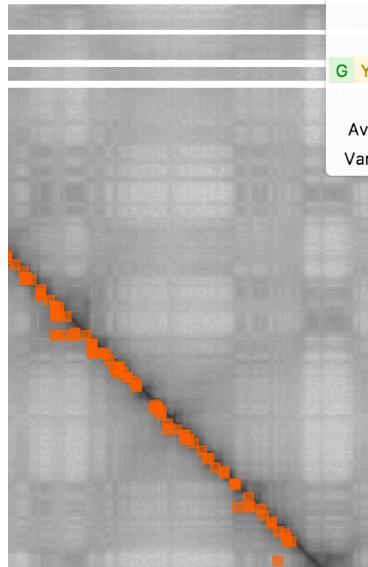
chr22:40,000,000



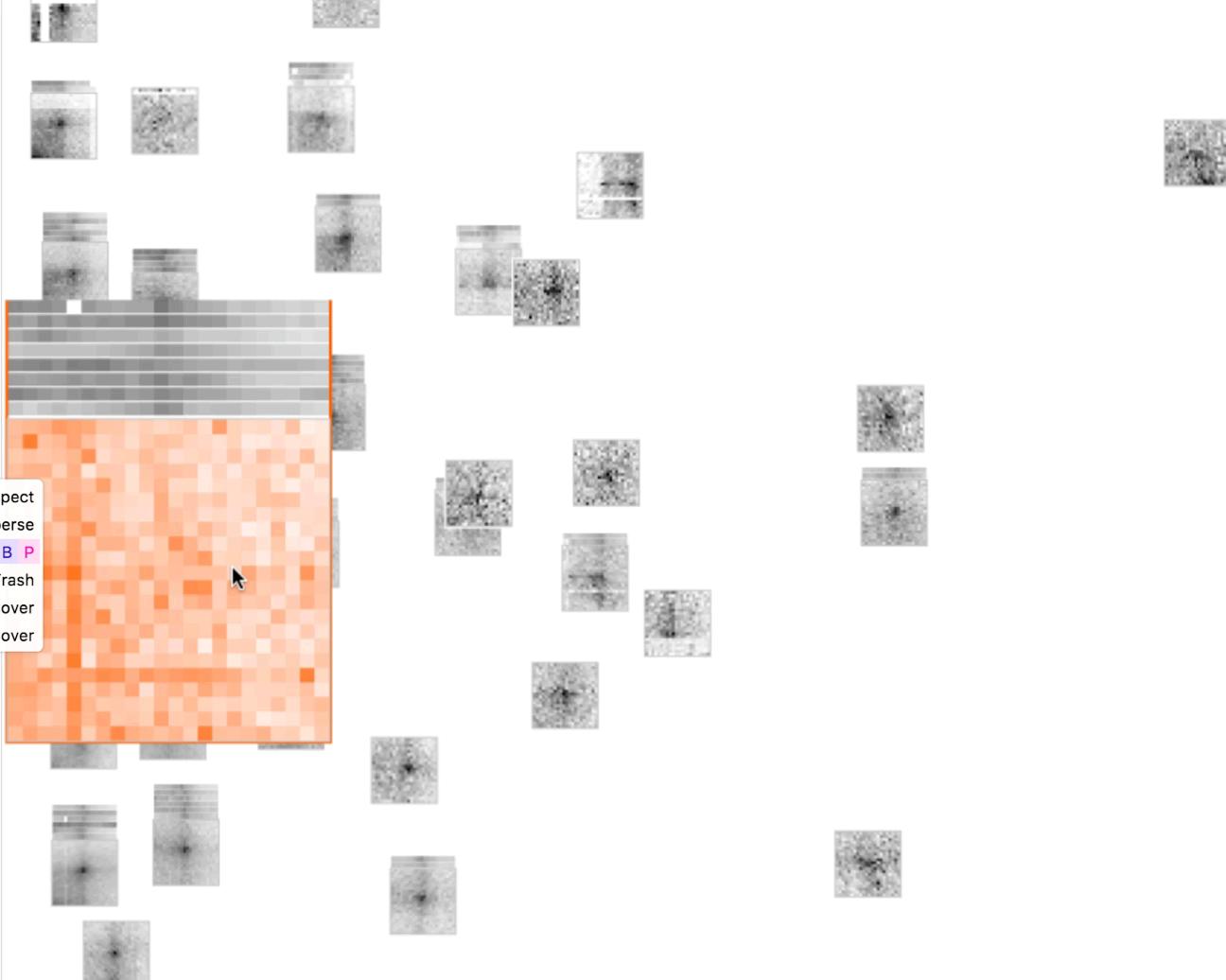
Noise



chr22:40,000,000



Inspect
Disperse
G Y C R B P
Trash
Average Cover
Variance Cover





SNIPPET MATRIX LINKING

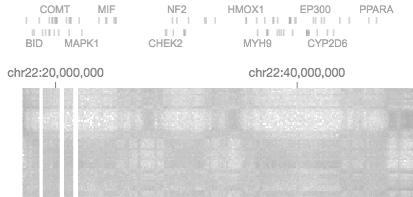


Correlation of features in their context

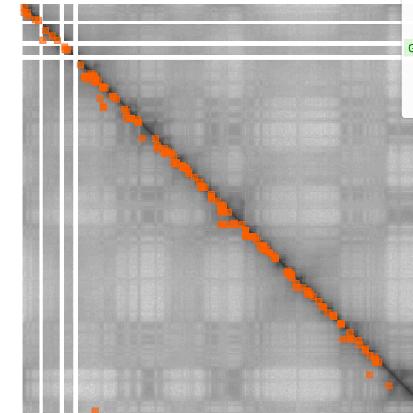


MATRIX

chr22:40,000,000



chr22:20,000,000



chr22:40,000,000



SNIPPETS

105000

0.222

0.094

2180000

Distance to diagonal

1820000

STA.

ARRANGE: Distance to diagonal × Noise ×

GRayscale HIGHL. SNIPP. DETAILS FADE OUT

Cluster MORE

MATRIX

SOD1	COL18A1
RUNX1	ITGB2

chr21:40,000,000

COMT	MIF	NF2	HMOX1	EP300	PRARA
BID	MAPK1	CHEK2	MYH8	CYP2D6	TYMP

chr22:20,000,000

SNIPPETS

105000

Distance to diagonal

1820000



SNIPPET CLUSTERING

Interactive Subgroup Exploration

chr22:26,185,136-32,021,553 & chr22:28,091,635-31,271,609 [offset 0,0:0,0]

SEZ6L, CRYBA4
TTC28-AS1
MN1
TTC28, CHEK2
AP1B1
LIF
EIF4EN1

[16k]

chr22:28,000,000

chr22:30,000,000

chr22:32,

chr22:230,000,000

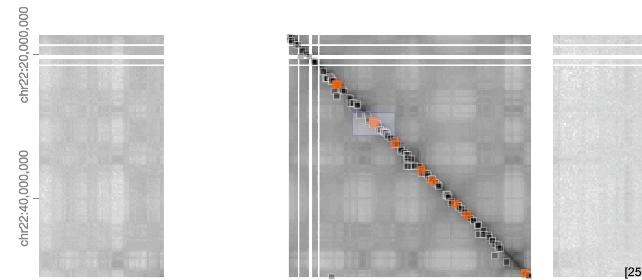
ARRANGE: Distance to diagonal × Noise ×

GRayscale HIGH-SNIPP DETAILS FADE OUT

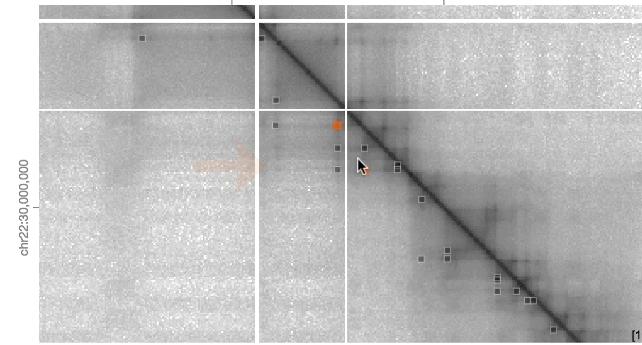


Cluster MORE

MATRIX



chr22:26,185,136-32,021,553 & chr22:28,091,635-31,271,609 [offset 0,0:0,0]



GRayscale Highl. Snipp. Details Fade Out

SNIPPETS

105000

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

0.22

MATRIX



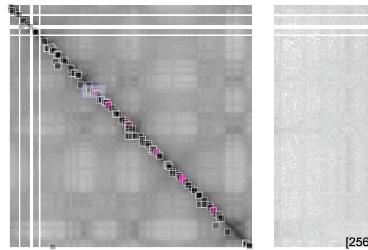
SNIPPETS

105000

0.222

Distance to diagonal

1820000



Noise

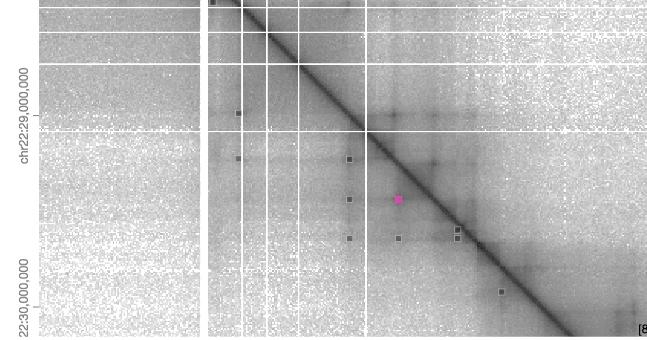
0.094

Distance to diagonal



8k

Distance to diagonal



HIGHL. SNIPP. DETAILS FADE OUT

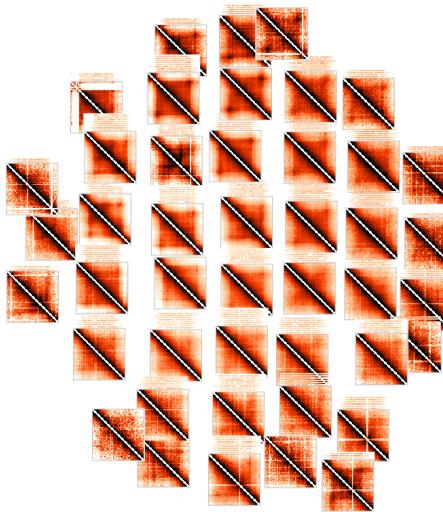
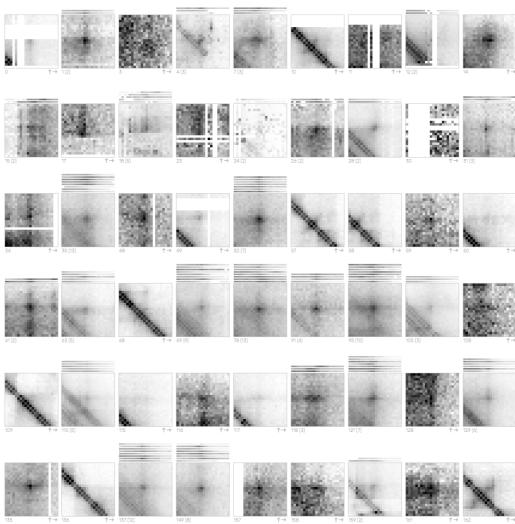


Cluster

ARRANGE: Distance to diagonal × Noise ×

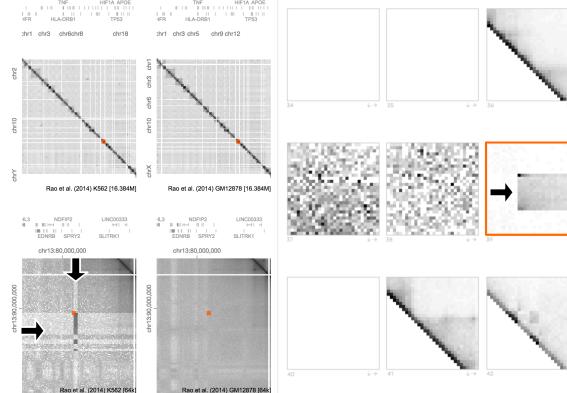
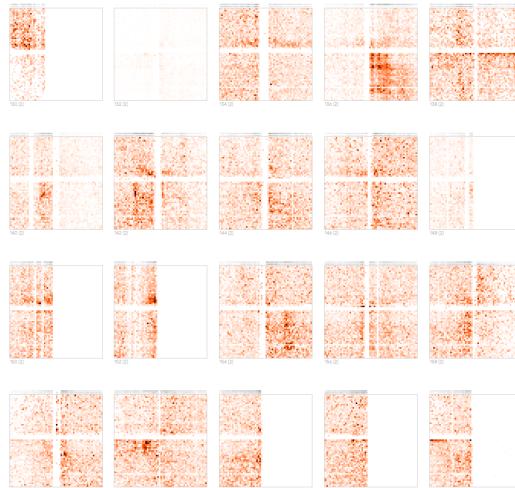
Loops

AVERAGES
SUBGROUP
FILTERING



Domains

AVERAGES
RESCALED
CLUSTERING



Structural Variation

EXPLORATION
PAIRWISE
COMPARISON

Use Cases

- Studying Hi-C features (one pattern type)

E.g.: Loops, TADs, compartments, ...

- Studying other genomic features (many pattern types)

E.g.: Genes, motifs, protein-binding sites, ...

- Compare locations

E.g.: Treatments, samples, time

A man with a beard and mustache is lying on a brown leather couch, looking very stressed. He has his head in his hands and is holding a stack of papers. He is wearing a purple striped shirt and patterned pajama bottoms. The background shows a window with a view of trees.

How do I get my BEDPE
files in
there...

A man with a beard and mustache is lying on a brown leather couch, looking very stressed. He has his head in his hands and is holding a stack of papers. He is wearing a purple striped shirt and patterned pajama bottoms. The background shows a window with a view of trees.

How do I get my BEDPE
files in
there...

Requirements

1. Multi-resolution cooler file
2. BED(PE)-like set of 2D regions (incl. derived metrics)
3. HiGlass server
4. A modern web browser (Chrome or Firefox)

Installation

1. Open hipiler.higlass.io. Done 🎉

Load loci into HiPiler

1. Create or convert BEDPE* to CSV

> Fast but predefined HiGlass view

2. Create a view config

> Slow but fully customizable HiGlass view

BEDPE TO CSV

REQUIRED
USEFUL
NUMERICAL
CATEGORICAL

<u>chrom1</u>	<u>start1</u>	<u>end1</u>	strand1	<u>chrom2</u>	<u>start2</u>	<u>end2</u>	strand2	<u>dataset</u>	<u>zoomOut_Level</u>	server	coords	pVal	<u>_group</u>
22	25000	45000	+	22	25000	45000	+	rao-gm12878-14	2	higlass.io	hg19	0.897	WT
22	25000	45000	+	22	25000	45000	+	rao-k562-14	2	higlass.io	hg19	0.833	T1
17	25000	45000	+	21	125000	145000	+	rao-gm12878-14	1	higlass.io	hg19	0.971	L1

BEDPE TO CSV

REQUIRED
USEFUL
NUMERICAL
CATEGORICAL

<u>chrom1</u>	<u>start1</u>	<u>end1</u>	strand1	<u>chrom2</u>	<u>start2</u>	<u>end2</u>	strand2	<u>dataset</u>	<u>zoomOut_Level</u>	server	coords	pVal	<u>_group</u>
22	25000	45000	+	22	25000	45000	+	rao-gm12878-14	2	higlass.io	hg19	0.897	WT
22	25000	45000	+	22	25000	45000	+	rao-k562-14	2	higlass.io	hg19	0.833	T1
17	25000	45000	+	21	125000	145000	+	rao-gm12878-14	1	higlass.io	hg19	0.971	L1



From higlass.io
(or your own instance)



Defined by you

A Peak Into the Future

(in about one week...)

Create View Config for HiPiler

1. Create or convert BEDPE* to JSON
2. Define how features should be cut out
3. Create HiGlass view for the matrix

HiPiler

View Config

```
{  
  "fgm": {  
    "fragmentsServer": "http://higlass.io/",  
    "fragments": [ ... ],  
    "fragmentsDims": 20,  
    "fragmentsPercentile": 100,  
    "fragmentsPadding": 0,  
    "fragmentsIgnoreDiags": 0,  
    "fragmentsNoBalance": false,  
    "fragmentsPrecision": 2,  
    "fragmentsNoCache": false,  
  },  
  "hgl": { ... }  
}
```

HiPiler View Config

```
{  
  "fgm": { // Defines snippets view  
    "fragmentsServer": "http://higlass.io/",  
    "fragments": [ ... ],  
    "fragmentsDims": 20,  
    "fragmentsPercentile": 100,  
    "fragmentsPadding": 0,  
    "fragmentsIgnoreDiags": 0,  
    "fragmentsNoBalance": false,  
  },  
  "hgl": { ... } // Defines HiGlass view  
}
```

HiPiler View Config

```
{  
    "fgm": {  
        "fragmentsServer        server  
        "fragments        "fragmentsDims": 20, // Number of bins  
        "fragmentsPercentile": 100, // Upper percentile capping  
        "fragmentsPadding": 0, // Padding relative to loci  
        "fragmentsIgnoreDiags": 0, // Num. of ignored diagonals  
        "fragmentsNoBalance": false, // Cooler balancing  
    },  
    "hgl": { ... }  
}
```

BEDPE JSON ARRAY

REQUIRED
NUMERICAL
_CATEGORICAL

[
 ["chrom1", "start1", "end1", "strand1", "chrom2", "start2",
 "end2", "strand2", "dataset", "zoomOutLevel", "corner-score",
 "U-var", "L-var", "U-sign", "L-sign", "_group"],
 ["22", 17425000, 17545000, "+", "22", 17425000, 17545000,
 "+", "rao-gm12878-1kbmr", 1, 0.91491, 0.061801, 0.033795,
 0.60558, 0.6278, 1],
 ["22", 17555000, 17645000, "+", "22", 17555000, 17645000,
 "+", "rao-k563-1kbmr", 1, 0.89306, 0.035257, 0.020245,
 0.54321, 0.69136, 1],
 ...
]
]

LOCI

BEDPE JSON ARRAY

REQUIRED
NUMERICAL
_CATEGORICAL

Pandas DataFrame:

```
json.dumps(  
    [list(df.columns)] + df.values.tolist()  
)
```

R Data Frame:

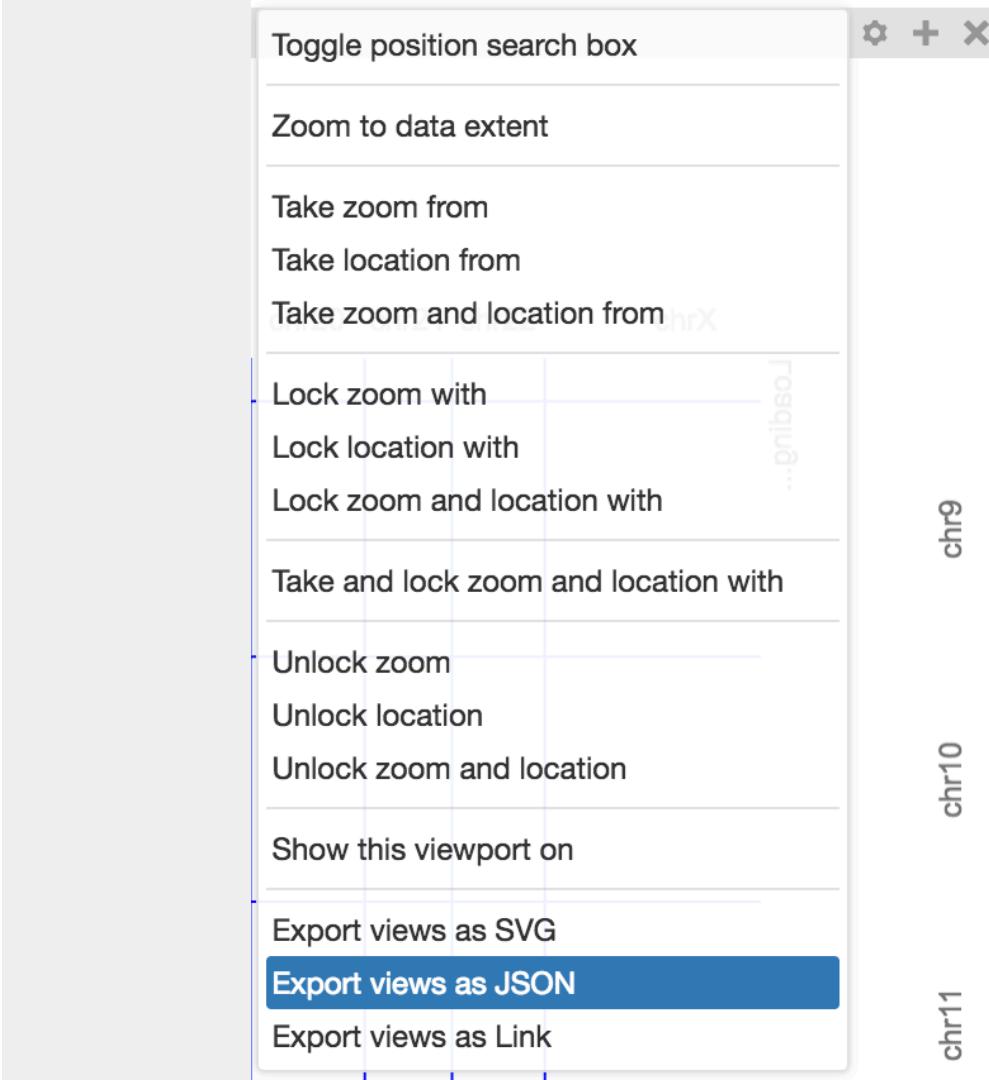
```
library(jsonlite)  
noquote(paste(  
    "[", toJSON(c(colnames(df), "name")), ",  
    substring(toJSON(df, dataframe='values'), 2),  
    sep=""  
)
```

HiGlass

View

Config

1 Row Only
Disable editing
(recommended)



Links

Examples:

<http://hipiler.higlass.io>

Example View Config:

<https://gist.github.com/flekschas/8b0163f25fd4ffb067aaba2a595da447>

Docs:

<https://github.com/flekschas/hipiler/wiki/Data#config-file>



HiPiler



- SLIDES:** github.com/hms-dbmi/hic-data-analysis-bootcamp
- DEMO:** hipiler.higlass.io
- PROJECT:** hipiler.lekschas.de
- DOCS:** hipiler.higlass.io/docs
- CODE:** github.com/flekschas/hipiler