Interact (and bigInteract) tracks in the UCSC browser

Kate Rosenbloom
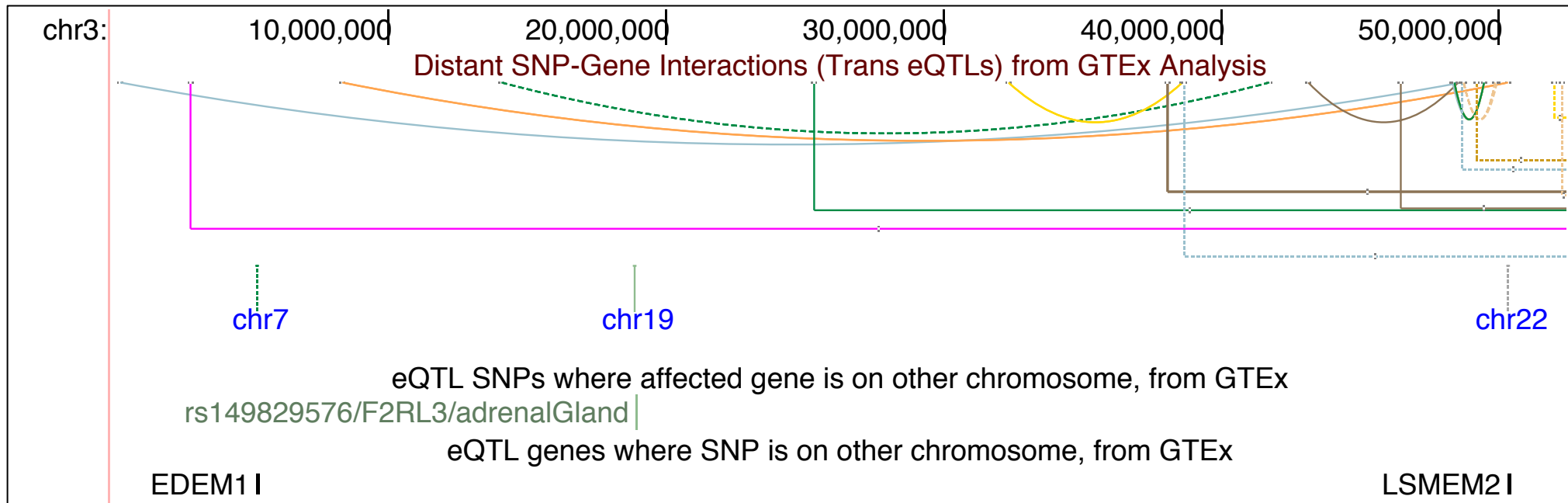
October 2018

UNIVERSITY OF CALIFORNIA SANTA CRUZ | Genomics Institute

UCSC

Motivation

- Show long-range pairwise interactions/ relationships

- Support cross-chromosome as well as same chromosome interactions

- Handle multiple data types:
  - Regulatory elements (e.g. SNP/gene)
  - Chromatin interactions (e.g. chiaPet)
  - Genomic rearrangements

Interact track features

- Curved connectors represent interactions with both endpoints on-screen. Curve type is configurable (bezier, ellipse)

- Dashed lines represent reverse direction

- Mouseover and click-through on endpoints and 'peaks' (with glyph)

- Vertical in lower band represents off-chromosome interaction. Labeled if space permits

- User-colored or grayscale coloring

Example: GTEx trans-eQTLs
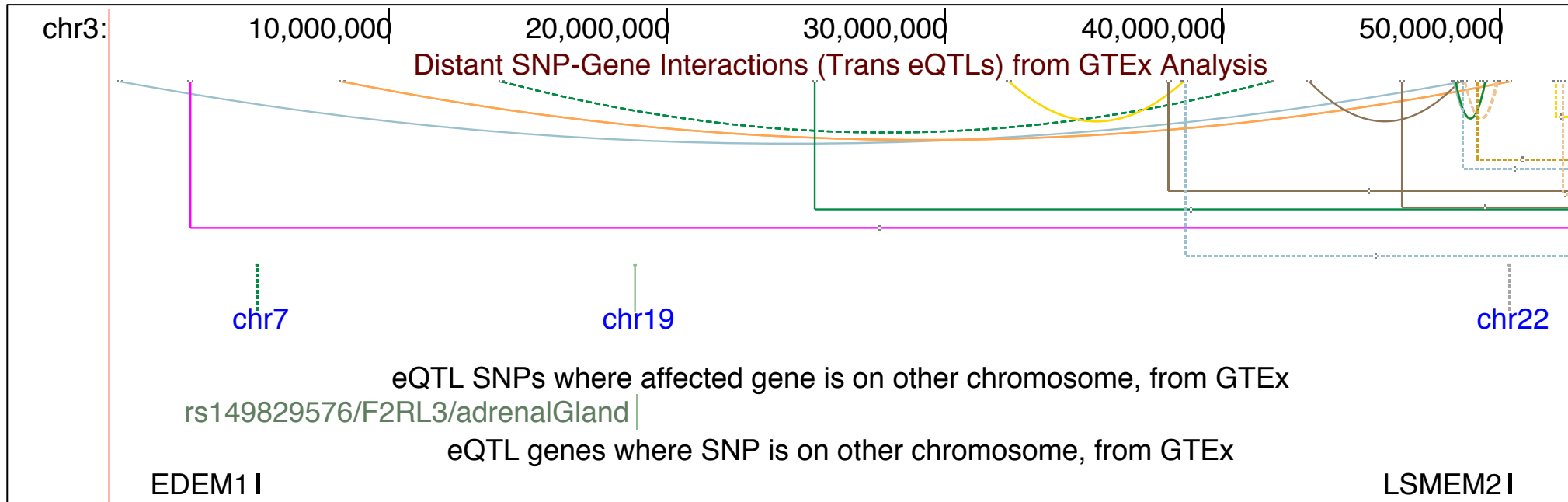
50 Mbp region of chromosome 3 where GTEx analysis identifies 15 significant long-distance eQTLs spanning the region, one variant affecting expression of a gene on another chromosome (F2RL3 on chromosome 19) and two genes (EDEM1 and LSMEM2) affected by variants on other chromosomes. Interactions are colored to indicate tissue.

Example:
chiaPET chromatin interactions
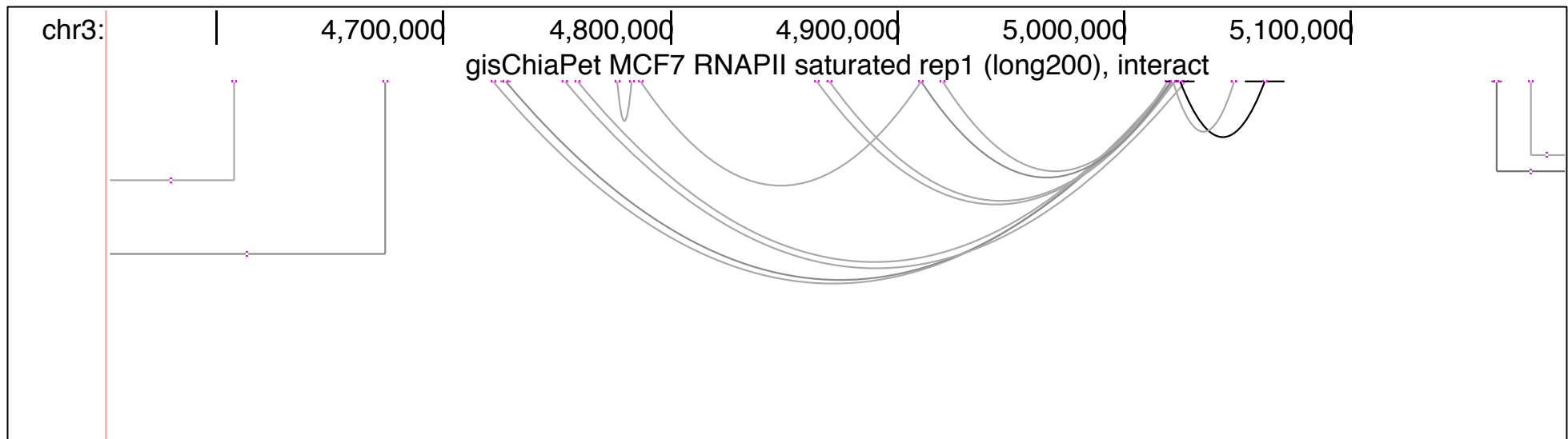
16 ChiaPet chromatin interactions identified in a 640 Kbp region of chromosome 3 in the MCF7 cell line (courtesy Genome Institute of Singapore). Interactions are displayed in gray-scale, with stronger interactions shown darker.

First native interact track (just released):
Genehancer

GeneHancer: genome-wide integration of enhancers and target genes in GeneCards

Contributors: Simon Fishelivich and Marilyn Safran, Genecards group, Weizmann Institute
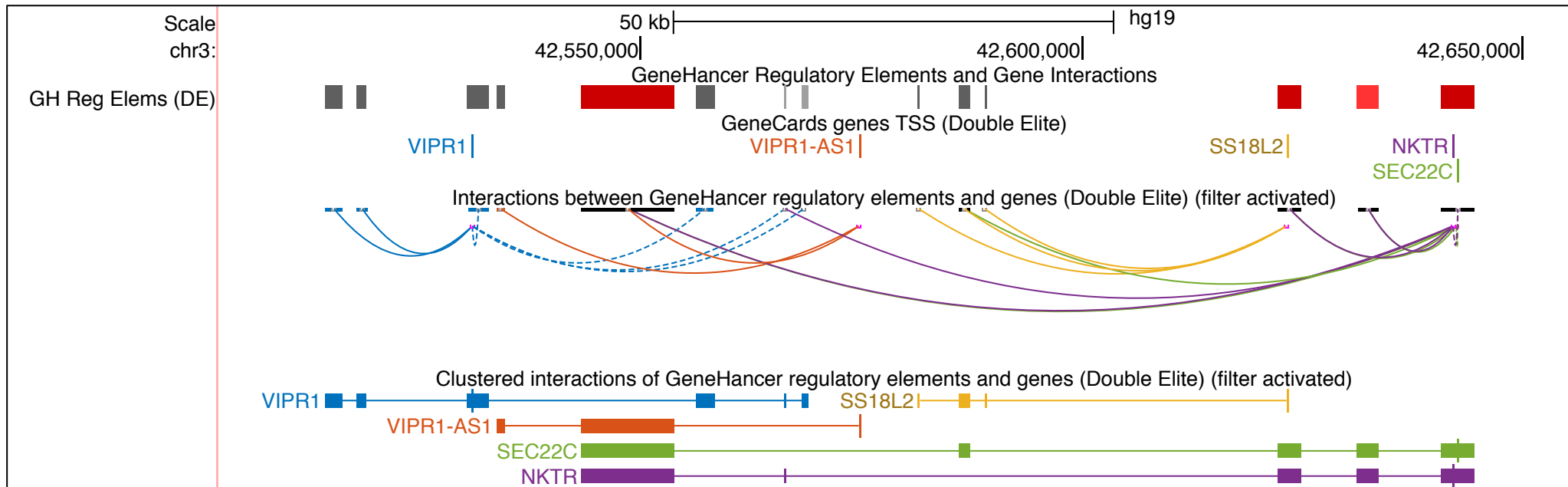

GeneHancer is a database of human regulatory elements (enhancers and promoters) and their inferred target genes, which is embedded in GeneCards, a human gene compendium. The GeneHancer database was created by integrating >1 million regulatory elements from multiple genome-wide databases. Associations between the regulatory elements and target genes were based on multiple sources of linking molecular data, along with distance.

https://genome-test.gi.ucsc.edu/cgi-bin/hgTracks?
hgS_doOtherUser=submit&hgS_otherUserName=Kate&hgS_otherUserSessionName=GeneHancer%20demo%20session

Interact track feature enhancements (Oct 2018)

1) New pack (and squish) visibilities and improved dense visibility, using linked feature display with endpoint labels
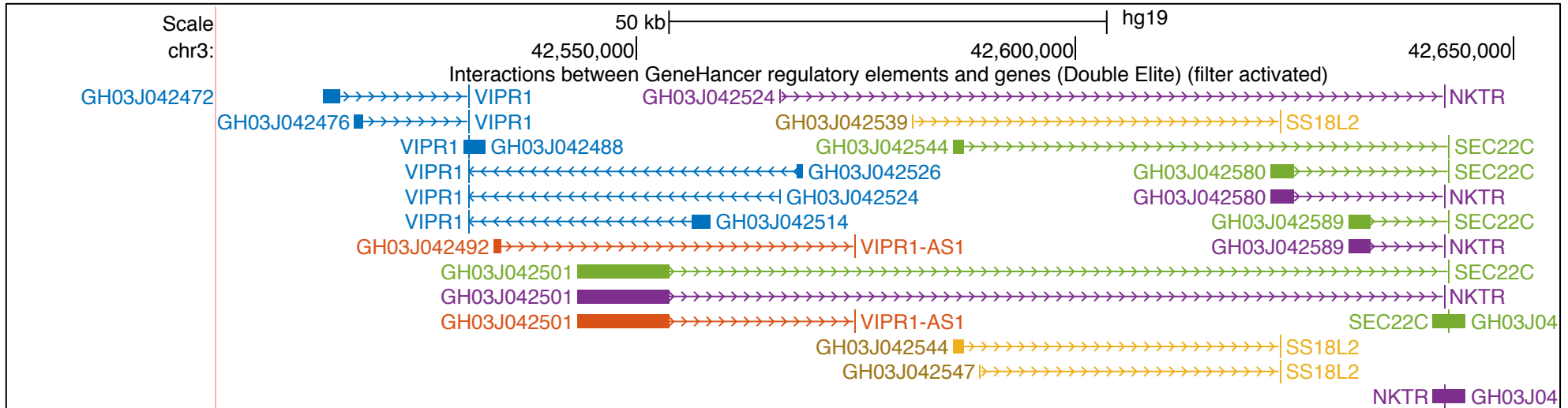
Pack visibility resolves overlapping items
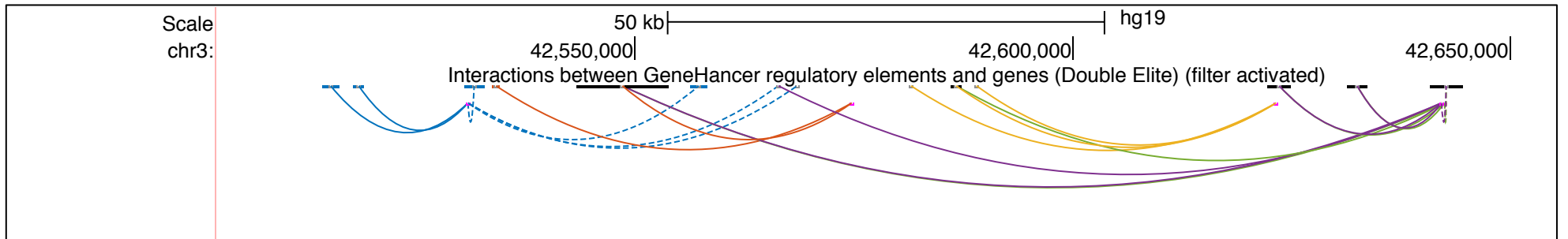
Compare to full visibility (black where colors overlap)
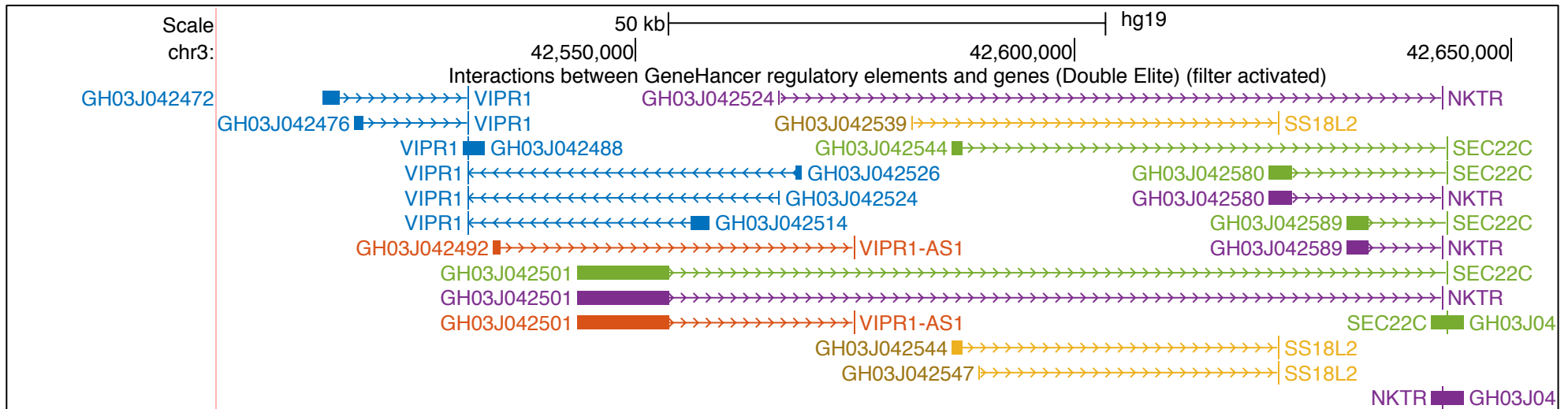
Interact track feature enhancements, cont.

2) Cluster view, groups items by source or target. *Activated by track setting.*
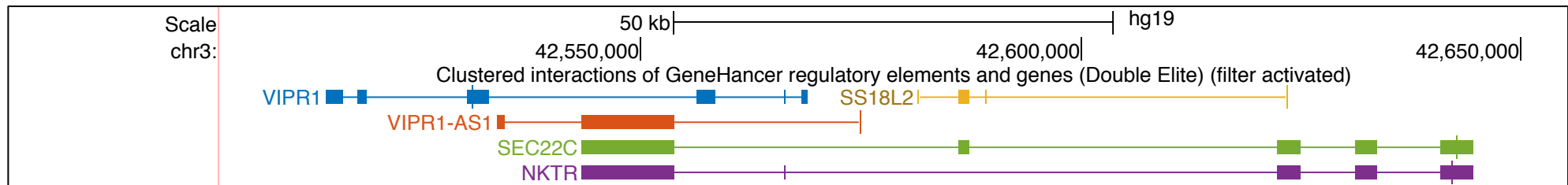
Interactions grouped by target, in pack mode

Compare to single interaction view in pack mode

Interact track feature enhancements, cont.

3) Inverted display. *Activated by track setting*.
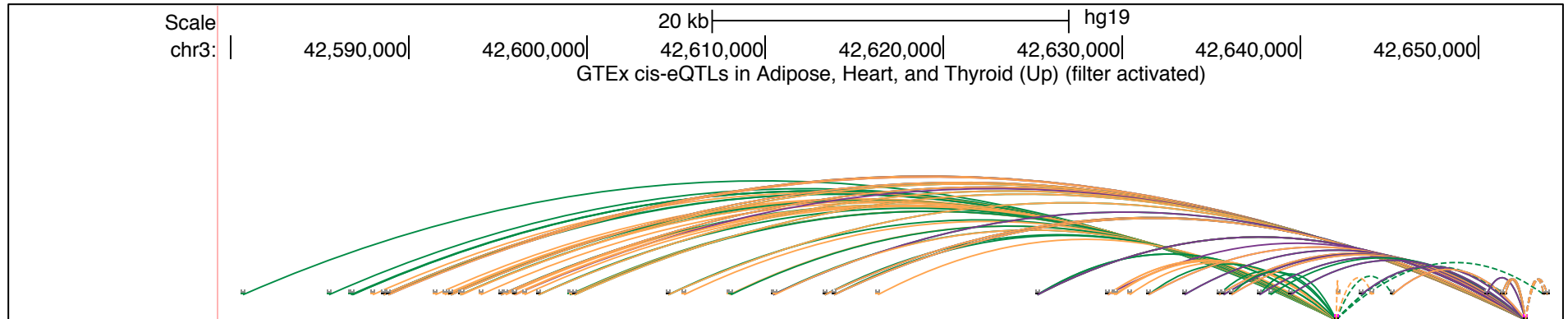
Inverted display ("hills")
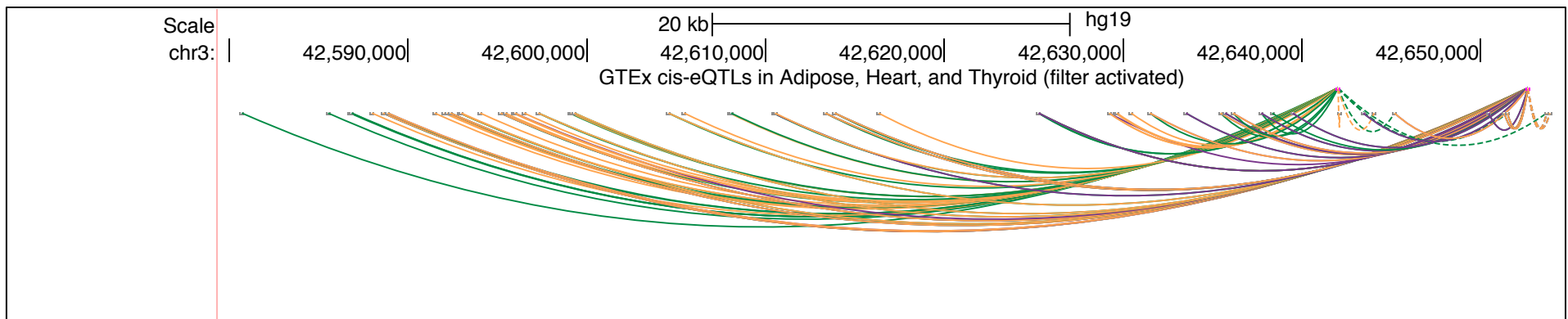
Compare to default display ("valleys")

Interact track feature enhancements, cont.

4) Offset source or target to separate visually. *Activated by track setting.*
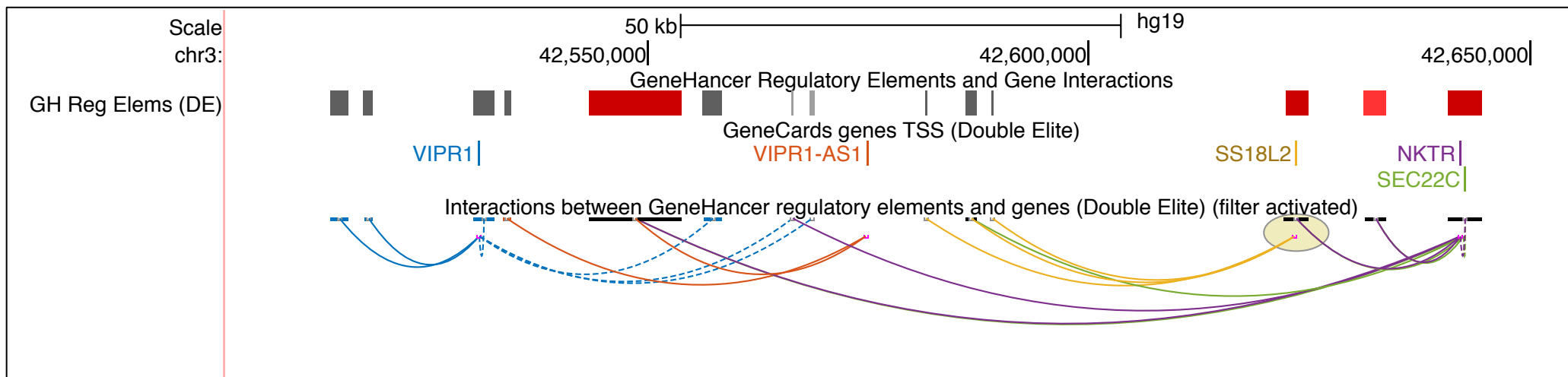
In this example, the regulatory elements (sources) are shown immediately below the track title, and the gene TSS (target) are shown on a horizontal some pixels lower in the image.

(The interaction track follows BED tracks of the regulatory elements and genes).

Interact track settings

bigInteract - Pairwise interaction display

type bigInteract

bigDataUrl <url/relativePath>

interactDirectional <true|offsetSource|offsetTarget|clusterSource|clusterTarget>

This setting is used when the interaction has an orientation (direction of effect). The offset setting shows the source (offsetSource) or target (offsetTarget) below the other end type; that is vertically displaced in the image. The interaction is drawn with dashed lines when the target region precedes the source region (reverse direction) in the genome.

The cluster setting collects all interactions with the same source (clusterSource) or target (clusterTarget) and displays each group as a single linked block display in the browser. This provides an alternate view of an interact file.

interactUp <true|false>

This setting flips the curved full visibility display so that the peak of the curves is 'up' (hills instead of valleys).

Additional settings defined in other sections are also available for displaying bigInteract tracks.

maxHeightPixels scoreMin spectrum,

Example of an interact/bigInteract track

```
track snpGeneInteractions
type bigInteract
interactDirectional true
maxHeightPixels 300:150:20
bigDataUrl http://...
```

GeneHancer Regulatory Elements and Gene Interactions (▲All Regulation tracks)

Maximum display mode: [hide ▲▼] [Submit] [Cancel] Reset to defaults

Select view (help):
Regulatory Elements ▼ [hide ▲▼]   **Gene TSS** [hide ▲▼]   **Interactions ▼** [full ▲▼]   **Clusters ▼** [pack ▲▼]

Interactions Configuration

➡ **Show interactions:** ⦿ all ◯ at least one end ◯ both ends in window

Track height: [100] pixels (range: 50 to 200, default: 100)

Draw mode: ⦿ curve ◯ ellipse ◯ rectangle

➡ ☑ Draw reverse direction interactions with dashed lines

Show only items with score at or above: [0] (range: 0 to 1000)

Select subtracks by set and view:

[+][−] All		**Double Elite**	**All**
View	**Set**	[+][−]	[+][−]
Regulatory Elements	[+][−]	☑	☐
Gene TSS	[+][−]	☑	☐
Interactions	[+][−]	☑	☐
Clusters	[+][−]	☑	☐

Clusters Configuration

Show interactions: ◯ all ◯ at least one end ⦿ both ends in window

➡ **Cluster by:** ◯ source ⦿ target

Show only items with score at or above: [0] (range: 0 to 1000)

List subtracks: ⦿ only selected/visible ◯ all (4 of 8 selected)   Top⇑

		Set↓1	View↓2	Track Name↓3	
☑	[hide] 🔧	Double Elite	Regulatory Elements	Enhancers and promoters from GeneHancer (Double Elite)	schema
☑	[hide]	Double Elite	Gene TSS	GeneCards genes TSS (Double Elite)	schema
☑	[hide] 🔧	Double Elite	Interactions	Interactions between GeneHancer regulatory elements and genes (Double Elite)	schema
☑	[hide] 🔧	Double Elite	Clusters	Clustered interactions of GeneHancer regulatory elements and genes (Double Elite)	schema

4 of 8 selected

GeneHancer trackDb

Single interactions

track ghInteraction
shortLabel Interactions
type bigInteract
interactDirectional offsetTarget
view c_I
viewUi on
maxHeightPixels 50:100:200
url https://www.genecards.org/cgi-bin/carddisp.pl?

       gene=$<geneName>
       &keywords=$<geneHancerIdentifier>
       &prefilter=enhancers#enhancers
urlLabel Interaction in GeneCards
visibility full
parent geneHancer

  track geneHancerInteractions
  shortLabel GH Interactions
  longLabel Interactions between GeneHancer regulatory elements and genes
  bigDataUrl /gbdb/$D/geneHancer/geneHancerInteractionsAll.$D.bb
  urlLabel Interaction in GeneCards
  parent ghInteraction off
  subGroups set=b_ALL view=c_I

Grouped interactions ("cluster view")

track ghClusteredInteraction
shortLabel Clustered Interactions
type bigInteract
interactDirectional clusterTarget
view d_I
url https://www.genecards.org/cgi-bin/carddisp.pl?

       gene=$<geneName>&keywords=
       $<geneHancerIdentifier>
       &prefilter=enhancers#enhancers
urlLabel Interaction in GeneCards
visibility pack
parent geneHancer

  track geneHancerClusteredInteractions
  shortLabel GH Clusters
  longLabel Clustered interactions of GeneHancer regulatory elements and genes
  bigDataUrl /gbdb/$D/geneHancer/geneHancerInteractionsAll.$D.bb
  urlLabel Interaction in GeneCards
  parent ghClusteredInteraction off
  subGroups set=b_ALL view=d_I

Implementation

- Based on longTabix track type
  - Extended drawing modes (curves, dashed)
  - New schema to support more display and data mining features, and more standard BED layout. (vs. WashU+Ensembl schema used in longTabix).
  - Additional enhancements: e.g. cross-chrom item layout, endpoint and peak mouseovers, pack mode, cluster view, inverted display

- Implemented for native tracks, custom tracks, and hubs

longTabix track type
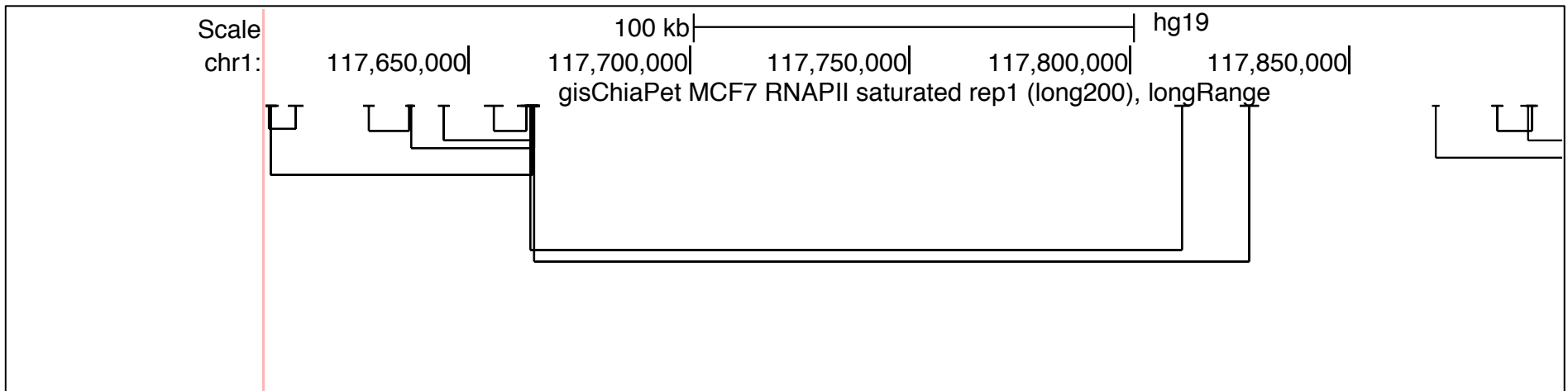
#chrom	chromStart	chromEnd	interactingRegion	id	strand
chr1	117602428	117606376	chr1:117608302-117612634,5	2155	+
chr1	117608302	117612634	chr1:117602428-117606376,5	2156	-

Format created at Wash U ("LongRange"). Somewhat extended by Ensembl:

http://wiki.wubrowse.org/Long
https://uswest.ensembl.org/info/website/upload/pairwise.html

Not documented at UCSC. Requires 2 rows/interaction. Interactions disappear when zoomed in, etc.   -> *Superceded by interact*

Interact schema

```
% cat interact.as

table interact
"Interaction between two regions"
    (
    string chrom;       "Chromosome (or contig, scaffold, etc.). For interchromosomal, use 2 records"
    uint chromStart;   "Start position in chromosome of lower region.
                                    For interchromosomal, set to chromStart of this region"
    uint chromEnd;     "End position in chromosome of upper region.
                                    For interchromosomal, set to chromEnd of this region"
    string name;        "Name of item, for display.  Usually 'name1/name2' or empty"
    uint score;         "Score from 0-1000."
    double value;       "Strength of interaction or other data value. Typically basis for score"
    string exp;         "Experiment name (metadata for filtering). Use . if not applicable"
    uint color;         "Item color, as itemRgb in bed9. Typically based on strength or filter"

    string sourceChrom;  "Chromosome of source region (directional) or lower region.
                                    For non-directional interchromosomal, chrom of this region."
    uint sourceStart;        "Start position in chromosome of source/lower/this region"
    uint sourceEnd;          "End position in chromosome of source/lower/this region"
    string sourceName;   "Identifier of source/lower/this region. Can be used as link to related table"
    string sourceStrand;  "Orientation of source/lower/this region: + or -.  Use . if not applicable"

    string targetChrom;   "Chromosome of target region (directional) or upper region.
                                    For non-directional interchromosomal, chrom of other region"
    uint targetStart;        "Start position in chromosome of target/upper/this region"
    uint targetEnd;          "End position in chromosome of target/upper/this region"
    string targetName;   "Identifier of target/upper/this region. Can be used as link to related table"
    string targetStrand;   "Orientation of target/upper/this region: + or -.  Use . if not applicable"
    )
```

NOTE: BigInteract usres are strongly encouraged to rename and describe fields to fit their data

Documentation

UCSC	Genomes	Genome Browser	Tools	Mirrors	Downloads	My Data	Help	About Us

interact and bigInteract Track Format

The interact (and bigInteract) track format displays pairwise interactions as arcs or half-rectangles connecting two genomic regions on the same chromosome. Cross-chromosomal interactions can also be represented in this format; the display shows the region on the currently viewed chromosome, with a vertical bar, labeled with the chromosome of the connected region (space permitting). For directional interactions such as SNP/gene, the interactions in the reverse direction are displayed as a dashed line or curve.

This format is useful for displaying functional element interactions such as SNP/gene interactions, and is also suitable for low-density chromatin interactions, such as ChIA-PET, and other use cases with a limited number of interactions on the genome. It is not suitable for high-density chromatin data such as Hi-C.

The interact format is available as a standalone plain text `bed5+13` format for use with smaller datasets as a custom track, and as a binary indexed format (bigInteract) suitable for track hubs and custom tracks. The bigInteract format provides more track customization features (i.e. schema customization), and is recommended for users who can use command-line tools and have web-accessible data storage. If you do not have web-accessible data storage, please see the **Hosting** section of the Track Hub Help documentation. Interact format files are converted to bigInteract files using the program `bedToBigBed` , run with the `-as` option to pull in a special **autoSql** (.*as*) schema file defining the fields of the bigInteract.

interact format definition

The following autoSql definition illustrates the basic schema supporting interact (and bigInteract) tracks.

```
table interact
"interaction between two regions"
    (
    string chrom;         "Chromosome (or contig, scaffold, etc.). For interchromosomal, use 2 records"
    uint chromStart;      "Start position of lower region. For interchromosomal, set to chromStart of this region"
    uint chromEnd;        "End position of upper region. For interchromosomal, set to chromEnd of this region"
    string name;          "Name of item, for display.  Usually 'sourceName/targetName/exp' or empty"
    uint score;           "Score (0-1000)"
    double value;         "Strength of interaction or other data value. Typically basis for score"
    string exp;           "Experiment name (metadata for filtering). Use . if not applicable"
    string color;         "Item color.  Specified as r,g,b or hexadecimal #RRGGBB or html color name, as in //www.w
    string sourceChrom;   "Chromosome of source region (directional) or lower region. For non-directional interchro
    uint sourceStart;     "Start position in chromosome of source/lower/this region"
    uint sourceEnd;       "End position in chromosome of source/lower/this region"
    string sourceName;    "Identifier of source/lower/this region"
    string sourceStrand;  "Orientation of source/lower/this region: + or -.  Use . if not applicable"
    string targetChrom;   "Chromosome of target region (directional) or upper region. For non-directional interchro
    uint targetStart;     "Start position in chromosome of target/upper/this region"
    uint targetEnd;       "End position in chromosome of target/upper/this region"
    string targetName;    "Identifier of target/upper/this region"
    string targetStrand;  "Orientation of target/upper/this region: + or -.  Use . if not applicable"

    )
```

Create your own interact track

Custom Track *file* with 4 interactions:

```
track type=interact visibility=full
# Interact                                        Source                        Target
# chrom  start     end       name    score value exp     color    chrom start     end       name strand chrom start     end       name strand
chr6    99790007  99798261  inter1  233   4.0   brain   255,0,0  chr6  99790007  99793559  rs1  .      chr6  99794572  99798261  gene1 +
chr6    99797231  99875440  inter2  175   3.2   blood   0,255,0  chr6  99797231  99801068  rs2  .      chr6  99873105  99875440  gene2 +
chr6    99834530  99843149  inter3  175   3.1   liver   0,0,255  chr6  99834530  99837680  rs3  .      chr6  99839323  99843149  gene3 −
chr6    99848067  99857519  inter4  175   3.5   kidney  0        chr6  99848067  99852034  rs4  .      chr6  99855098  99857519  gene4 +
```

First data row:

Track line:
```
track type=interact visibility=full
        description="Interact track example"
```

```
chr6       # chrom
99790007   # start
99798261   # end
inter1     # name
233        # score
4.0        # value
brain      # exp
255,0,0    # color
chr6       # sourceChrom
99790007   # sourceStart
99793559   # sourceEnd
rs1        # sourceName
.          # sourceStrand
chr6       # targetChrom
99794572   # targetStart
99798261   # targetEnd
gene1      # targetName
+          # targetStrand
```

Browser image:
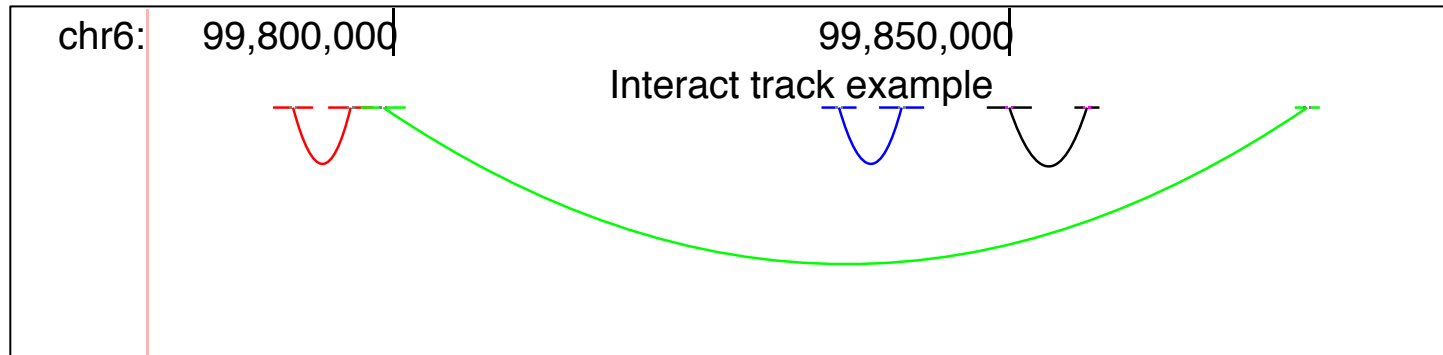
Format as bigInteract track, for a *track hub*

```
bedToBigBed —tab —type=bed5+13 —as=interact.as
      myData.inter.bed chrom.sizes myData.inter.bigBed
```

Many thanks!

- Jim, Max and Braney for design input
- Braney for engineering guidance through longTabix land
- Jairo and Lou for quality review
- And others who I have inadvertently omitted ;-(

UCSC