HiPiler: Visual exploration of large genome interaction matrices with interactive small multiples

Some background

Genome

Three billion characters long

Genome is 3D

You wish --

More like it :-/

What data are we dealing with?

Hi-C data

Interaction Matrix

Diagonally symmetric

Why: Regions of Interest (ROI)

Task
- View ROI's
- Examine pattern of ROI
- Compare patterns of different ROI's
- Preserve global context

Challenge
- Too many ROI's
- Need to zoom
- Often far away

How: HiPiler

Task abstraction (abridged)
1. Search for known patterns
2. Examine one instance
3. Compare instances

The How

Matrix

Snippets

Interaction

Context

Snippets layout

Layout (using user specified metrics/attributes)

+3D tSNE clustering
2D scatterplot
1D sorting

Reannonated from Fig. 7 of HiPiler paper

Snippet layout

Context (mainly HiGlass)

Snippets layout

Remade from Fig. 5 of HiPiler paper
**Layout: Multi-dimensional clustering**

- tSNE clustering
- Configurable

**Selection, aggregation, filtering**

- Lasso selection
- Combine snippets into a pile
- Delete snippet (pile)

**Aggregation**

- Pile cover
- In stack preview
- Inspection

**Stated Limitations**

- Only square ROI
- Fixed rows/cols order
- Scalability (beyond 2000 snippets)

**Evaluation**

- Five domain experts/users
- 1-2 hours interview
  - Training
  - Pre-prepared data
  - Own data

**Evaluation feedback**

- Good snippet/matrix linking
- Validated (behaviorally) task abstractions
- Easy to learn
- Big improvement on state-of-the-art

**Summary**

**What?**

- Interaction matrix

**Why?**

- Explore and validate

**How?**

- Snippets:
  - Linked
  - Laid out
  - Select/Aggregate/Filter

**Critique**

- Tool:
  - Open source
  - Nice aesthetic
  - Does what it says!

- Paper:
  - Easy to read
  - Great online resources (demo, video, slides, docs, etc.)

**Great!**

- Tool:
  - Open source
  - Nice aesthetic
  - Does what it says!

- Paper:
  - Easy to read
  - Great online resources (demo, video, slides, docs, etc.)

**Not so great :-(**

- Software:
  - Selection sucks, real bad
  - Slow (too much on the go computation)
  - Heavy on client-side memory
  - Unintuitive settings
  - Very buggy
  - Too much installation overhead (server)

**Thanks! Questions?**