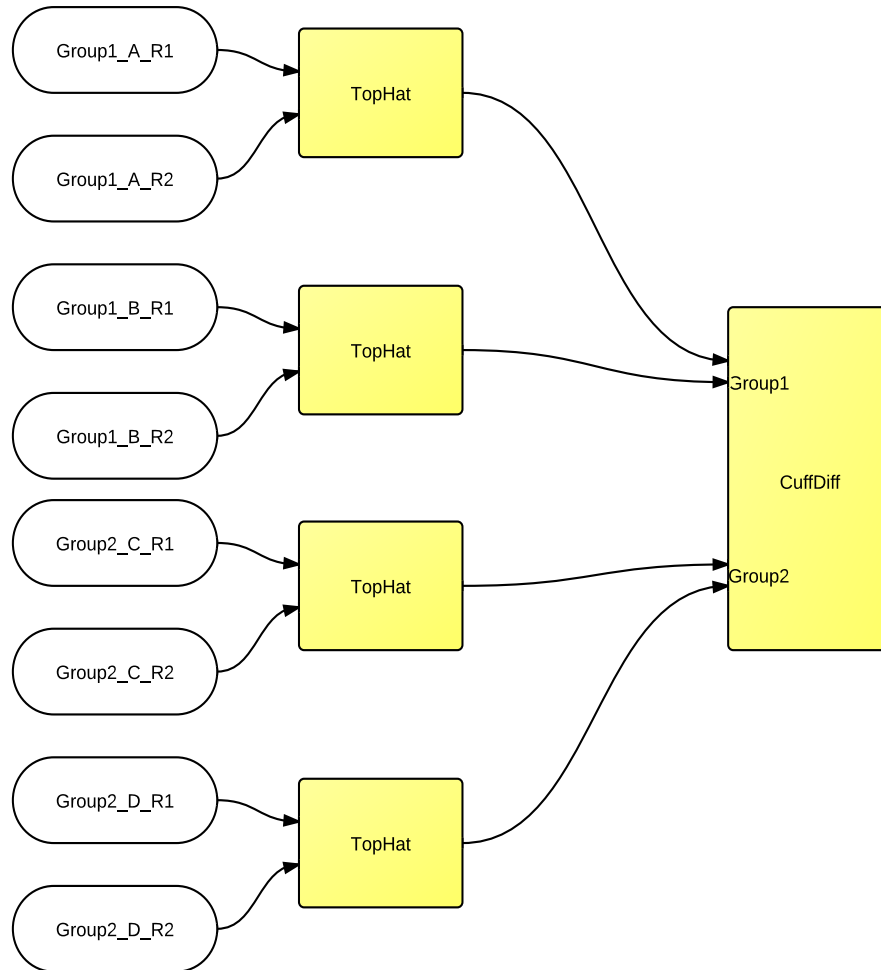


Reusable, scalable workflows
or
how do we do high-throughput
analysis with **less clicking**?



Scalability

2 conditions x 2 replicates x 2 paired reads
= 8 input files

3 conditions x 10 replicates x 3 lanes per sample
x 2 paired reads = 180 input files

Can (mostly) solve with API:

- Use histories API to get input data, and workflows or tools API to run TopHat
- Use tools API to add the Cuffdiff step with all the right intermediate files hooked up as inputs

We'd need to extend the API to fill in some gaps, e.g. if running workflows you can't use set-at-runtime parameters.

Can solve completely with UI... if you don't mind a LOT of mouse-clicking each time you run new data!