Activities in SVs, focusing on breakpoint characterization
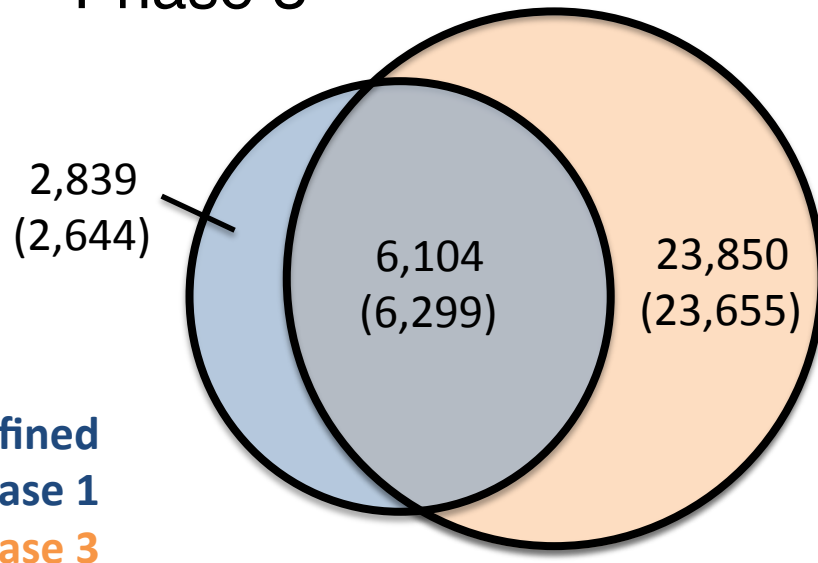
Mark Gerstein, Yale

Our Activities Related to SVs

- SV calling (eg Retroduplications)
- Functional enrichment
- **Breakpoints/Mechanism study**
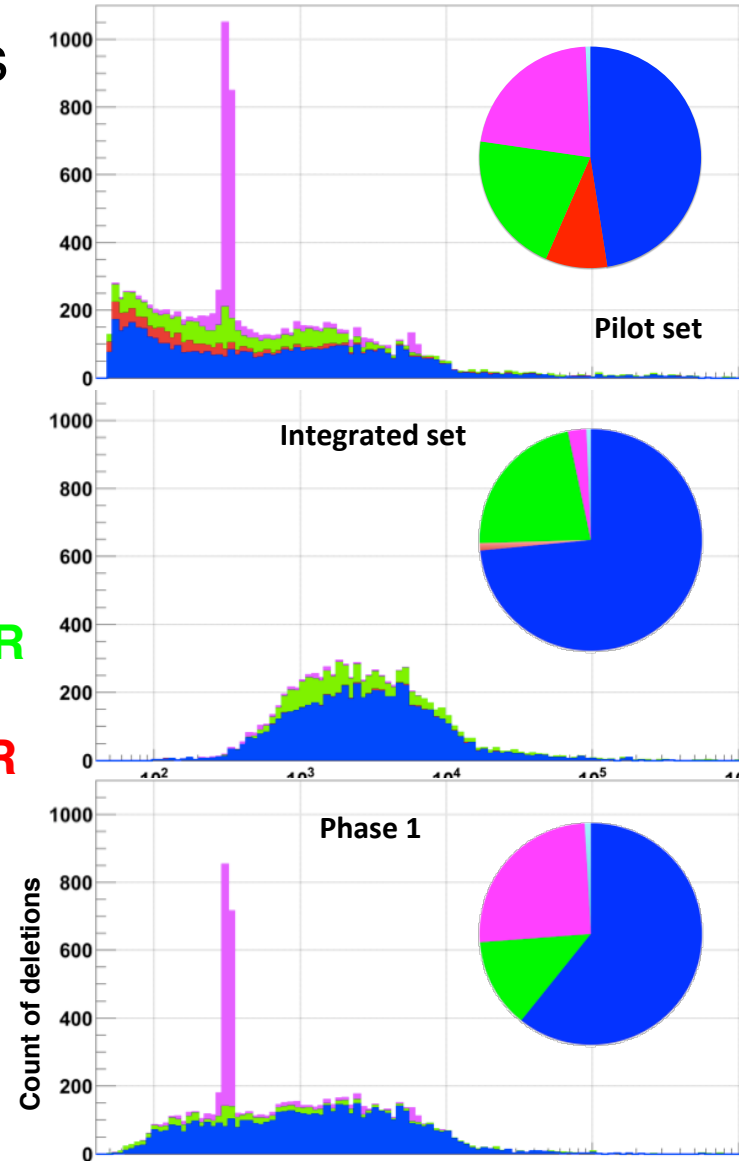
Breakpoint characterization in 1000G

- Breakseq #1 w/ ~2000 breakpoints [Lam et al. Nat. Biotech. ('10)]

- Pilot
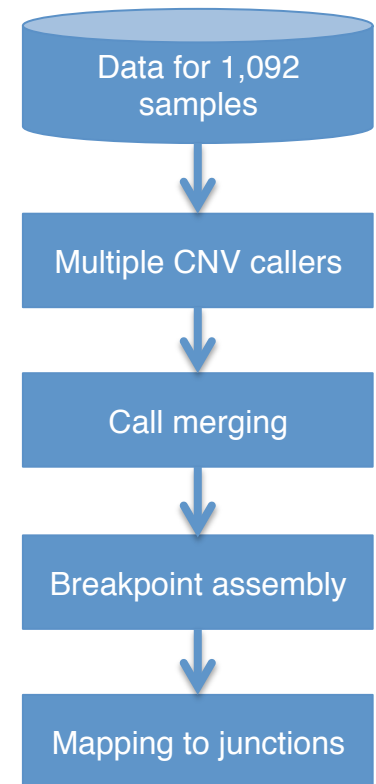
- Phase 1 "Integrated" & Phase 1 refined

- Phase 3

2,839
(2,644)

6,104
(6,299)

23,850
(23,655)

Refined
Phase 1
Phase 3

Exact match
Number in parentheses: >50% reciprocal match

TEI
NAHR
NH
VNTR

Pilot set

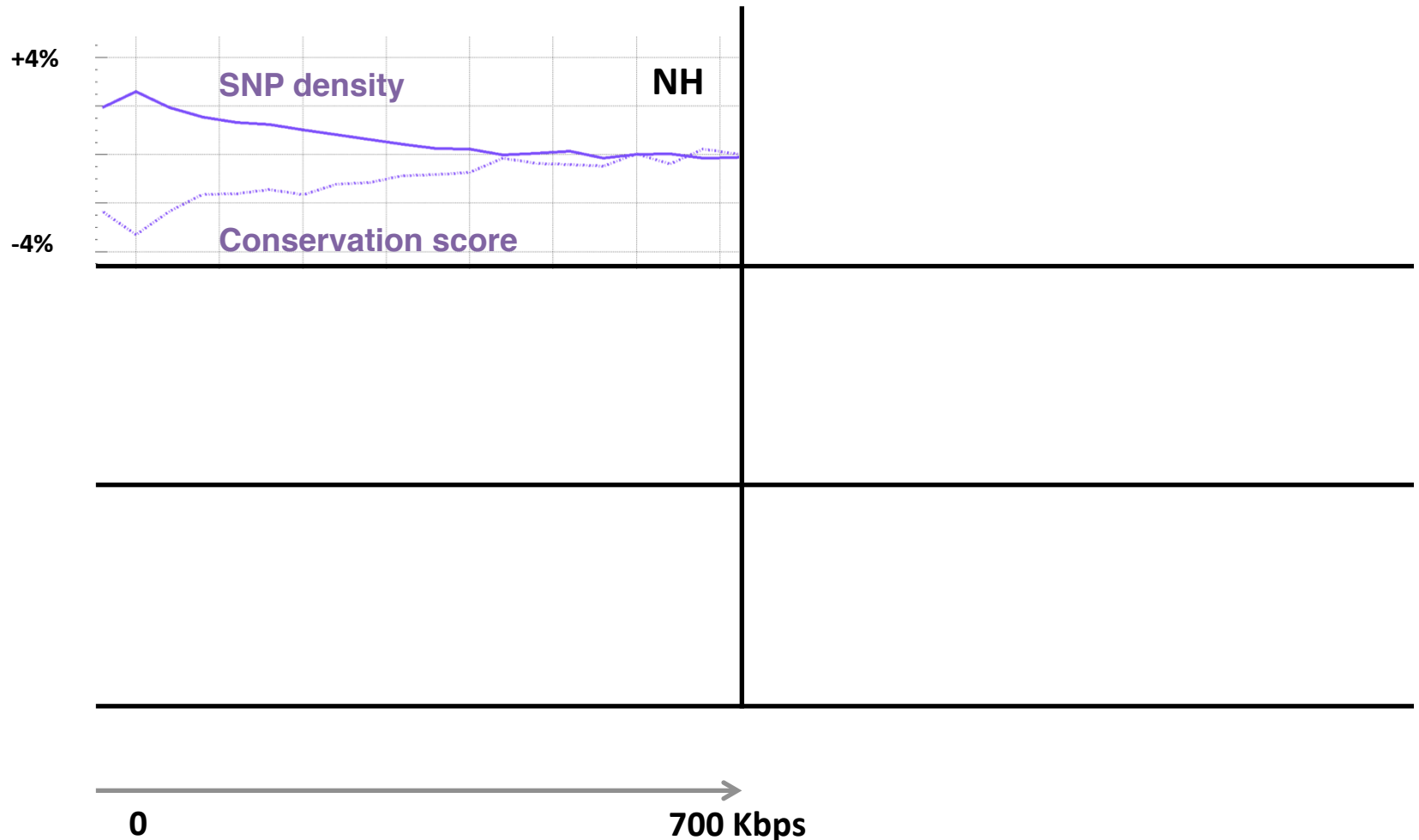Integrated set

Phase 1

Count of deletions

8,943 Deletion Breakpoints (Phase I Refined)

- FDR from IRS, PCR, and high-coverage trios
  - ~7% for site existence
  - 13% for site existence + sequence precision

Data for 1,092 samples

Multiple CNV callers

Call merging
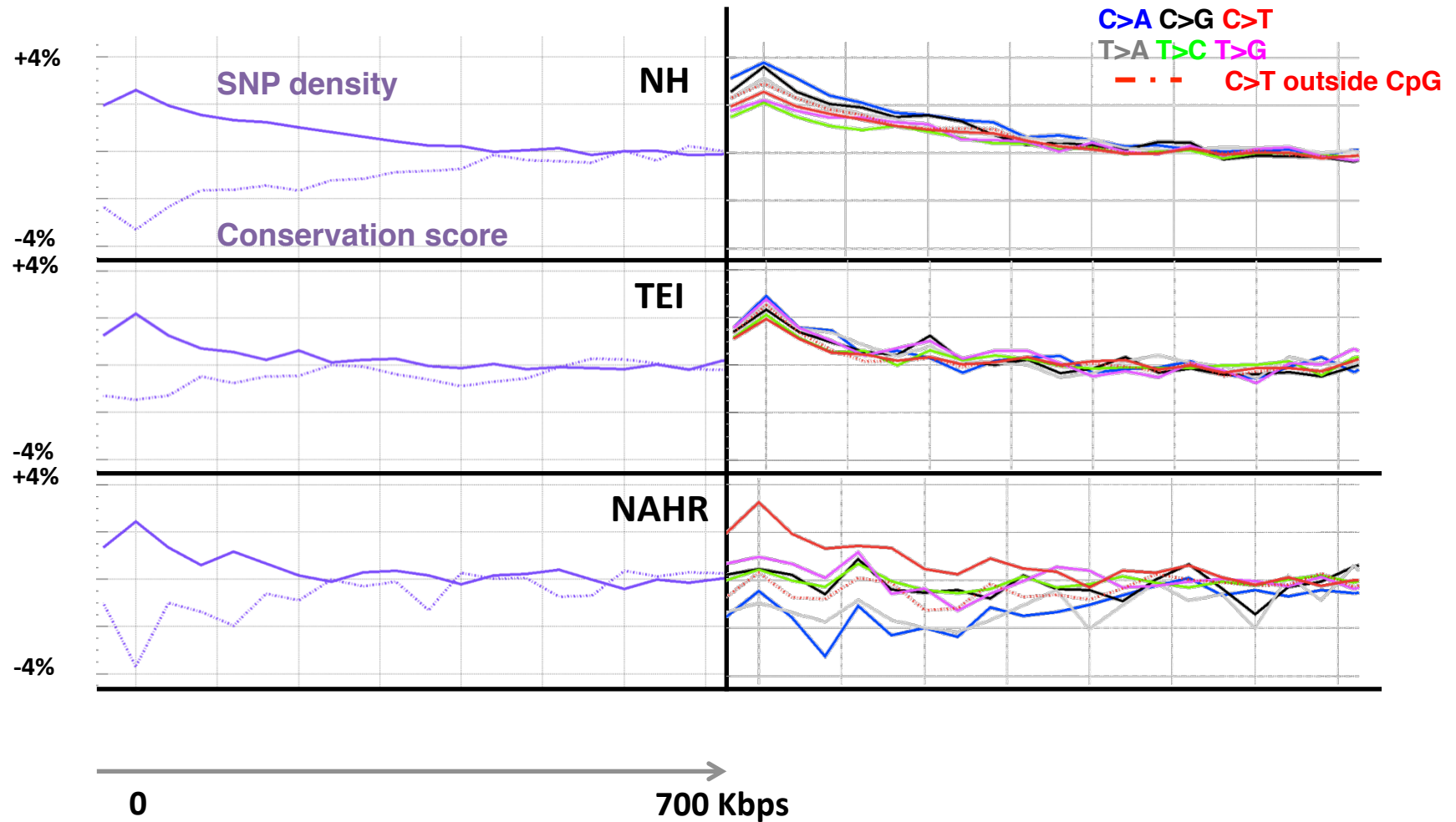
Breakpoint assembly

Mapping to junctions

Higher SNP Density and Relaxed Selection at NH Breakpoints

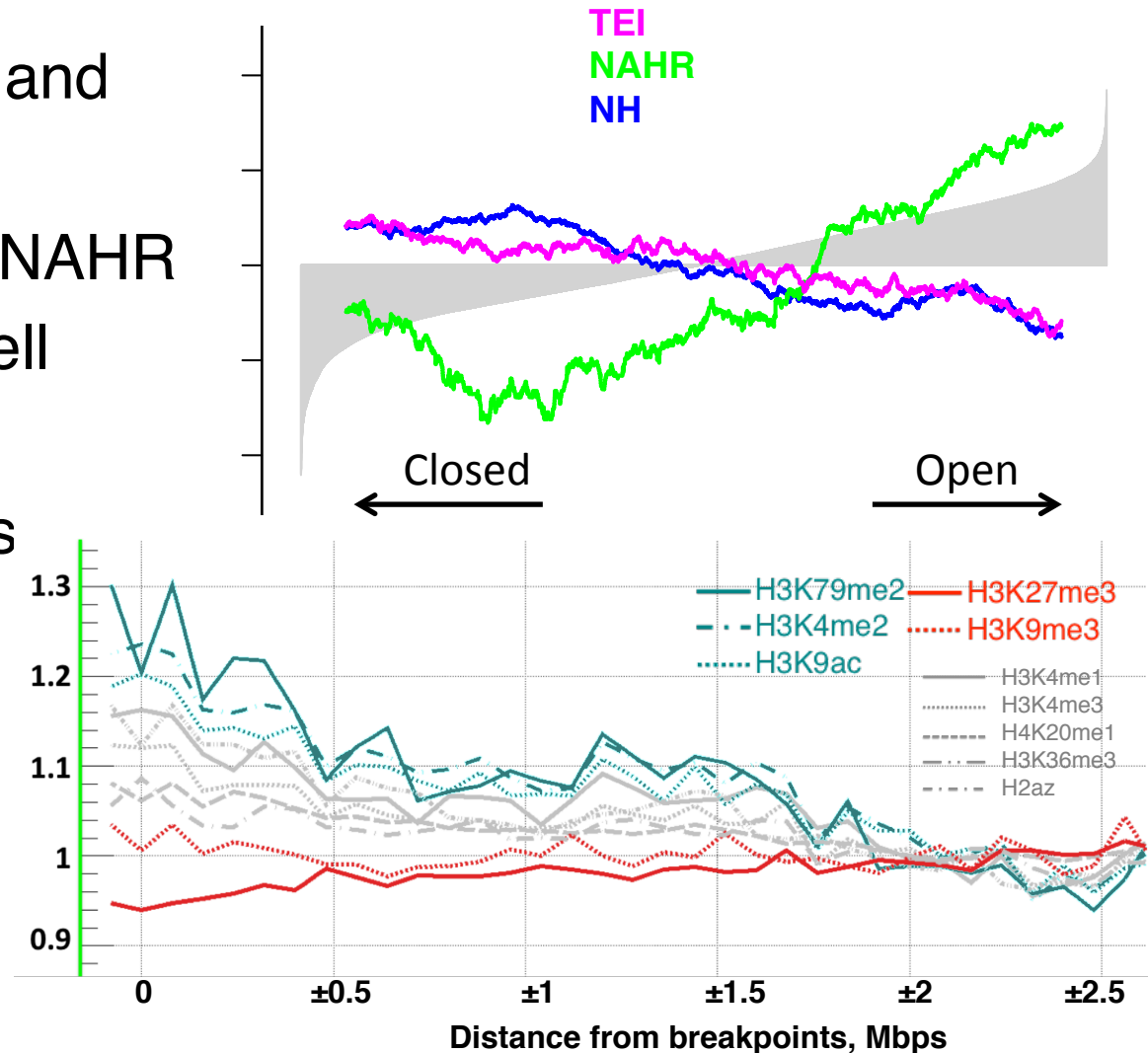# Higher SNP Density and Relaxed Selection at all Breakpoints
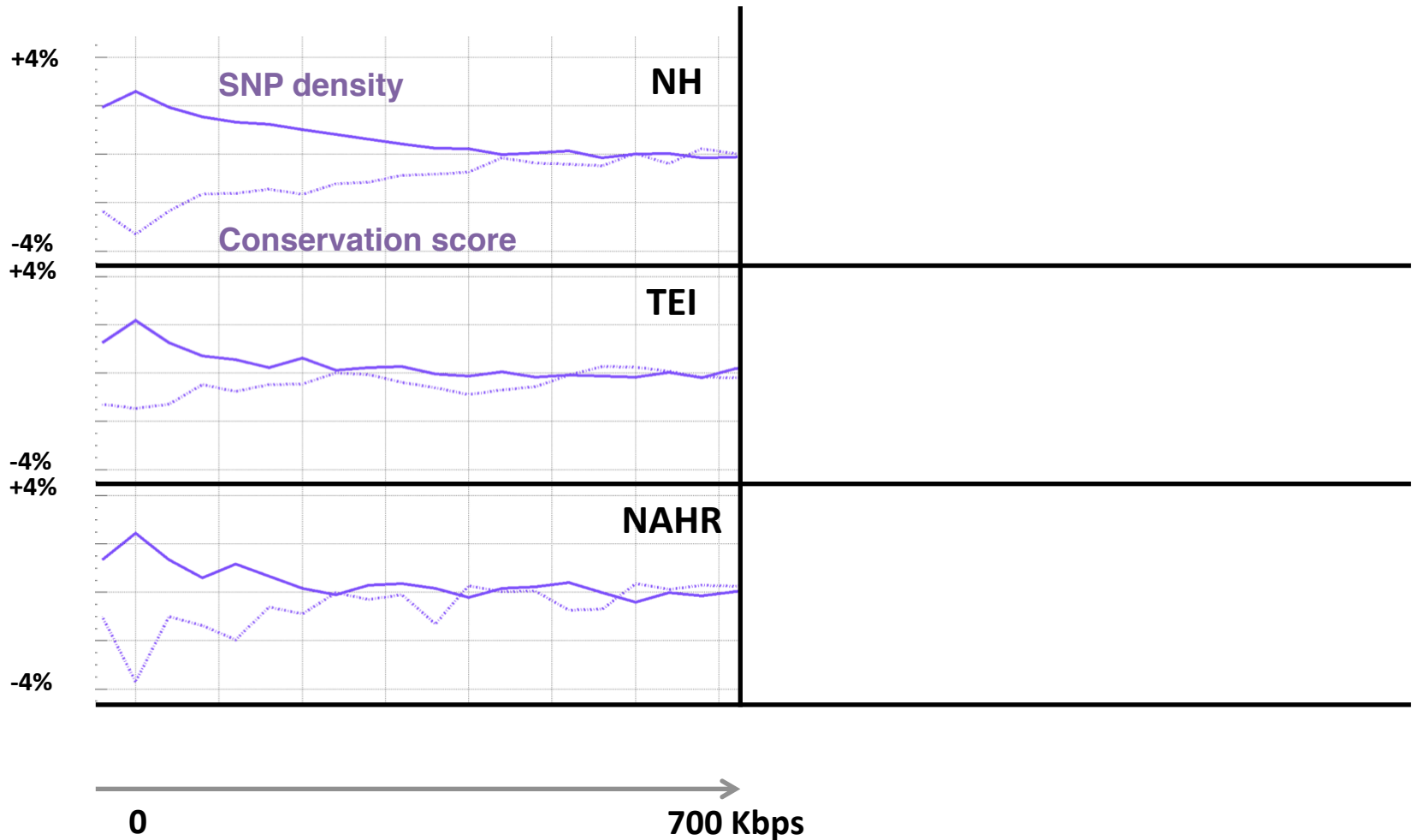
SNP Density at NAHR is Driven by High C>T

NAHR breakpoint are associated with open chromatin environment

- Supported by Hi-C and Histone modification

- Hypothesis: Some NAHR deletions occur w/o cell Replication
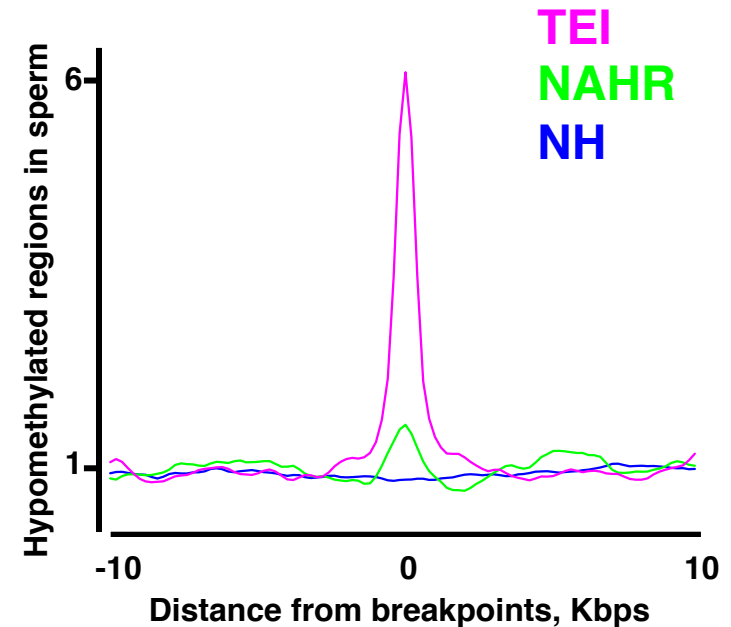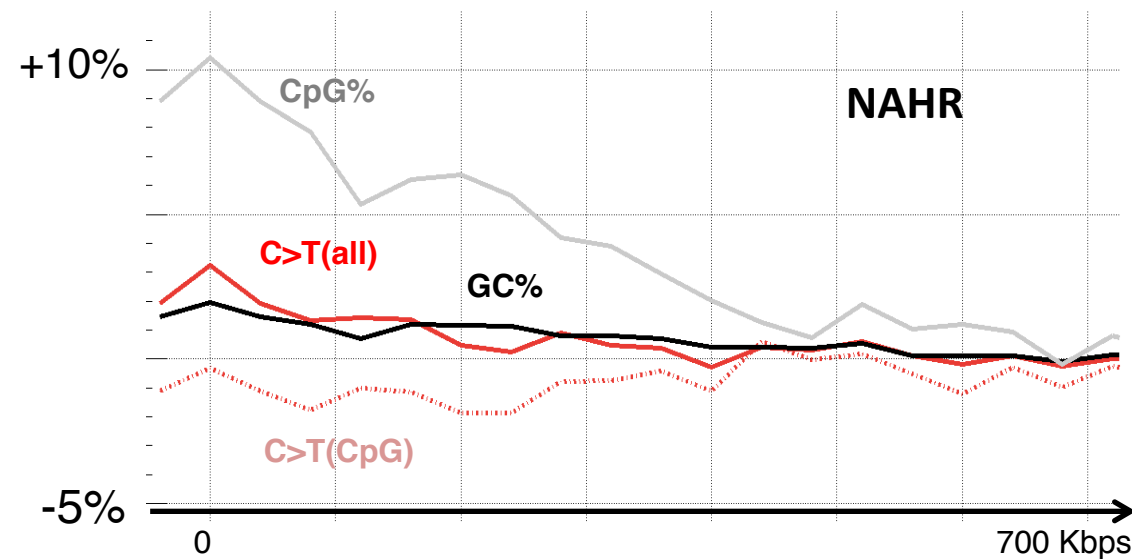
* H1 & GM12878 cells

Abyzov et al. 2015

Methylation pattern associated with breakpoints mechanisms

- Lower C>T in CpG around NAHR breakpoints
  - indicates lower methylation level in germline & embryonic cells
- Confirmed in male gamete

Micro-homologies Identified around Breakpoints

- Breakpoints have Micro-homologous sequences with the template sites.
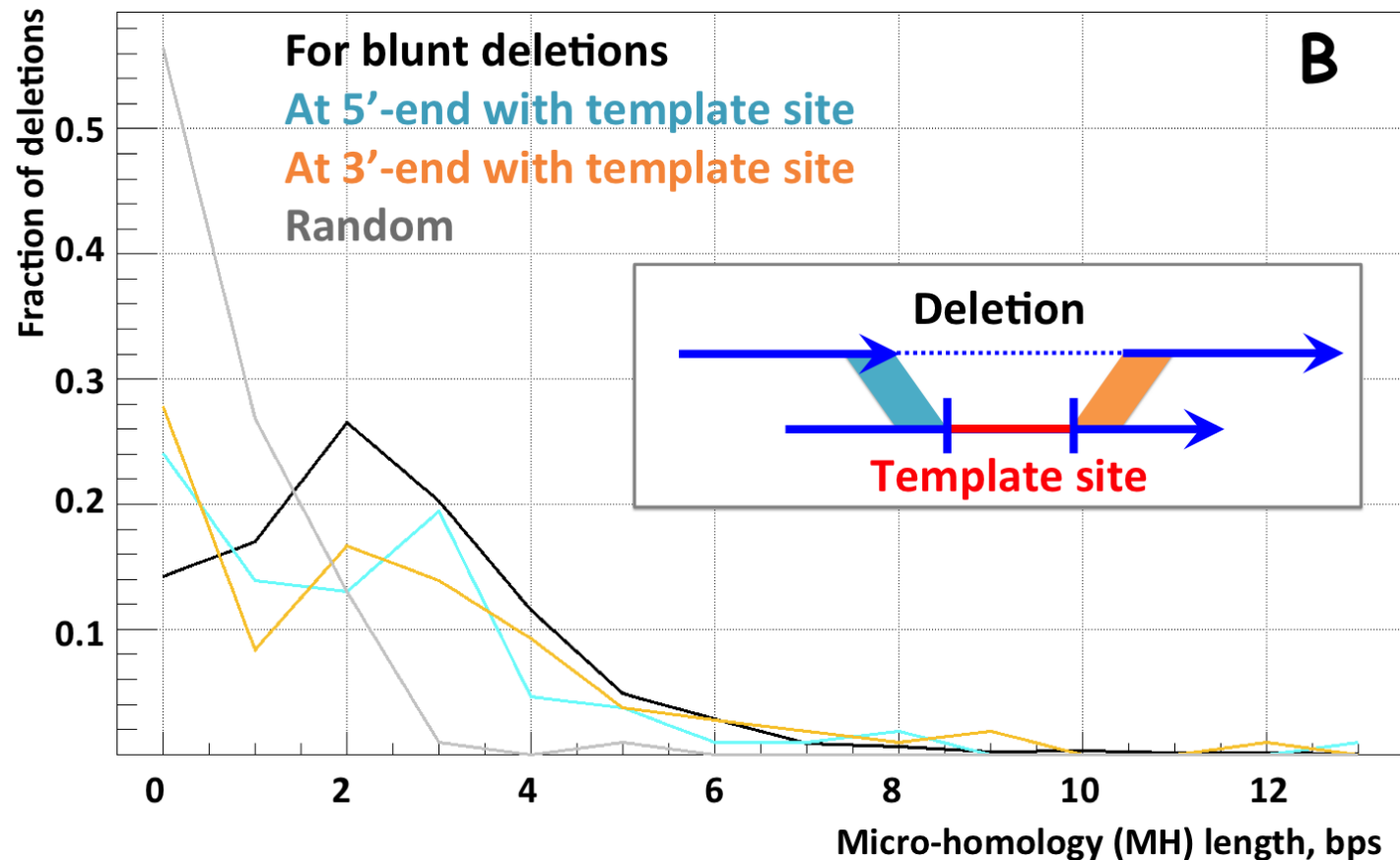
NH deletions are often coupled with micro-insertions

- Templates located at 2 characteristic distances from breakpoints, which tend to replicate late

- Suggests spatial & temporal configuration of DNA during template switching

More about breakpoints/ mechanisms

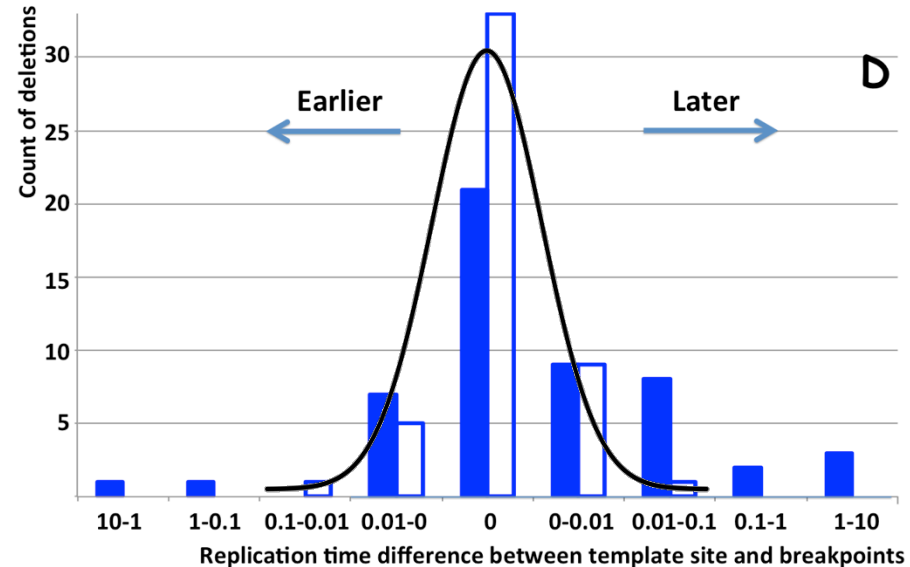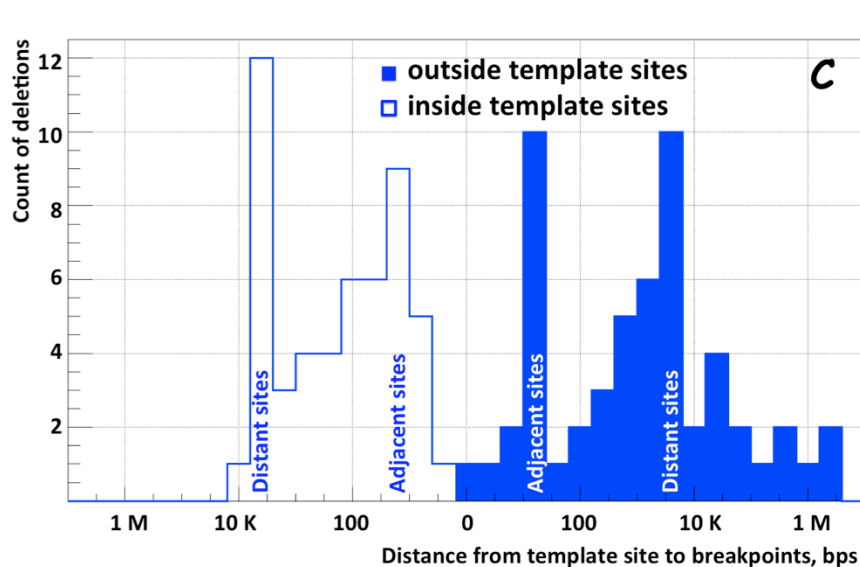- More accurate breakpoints and detailed investigation from the trios
- More mechanism assignment
  - Recombination effects on SV origin
    - Ectopic recombination mediated by LTR/Alu/LINE1
    - Deletions close to recombination hotspots (population specific)
  - BreakSeq2 - developing a more accurate mechanism pipeline taking advantage of the longer reads

- Using high-res. Hi-C data to further investigate the relationship of high-order chromatin structure and mechanisms of bkpts
  - Mechanisms of bkpts; Special focus on NAHR
  - micro-insertion template (potential tool to automate template finding/micro-homology identification)

More Functional Characterization of SVs

- More functional enrichment
  - Detect regions with high/low variations in the trio child compared to the parents, and identify potential involved functions; compare the trend in multiple trios

- Functional region biased bkpts discovery

- SV-eQTL
  - Limited by sample size if only look at trios
  - Instead, pick out the eQTLs identified from populations, check whether they are significant in the trios, and why (might not have exactly matching populations)

More SV calling & retrodups

- Autonomous/non-autonomous mobile elements insertion polymorphism
  - Repetitive elements mapping strategies
  - Transgenerational presence/absence dimorphism

- More retroduplications - compare retroduplications in parents and child:
  - Inherited retroduplications
  - Newborn retroduplications in the child
  - Missing heritability
  - Parent-of-origin

- CNVnator refinements

- Use HiC data to investigate influence of SVs on the association between the genomic elements and its target region.
  - Specifically look into SV hotspot region

Acknowledgements

- Refined Phase 1 Breakpoints Analysis
Alexej Abyzov, Shantao Li,
Daniel Rhee Kim, Marghoob Mohiyuddin, Adrian Stuetz, Nicholas F. Parrish, Xinmeng Jasmine Mu, Wyatt Clark, Ken Chen, Matthew Hurles, Jan Korbel, Hugo Y.K. Lam, Charles Lee

- Other SV participants
  - **Y Zhang, J Zhang, F Navarro, S Kumar**

Info about content in this slide pack

- Breakpoints analysis was from Abyzov et al. Nat. Comm. ('15, in press)
- General PERMISSIONS

  – This Presentation is copyright Mark Gerstein, Yale University, 2015.

  – Please read permissions statement at http://www.gersteinlab.org/misc/permissions.html .

  – Feel free to use slides & images in the talk with PROPER acknowledgement (via citation to relevant papers or link to gersteinlab.org).

  – Paper references in the talk were mostly from Papers.GersteinLab.org.

- For SeqUniverse slide, please contact Heidi Sofia, NHGRI

- PHOTOS & IMAGES. For thoughts on the source and permissions of many of the photos and clipped images in this presentation see http://streams.gerstein.info .

  – In particular, many of the images have particular EXIF tags, such as kwpotppt , that can be easily queried from flickr, viz: http://www.flickr.com/photos/mbgmbg/tags/kwpotppt