SCALPEL

MICRO-ASSEMBLY APPROACH TO DETECT INDELS WITHIN EXOME-CAPTURE DATA

Giuseppe Narzisi, PhD

Schatz Lab

Cold Spring Harbor Laboratory

Outline

① Scalpel micro-assembly pipeline

② Large-scale validation experiment

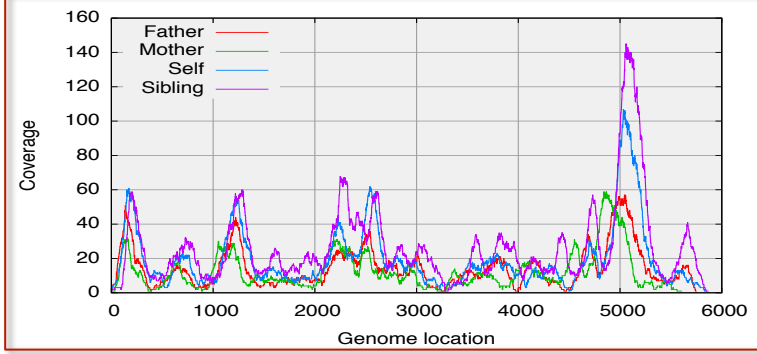
③ De novo/Transmitted mutations in Autism

SCALPEL

Micro-assembly pipeline

The detection challenge

Coverage



Irregularity in capture efficiency near the edges of the coding region

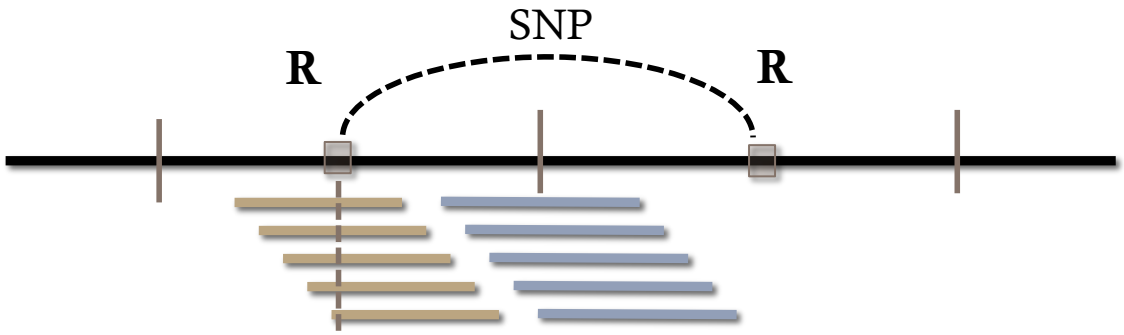
Mapping errors

SNP or Indel?

```
..TTGAATTAGCCTTGGTGAATTGAGCCTT....TTTAG--------AGTGC..
    ||||  ||||     ||||  |||||     |||||          |||||
  GAATGAGCC        GAAT-GAGCC     TTTAGAATAGGC     |||||
                                        ATAGGCGAGTGC
```

Long insertion

Repeats

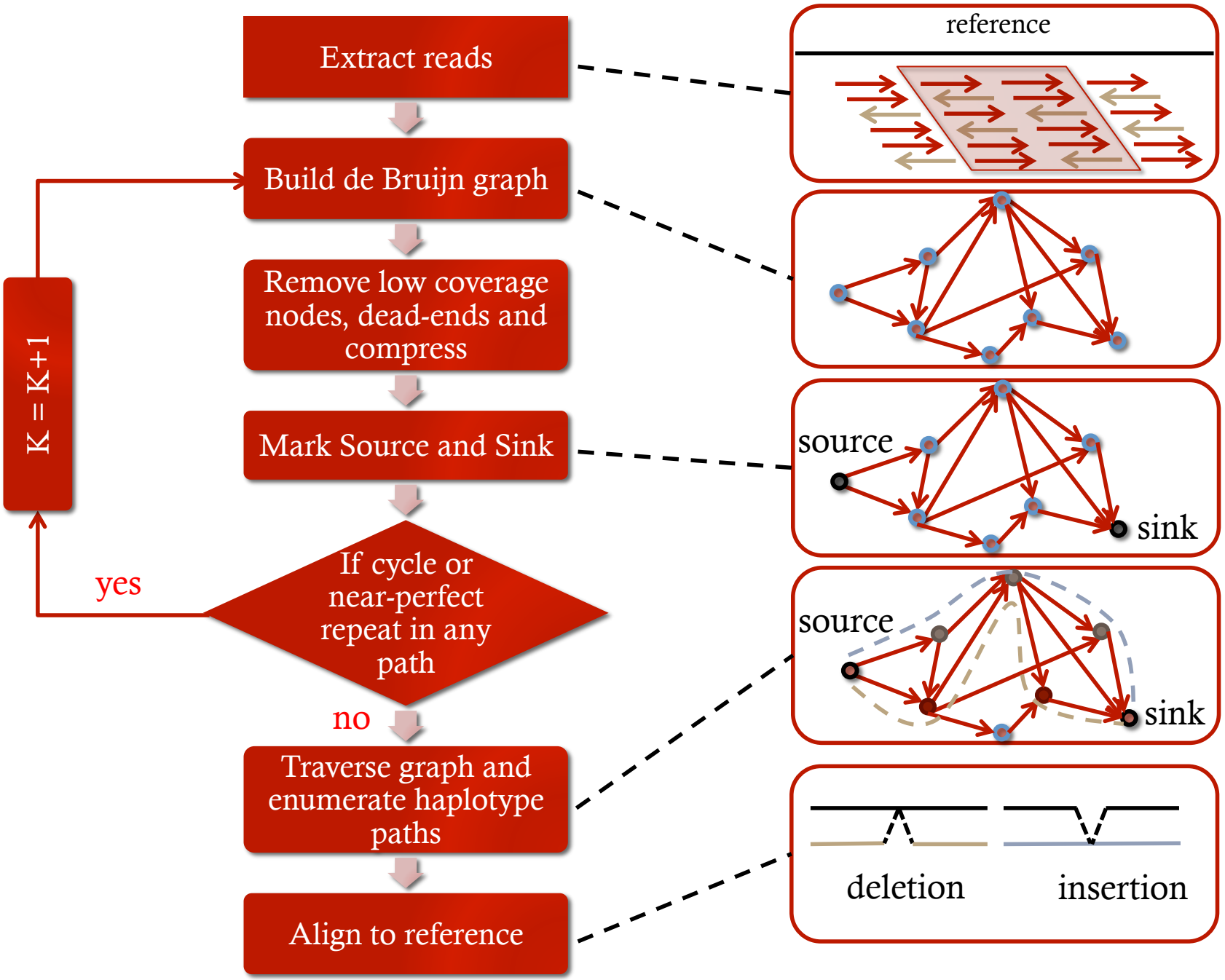


Scalpel

- Novel DNA sequence **micro-assembly** pipeline to detect mutations within exome-capture data.

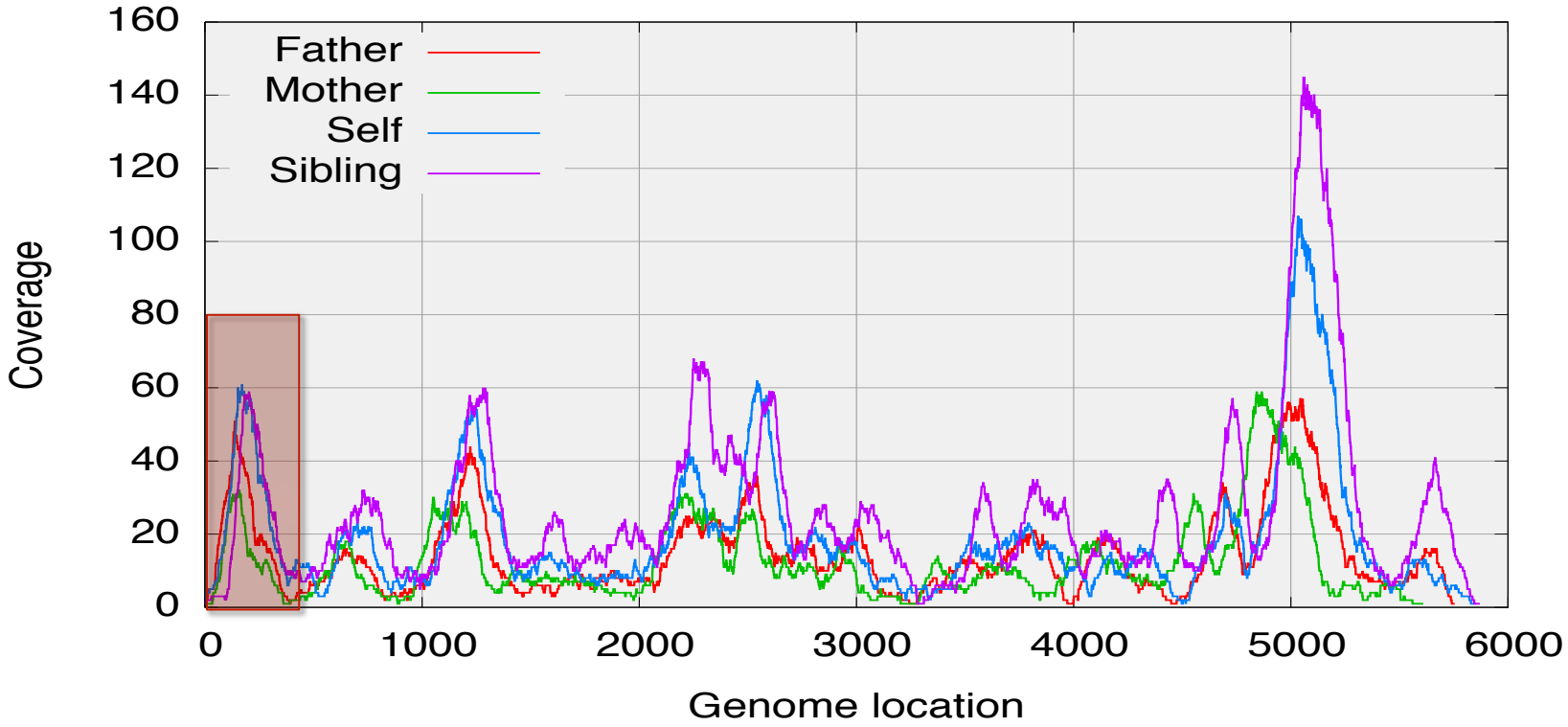
Whole-Genome assembly	Micro-assembly
Large scale genome structure	Detect genome variations
Genotypic	Haplotypic (Hom/Het state)
Heuristics to optimize resources (Time and Space)	Feasible to perform exhaustive search

- Features:
  1. Self-tuning k-mer.
  2. On-the-fly repeat composition analysis.
  3. Family pedigree: joint analysis of family members to detect de novo and transmitted mutations.


Extract reads
Build de Bruijn graph
Remove low coverage nodes, dead-ends and compress
Mark Source and Sink
If cycle or near-perfect repeat in any path
yes
K = K+1
no
Traverse graph and enumerate haplotype paths
Align to reference
reference
source
sink
source
sink
deletion
insertion

Walking along the exome

- Extraction, assembly, alignment and INDEL detection performed in **overlapping windows** along the exon.
  1. Localized assembly (smaller graph).
  2. Minimize problem with coverage drops.
  3. Distributed approach.



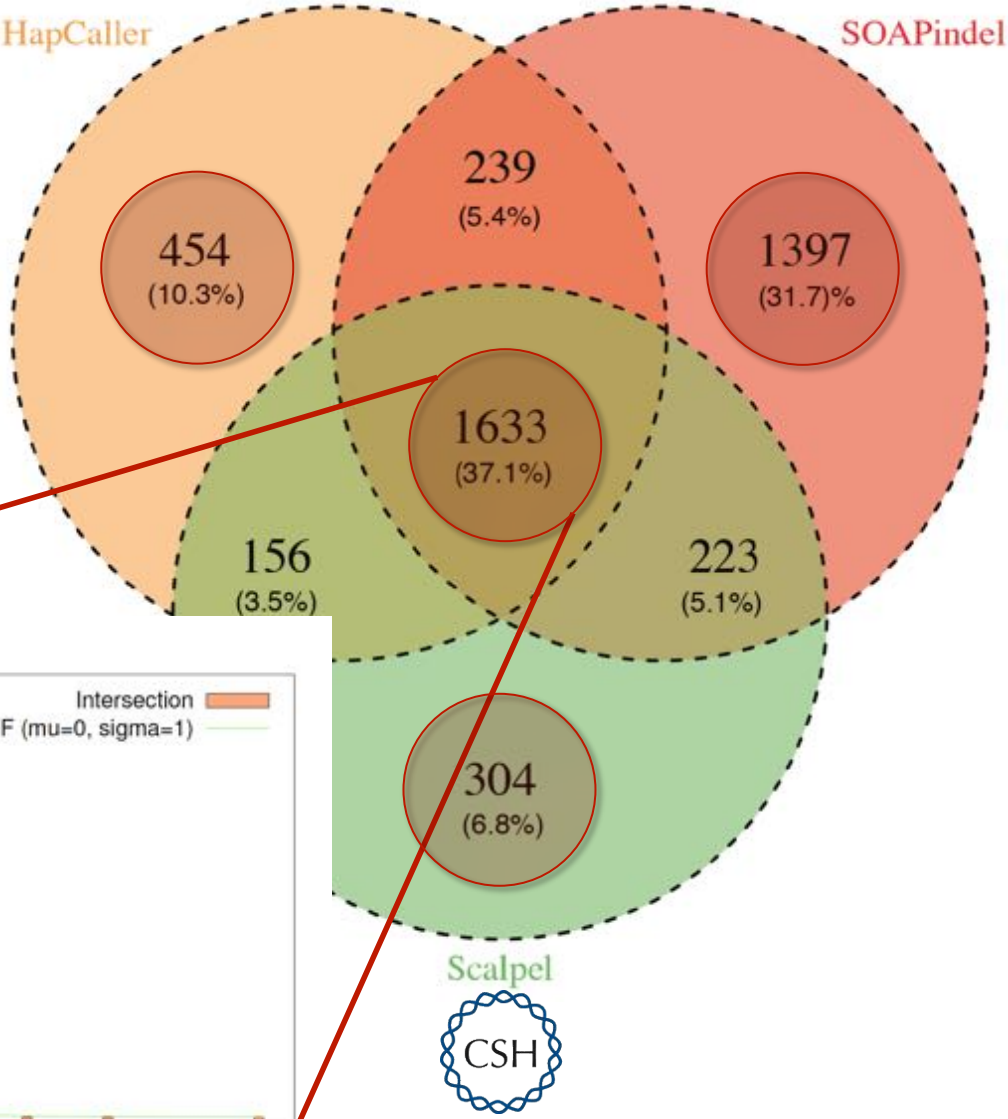
LARGE SCALE EXPERIMENT

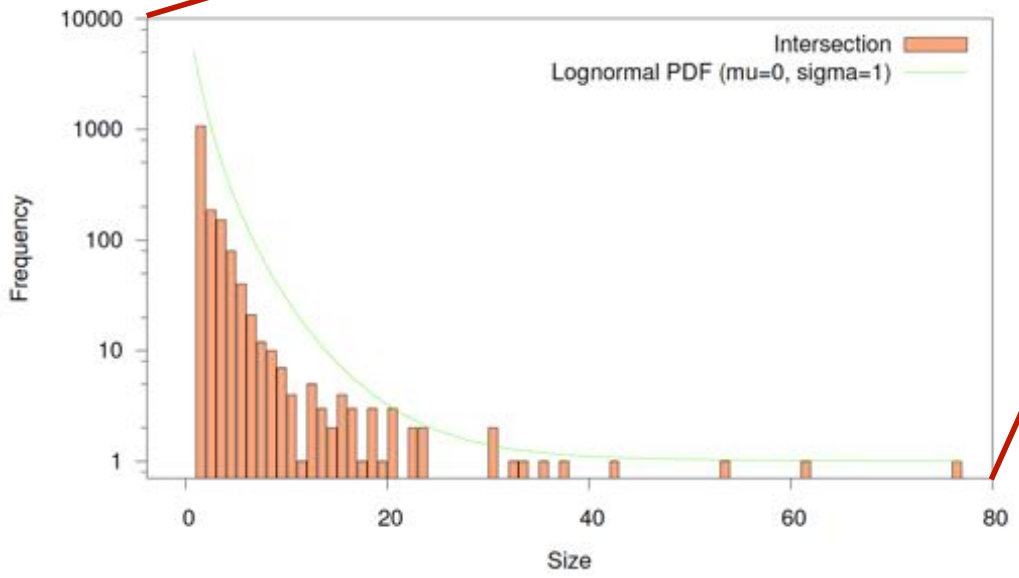
Re-sequencing of 1000 INDELs

INDELs in one Exome

Individual affected by Attention Deficit/ Hyperactivity Disorder (ADHD)

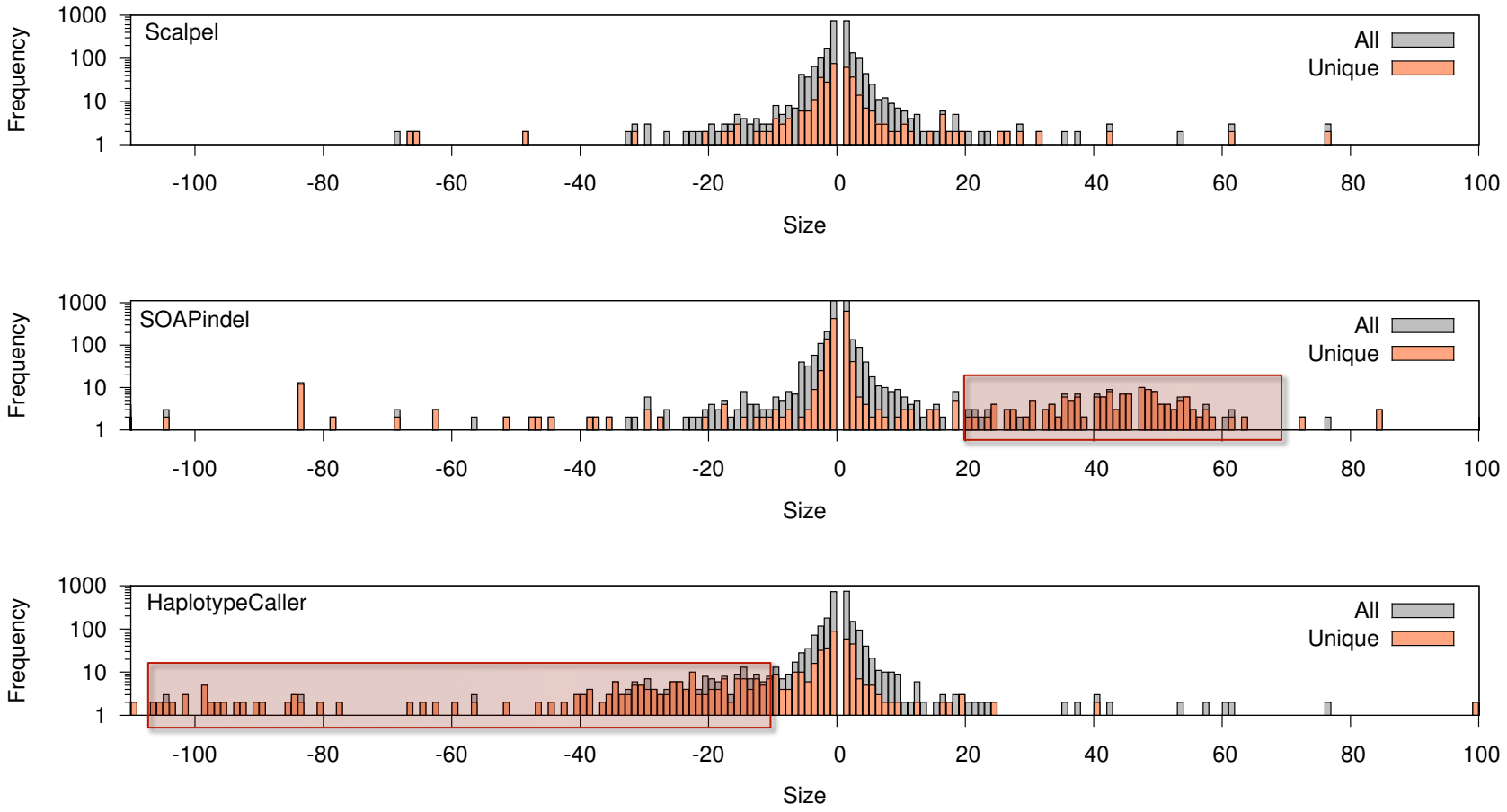
Captured using **Agilent SureSelect v.2** and sequenced on the **Illumina** platform.





quality of INDELs specific to each
sensitivity or poor specificity ??

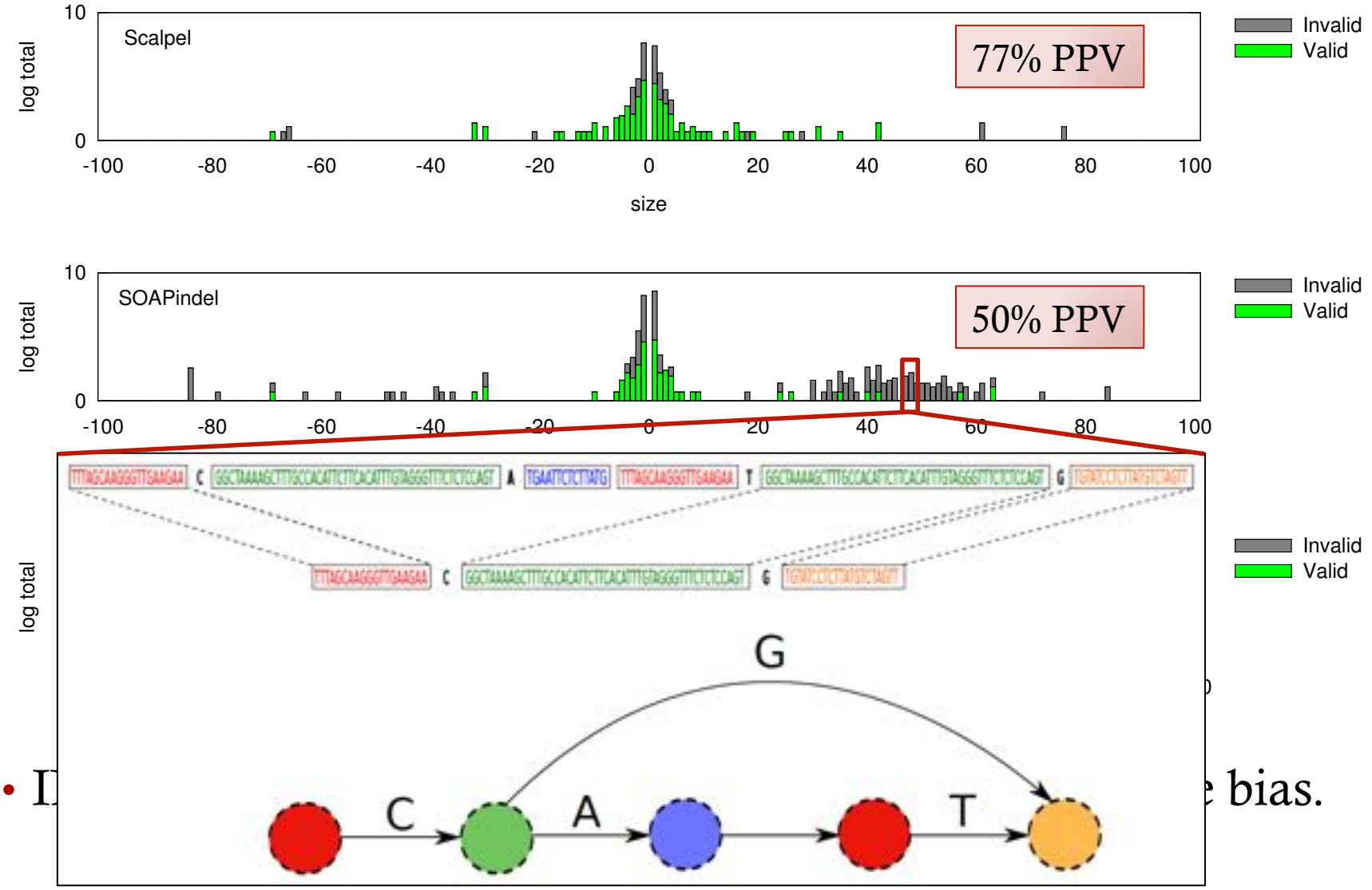
Focus on size distribution



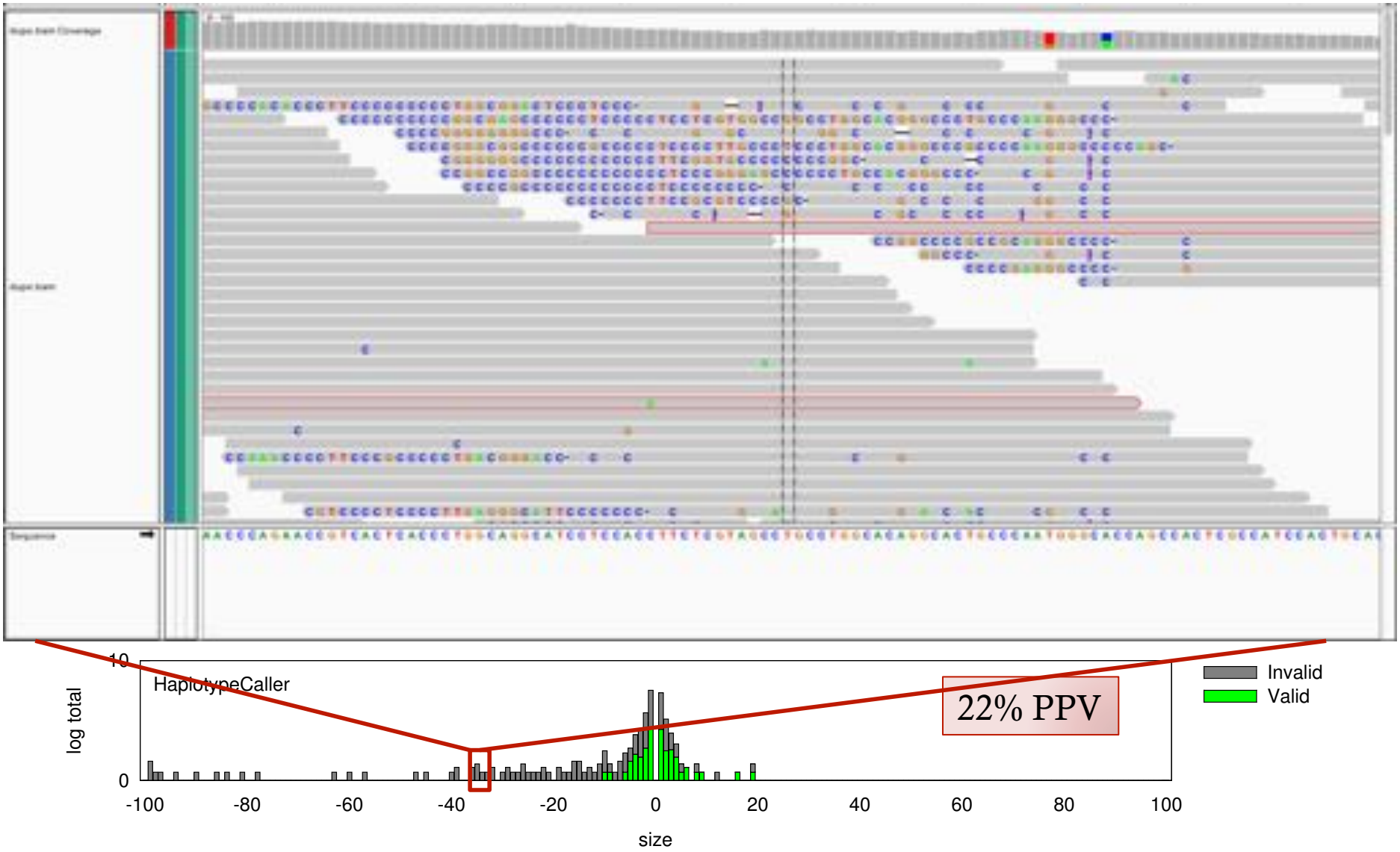
- **Bias** towards deletions (for HaplotypeCaller) or insertion (for SOAPindel).
- Scalpel instead shows a **well-balanced** distribution between insertions and deletions

Validated INDELs

specific to each pipeline



- I… …e bias.



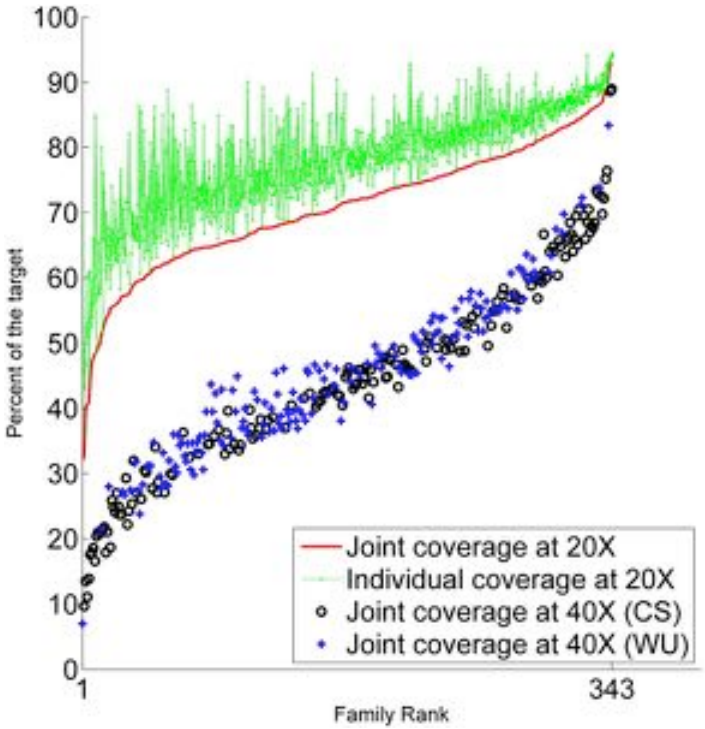
- INDELs not passing validation correlate well with size bias.

DE NOVO MUTATIONS IN AUTISM

Simons Simplex Collection

Simons Simplex Collection

- ~2700 families.
- Quad: two parents, one affected child and one unaffected child.
- NimbleGen SeqCap EZ Exome v2.0 (36 Mb).
- Illumina HiSeq: ~93bp reads after removing barcodes.



Three major studies reporting strong enrichment for de novo gene killing mutations in autistic kids:

① CSHL: Iossifov et al. (2012) **Neuron**. 74:2 285-29

② Yale: Sanders et al. (2012) **Nature**. 485, 237–241.

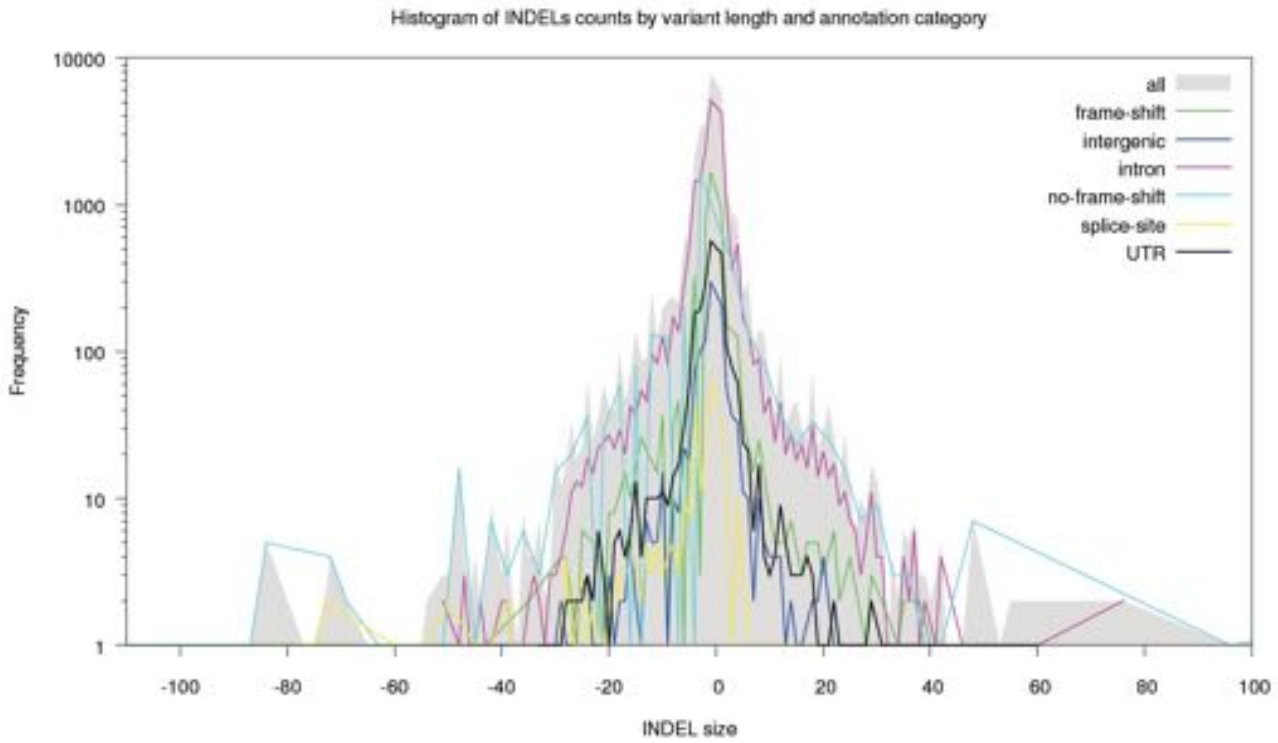
③ WashU: O'Roak et al. (2012) **Nature**. 485, 246–250.

INDELs in 593 families

Database with > 3 million INDELs

Increased power to detect insertions.

Subdivide by annotation category.



Goal: discover significant biology that was impossible to measure a few year ago

De novo INDELs in Autism

593 families: 343 CSHL, 200 StateLab, and 50 EichlerLab

INDEL effect	Aut	Sib	Aut M	Aut F	Sib M	Sib F	Total
Frame shift	35	16	25	10	12	4	51
Intron	13	16	11	2	6	10	29
Intergenic	2	0	2	0	0	0	2
No frame shift	4	5	4	0	1	4	9
Splice-site	2	0	2	0	0	0	2
UTR	2	2	2	0	0	2	4
Total	58	39	46	12	19	20	97

De novo INDELs that are likely to severely disrupt the encoded protein are significantly more abundant in affected children than in unaffected siblings

CONCLUSION

Conclusions

- Scalpel: highly accurate tool to detect de novo, transmitted, and somatic INDELs.
- Errors of current detection software explained by a large-scale (1000 INDELs) re-sequencing experiment.
- Population wide analysis: de novo INDELs in Autism.

Acknowledgment

Michael C. Schatz

Michael Wigler

Gholson J. Lyon

Ivan Iossifov

ADHD project

- Jason O'Rawe
- Yiyang Wu

Autism project

- Dan Levy
- Michael Ronemus
- Yoonha Lee
- Zihua Wang
- Ewa Grabowska
- Peter Andrews
- Mitchell Bekritsky
- Jude Kendall

SFARI SIMONS FOUNDATION AUTISM RESEARCH INITIATIVE

THANK YOU

Email: gnarzisi@cshl.edu