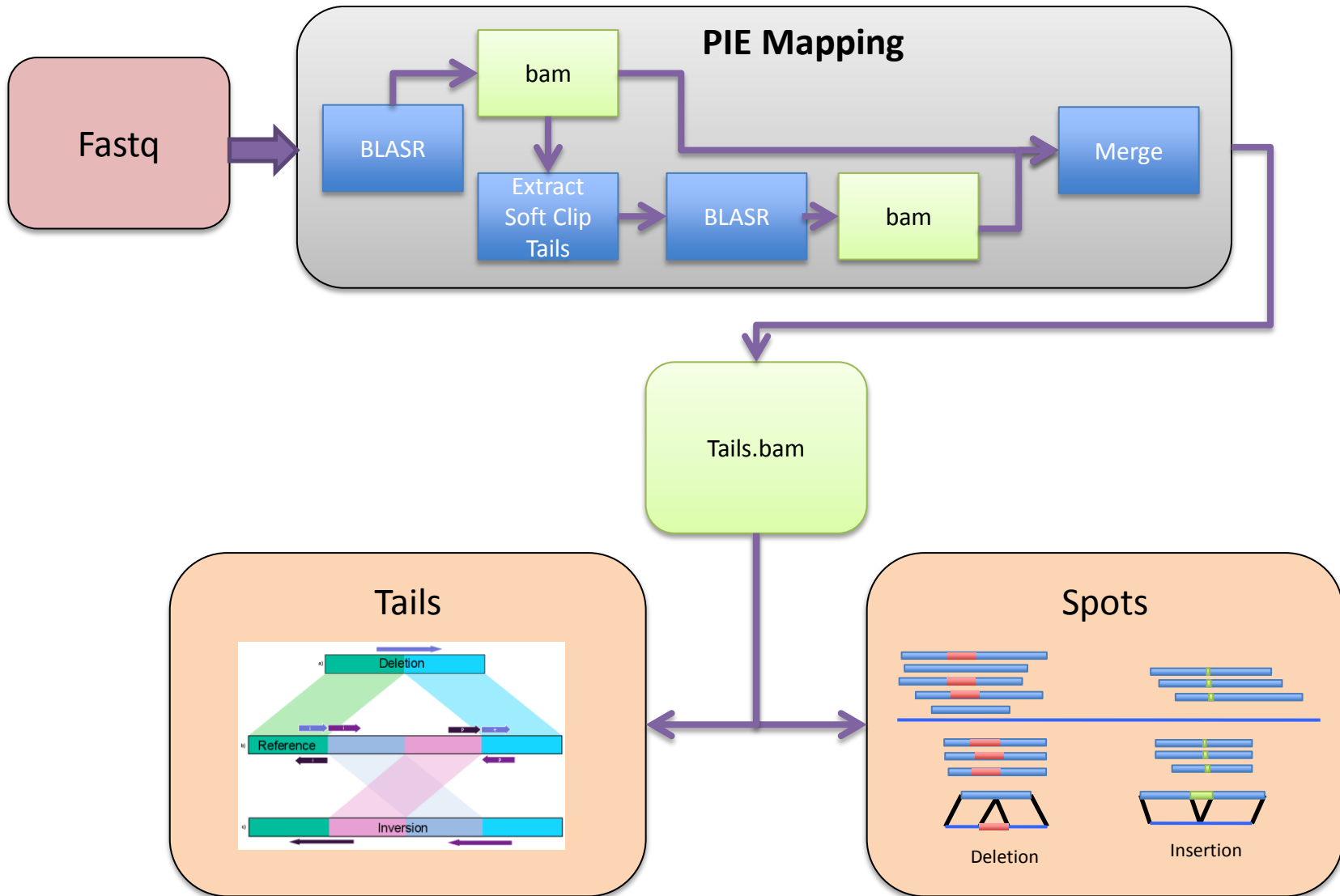


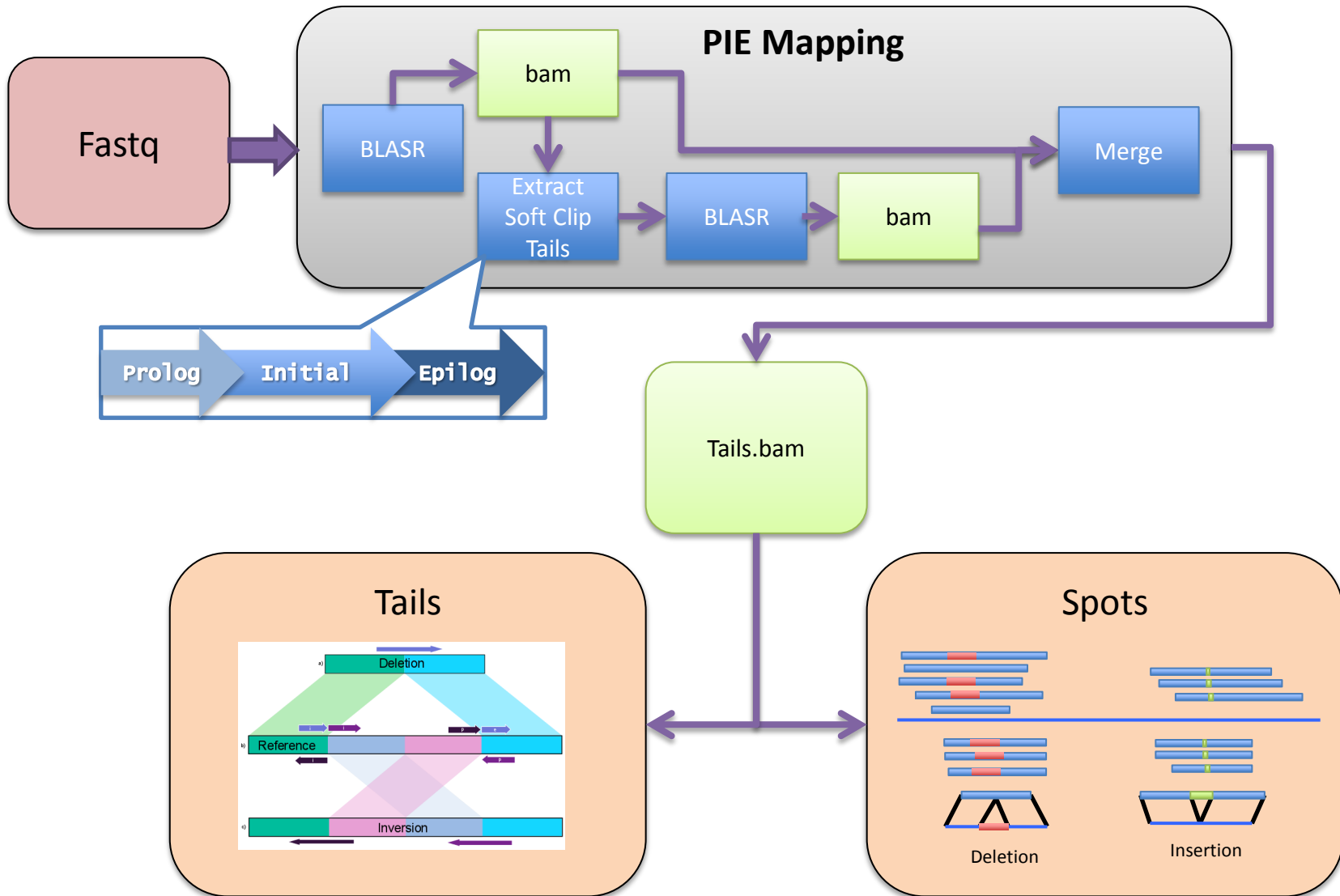
PBHoney – Detecting SVs with Long-Read Sequencing

Adam C. English
Human Genome Sequencing Center
Baylor College of Medicine

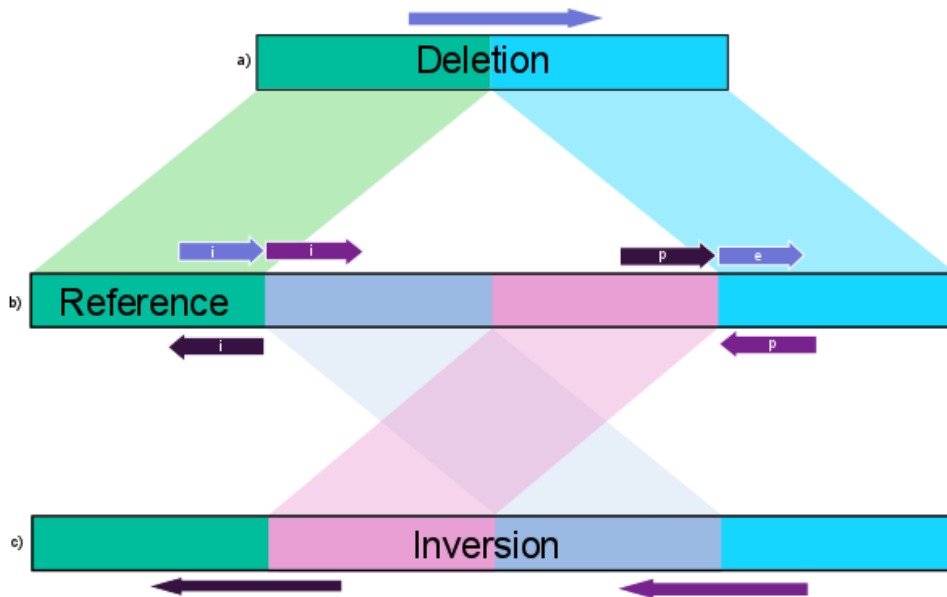
PBHoney Workflow



PBHoney Workflow



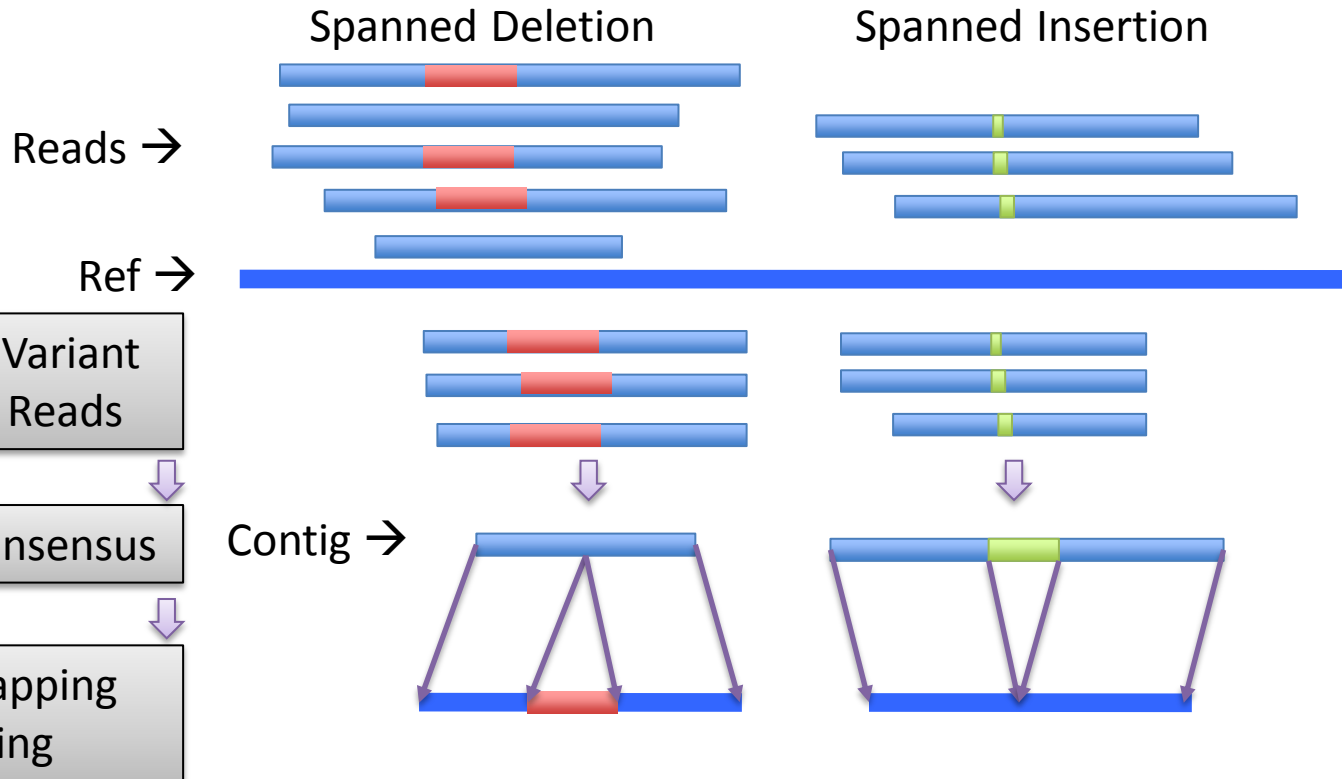
Honey Tails



BestN Tail Rank	Count	Percent
2	3,558	9.85%
3	360	1.00%
4	93	0.26%
5	72	0.20%
6	65	0.18%
7	37	0.10%
8	22	0.06%
9	17	0.05%
10	9	0.02%
11	10	0.03%
12	14	0.04%
13	3	0.01%
14	4	0.01%
15	2	0.01%
16	7	0.02%
17	18	0.05%
18	10	0.03%
19	12	0.03%
20	46	0.13%
Ranked	4,359	12.07%
UnRanked	31,754	87.93%
Total	36,113	



Honey Spots



- Full Insertion/Deletion Sequence recovered
- Better Breakpoints
- Var/Alt Supporting Counts with Bayes estimation gives GT
- Multiprocessing with Consumer/Worker design
- Faster (40x PacBio data on 8 cores 48Gb mem – 55 Hours)

CHM1 hydatidiform mole



licine

Totals	Total Loci		Assembled Loci		~Specificity		~Sensitivity	
	DELs	INs	DEL	INs	DEL	INs	DEL	INs
Illumina Loci	7,399	73,816	2,582	1,508	34.90%	2.04%	78.94%	26.98%
Chaission et al. Loci	3,989	7,321	2,192	3,847	54.95%	52.55%	67.01%	68.83%
HON Loci	3,565	4,856	2,615	3,309	73.35%	68.14%	79.94%	59.21%

Parliament & HS1011

Diploid Human SVs from Illumina HiSeq + PacBio WGS + Other

Final High Quality Call Set: 9,777 SVs

5,044 deletions, 4,463 insertions, and 270 inversion

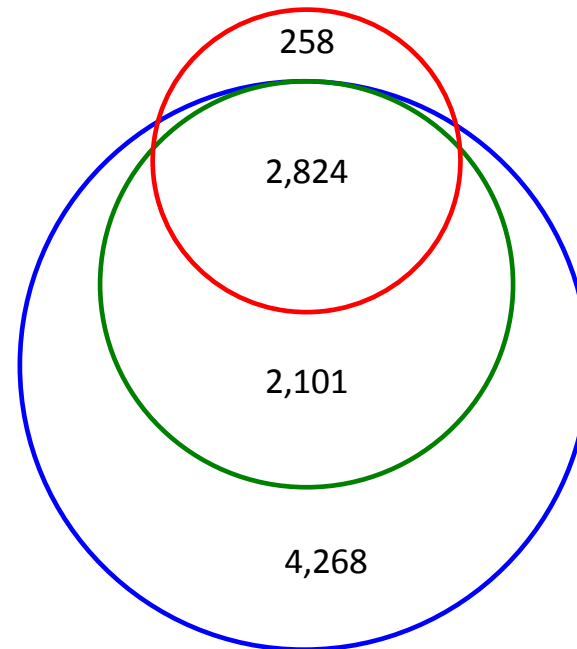


Source	Method	Data	Total Calls	Total Loci	Assembled Loci	Multi Source Loci	Force Called Loci	Solo Assembled	Solo Forced
Illumina HiSeq Total			27,778	17,765	3,751	788	814	979	622
PBHoney	Local Error and Tail Mapping	PacBio RS	10,611	10,340	5,883	483	0	3,792	0
Other			7,453	7,088	922	1,048	219	167	219
Total			45,842	31,184	7,733	1,133	1,033	4,897	841

**Total ILLOnly
3,082**

**Total ILLHyb
4,925**

**Total PB-ILL
9,193**



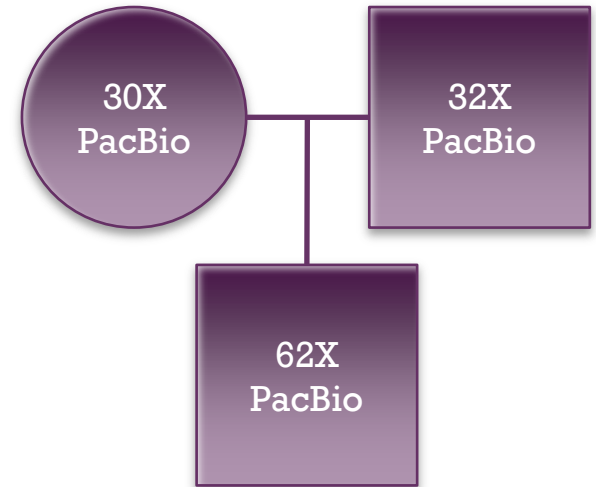
Ashkenazi Trio



■ Mendelian Consistency of Honey Spots

Merged Honey Spots by SVType and Max Distance 50bp - No Filtering

- 1/1, 1/1 in parents: 3914
- 1/1 in proband: 3526 (90.1%)
- 0/1 in proband: 295 (7.5%)
- 0/0 in proband: 93 (2.4%)
- 1/1 in proband: 9104
- 0/1, 0/1 or 1/1 in parent(s): 6550 (71.9%)
- 0/1, 0/1 in parents: 1695
- 0/0 in proband 414 (24.4%)
- 0/1 in proband 924 (54.5%)
- 1/1 in proband 357 (21.1%)



NA12878



- 2,676 – High Quality Deletions from NIST ('svclassify: a method to establish benchmark structural variant calls' - Parikh et al.)
- PBHoney – 7,514 Deletions
- Comparison
 - 2,447 intersection with reciprocal overlap of 75%:
 - 89 have Overlap and Size Difference ≤ 200 bp
 - 5 less than 200bp away and Size Difference ≤ 200 bp
- **132 (4.9%) of NIST HQ Missing**

Acknowledgements

NGSI

Jesse Farek

Will Salerno

Ziad Khan

Simon White

Narayanan Veeraraghavan

Ted Chang

Richard Gibbs

Eric Boerwinkle

Donna Muzny

Wet-Lab

Yi Han

Vanesa Vee

Parliament

Oliver Hampton

Andrew Carroll (DNANexus)

Niranjan Shekar (Spiral Genetics)

Aniko Sabo

Kim C Worley

