Efficient detection of novel sequence insertions using Linked-Read sequencing: Supplementary Tables and Figures

Dmitry Meleshko[1,2], Rui Yang[1], Patrick Marks[3], Stephen Williams[3], and Iman Hajirasouliha[2,4,*]

[1]Tri-Institutional Computational Biology & Medicine Program, Weill Cornell Medicine of Cornell University, NY, 10021, USA
[2]Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine of Cornell University, NY, 10021, USA
[3]10x Genomics Inc., Pleasanton, California, 94566, USA
[4]Englander Institute for Precision Medicine, The Meyer Cancer Center, Weill Cornell Medicine, NY, 10021, USA
*Corresponding author

Method	TP	FP	Missed	Sensitivity	Precision	F1
Coverage = 13X						
Novel-X	0	0	0	0	0	0
PopIns	134	2610	1866	0.07	0.05	0.06
NUI	1362	3	638	0.68	0.998	0.81
Coverage = 26X						
Novel-X	788	13	1199	0.40	0.983	0.565
PopIns	504	2775	1496	0.25	0.182	0.211
NUI	1725	2	275	0.863	0.999	0.923
Coverage = 39X						
Novel-X	1439	14	561	0.720	0.990	0.834
PopIns	779	2342	1221	0.390	0.250	0.304
NUI	1754	2	246	0.88	0.999	0.934
Coverage = 52X						
Novel-X	1700	4	300	0.85	0.998	0.918
PopIns	1011	2261	989	0.505	0.309	0.384
NUI	1746	3	254	0.873	0.998	0.931

Table S1: Comparison of Novel-X, PopIns and Supernova/paftools performance on downsampled data.

$\mathcal{U}$ assembly	Insertion reassembly	TP	Missed	FP
Velvet	SPAdes	1789	211	1
SPAdes	SPAdes	1680	320	5
Velvet	Supernova	1150	850	2
Supernova	SPAdes	1705	295	7
Velvet	Velvet	1386	614	0

Table S2: Performance of different assembly strategies on simulated data.