

Supplemental Information:

Summary of STRs target genotyping protocols

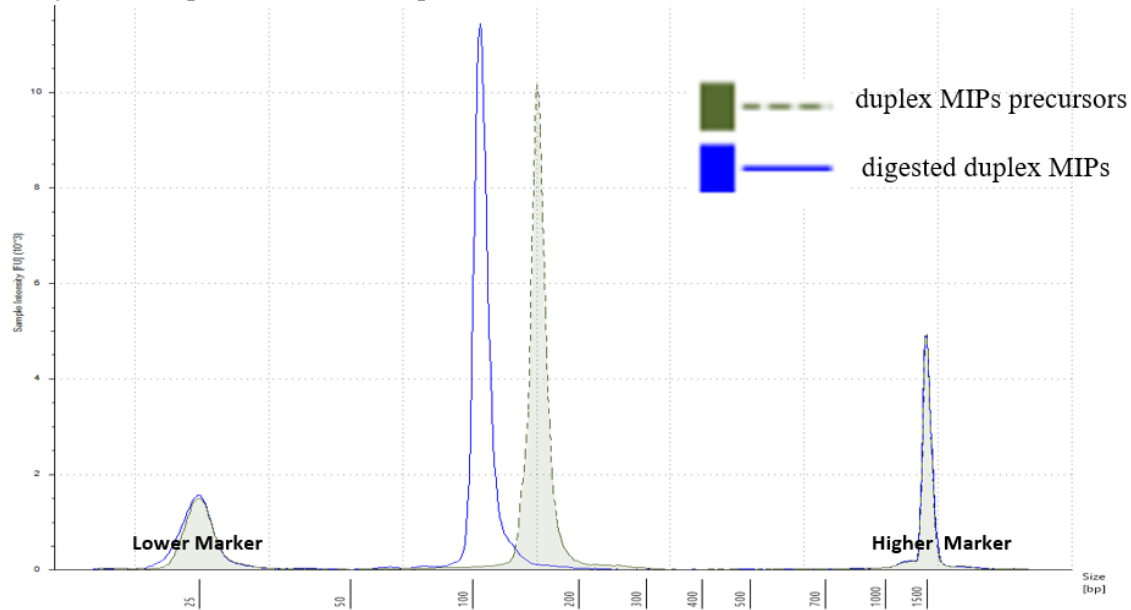
As listed in Supplemental Table1, several protocols have been developed for target genotyping STRs. Four of them were focus on bulk DNA: CODIS¹ was based on multiplex PCR and capillary electrophoresis, Guilmatre et al ² was based on array capture and NGS, Jorge Duitama et al³ was based on RNA probe capture and NGS, Carlson et al⁴ was based on MIPs and NGS. Two of them were focus on single cell WGA DNA, Shlush et al⁵ was based on multiplex PCR and capillary electrophoresis, Biezuner et al⁶ was based on Access Array and NGS.

Tissue	Template	Target Enrichmet	Calling Method	Majority STR Type	Targets	Purpose	Refs
human blood	Bulk	multiplex PCR	Capillary Electrophoresis	hexa-	~20	Forensic	1
Human	Bulk	Array capture	Next Generation Sequencing	all types	7851	Mutation Discovery	2
human blood	Bulk	RNA Probes	Next Generation Sequencing	tri- and longer	10764	Mutation Discovery	3
A.thaliana	Bulk	MIPs	Next Generation Sequencing	tri- and hexa-	102	Evolution phylogeny	4
human leukemia	scWGA	multiplex PCR	Capillary Electrophoresis	di-	128	Lineage Reconstruction	5
human cancer	scWGA	Access Array	Next Generation Sequencing	di-	~2000	Lineage Reconstruction	6
human cancer/normal	scWGA	duplex MIPs	Next Generation Sequencing	di-, mono-	~10,000	Lineage Reconstruction	

Supplemental Table1. STR capture methods summary

Quality control step used in duplex MIPs preparation.

The size of duplex MIPs precursor is ~150bp. Duplex MIPs precursors were first amplified by 20 cycle PCR and further digested by MlyI (NEB) in order to create a ready-to-run duplex MIPs. Expected size for precursor amplification product is ~150bp and following digestion, the size of ready-to-run duplex MIPs is ~105bp.



Supplemental Figure1: The size of duplex MIPs precursor and the digested duplex MIPs | The dashed green peak in the middle is duplex MIPs precursors; the solid blue peak in the middle is duplex MIPs.

An example of Unique Molecular Identifier (UMI) read counts in the MiseqR33

Samples from MiseqR33 was analyzed. All 64 different UMIs were detected in all the samples. The sample with barcodes number 743 from MiseqR33 was shown as an example. The reads mapping to their reference targets were collected and UMIs were counted by reads contained this UMI. Counts ranking from high to low by UMI compositions was shown in Sup. Fig2. The counts of UMI bases of this sample: 'T': 17658, 'A': 11363, 'G': 10063, 'C': 8060, 'N': 28, biased towards 'T'.

The calibration of duplex MIPs pipeline.

Three major steps: hybridization, gap-filling, digestion in the MIPs capture pipeline were calibrated in 18 different conditions. Hybridization were tested in 2, 4, 18 hours; Gap filling were tested in 1, 2, 4 hours; while the Digestion in 1, 2 hours(Data from MiseqR31, MiseqR32, and MiseqR33).

ProbeType	DNA	Hyb(hr)	Gap(hr)	Dig(hr)	TotalReads	Total Success	Success Rate	Loci>0	Loci>4	Loci>9
OM6	Hela	2	1	1	91805	18568	20%	6568	854	121
OM6	Hela	2	1	1	115167	23632	21%	7293	1322	243
OM6	Hela	2	1	2	121728	71250	59%	8892	4125	1770
OM6	Hela	2	1	2	114540	71036	62%	9229	4508	1960
OM6	Hela	2	2	1	199923	39214	20%	8365	2536	694
OM6	Hela	2	2	1	195185	79740	41%	9451	4911	2267
OM6	Hela	2	2	2	100563	56274	56%	8787	3641	1337
OM6	Hela	2	2	2	88212	51594	58%	8605	3247	1098
OM6	Hela	2	4	1	151143	48412	32%	8854	3198	997
OM6	Hela	2	4	1	141481	45520	32%	8390	3000	902
OM6	Hela	2	4	2	157111	84307	54%	9506	5147	2480
OM6	Hela	2	4	2	129168	88406	68%	9498	5333	2611
OM6	Hela	4	1	1	212479	111956	53%	10162	6138	3348
OM6	Hela	4	1	1	234372	133546	57%	10269	6808	4101
OM6	Hela	4	1	2	129933	52523	40%	8995	3295	1127
OM6	Hela	4	1	2	141878	62774	44%	9369	4097	1566
OM6	Hela	4	2	1	291192	151906	52%	10468	7360	4635
OM6	Hela	4	2	1	261932	154769	59%	10503	7442	4729
OM6	Hela	4	2	2	2279390	960410	42%	8474	8086	7674
OM6	Hela	4	2	2	158861	119662	75%	10064	6275	3624
OM6	Hela	4	4	1	258732	93063	36%	10062	5689	2785
OM6	Hela	4	4	1	175854	107480	61%	10156	6287	3512
OM6	Hela	4	4	2	207550	156801	76%	10395	7339	4781
OM6	Hela	4	4	2	146975	112963	77%	10028	6267	3519
OM6	Hela	18	1	1	108935	75979	70%	9946	5124	2297
OM6	Hela	18	1	1	281556	218901	78%	10831	8540	6092
OM6	Hela	18	1	2	229945	82983	36%	9935	5247	2571
OM6	Hela	18	1	2	161878	80571	50%	9948	5148	2376
OM6	Hela	18	2	1	112089	80908	72%	10092	5458	2587
OM6	Hela	18	2	1	191178	154354	81%	10649	7833	5016
OM6	Hela	18	2	2	97018	39422	41%	8628	2692	893
OM6	Hela	18	2	2	111756	57099	51%	9508	4006	1576
OM6	Hela	18	4	1	105243	87278	83%	10100	5780	2814
OM6	Hela	18	4	1	240644	200976	84%	10795	8679	6224
OM6	Hela	18	4	2	223929	95769	43%	10204	6009	3099
OM6	Hela	18	4	2	183300	145216	79%	10607	7781	4816

Supplemental Table 3. Calibration of duplex MIPs process: Hyb -Gap-Dig | Hyb means hybridization, the first step in duplex MIPs capture protocol. Gap means gap filing, the second step. Dig is the third step, linear DNA digestion. Green highlighted the protocol we chosen as standard. The success rate was calculated as: mapped reads/total reads. The loci captured were defined as loci that has at least 1 mapped read.

BluePippin Size (bp)	Name	TotalReads	Total Success	Success Rate	Loci >0 Captured
300	W151020 p2-C9	61816	57226	92.6%	7944
240-340	W151020 p2-C9	144518	130517	90.3%	9783
270-310	W151020 p2-C9	87924	82158	93.4%	8791
300	H1- 090215-B3	87665	83768	95.6%	3075
240-340	H1- 090215-B3	164359	155680	94.7%	4046
270-310	H1- 090215-B3	122252	117106	95.8%	3574
300	H1- 090215-B3	85631	81251	94.9%	2891
240-340	H1- 090215-B3	178585	168311	94.2%	3985
270-310	H1- 090215-B3	123557	117945	95.5%	3411
300	H1- 090215-B6	129546	123914	95.7%	5568
240-340	H1- 090215-B6	387493	368533	95.1%	6850
270-310	H1- 090215-B6	213020	203982	95.8%	6209
300	H1- 090215-E9	114190	109002	95.5%	5194
240-340	H1- 090215-E9	460648	436100	94.7%	6728
270-310	H1- 090215-E9	137124	131168	95.7%	5569
300	H1- 090215-A1	77196	73307	95.0%	5230
240-340	H1- 090215-A1	154987	146026	94.2%	6327
270-310	H1- 090215-A1	120505	114812	95.3%	5882
300	H1- 090215-F5	14620	13535	92.6%	3706
240-340	H1- 090215-F5	184932	170304	92.1%	7744
270-310	H1- 090215-F5	22392	20930	93.5%	4533
300	PC2	12488	11149	89.3%	5004
240-340	PC2	95192	81596	85.7%	10347
270-310	PC2	9078	8208	90.4%	4454

Supplemental Table 4. Calibration of Sequencing Library Size Selection | PC2 was bulk DNA; all the other samples were single cell WGA DNA

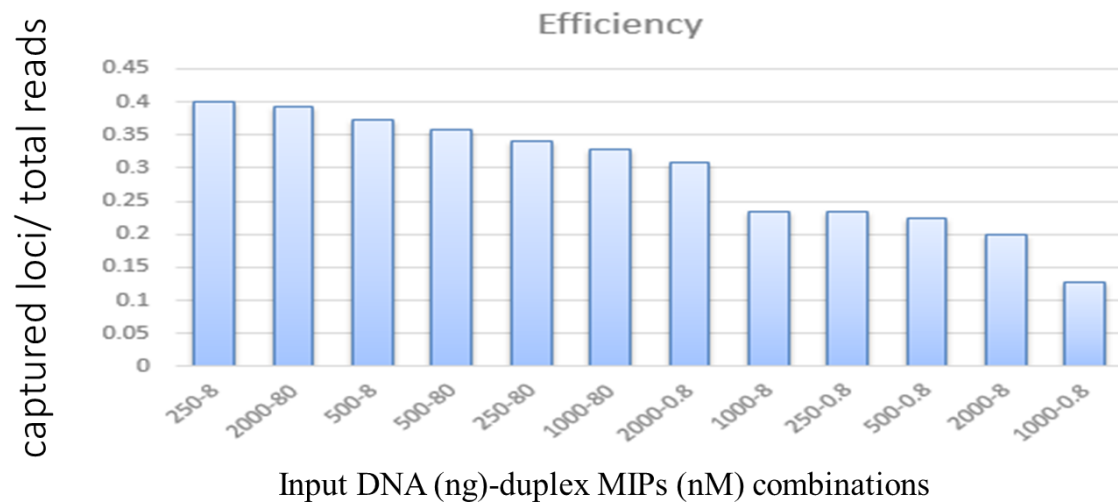
Calibration the impact of ratio between MIPs concentration and template DNA amount

		1	2	3	4	5	6	7	8	9	10
Total Reads	probe	80nM	80nM	8nM	8nM	0.8nM	0.8nM	0.08nM	0.08nM	0nM	0nM
	HeLa										
A	2000	158	12655	33748	29465	14111	2383	7146	259	51	90
B	1000	20166	15217	26525	24094	51035	56023	2941	3043	58	87
C	500	15363	13718	13075	11051	36246	17843	1837	1086	34	71
D	250	18666	12935	9976	10861	24819	20520	2364	1895	41	54
E	100 ng	4959	6063	1182	1643	3682	4032	380	383	17	30
F	10 ng	175	494	2989	87	973	595	66	89	27	35
G	1 ng	76	292	2103	32	172	261	28	26	26	28
H	0.1 ng	48	351	1126	1356	144	389	38	51	27	78
I	0.01 ng	122	56	1607	408	199	129	65	37	136	87
J	0	137	49	439	240	67	187	35	22	65	51

Supplemental Table 5. Total Reads of the calibration of DNA, duplex MIPs ratio.

		1	2	3	4	5	6	7	8	9	10
Loci >0	probe	80nM	80nM	8nM	8nM	0.8nM	0.8nM	0.08nM	0.08nM	0nM	0nM
	HeLa										
A	2000	127	4984	6386	6241	4346	1383	2866	170	27	47
B	1000	6102	5422	6019	5815	6784	6916	1645	1644	20	39
C	500	5332	5041	4670	4288	6268	4937	1076	685	10	25
D	250	5765	4824	4109	4221	5432	5100	1242	1148	11	18
E	100 ng	2614	3177	883	1056	1604	1970	231	240	12	14
F	10 ng	60	309	936	35	461	293	35	46	11	21
G	1 ng	11	236	212	15	38	59	10	10	14	15
H	0.1 ng	13	274	30	20	18	38	23	18	13	29
I	0.01 ng	46	20	40	23	45	18	19	25	68	51
J	0	31	17	33	42	23	13	15	14	35	30

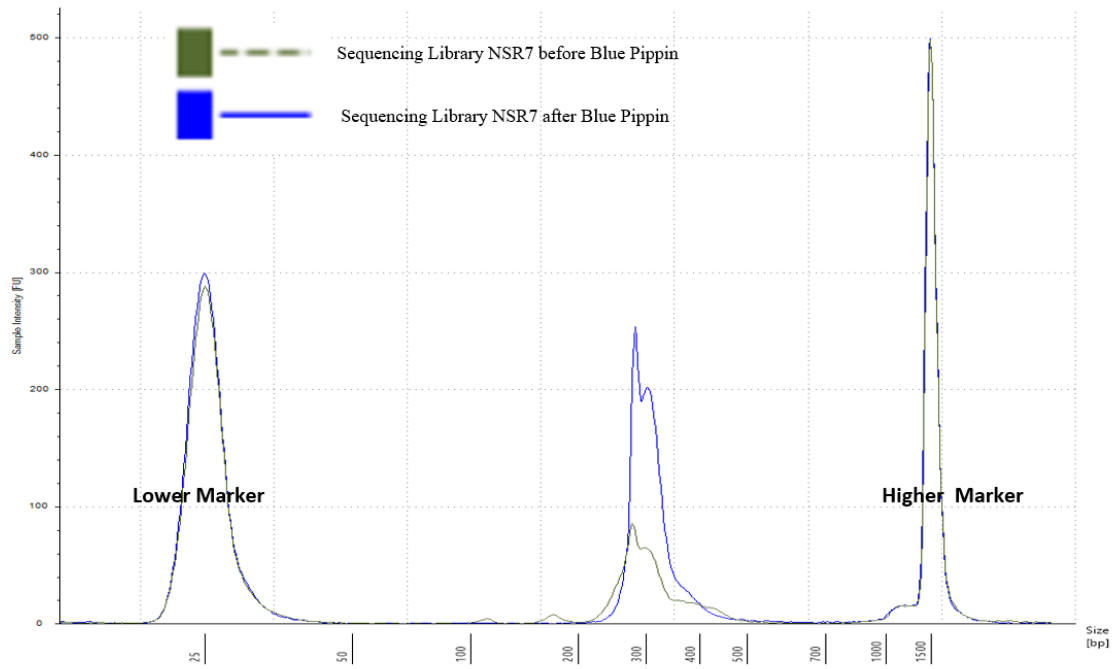
Supplemental Table 6. Captured loci of the calibration of DNA, duplex MIPs ratio.



Supplemental Figure 4. Efficiency comparison between different probes: template ratio | Efficiency was calculated by captured loci/ total reads as show in Supplemental Table 5 and 6.

Sequencing library quality control

As a sequencing library quality control step, Tape Station was applied to the libraries before and after Blue Pippin. 240~340bp range size selection setting was used on 2% V1 cassette. Two side product peaks were removed by BluePippin as shown below.



Supplemental Figure 5. Library Quality Control: Tape Station before Blue Pippin and after Blue Pippin

General cost of duplex MIPs capture pipeline was listed below.

The cost was calculated by 200 cells/run, WGA cost and sequencing run cost were not included.

Reagents	Cat.No	Cost(\$)	Total Volume(ul or reactions)	(ul) Volume per Reaction	(\$ Cost per Reaction
duplex MIPs	Home made	2200	9400000	1	0.000234043
Betaine solution 5M	B0306 1VL Sigma	49	1500	4	0.13
++Phusion High-Fidelity DNA Polymerase - 500 units	NEB-M0530L	424	250	0.4	0.68
Ampligase 10X Reaction Buffer 5ml	A1905B EPICENTRE	66	5000	2	0.03
Ampligase DNA Ligase W/O Buffer 10,000U	A3210K EPICENTRE	693	2000	1	0.35
Exonuclease I (E.coli) - 15,000 units	NEB-M0293L	268	750	0.175	0.06
Exonuclease III (E.coli) - 25000 units	NEB-M0206L	236	250	0.18	0.17
++ RecJf - 1,000 units	NEB-M0264L	272	167	0.1	0.16
Exonuclease T - 1,250 units,	NEB-M0265L	280	250	0.08	0.09
T7 Exonuclease - 5,000 units,	NEB-M0263L	248	500	0.4	0.20
Lambda Exonuclease	M0262L	268	1000	0.02	0.01
NEBNext Ultra II Q5 MasterMix	NEB-M0544L	395	12500	10	0.32
MinElute PCR Purification Kit (250) '	QIAGEN 28006	594	250reactions	2reaction/Run	0.02
Qubit® dsDNA HS Assay Kit,	Q32854	269	500reactions	2reaction/Run	0.01
Agencourt Ampure XP Beads	BeckmanCo ulter A63881	1485	600000	16	0.04
2% Agarose, dye-free, w/ internal standards, BluePippin, 100 - 60,	BDF2010	475	50reactions	1reaction/Run	0.05
TapeStation Screen Tap	5067-5582	211	112 reactions	2reaction/Run	0.02
TapeStation Reagents	5067-5583	90.33	112 reactions	2reaction/Run	0.01
				Consumable	3
	Initial Cost	8523.33		Cost per Cell	5.33

Supplemental Table 8. The cost of duplex MIPs capture pipeline.

The scalability of duplex MIPs

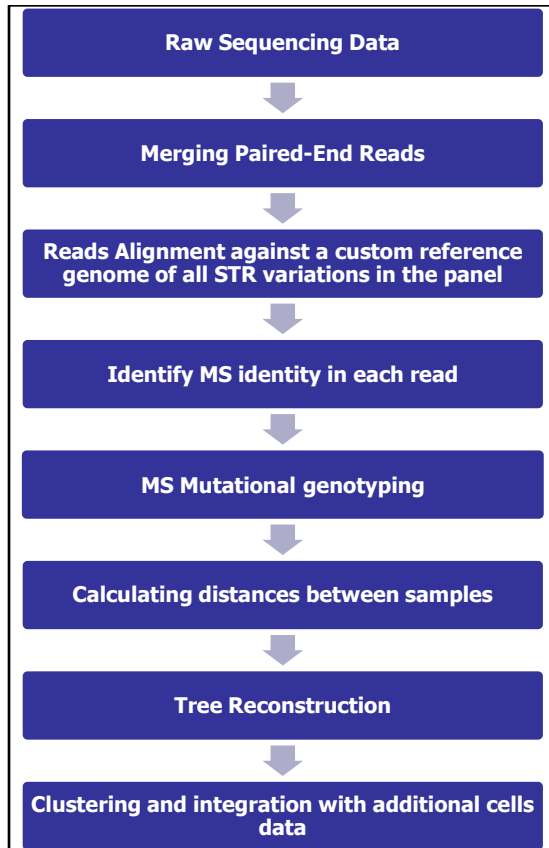
Shown together with AA pipeline, the cost trend of duplex MIPs while scaling up.



Supplemental Figure 6. Cost and Scalability between Access Array and duplex MIPs

A schematic diagram of the computational pipeline

A new mapping strategy was replaced the one in our previous work⁶. Reads were aligned against a custom reference genome of all possible STR variations in the panel. This improved the computing efficiency. All the source code was available in <https://github.com/ofirr/clineage>



Supplemental Figure 7. A schematic diagram of the computational pipeline

1. Bruce Budowle, T.R.M., Stephen J. Niezgoda and Barry L. Brown CODIS and PCR-Based Short Tandem Repeat Loci: Law Enforcement Tools.
2. Guilmatre, A., Highnam, G., Borel, C., Mittelman, D. & Sharp, A.J. Rapid multiplexed genotyping of simple tandem repeats using capture and high-throughput sequencing. *Hum Mutat* **34**, 1304-1311 (2013).
3. Duitama, J. et al. Large-scale analysis of tandem repeat variability in the human genome. *Nucleic Acids Res* **42**, 5728-5741 (2014).
4. Carlson, K.D. et al. MIPSTR: a method for multiplex genotyping of germline and somatic STR variation across many individuals. *Genome Res* **25**, 750-761 (2015).

5. Shlush, L.I. et al. Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. *Blood* **120**, 603-612 (2012).
6. Biezuner, T. et al. A generic, cost-effective, and scalable cell lineage analysis platform. *Genome Res* **26**, 1588-1599 (2016).