Microsatellite Instability & Tumor Heterogeneity

Jack DiGiovanna, PhD

GM, NGS Applications; Seven Bridges

Isaac M. Neuhaus, PhD

Director, Computational Genomics, Bristol-Myers Squibb

Stefan Kirov, PhD

Principal Research Scientist, Bristol-Myers Squibb

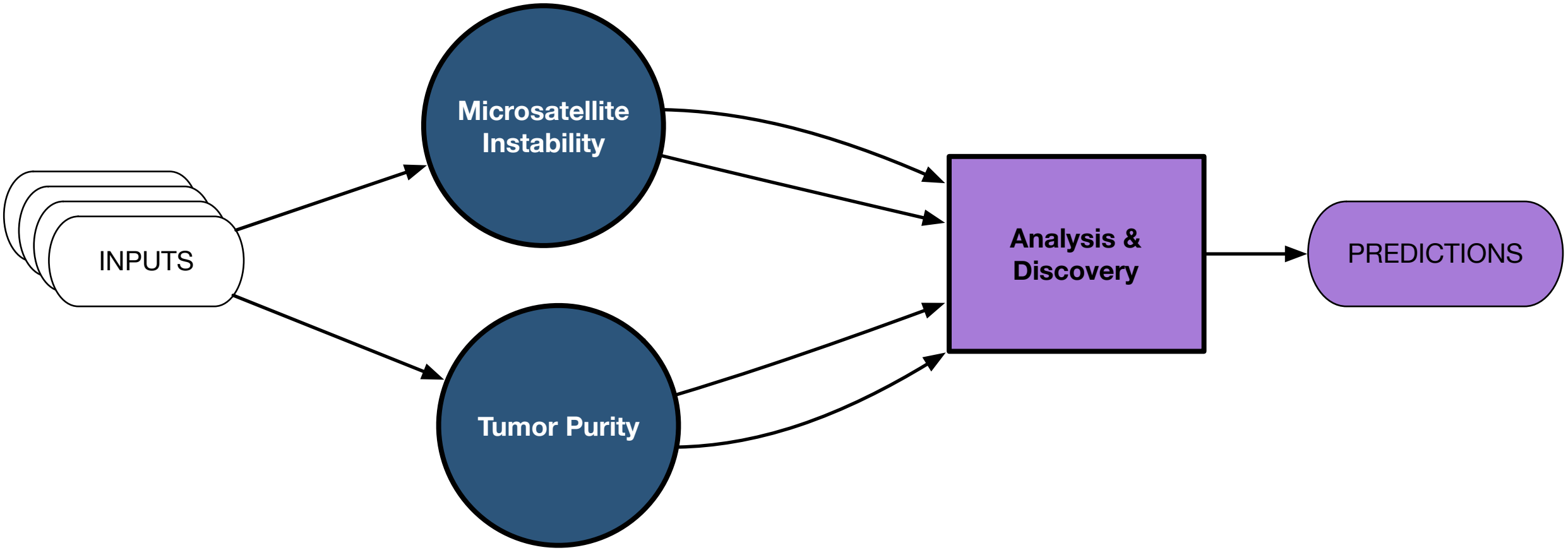
Danielle Greenawalt, PhD

Head, Early Stage Translation Bioinformatics Oncology, Bristol-Myers Squibb

SevenBridges

Bristol-Myers Squibb

Predict immuno-oncology outcomes



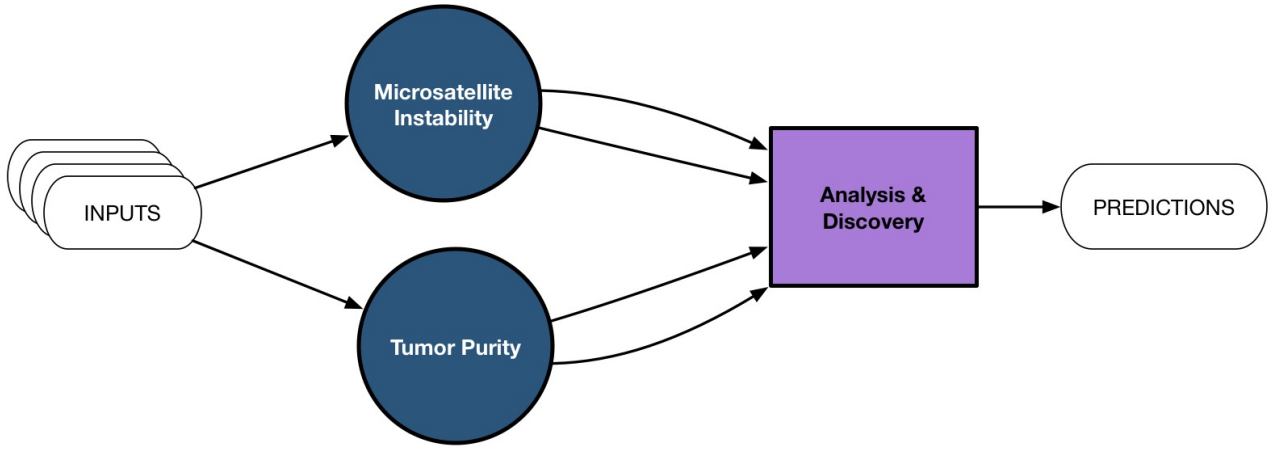
Overview

Microsatellite instability

Tumor heterogeneity

Combining results

Comparison to clinical testing



Microsatellites

Short tandem repeats of 1 to 6 base-pairs

Often site of DNA replication slippage

Microsatellites have high germline mutation rates (10^{-4} to 10^{-2} per locus per generation)

- 200x > CNVs
- 200,000x > SNPs

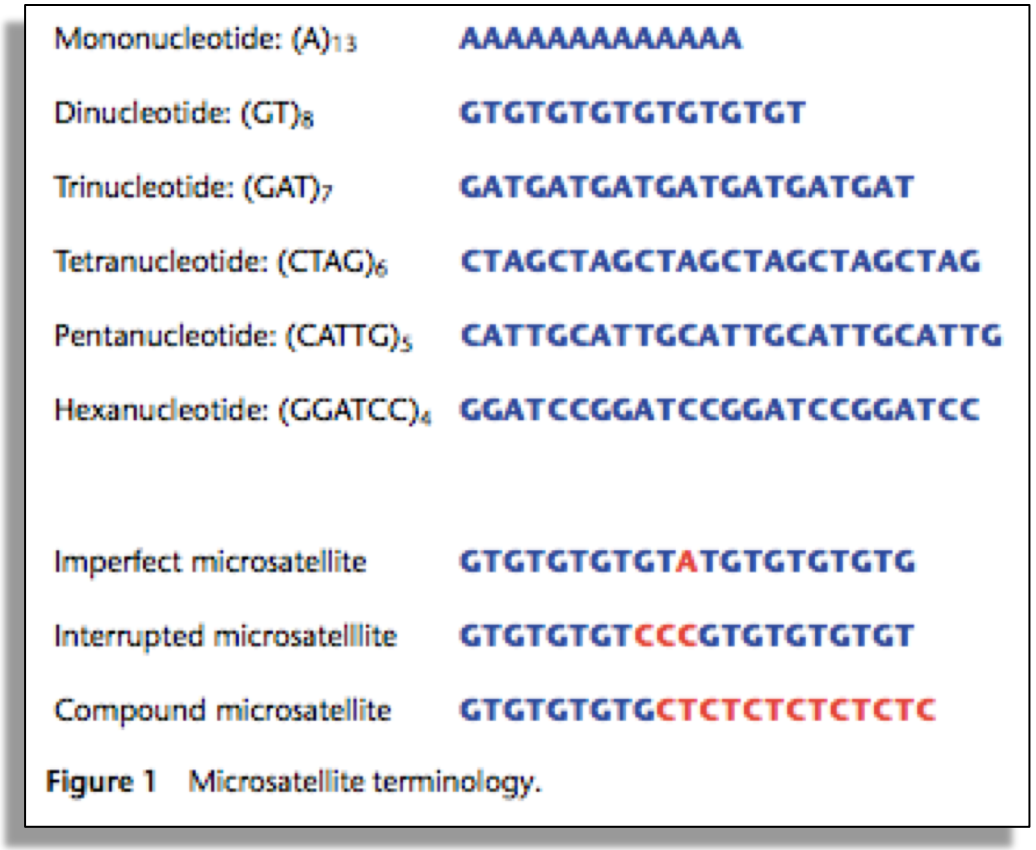


Figure 1 Microsatellite terminology.

image credit: *Schlotterer & Harr, Encyclopedia of Life Sciences, 2001*

*Subramanian et al. (2003) Genome Biol. **4**: R13*
*Ananda et al. (2012) Genome Biol. Evol. **5**(3): 606-620*

SevenBridges

Bristol-Myers Squibb

Microsatellite instability (MSI)

DNA mismatch repair (MMR) that is not functioning properly results in genetic hypermutability – manifests as MSI

- Key components include MSH1, PMS1-8, MSH2/3, PARP, BRCA, ERCC2
- Can be observed via PCR assays or whole exome/whole genome sequencing

FDA has approved treatments for MMR deficient (or MSI high tumors)

This makes MSI status useful patient stratification biomarker

- FDA recently approved a companion diagnostic test (F1CDx)

Kim and Park (2014) Cancer Res. ***74****(22): 6377-6382*
Salipante et al. (2014) Clin. Chem. ***60****(9): 1192-1199*

MSI detection challenges

Aligners can reject large INDELs

DNA amplicons generated by PCR can lead to false positives due to replication ‘slippage

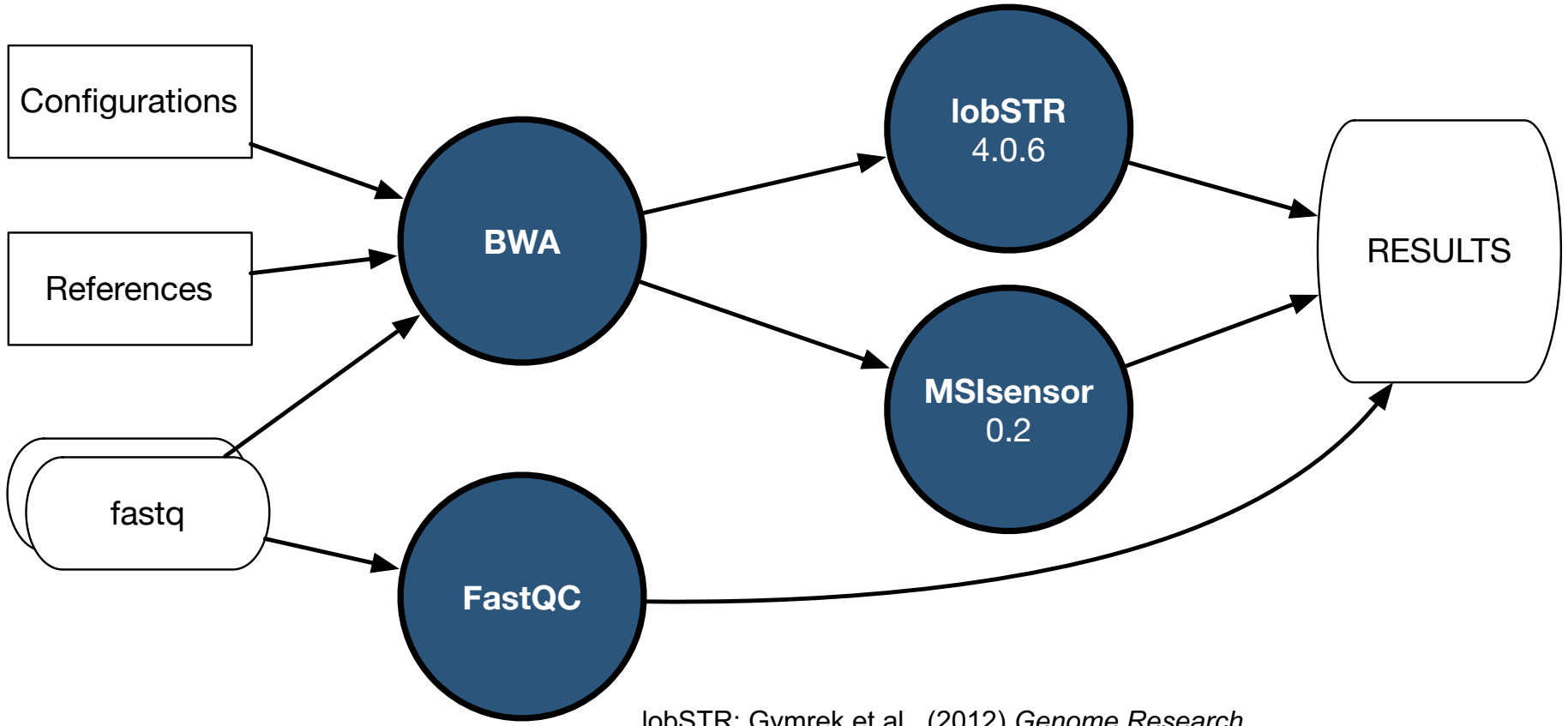
There are several applications that can call MSI status from NGS data

– MSIsensor, mSINGS, MANTIS, MOSAIC, lobSTR

MSI profiling

lobSTR 4.0.6 – variant calling of MSI mutations

MSIsensor – tests statistical difference in STR loci length distributions



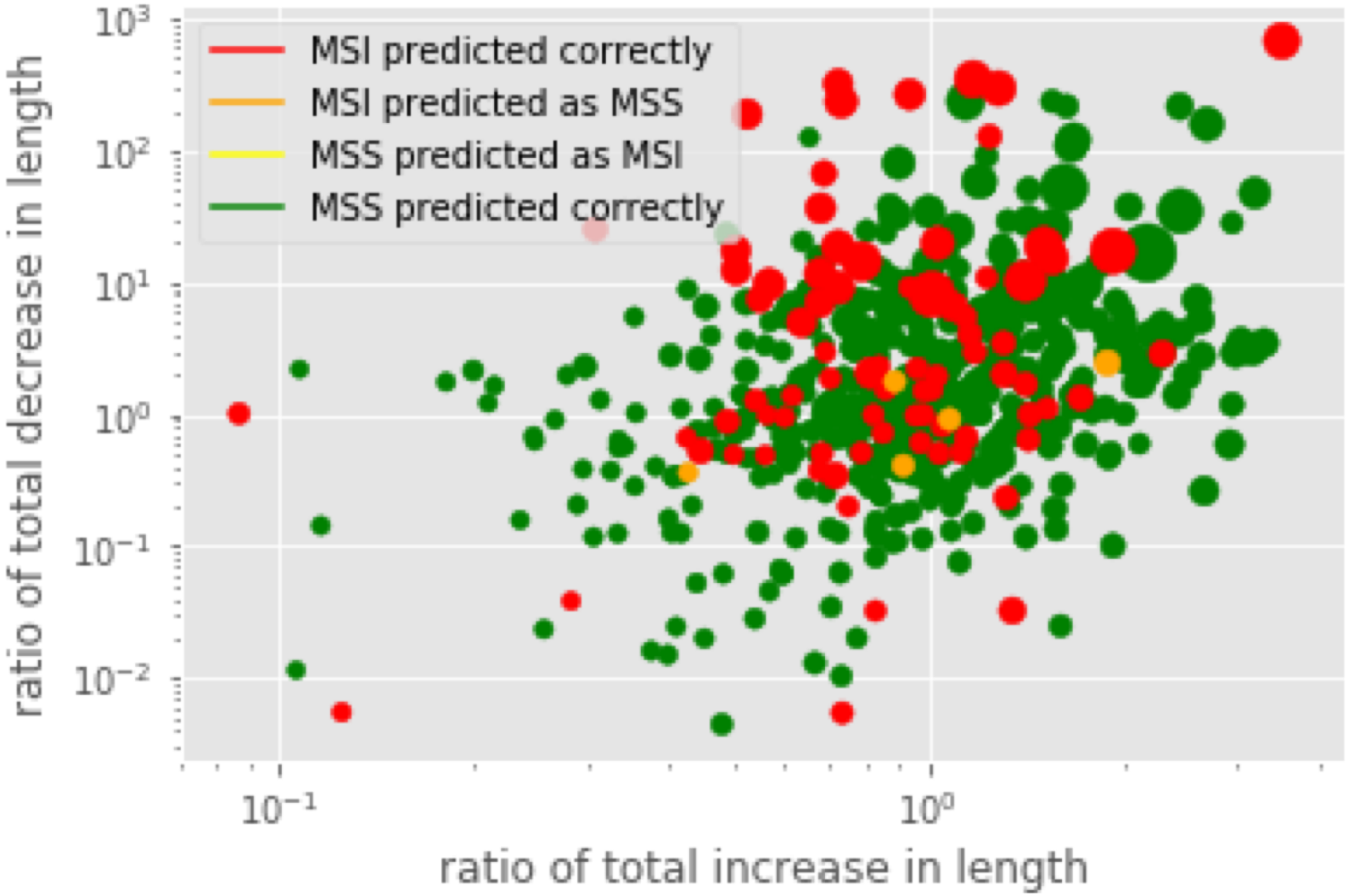
lobSTR: Gymrek et al., (2012) *Genome Research*
MSIsensor: Niu *et al.* (2014) *Bioinformatics*
WF: Tesic, & Kalinic (2017) *ISMB*

Validating predictions (lobSTR)

Colorectal adenocarcinomas with MSI status (n = 592)

Support Vector Machine to classify MSI vs MSS

- Precision = 100%
- Recall = 90.7%



SevenBridges

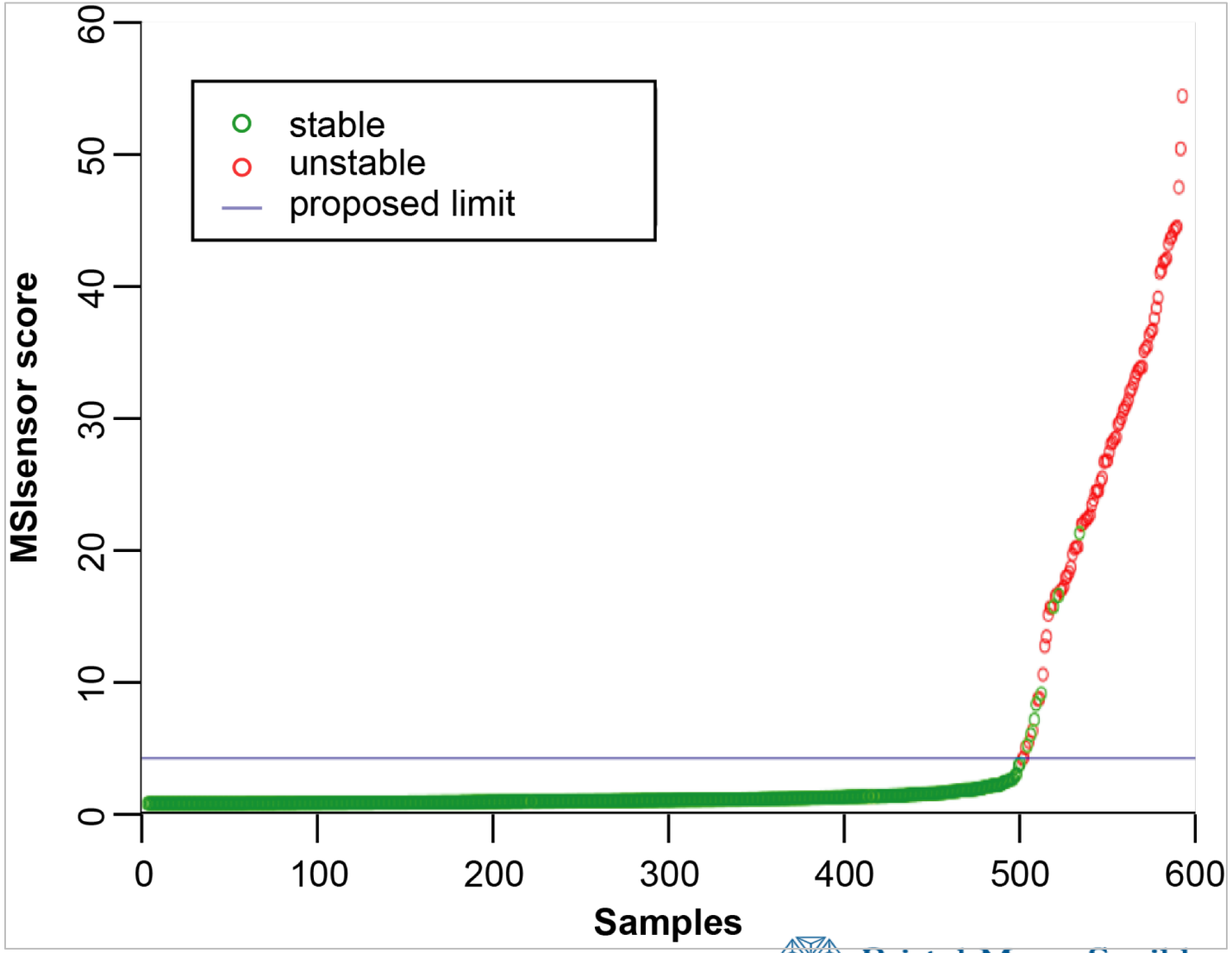
Bristol-Myers Squibb

Validating predictions (MSIsensor)

Colorectal adenocarcinomas with MSI status (n = 592)

Compared output with published threshold

- Precision = 91.1%
- Recall = 95.4%



SevenBridges

Bristol-Myers Squibb

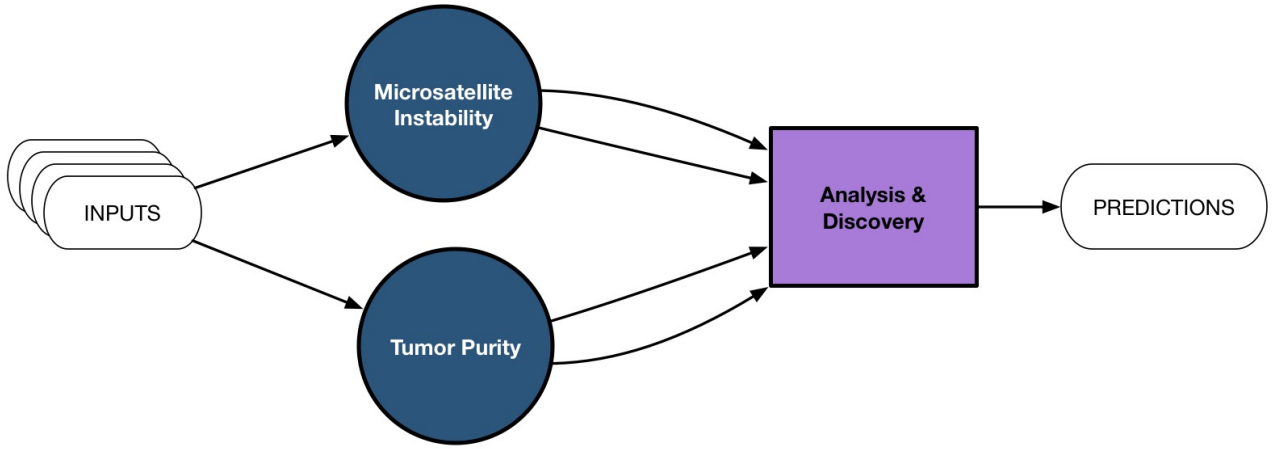
Overview

Microsatellite instability

Tumor heterogeneity

Combining results

Comparison to clinical testing



Tumor Heterogeneity (purity)

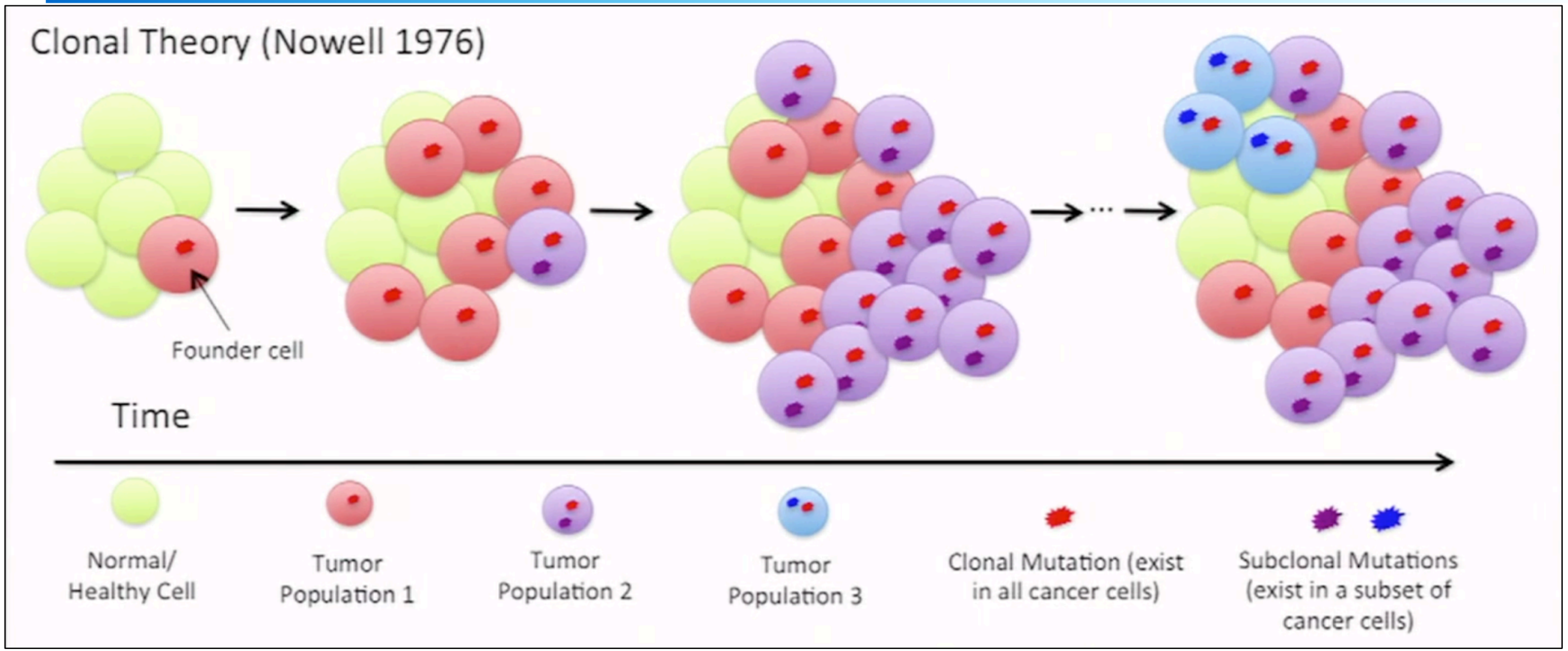


image credit: *Nowell (1976) Science*

Clones have different potentials to metastasize

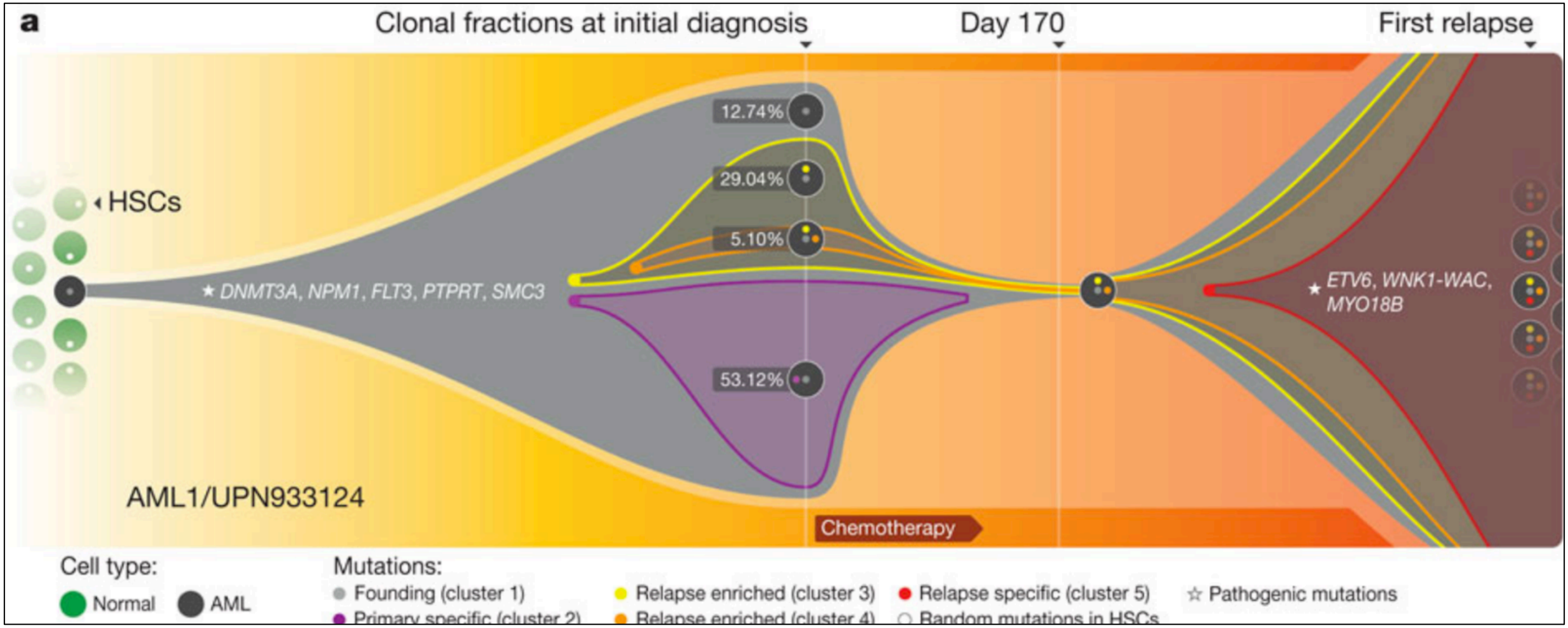
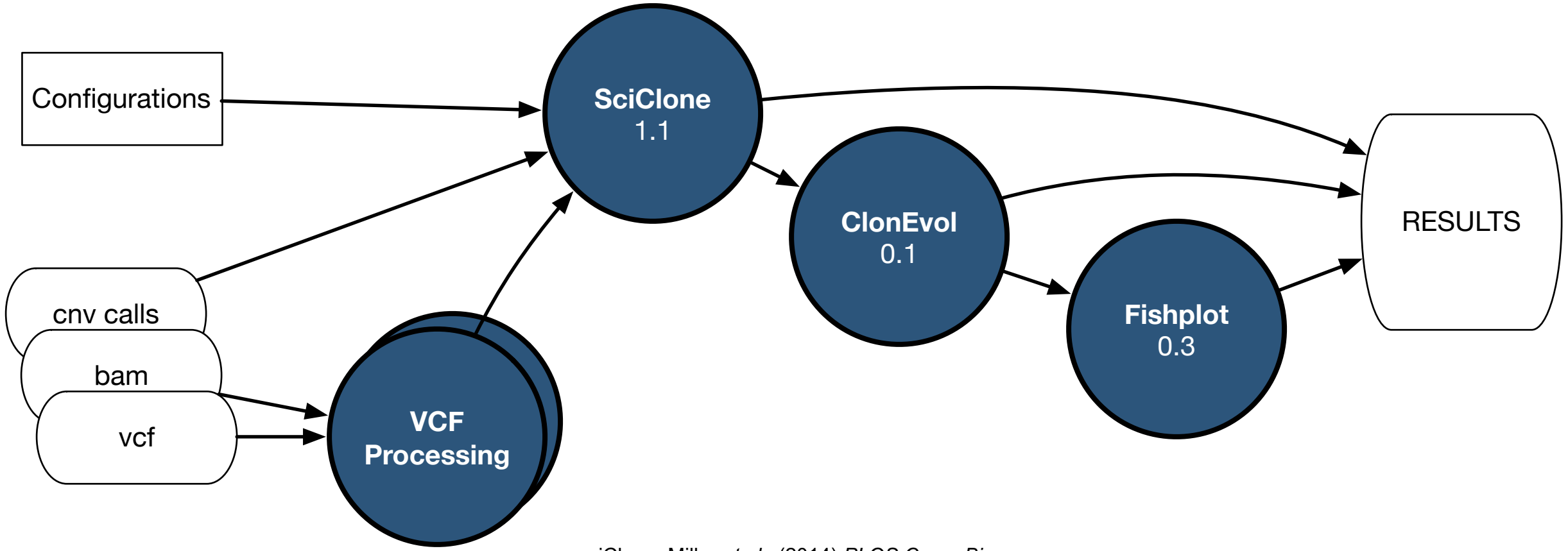


image credit: *Ding et al (2012) Nature*

Heterogeneity (purity) workflow

SciClone 1.1 – predicts heterogeneity from clustering AF

Fishplot – plots of clonal evolution



sciClone: Miller *et al.,* (2014) *PLOS Comp Bio*
WF: Sipetic (2017) *ISMB*

SevenBridges

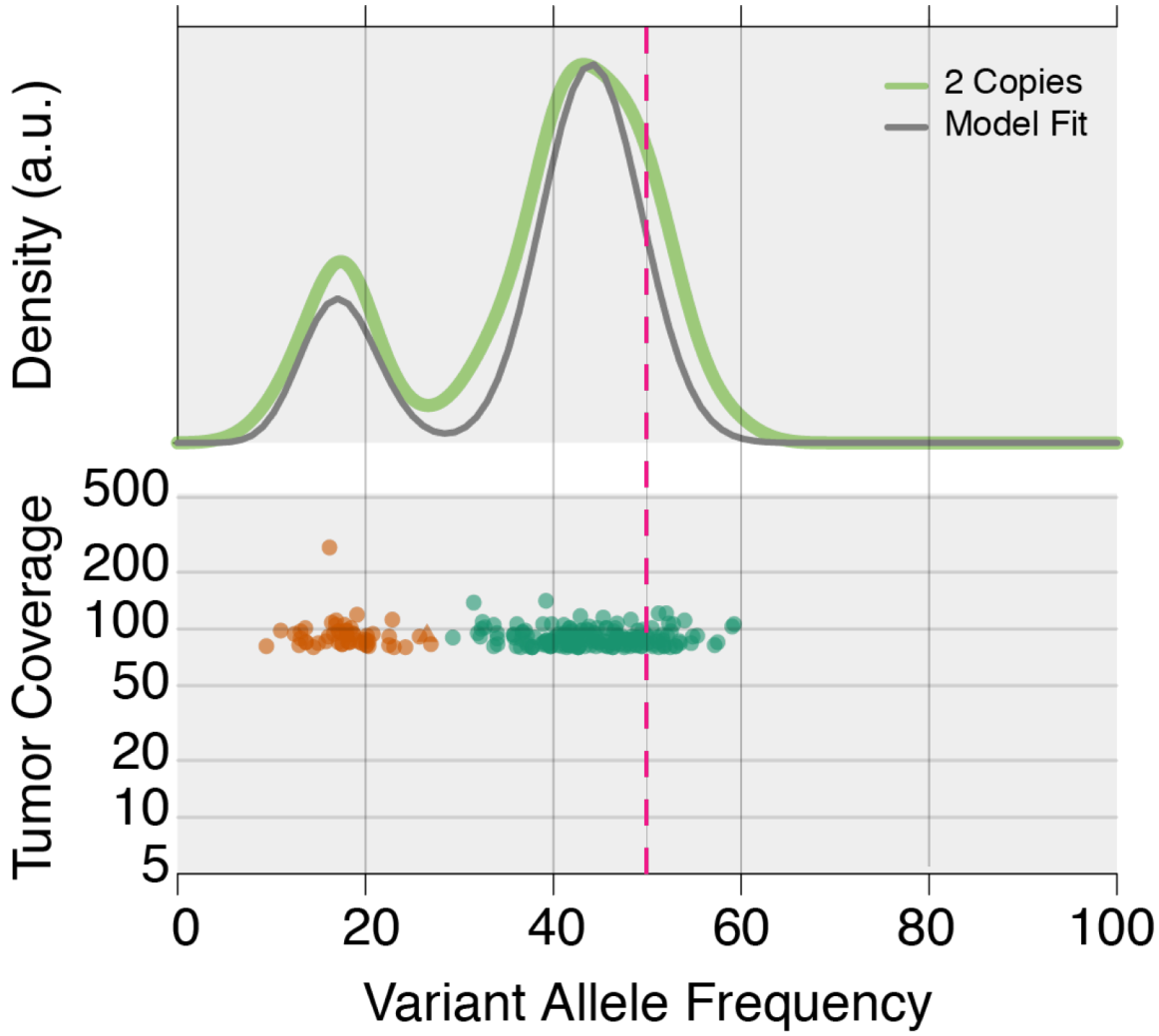
Bristol-Myers Squibb

Heterogeneity validation

ICGC-TCGA DREAM Somatic Mutation Calling - Tumor Heterogeneity Challenge

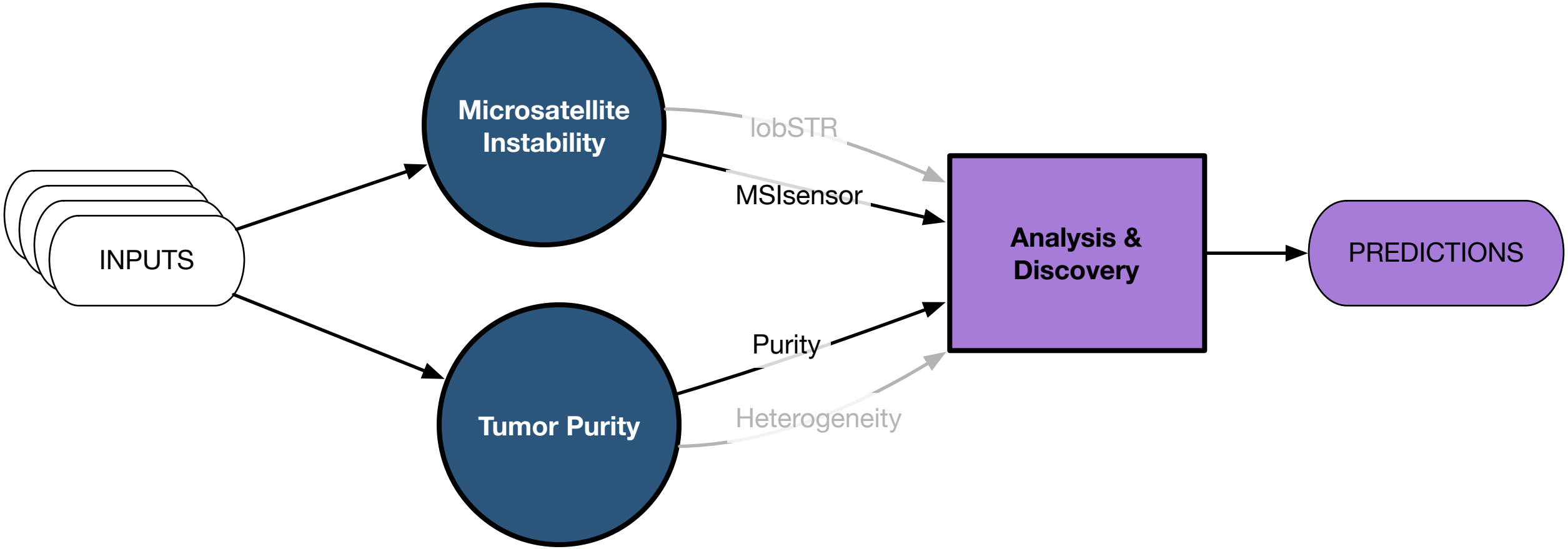
- *Purity* estimate within 5%
- 5 of 6 *groups of tumor cells with similarly descriptive mutations* (clusters) predicted

	Purity		Number of clusters		Cluster info	
	Truth	SBG	Truth	SBG	Truth	SBG
Tumor 1	70%	64.9%	4	3	1: 713 (70%) 2: 1,483 (45%) 3: 1 (13%) 4: 1,285 (10%)	1: 729 (65%) 2: 1,923 (43%) 3: 830 (15%)
Tumor 2	90%	88.3%	2	2	1: 186 (90%) 2: 54 (36%)	1: 186 (88%) 2: 54 (35%)



SevenBridges

Bristol-Myers Squibb


Microsatellite Instability
lobSTR
MSIsensor
INPUTS
Analysis & Discovery
PREDICTIONS
Purity
Tumor Purity
Heterogeneity

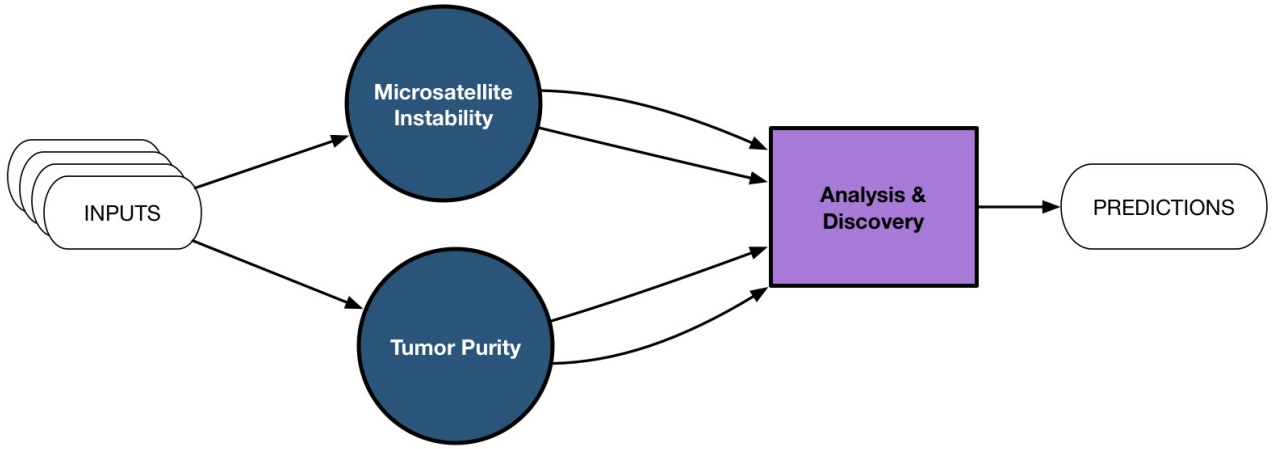
Overview

Microsatellite instability

Tumor heterogeneity

Combining results

Comparison to clinical testing

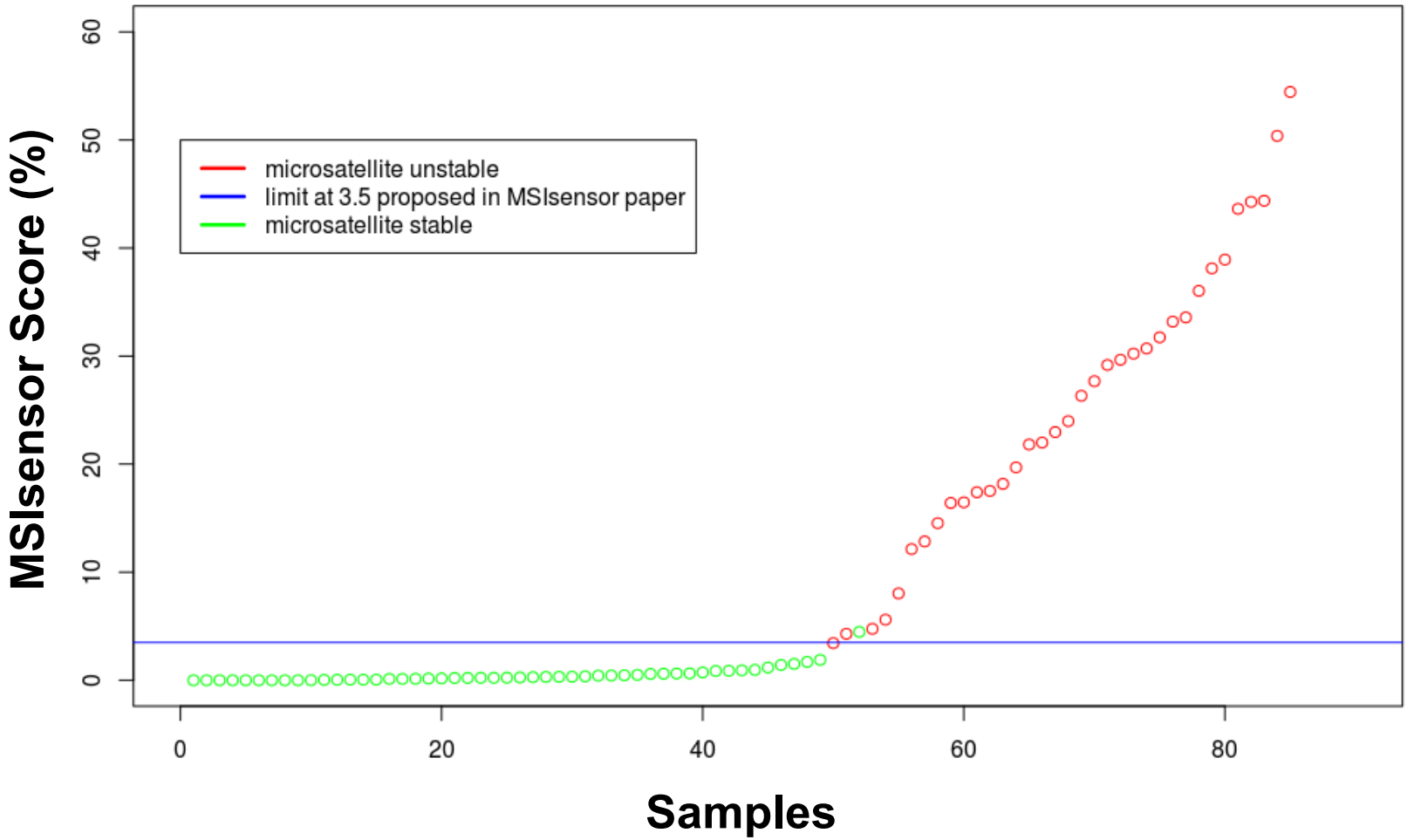


Benchmarking MSIsensor

Dataset of 85 legacy CRC samples as a truth set

MSI, as many other DNASeq measurements, relies on somatic mutations

Accurate in samples with high to moderate tumor content



SevenBridges

Bristol-Myers Squibb

Benchmarking SciClone

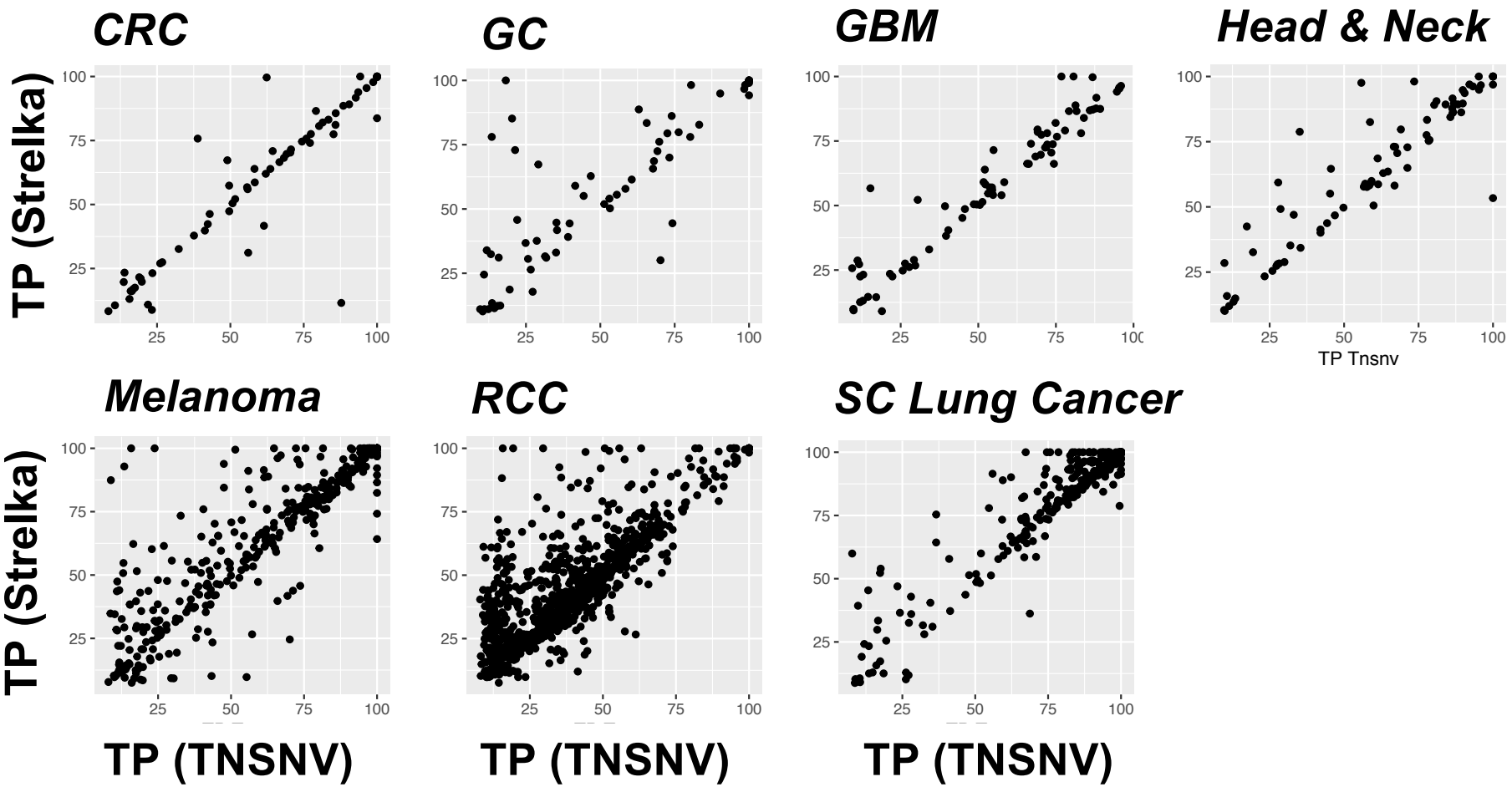
TNSNV vs Strelka

SciClone utilizes the CNV/LOH neutral parts of the genome and quantification of variant allele frequencies within those regions. The variant data is used to build a graph with the centroid being the germ line.

In order to filter samples we use the centroid for estimating the purity

Purity can be over >100% in the raw output - would indicate lack of germ cells DNA presence

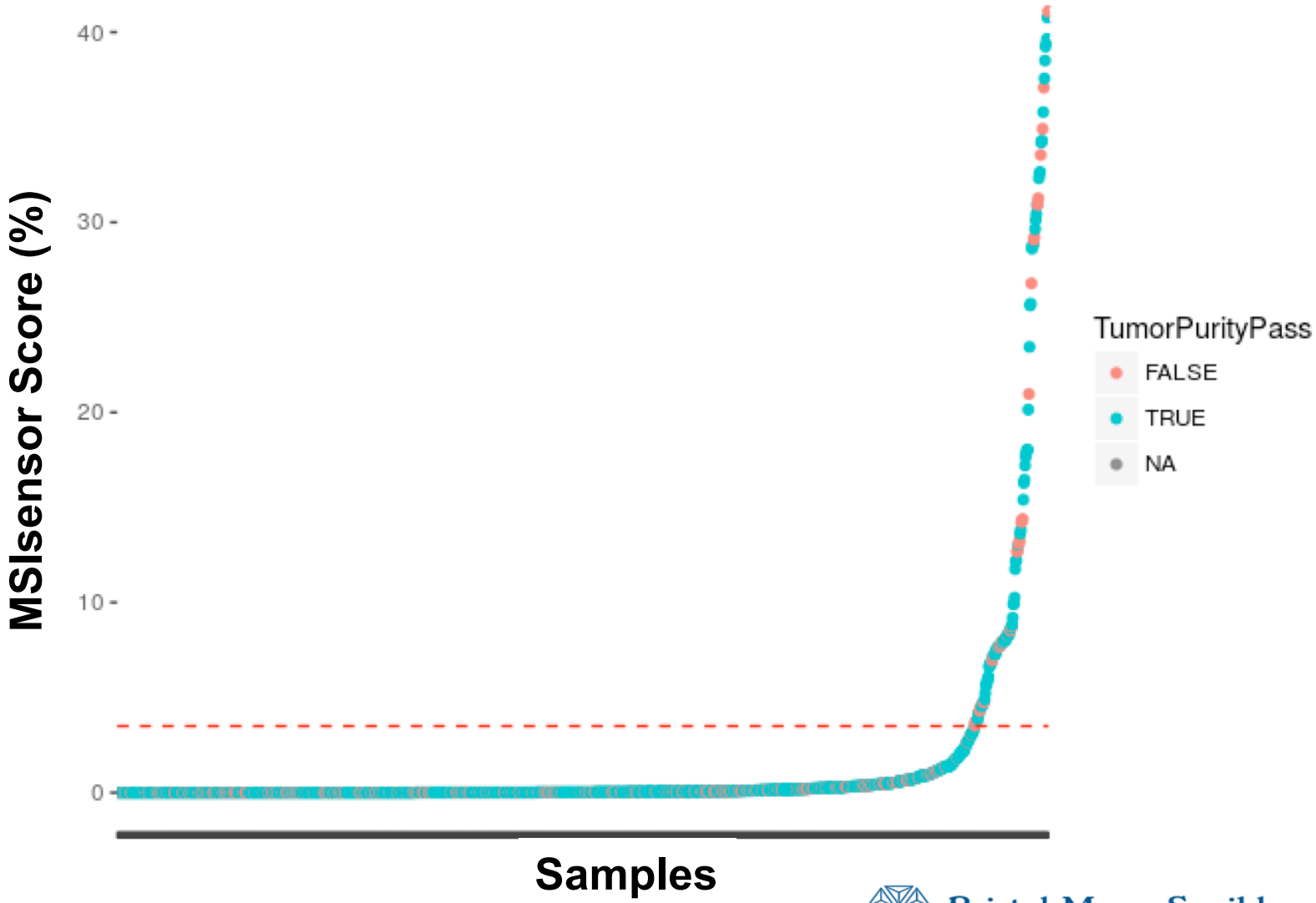
Occasionally these may disagree. Currently we use the mean of both methods and we are in the process of optimizing the approach.



Bristol-Myers Squibb

Clinical MSI data

- **Evaluated MSI status across a number of clinical studies (CRC, RCC, melanoma, etc.)**
- **Distribution of MSI values for 1496 pan-tumor samples**
  - TumorPurityPass is TRUE if purity > 20%



SevenBridges

Bristol-Myers Squibb

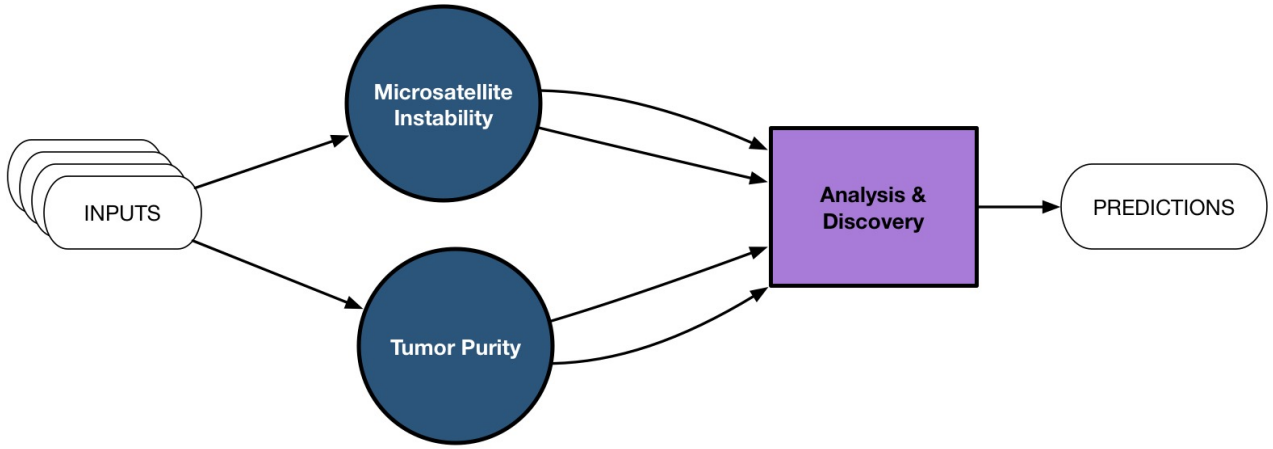
Overview

Microsatellite instability

Tumor heterogeneity

Combining results

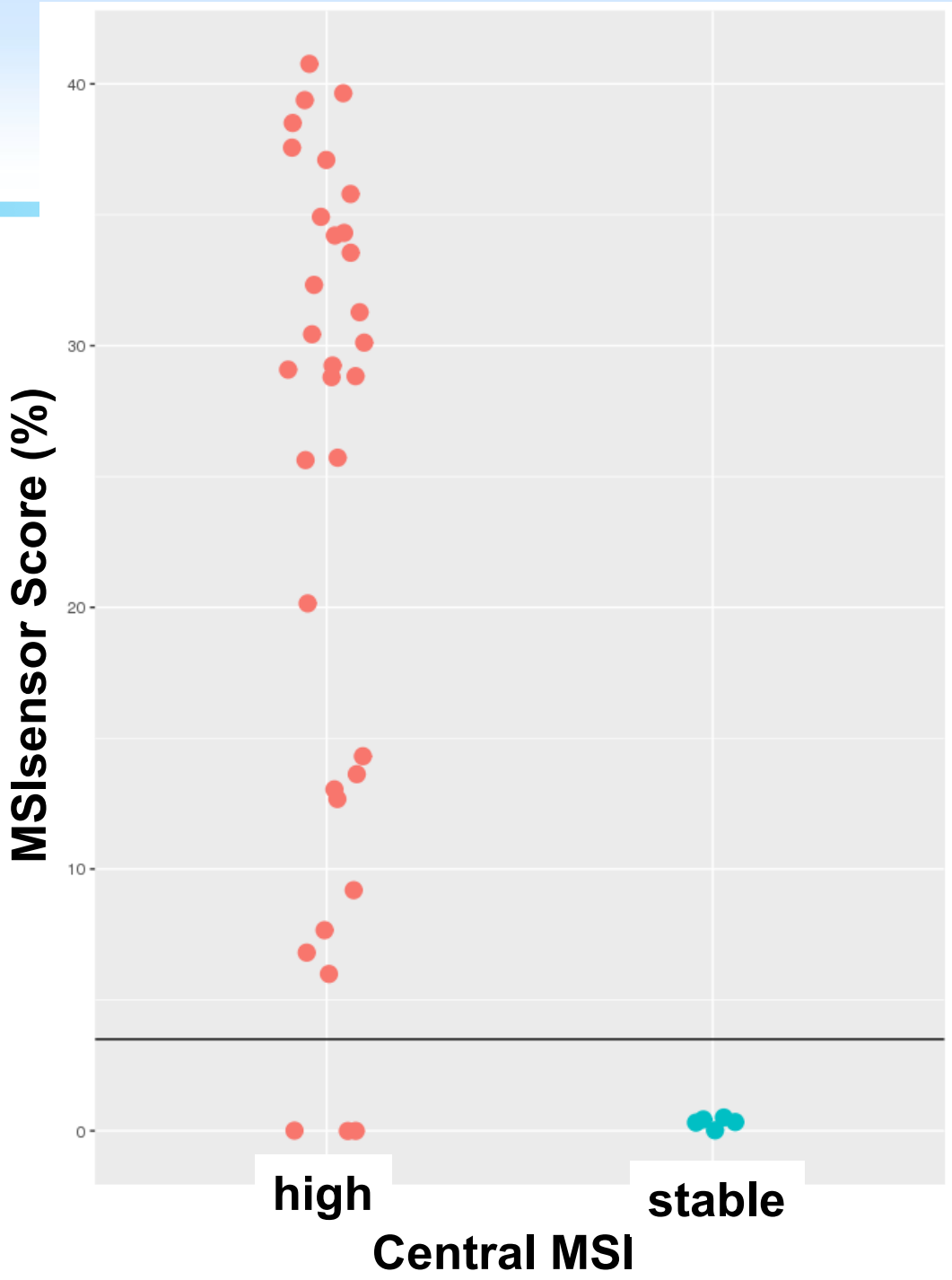
Comparison to clinical testing



MSIsensor vs Lab Results

- 38 samples with PCR or IHC based MSI value available
- Compared to MSISensor using a cutoff score of 3.5

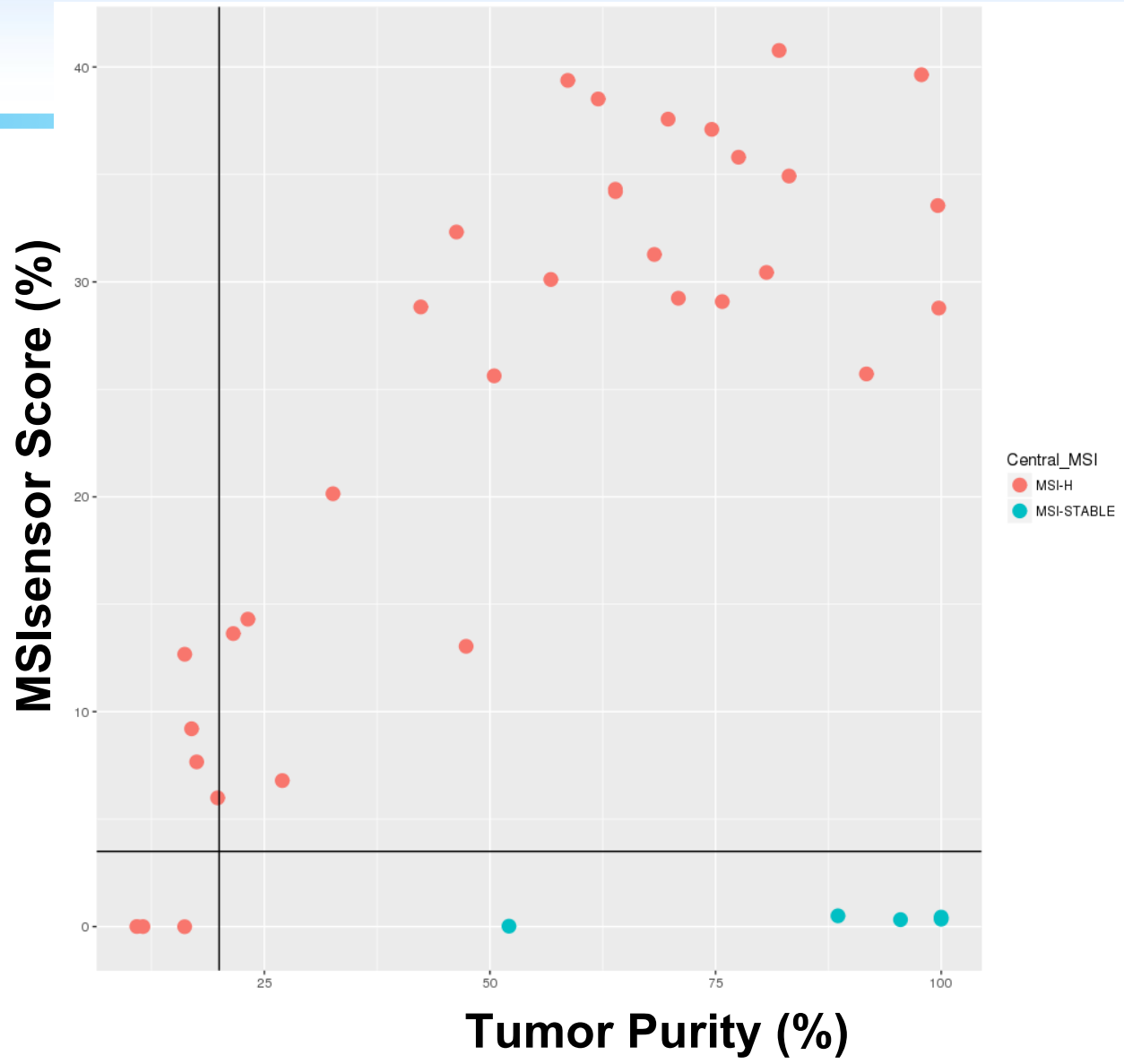
MSI Sensor	*Lab Results*		
		MSI-H	MSI-Stable
	MSI-H	30	0
	MSI-Stable	**3**	5



SevenBridges

MSI Sensor vs Tumor Purity

- Samples that were classified incorrectly by MSIsensor have low tumor purity
- Removing samples with <20% tumor purity allows us to generate concordant calls for all samples



SevenBridges

Bristol-Myers Squibb

Summary

Accurate and robust workflows for MSI and tumor heterogeneity

We have established best practices for MSI status calling based on WES data

- We use *SciClone* to evaluate tumor purity from WES data and remove samples with tumor content less than 20%
- We then utilize *MSIsensor* and call MSI status based on recommended threshold

Currently we are using these guidelines to deliver MSI status for a number of clinical and preclinical studies across multiple indications