

# 干实验过程的方法建立及性能确认

张诗乐 Ph.D, Illumina Inc. Nov 2018



# AGENDA

General Overview – Oncology Strategy

Preamanalytical Validation

Analytical Validation

Genomic biomarkers for Immunotherapy



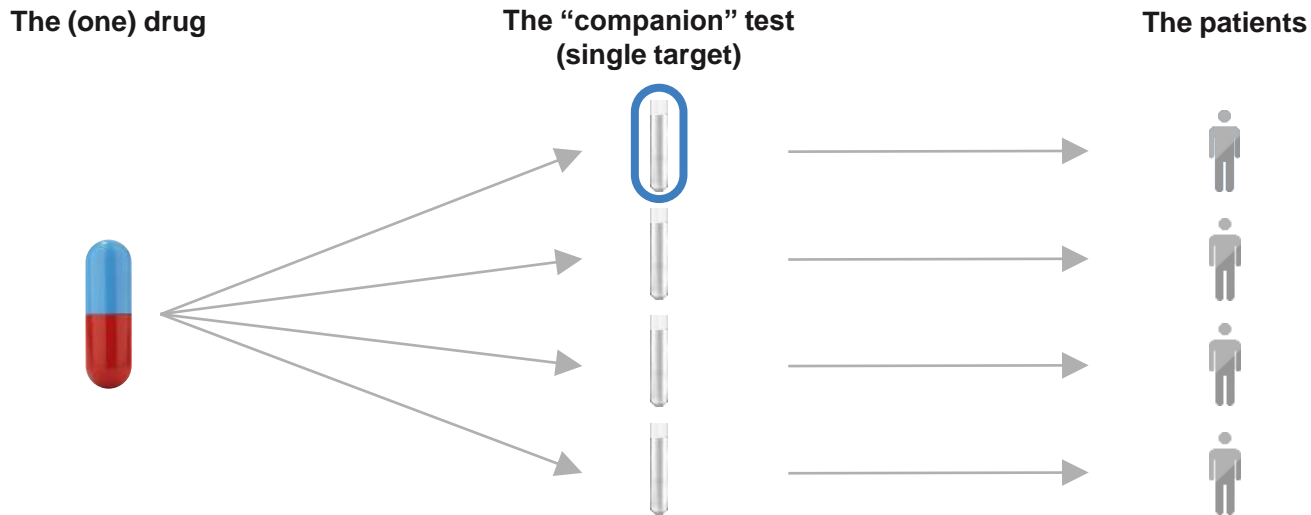
# The Cancer Patient Journey

*Our strategic focus on the Therapy Selection and Monitoring Segments*

			
<b>Hereditary Factors</b>	<b>Screening</b>	<b>Therapy Selection</b>	<b>Monitoring</b>
 <b>Germline</b>	 <b>cfDNA</b>	  <b>Tissue, cfDNA</b>	 <b>cfDNA</b>

# Drug-Centered Oncology Rx

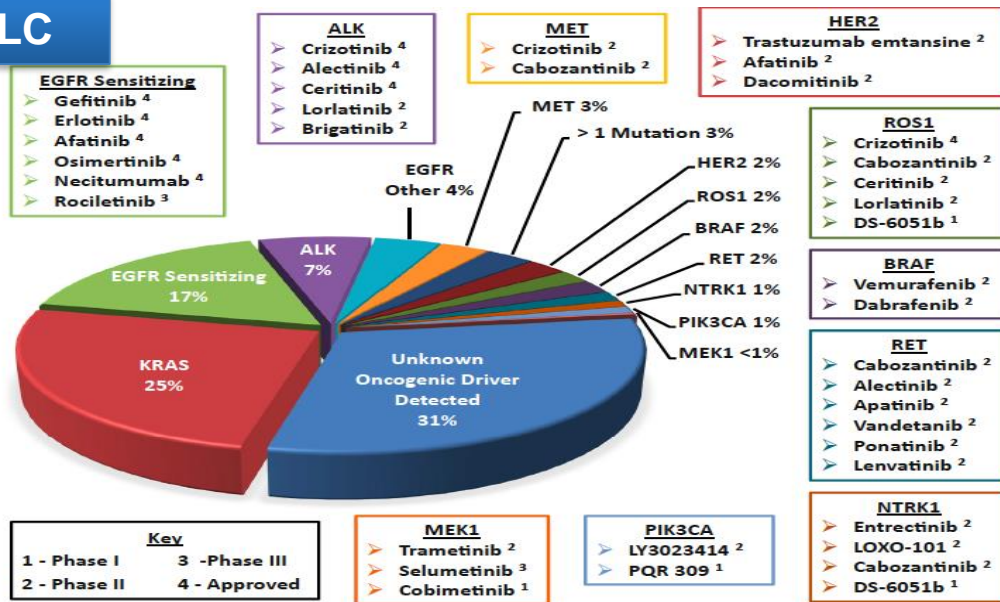
*Traditional approach*



# Too many questions for a small amount of tissue

*Comprehensive testing can save time, sample and impact patients lives earlier*

## NSCLC



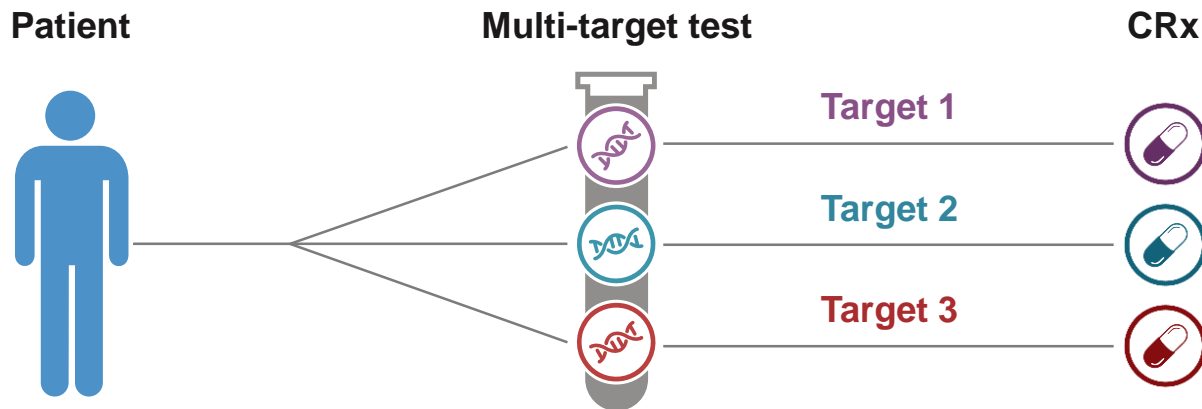
Frequency of molecular aberrations in various driver oncogenes in lung adenocarcinomas and current available drugs against these oncogenic proteins.

**+ New Markers: TMB, MSI**

What if, instead of multiple tests, you could perform a **single test** that looked at **all relevant biomarkers** at the same time?

# In Progress: Precision Oncology Treatment

*From companion diagnostics to companion therapeutics*



# Companion Diagnostic Development

*Partnering to Power Oncology Precision Medicine*

Companion Diagnostic product (under development)  
based on TruSight™ Tumor 170 content

Companion Diagnostic product (under development)  
based on TruSight™ Oncology 500 content



**Loxo to expand oncology menu  
for NextSeqDx**

**Bristol-Myers Squibb to expand  
oncology menu for NextSeqDx**



# TruSight™ Tumor 170 | TruSight™ Oncology 500

*Multi Biomarker Panels for Pan-Cancer Analysis*



CONTENT	TST170	TSO500
Genes	170	523
Size	0.5Mb	~2Mb
SNVs	✓	✓
InDels	✓	✓
CNVs	✓	✓ *
Fusions	✓	✓ ∞
Splice Variants	✓	✓ ∞
MSI	✓ +	✓
TMB	x	✓

\* CNVs calling available after launch  
+ Software solution for MSI calls coming soon  
∞ fusions and splice calls enabled by combining TSO500 DNA product with the RNA workflow from TST170



# AGENDA

General Overview – Oncology Strategy

**Prealanalytical Validation**

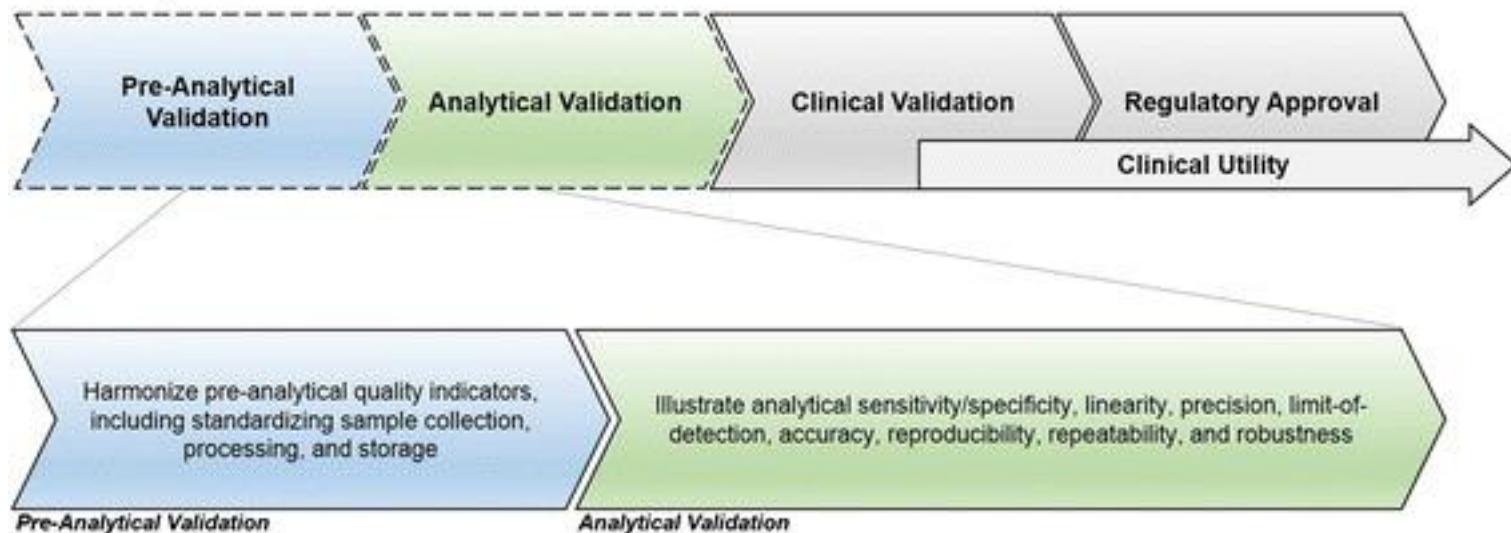
Analytical validation

Genomic biomarkers for Immunotherapy



# The biomarker development process

## *CDx Development*



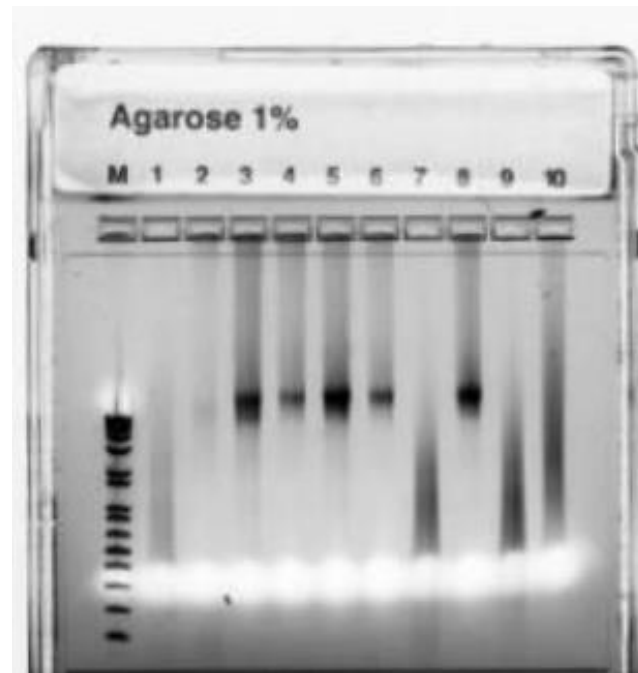
Masucci et al. (2016)

# FFPE DNA

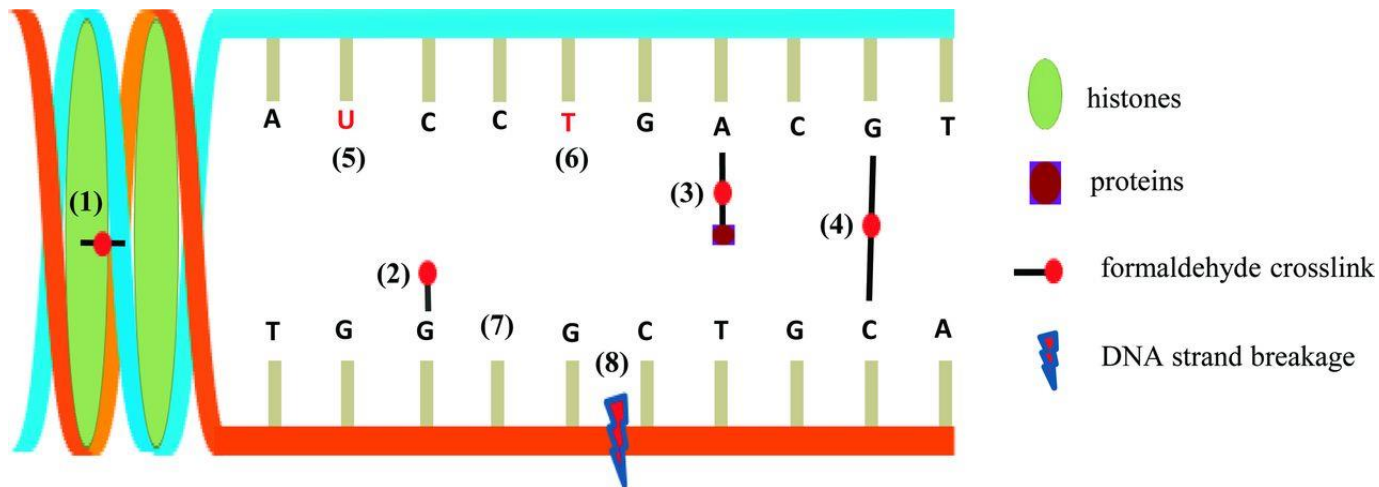
- **Pathological review to identify tumor-rich areas for microdissection**
  - Tumor purity information is important for interpretation of the results
  - Limit of detection at 5% variant allele frequency would require tumor purity >10%
- **DNA quantification**
  - Spectrophotometry (such as NanoDrop): Presence of nucleotides; could be single-stranded, highly degraded DNA
  - Fluorometric-based methods (such as Qubit or PicoGreen): Double-stranded DNA, however could be variable fragment size
  - Quantitative RT-PCR is preferable to quantify the amount of amplified templates
    - qPCR reaction determines the fragmentation status and amplification potential of FFPE samples. The status and potential are compared to ACD1 control DNA to calculate a  $\Delta Cq$  value for each sample
    - The higher the  $\Delta Cq$  of the samples, the lower the quality and higher the amount of input DNA required for sequencing

# FFPE Nucleic Acids

- **Highly variable within a set**
- **DNA damage**
  - Higher duplicate reads in FFPE libraries
  - Chimeric fragments are more common in FFPE samples
  - Smaller insert size in FFPE samples
  - GC bias is more severe in FFPE samples
  - **Noisier variant calling results**
  - Noisier CNV results



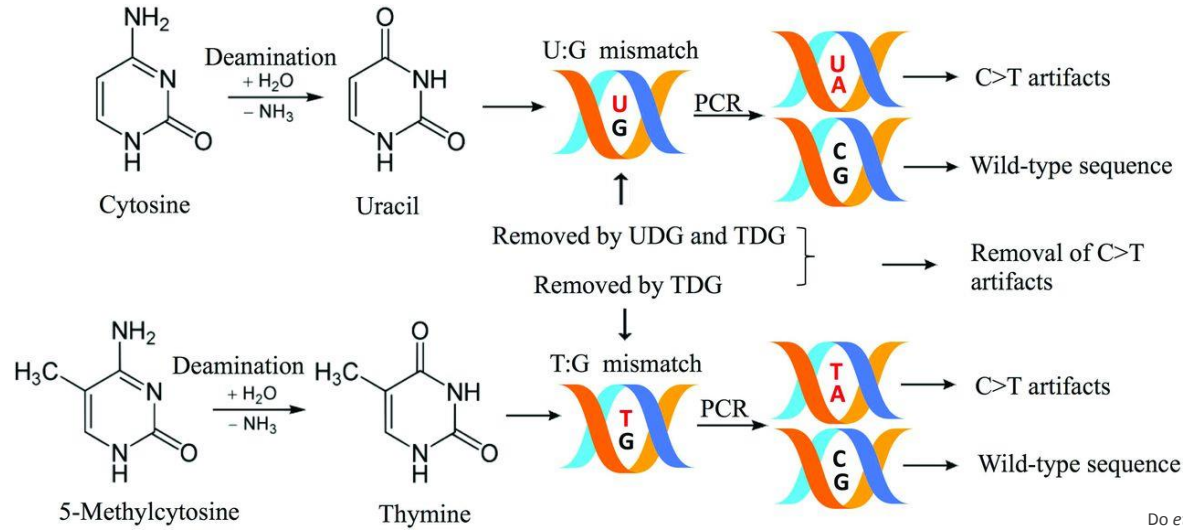
# Types of DNA damage in FFPE tissues



Do et. al. *Clinical Chemistry* (2014)

# Deamination of Cytosine bases

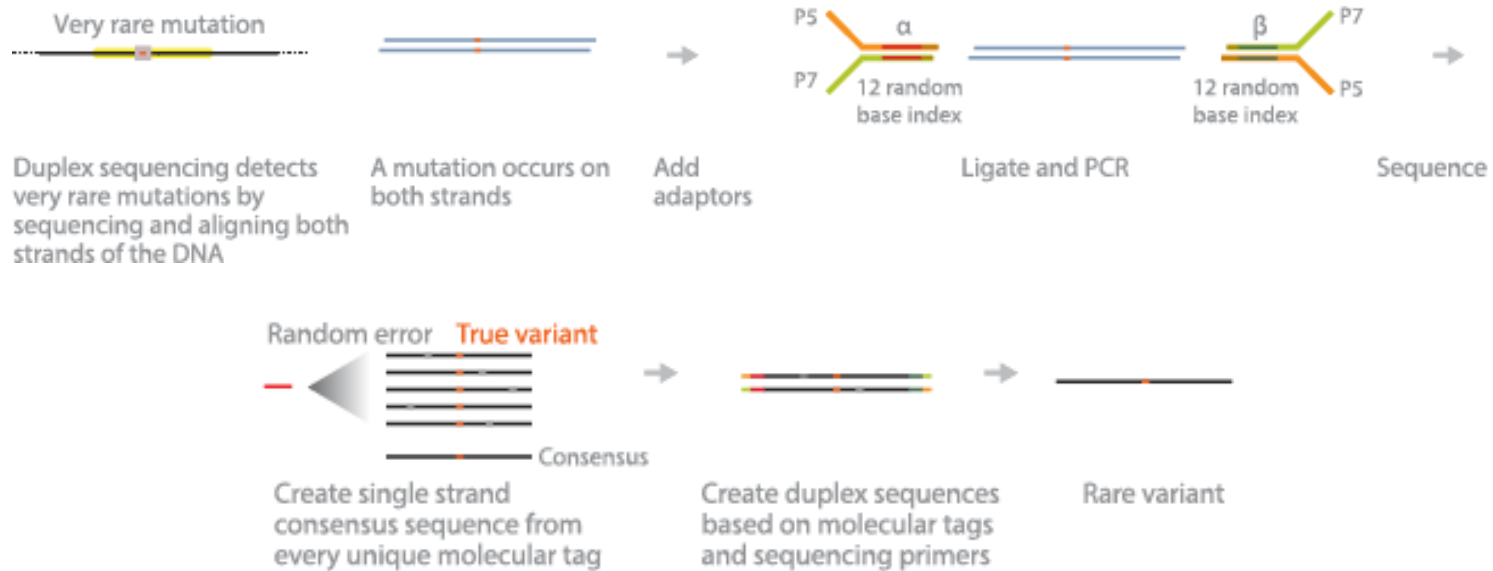
- Hydrolytic deamination of cytosine bases: C → U
- Result in artificial C:G > T:A SNVs
- Prevalent in CpG dinucleotide



Do et. al. Clinical Chemistry (2017)



# Duplex sequencing



## Bioinformatics:

Collapsing reads based on duplex sequences

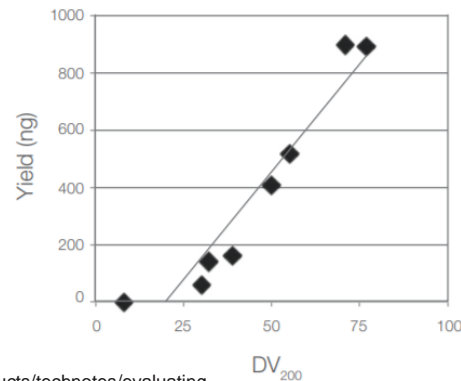
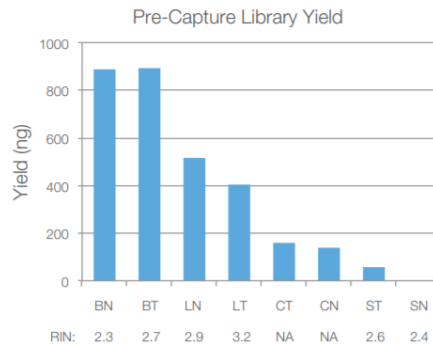
Build error model based on duplex information

<https://www.illumina.com/science/sequencing-method-explorer/kits-and-arrays/duplex-sequencing.html>

# FFPE RNA

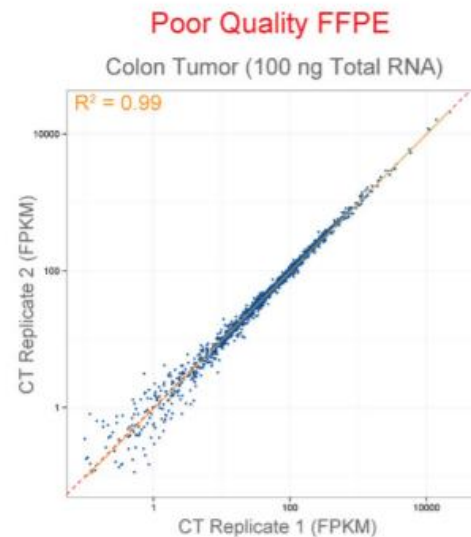
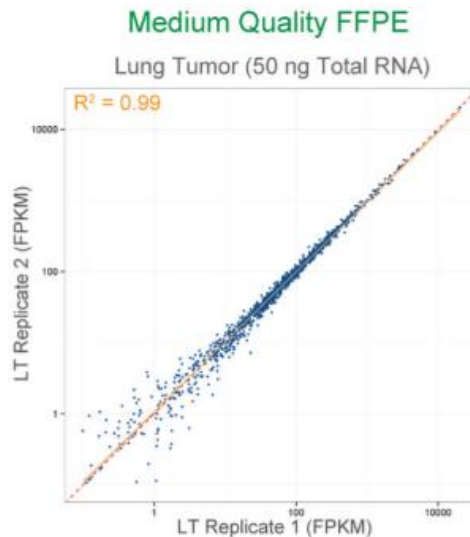
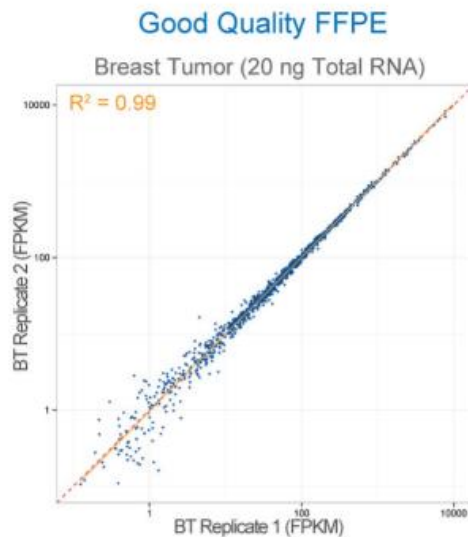
- **Traditional quality assessment: RNA Integrity Number (RIN) from RNA electropherogram traces (e.g. Bioanalyzer traces)**
- **DV200 metric**
  - Measure percentage of RNA >200 nucleotides
  - DV200 >30% ensures degraded RNA fragment meet the requirement for efficient target capture
- **DV200 can be calculated from a Bioanalyzer trace by performing a Smear Analysis**

Sample	RIN	DV <sub>200</sub> *
Breast Normal	2.3	77
Breast Tumor	2.7	71
Lung Normal	2.9	55
Lung Tumor	3.2	50
Colon Normal	N/A	32
Colon Tumor	N/A	39
Stomach Tumor	2.4	30
Stomach Normal	2.6	8



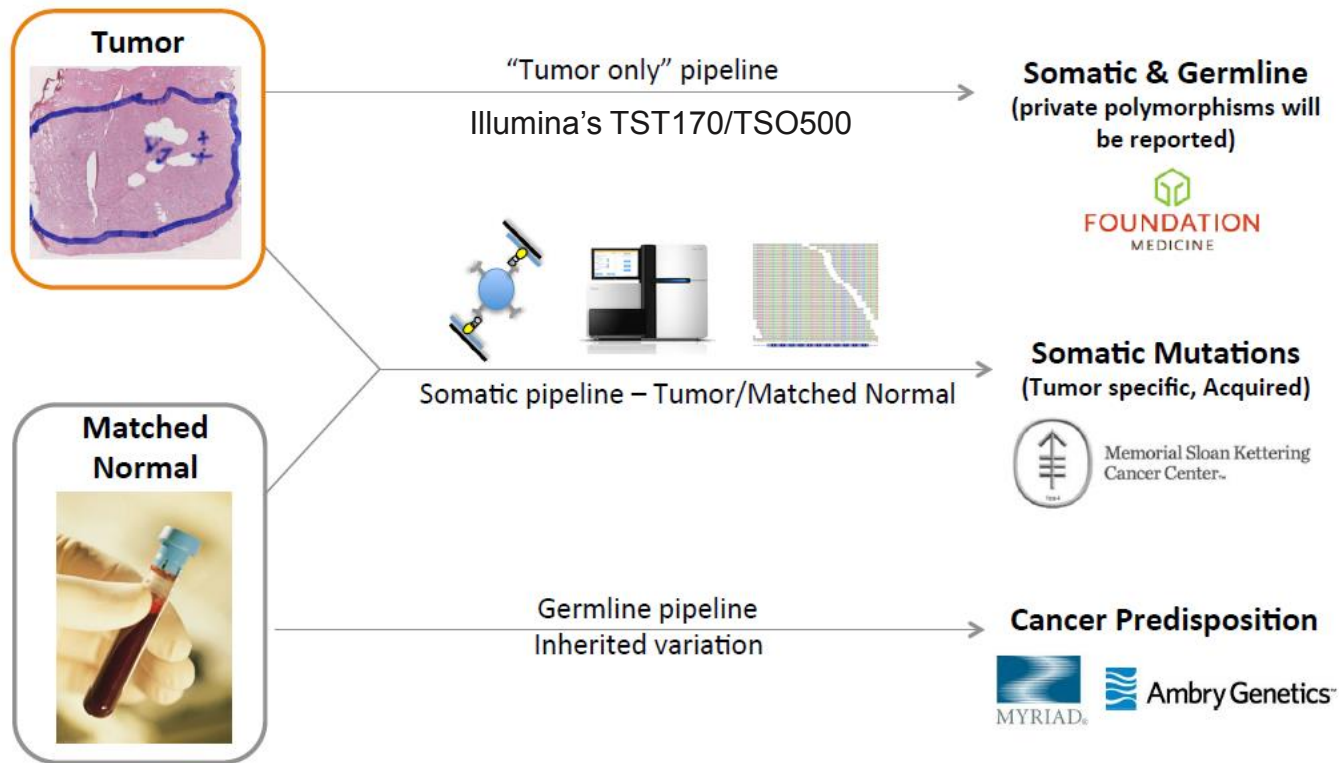
<https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/evaluating-rna-quality-from-ffpe-samples-technical-note-470-2014-001.pdf>

# High concordance of gene expression between replicates across FFPE samples of varied quality



<https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/evaluating-rna-quality-from-ffpe-samples-technical-note-470-2014-001.pdf>

# Current NGS Oncology Market



# Tumor/normal vs. tumor only

- **Tumor/normal**

- To distinguish somatic mutations from germline variants
- To capture false positive calls caused by poorly behaved genomic regions
- To confirm germline variants that are associated with tumor development (BRCA genes, MMR genes etc)

- **Tumor-only**

- Matched normal specimens are not routinely collected in clinical practice
- Tumor/normal analysis has significant cost increase even if normal is sequenced at a lower depth

Practical consideration  
in patient management

# Build robust pipeline for tumor only workflow

- **Tumor/normal**

- To distinguish somatic mutations from germline variants
- To capture false positive calls caused by poorly behaved genomic regions
- To confirm germline variants that are associated with tumor development (BRCA genes, MMR genes etc)

- **Tumor-only**

- Removal of common germline variants documented in public databases, e.g. dbSNP, 1000G, ExAC and gnomAD

Tumor-only analyses			
	Candidate mutations per tumor	Candidate mutations per tumor after removal of common germline variants (dbSNP)	Alterations in actionable genes after dbSNP filter
Total alterations	1401	382	2.41
True positives (somatic)	135 (10%)	133 (35%)	1.61 (67%)
False positives (germline)	1266 (90%)	249 (65%)	0.80 (33%)

WES analysis in Jones et al. *Science TM*, 2015

- More population sequencing datasets will help or one-time germline WGS at birth or healthy

Cohort  
Normal

Baseline  
Normal



# Build robust pipeline for tumor only workflow

- **Tumor/normal**

- To distinguish somatic mutations from germline variants
- To capture false positive calls caused by poorly behaved genomic regions
  - If a matched normal is from blood or saliva, it could not capture the systematic noise from sample preservation and nucleic acid extraction
  - Adjacent non-tumor tissues might be hard to collect in clinical practice
- To confirm germline variants that are associated with tumor development (BRCA genes, MMR genes etc)

- **Tumor-only**

- Reference non-tumor tissues undergone the same process would capture majority if not all systematic bias
- Use the systematic biases in reference samples as background filter

Pipeline  
Normal

# Build robust pipeline for tumor only workflow

- **Tumor/normal**

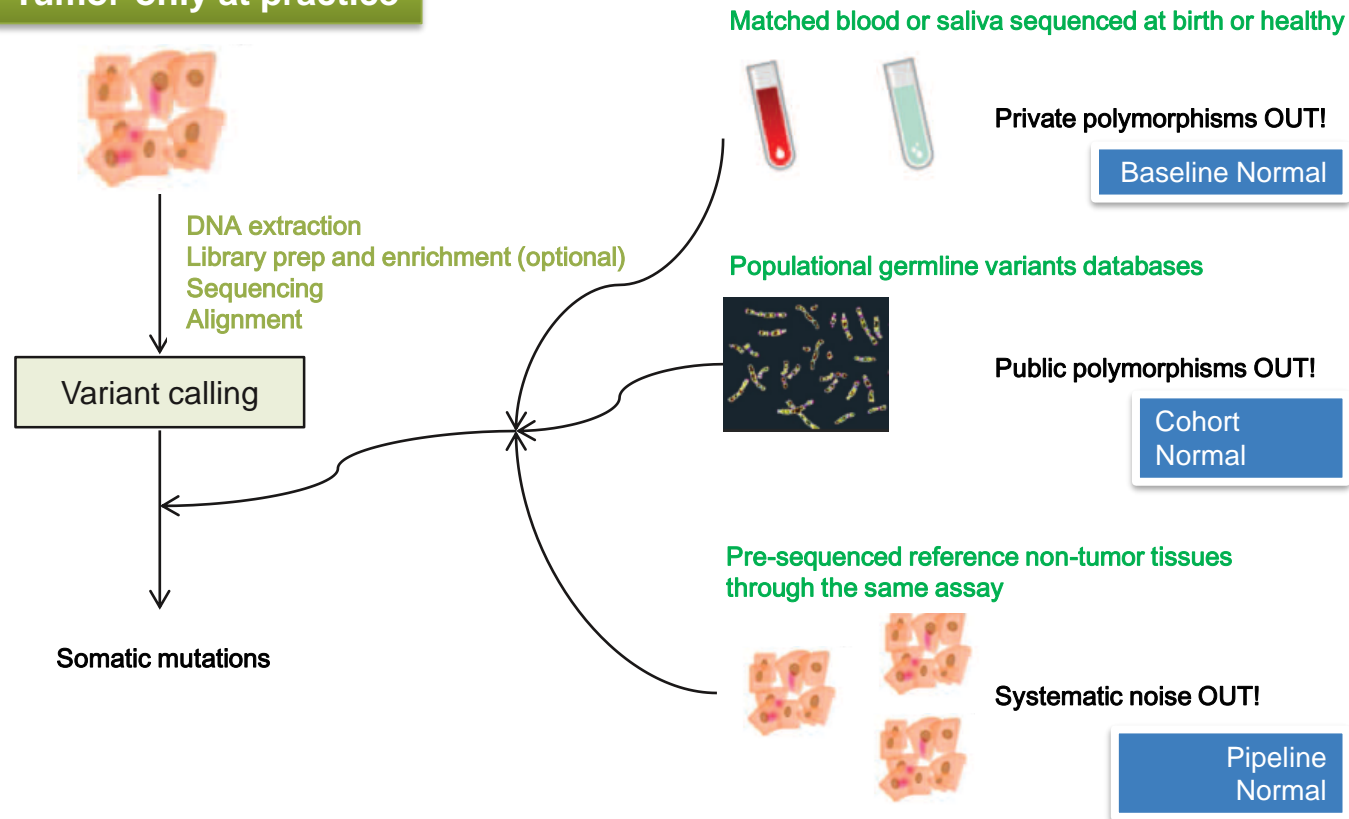
- To distinguish somatic mutations from germline variants
- To capture false positive calls caused by poorly behaved genomic regions
- To confirm germline variants that are associated with tumor development (BRCA genes, MMR genes etc)

- **Tumor-only**

- Serves as a screening test for pathogenic germline variants
- Once suspect germline variants are identified in the tumor sample, then confirm with germline sequencing

# Recommendation for tumor-only workflow

## Tumor-only at practice



# AGENDA

General Overview – Oncology Strategy

Preanalytical validation

**Analytical Validation**

Genomic biomarkers for Immunotherapy



# FDA's approach to regulatory oversight for NGS in vitro diagnostic (IVD) tests

The screenshot shows the FDA's official website with a dark blue header. The header includes the FDA logo, the text 'U.S. FOOD & DRUG ADMINISTRATION', and a search bar. Below the header is a navigation bar with links to Home, Food, Drugs, Medical Devices, Radiation-Emitting Products, Vaccines, Blood & Biologics, Animal & Veterinary, Cosmetics, and Tobacco Products. The main content area is titled 'News & Events' and features a breadcrumb trail: Home > News & Events > Newsroom > Press Announcements. The headline of the news release is 'FDA finalizes guidances to accelerate the development of reliable, beneficial next generation sequencing-based tests'. Below the headline are social media sharing buttons for Facebook, Twitter, LinkedIn, Pinterest, Email, and Print. The release is dated 'April 12, 2018' and is marked as 'For Immediate Release'. A 'Summary' section begins with the text: 'The agency is leveraging new tools and policies to advance the creation of innovative genetic and genomic-based tests and help ensure the validity of their results'. On the right side of the page, there are sections for 'Inquiries' (Media and Consumers), 'Related Information' (linking to 'FDA: Use of Public Human Genetic Variant Databases to Support Clinical Validity for Genetic and Genomic-Based In'), and a 'Search FDA' bar at the top right.

**FDA News Release**

## FDA finalizes guidances to accelerate the development of reliable, beneficial next generation sequencing-based tests

SHARE | TWEET | LINKEDIN | PIN IT | EMAIL | PRINT

**For Immediate Release**      April 12, 2018

**Summary**

The agency is leveraging new tools and policies to advance the creation of innovative genetic and genomic-based tests and help ensure the validity of their results

**Inquiries**

**Media**

✉ Tara Rabin  
☎ 240-402-3157

**Consumers**

☎ 888-INFO-FDA

**Related Information**

- FDA: Use of Public Human Genetic Variant Databases to Support Clinical Validity for Genetic and Genomic-Based In

“...assure the safety and effectiveness (which, for IVDs, generally means a reasonable assurance of analytical and clinical validity) of these test...”

# Accuracy

- **Positive percent agreement (PPA) :  $TP/(TP+FN)$** 
  - Ability to correctly identify variants
- **Negative percent agreement (NPA) :  $TN/(TN+FP)$** 
  - Ability to correctly identify wt bases
- **Technical positive predictive value (TPPV) :  $TP/(TP+FP)$** 
  - Particular import if variant confirmation will not be performed
- **Minimum acceptable overall and target thresholds for PPA, NPA, and TPPV point estimate and lower bound of 95% confidence interval (CI) of an NGS-based test should be predefined and reported for each variant type claimed by the test**



# No calls in the NGS assay

- **No calls** or **invalid calls** describe the result where a base call is not made, which can be a result of a number of factors
  - the base level performance not meeting predefined thresholds for quality, resulting in insufficient data to make a variant call.
- **No calls and invalid calls should not be used in PPA, NPA, or TPPV calculations**, but should be individually documented as part of the accuracy study results.
- **Minimum acceptable values for the number of no calls or invalid calls will depend on indications for use and test design.**

# Accuracy calculation

		Comparator Method		Total
		Positive	Negative	
Test	Positive	A	B	A+B
	Negative	C	D	C+D
	No calls or invalid calls	E	F	E+F
	Total	A+C+E	B+D+F	N

- Percent of no calls or invalid calls :  $(E+F)/N$  along with a 95% two-sided confidence interval
- PPA :  $A/(A+C)$
- NPA:  $D/(D+B)$
- TPPV:  $A/(A+B)$

# Precision (Reproducibility and Repeatability)

- **Reproducibility: NGS tests involves measuring test variability under a variety of specified condition**
  - Samples
  - Runs
  - Reagent lots
  - Days
  - Operators
- **Repeatability: measuring test result variability when using the same operators, the same measuring system, the same operating conditions and the same location, and replicating measurements on the same or similar objects over a short period of time**
- **The percentage of no calls or invalid replicates should be also reported**

# Limit of Detection (LoD)

- Establish and document the **input DNA range**, that will enable the test to provide expected results in 95% of test runs
- Establish and document that the **level of invalid calls** or **no call** results is acceptable
- Establish and document the lower LoD for each **variant type**, and **in different sequence context** included in the test's indication for use
- The lower LoD is calculated as the lowest concentration of analyte at which **at least 95% of positive calls**
- An upper LoD should be established and documented

# Analytical Specificity

**Analytical specificity relates to the ability of a test to measure solely the intended analyte.**

- **Interference**

- Identify and document any interfering substances that might reduce the ability to amplify or sequence

- **Cross-reactivity**

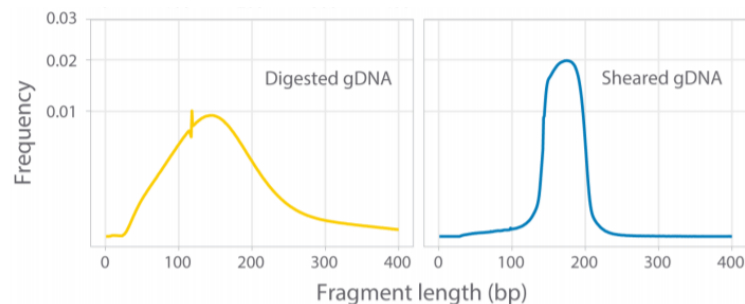
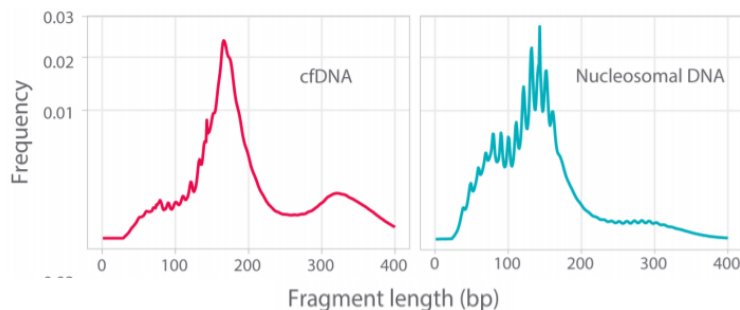
- Access and document the potential for cross-reactivity of known cross-reactive alleles and homologous regions (e.g., pseudogenes), based on the targets that will be interrogated by the test

- **Cross-contamination**

- Develop, validate and document methods to detect carryover or cross-contamination between patient specimens or samples

# Reference materials

- **Synthetic mimic libraries, cell lines, synthetic vectors, company or WHO and NIST reference**
  - NA12878 from NIST: Genome in a Bottle (GIAB) Consortium
  - FFPE tissue-based controls are available from companies
    - Horizon Dx, Acrometrix (Thermo Fisher, Inc), Seracare Life Science
- **Other concerns for reference materials**





# AGENDA

General Overview – Oncology Strategy

Preamanalytical validation

Analytical Validation

Genomic biomarkers for immunotherapy



# Companion Diagnostic Development

*Partnering to Power Oncology Precision Medicine*

Companion Diagnostic product (under development)  
based on TruSight™ Tumor 170 content



**Loxo to expand oncology menu  
for NextSeqDx**



Companion Diagnostic product (under development)  
based on TruSight™ Oncology 500 content



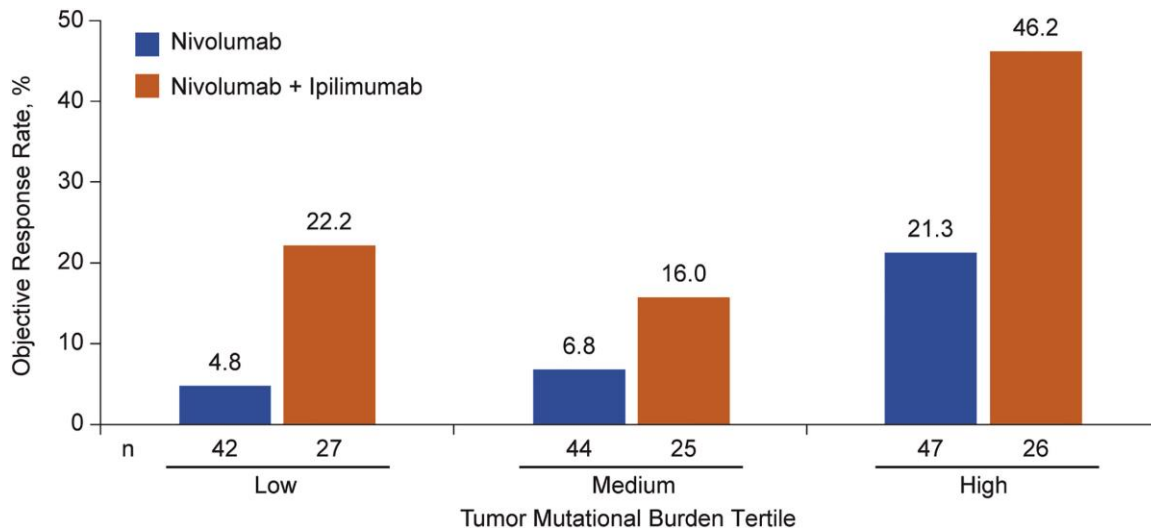
**Bristol-Myers Squibb to expand  
oncology menu for NextSeqDx**

# Cancer Immunotherapy Biomarker Advancement

*Tumor Mutation Burden correlation to therapy outcomes*

Not every tumor responds well to IO therapy. High Tumor Mutation Burden has been linked to better responses

Results from Clinical Trials demonstrate the link between High Tumor Mutation Burden and better response to immunotherapies.



**TUMOR MUTATION BURDEN (TMB):** The measurement of mutations carried by tumour cells. A new predictive marker being used in response to IO therapies

# TruSight™ Tumor 170 | TruSight™ Oncology 500

*Multi Biomarker Panels for Pan-Cancer Analysis*



CONTENT	TST170	TSO500
Genes	170	523
Size	0.5Mb	~2Mb
SNVs	✓	✓
InDels	✓	✓
CNVs	✓	✓ *
Fusions	✓	✓ ∞
Splice Variants	✓	✓ ∞
MSI	✓ +	✓
TMB	x	✓

\* CNVs calling available after launch  
+ Software solution for MSI calls coming soon  
∞ fusions and splice calls enabled by combining TSO500 DNA product with the RNA workflow from TST170

# TruSight™ Oncology 500 Content

*Pan-cancer DNA + RNA\* analysis: Consolidate multiple tests into one*

## Single Assay – 523 Genes

Detection of currently relevant DNA & RNA variants for multiple types of cancer. Some key biomarkers included:



Lung	Melanoma	Colon	Ovary	Breast	Gastric	Bladder	Myeloid	Pan-Cancer
AKT1	BRAF	AKT1	BRAF	AKT1	BRAF	MSH6	ALBL1	MSI
ALK	CTNNB1	BRAF	BRCA1	AR	KIT	PMS2	ASXL1	NTRK
BRAF	GNA11	HRAS	BRCA2	BRCA1	KRAS	TSC1	CALR	
DDR2	GNAQ	KRAS	KRAS	BRCA2	MET		CEBPA	
EGFR	KIT	MET	PDGFRA	ERBB2	MLH1		ETV6	
ERBB2	MAP2K1	MLH1	FOXL2	FGFR1	PDGFRA		EZH2	
FGFR1	NF1	MSH2	TP53	FGFR2	TP53		FLT3	
FGFR3	NRAS	MSH6		PIK3CA			GATA2	
KRAS	PDGFRA	NRAS		PTEN			IDH1	
MAP2K1	PIK3CA	PIK3CA					IDH2	
MET	PTEN	PMS2					JAK2	
NRAS	TP53	PTEN					KIT	
PIK3CA		SMAD4					MPL	
PTEN		TP53					NPM1	
RET							RUNX1	
TP53							SF3B1	
TMB							SRSF2	
							TP53	

\* The recommended product to evaluate DNA + RNA variants is the TSO500 DNA+RNA bundle (PN: 20028216) – commercially available in January -2019

**For Research Use Only. Not for use in diagnostic procedures.**

**illumina**

# TruSight™ Oncology 500

*Premium tumor profiling solutions; Include all genes covered by FoundationOne and MSK-IMPACT*

## Premium Content

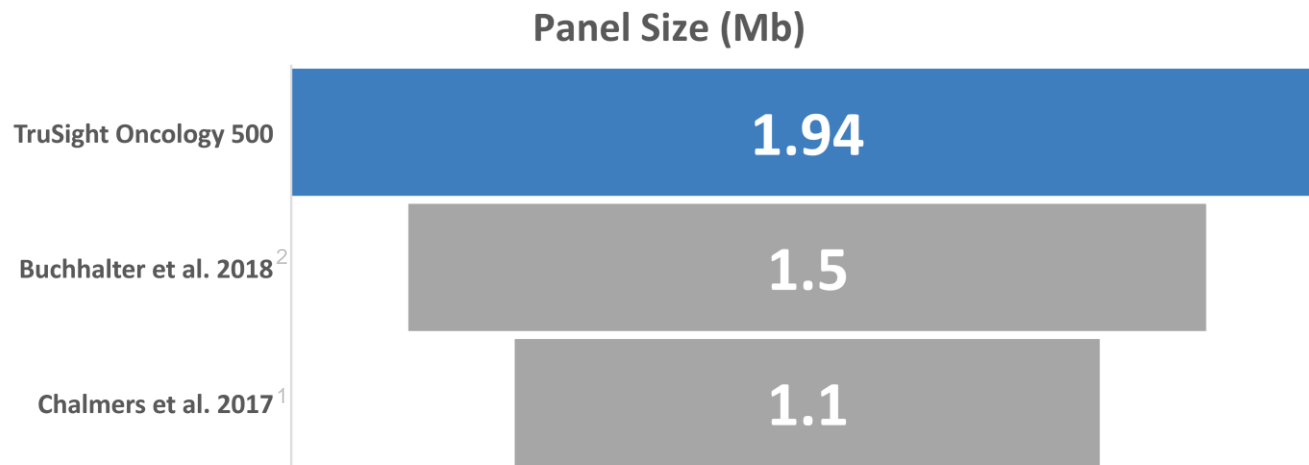
- Our assays are **comprehensive**, with **high coverage of relevant biomarkers**, such as the ones present in **NCCN Guidelines** and **Clinical Trials**.
- Enrichment enables **better detection of novel** fusions and splice variants.
- **TSO500 can detect both TMB+MSI**. Detecting both markers is key in immuno-oncology

	TSO500 DNA+RNA
Total # of Genes	523
SNVs, InDels	523
Amplifications	59
Fusions	55
Splice Variants	3
TMB	YES
MSI	YES
Presence in Guidelines	42
Presence in Clinical Trials	>1,600

# TruSight™ Oncology 500 | Highly Accurate TMB Scoring

*Panel size of 1.94 Mb surpasses minimum requirements*

Recent Evidence shows that TMB can accurately calculated by targeting  $\geq 1.5$  Mb of coding genome.<sup>2</sup>  
Larger Panels correlate with greater sensitivity and have increased correlation with WES.<sup>1,2</sup>



1. Chalmers et al, *Genome Medicine* 2017 9:34;

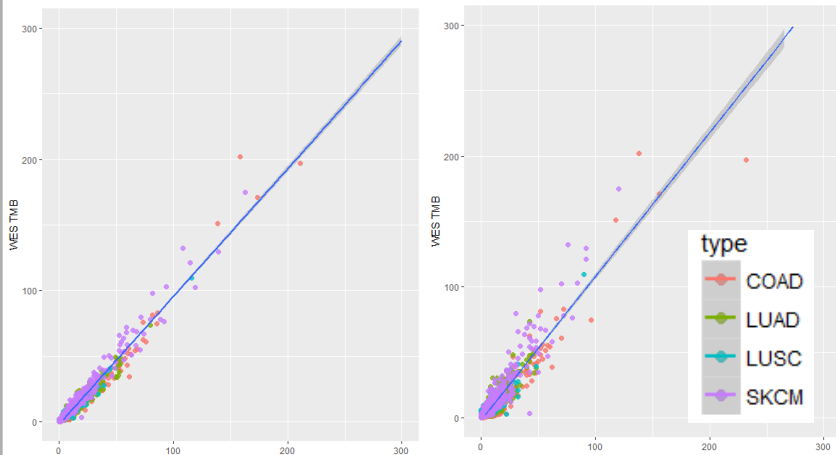
2. Buchhalter et al., *Int J Cancer*. 2018 Sep 21. doi: 10.1002/ijc.31878.

**For Research Use Only. Not for use in diagnostic procedures.**

# TruSight™ Tumor 170 | TruSight™ Oncology 500

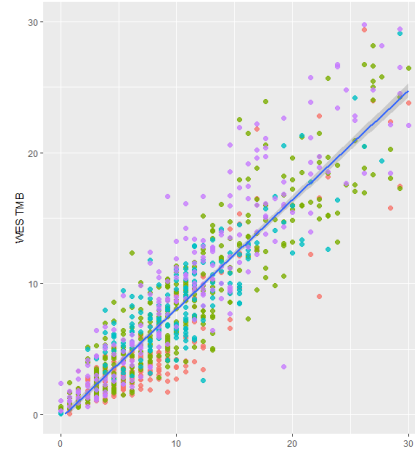
*When it comes to TMB detection, size matters*

WES vs. TSO500  $R^2 = 0.97$     WES vs. TST170  $R^2 = 0.89$

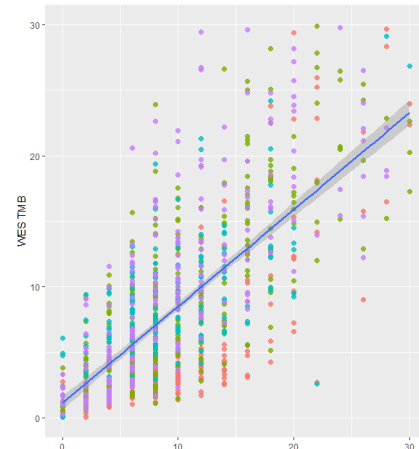


All samples

WES vs. TSO500  $R^2 = 0.84$



WES vs. TSO170  $R^2 = 0.51$

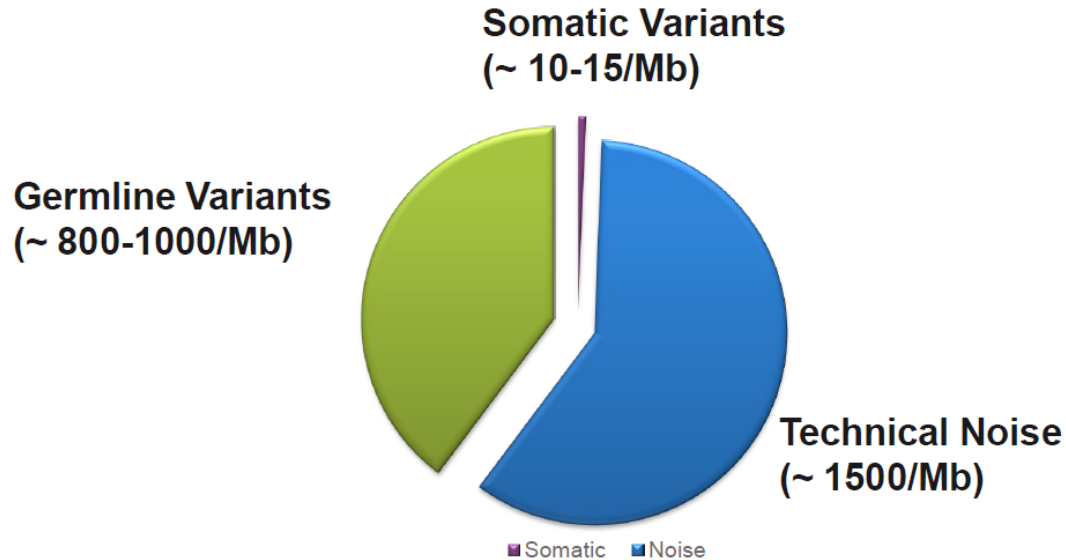


Samples in TMB  
range between 0-30

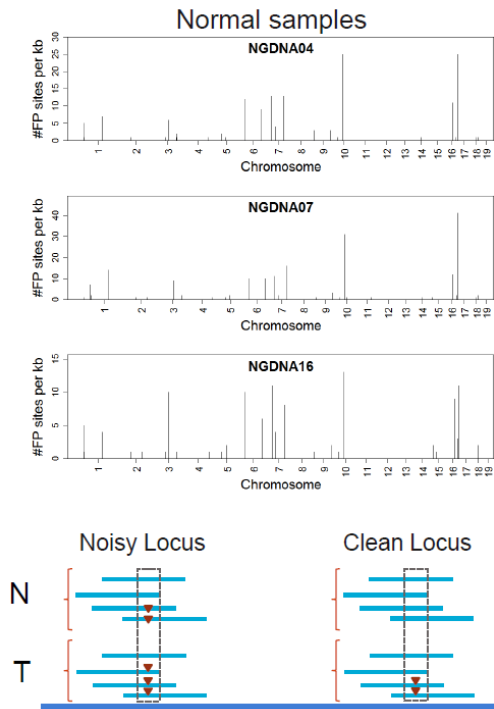


# TMB in tumor only variant calling

- ▶ TMB in tumor only setting is a “needle in a haystack” problem
- ▶ Main challenge is to effectively remove technical noise (FFPE artifacts) and germline variants



# False positive removal with normal baseline



- ▶ Systematic noisy regions exist across samples
- ▶ These usually represent low complexity regions in the genome
- ▶ A cohort of normal FFPE samples can capture such recurrent error, leading to improved specificity
- ▶ This approach removes 80-90% of FPs in a given sample
- ▶ However, sample specific noise can not be removed by normal baseline

# Deamination contributes heavily to FPs

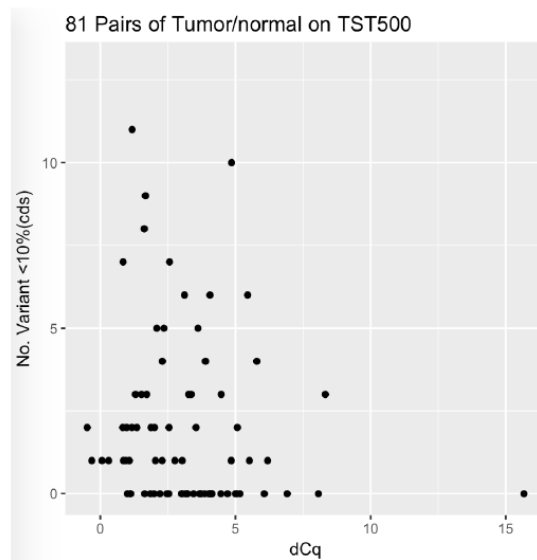
	No. FP (<10%)													
NA21781.genome.vcf	0	0	0	0	0	0	1	0	0	0	0	0	2	0
NA21731.genome.vcf	0	0	0	0	0	0	0	1	1	1	0	0	3	0
NA21730.genome.vcf	0	2	0	0	1	0	1	0	1	0	0	2	1	0
NA21687.genome.vcf	0	0	0	3	0	0	0	0	0	0	0	0	8	0
NA21677.genome.vcf	0	0	0	0	0	1	0	0	1	0	1	1	1	0
NA21660.genome.vcf	0	0	0	0	0	0	0	0	0	0	0	1	0	0
NA21070.genome.vcf	0	0	0	0	1	0	1	0	0	0	0	0	0	0
NA20431.genome.vcf	1	0	0	0	1	0	0	0	0	0	0	3	0	0
2135.genome.vcf	1	10	1	1	1	19	23	0	1	5	4	4	8	1
2127.genome.vcf	2	11	2	4	2	14	16	1	3	4	11	2	6	1
2112.genome.vcf	2	23	5	3	7	18	16	5	1	1	14	7	14	1
2111.genome.vcf	4	18	4	4	3	11	14	3	2	3	15	1	5	2
1955.genome.vcf	5	35	3	10	8	32	37	4	3	8	29	6	9	2
1954.genome.vcf	12	44	6	11	8	41	38	3	4	9	36	6	14	2
1945.genome.vcf	2	29	5	3	6	23	21	5	1	1	34	4	8	3
1944.genome.vcf	7	18	3	1	6	22	18	2	0	3	23	3	5	0
	A_C	A_G	A_T	C_A	C_G	C_T	G_A	G_C	G_T	T_A	T_C	T_G	Indel	MNV

High quality samples

Low quality samples

- After removing site specific errors, No. FP is ~5-20/Mb in High quality sample and 200-400/Mb in Low quality sample
- No. FP is associated with sample quality and strong deamination signature leads to **C:G->T:A** (deamination of C) or **A:T->G:C** (deamination of A) variants

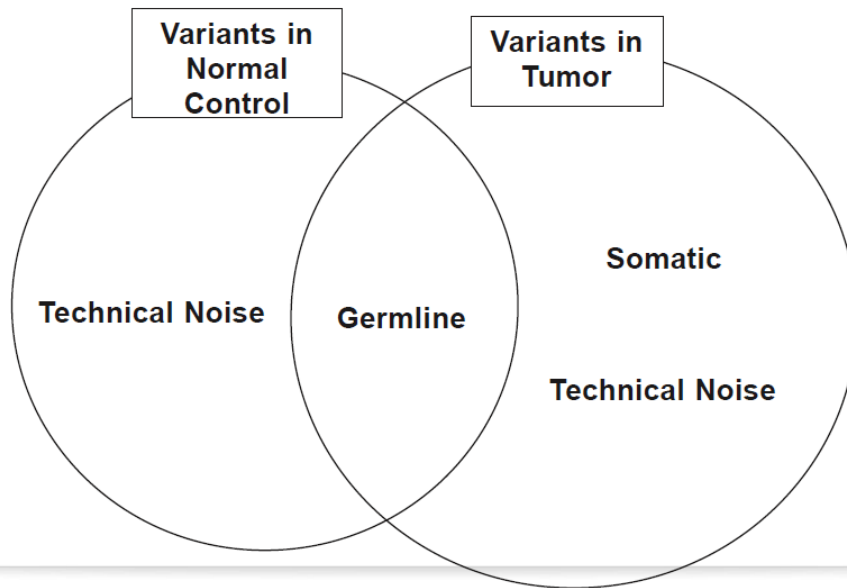
# Specificity evaluation on FFPE



- All samples pass 99.999% specificity
- Specificity is not correlated with sample quality (dCq):
  - Very important not to bias TMB measurement

# Germline filtering – the challenge with tumor only workflow

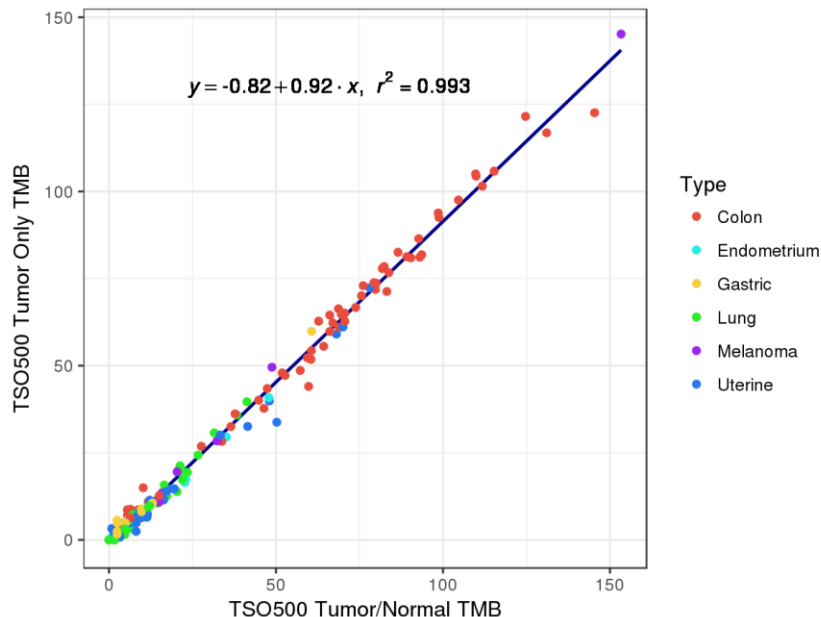
- ▶ If both tumor and normal samples are available, we can use the variant calling results from the normal sample to filter out germline variants.
- ▶ There are ~800/Mb germline variants in each individual, some are commonly observed in population, some are rare.



# TruSight™ Oncology 500 | Highly Accurate TMB Scoring

*Combining panel size with proprietary pipeline for high precision*

## Tumor-Only TMB Scoring



- 170 FFPE Samples tested: colon, endometrium, gastric, lung, Melanoma, Uterine
- Tumor/Normal compared to tumor-only
- Illumina's germline filtering algorithm is key to enable tumor-only analysis
- 99.3% correlation ( $R^2$ )

Illumina data on file, 2018

**For Research Use Only. Not for use in diagnostic procedures.**

**illumina®**

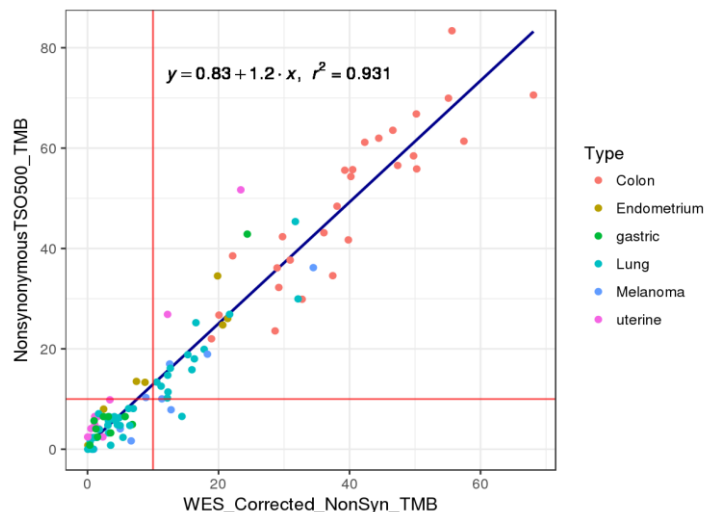
# TruSight™ Oncology 500 | Highly Accurate TMB Scoring

*TMB Detection pipeline maximizes sensitivity, specificity and accuracy*

Bioinformatic Attribute	TSO500
Includes non-syn + syn SNVs	✓
Includes InDels	✓
5% LOD	✓
Enrichment Chemistry	✓
FFPE and Seq Error Artifact Removal	✓
500 – 1,000x false positive reduction	✓
Removal of variants in low confidence regions	✓
Removal of driver mutations	✓
Germline filtering using variant databases, allele frequency and machine learning	✓

# TruSight™ Oncology 500 | Highly Accurate TMB Scoring

*Combining panel size with pipeline for high precision*



Tumor/Normal whole exome sequencing was compared at Illumina to the tumor only TSO 500 workflow, across 108 FFPE tissue samples. TMB results showed high correlations between the two assays.

## Performance at TMB threshold of 10mut/Mb

	10 mutation/Mb cutoff
Positive % Agreement	94.7%
Negative % Agreement	96.1%
Overall % Agreement	95.4%

	WES T/N TMB High	WES T/N TMB Low
TSO500 TMB High	54	3
TSO500 TMB Low	2	49

Illumina data on file, 2018

For Research Use Only. Not for use in diagnostic procedures.

illumina®

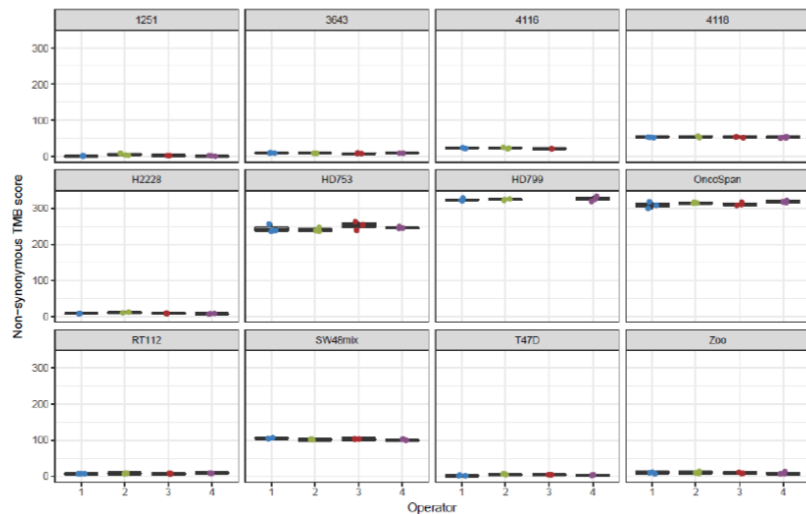


# Reproducibility: TMB

## Reproducibility TMB score

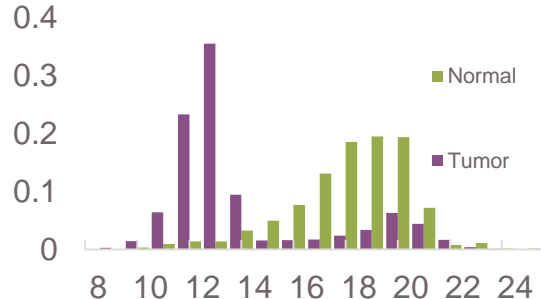
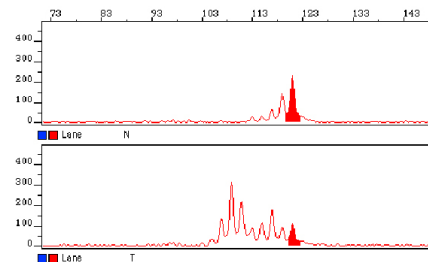
- FFPE and cell line DNA
- 12 replicates tested across 4 operators to determine TMB score

### Robust TMB scoring across operators



# Challenges for MSI calling for tumor only workflow

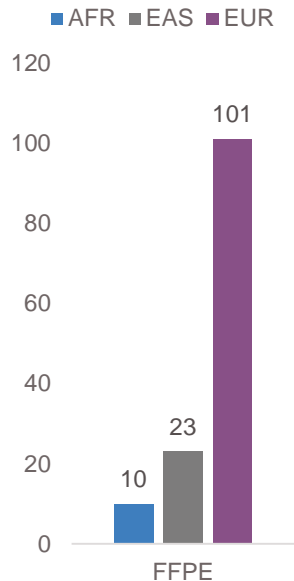
- **Distribution of multiple alleles instead of a single allele in repetitive regions**
  - Probe synthesis
  - PCR (polymerase slippage)
  - Enrichment bias (AT drop, TC drop)
  - Sequencing error
  - Mapping and variant calling
- **Lack of prior knowledge**
  - Need to rely on noncanonical sites
- **Tumor only workflow**
  - Genetic diversity for normal samples
  - Ethnicity specific sites
  - Tumor samples contain mixed tumor and normal cells
- **A robust metric to capture the distribution shift for each MSI site**



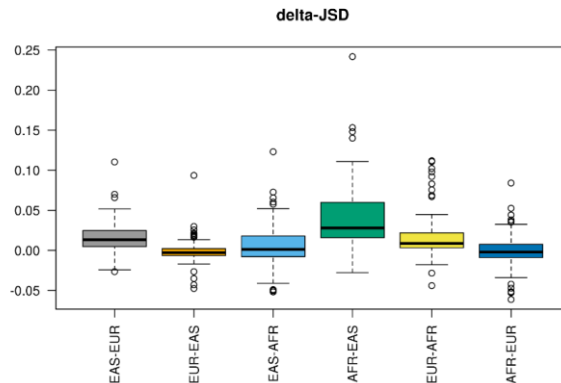
# Enable Tumor Only Testing with baseline samples

- **Use 61 normal samples for baseline distribution of same MSI sites**
- **Jensen-Shannon Distance between baseline samples**
  - Pairwise calculation for any two BL samples
    - discrete distribution of relative alleles :  $BL1 = \Pr [X=x]$  ;  $BL2 = \Pr [X=x]$
    - $JS <- 0.5 * (\text{sum}(BL1 * \log(BL1 / m)) + \text{sum}(BL2 * \log(BL2 / m)))$ ;  $m <- 0.5 * (BL1 + BL2)$
    - $d1 <- \text{sqrt}(JS)$
- **Jensen-Shannon Distance between the test sample to Baseline samples**
  - Pairwise calculation for one BL sample and the test sample
    - $BL1 = \Pr [X=x]$  ;  $T = \Pr [X=x]$
    - $JS <- 0.5 * (\text{sum}(BL1 * \log(BL1 / m)) + \text{sum}(T * \log(T / m)))$ ;  $m <- 0.5 * (BL1 + T)$
    - $d2 <- \text{sqrt}(JS)$
- **Comparison between two Jensen-Shannon Distance distributions**
  - One sided t-test:  $d1 < d2$ ,  $p < 0.01$
  - $d2 - d1 \geq 0.1$
- **Final score:**
  - Number of unstable sites / number of total sites tested
  - Similar to MSI-PCR

# Identify MSI sites with ethnicity bias

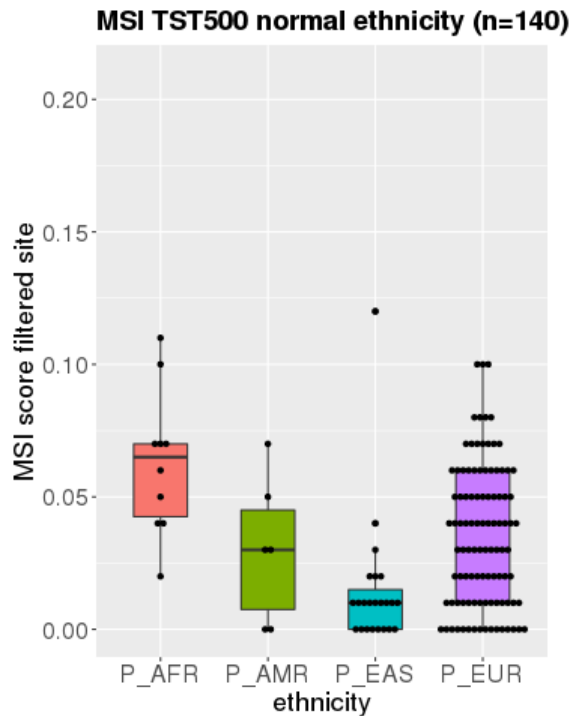
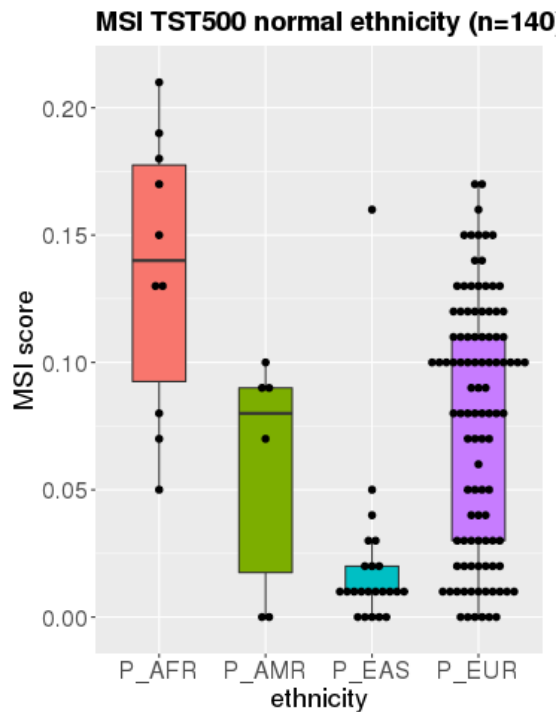


$$\Delta JSD = \text{avg}(JSD_{\text{between}}) - \text{avg}(JSD_{\text{within}})$$

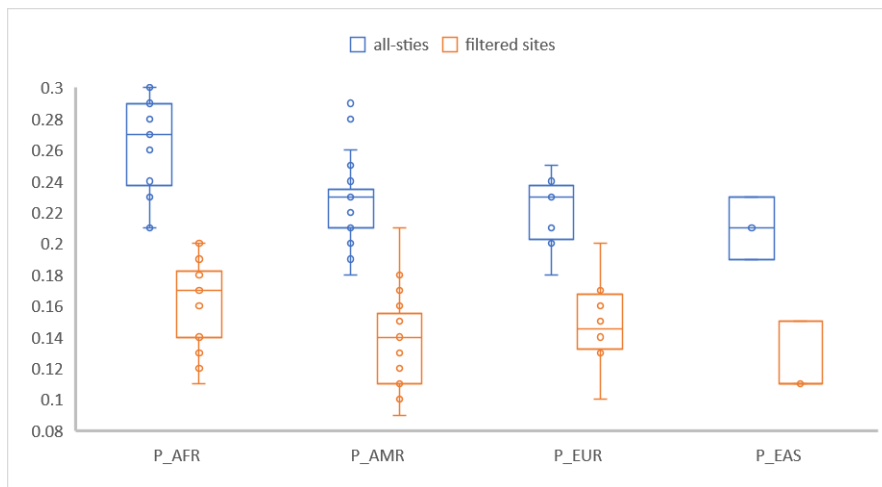


- 175 sites with at least 20 supporting reads for minimum 5 samples of each ethnicity group
- Identified 45 sites with  $\geq 0.1$   $\Delta JSD$  based on pair-wised comparison between three ethnicity groups

# Normal sample MSI-score with



# Use cell line sample as a validation cohort



- For filtered sites MSI score, the highest value is sample IHW09101 (MSI score: 0.21)
- This is annotated with Warao SA Indian ethnicity

- 10 IHW and 48 coriell
- P\_AFR (n=22)
- P\_AMR (n=25)
- P\_EUR (n=8)
- P\_EAS (n=3)

## Warao



A Warao family in their canoe.

### Total population

(20,000)

### Regions with significant populations

Venezuela, Guyana, Suriname, Trinidad and Tobago

### Languages

Warao language

### Religion

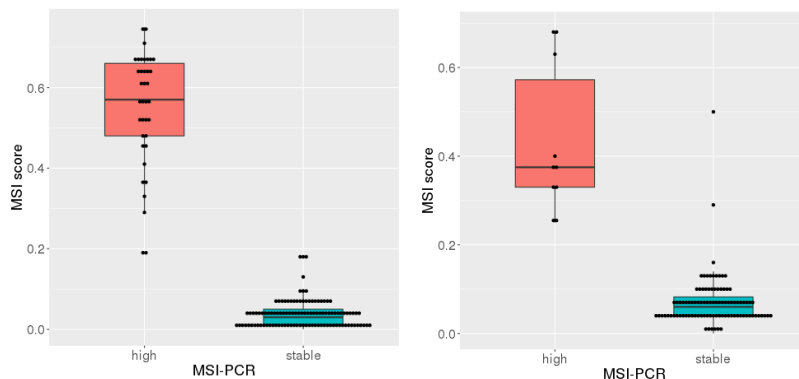
Traditional beliefs, Christianity



# TruSight™ Oncology 500 | Highly Accurate MSI Detection

*Tumor-only MSI workflow, Precise MSI detection*

## Tumor-Only MSI Detection



	Training set	Validation set
PPA	97.3% (85.8%-99.9%)	100% (69.2%-100.0%)
NPA	100% (96.7%-100.0%)	98.0% (92.7%-99.8%)
OPA	99.3% (96.3%-100.0%)	98.1% (93.4%-99.8%)

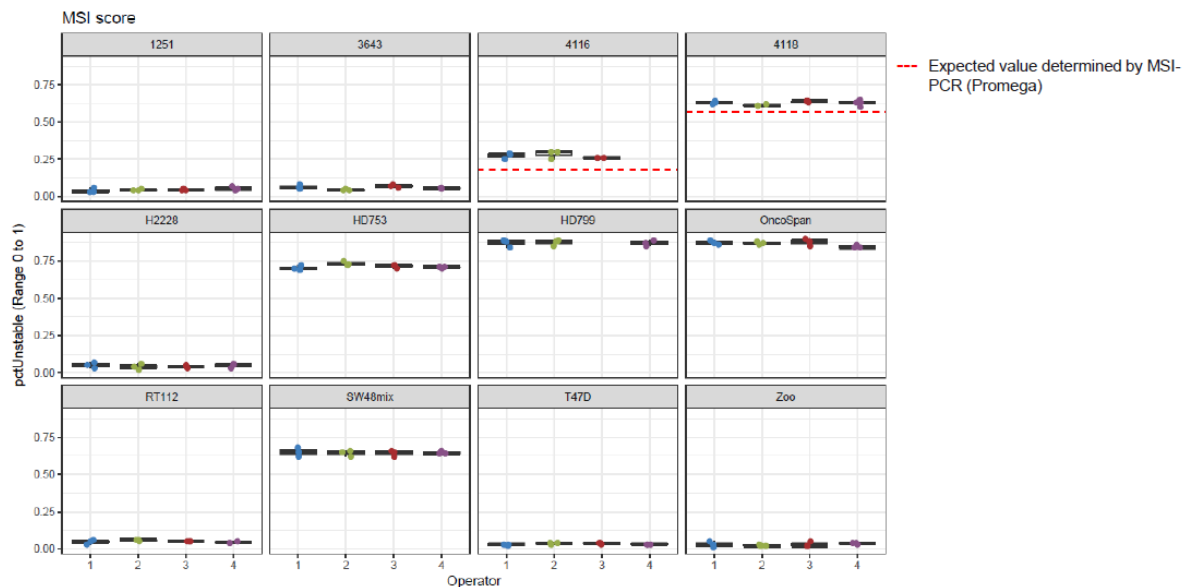
Illumina data on file, 2018

# Reproducibility: MSI

## Reproducibility MSI status

- FFPE and cell line DNA
- 12 replicates tested across 4 operators to determine MSI score

### Robust MSI scoring for all samples across operators

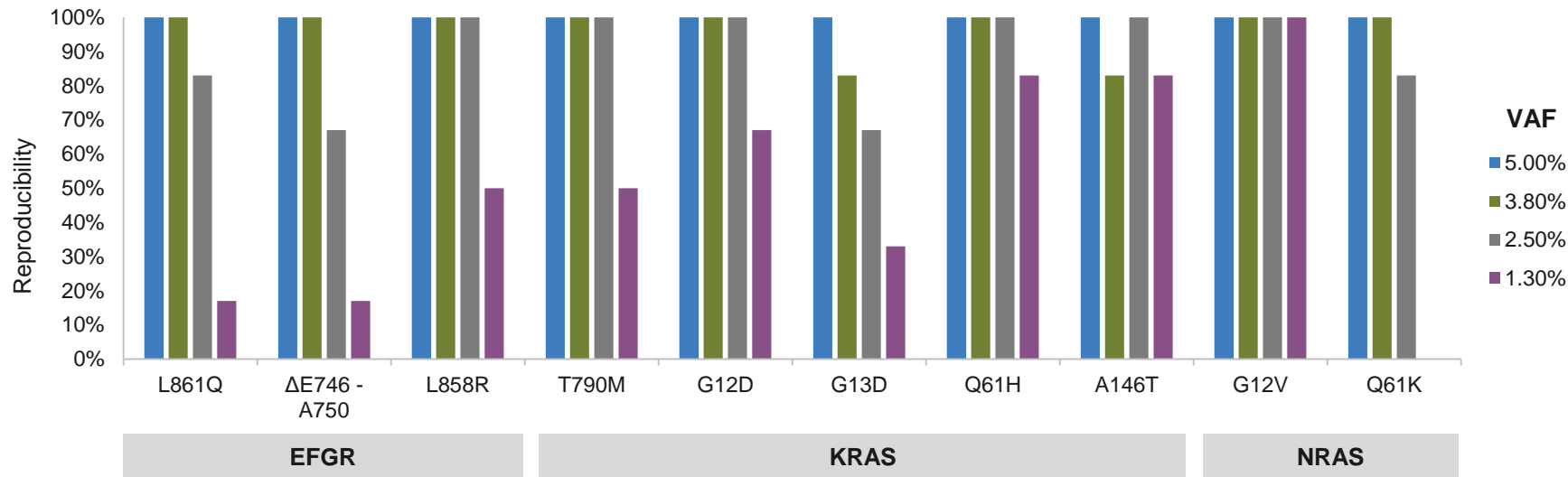


For Research Use Only. Not for use in diagnostic procedures.



# TruSight™ Oncology 500 | Highly Accurate Small Variants Detection

*Precise Assay chemistry combined with sophisticated variant calling pipeline*

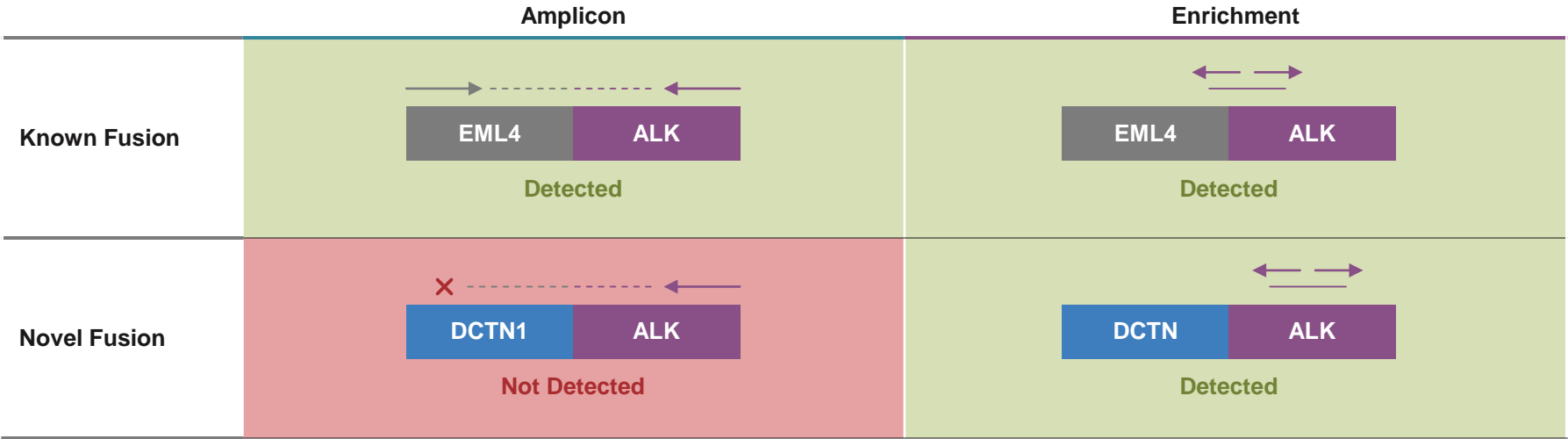


FFPE cell line samples were tested at 40 ng input and VAF ranges of 1.3% - 5.0%, 6 replicates each. VAFs with  $\geq 95\%$  sensitivity are highlighted in green, while VAFs below 95% sensitivity are in yellow.

Illumina data on file, 2018

# TruSight™ Oncology 500 | Ability to Detect Novel Fusions

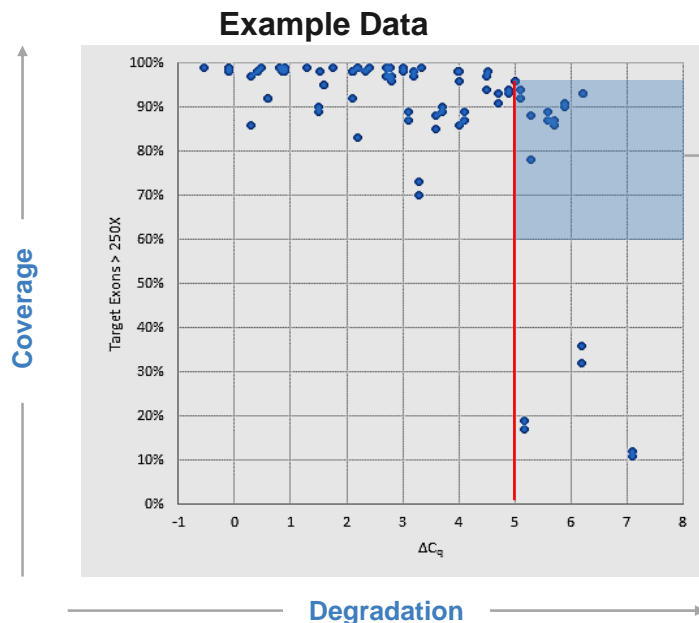
*Enrichment based assay combined with probe design algorithm*



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624911/>

# TruSight™ Oncology 500 | Tolerance to Sample Degradation

*Enrichment based assay can generate yield from highly degraded samples*



Our enrichment based assay and our design algorithm, enable our assays to generate results from even highly degraded samples (fragments <100bp), increasing sensitivity.

Ability to generate data even from highly degraded samples due to enrichment technology. Amplicon technologies tend to have more difficulty and have less sensitivity to such samples

Illumina data on file, 2018

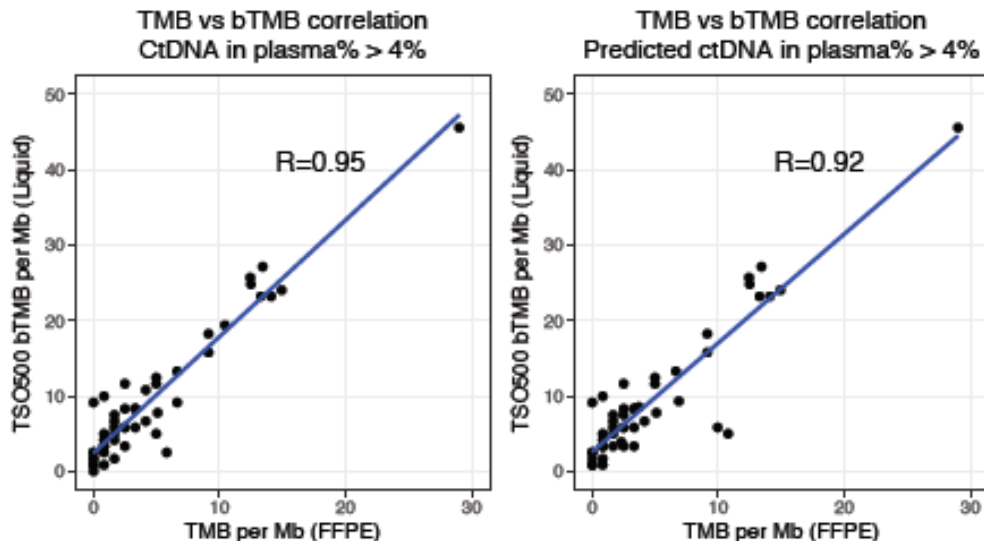
For Research Use Only. Not for use in diagnostic procedures.

illumina®

# TruSight™ Oncology 500

*Future-Proof: UMIs incorporated to enable Liquid Samples analysis  
(under development)*

## TMB evaluation from Liquid Samples (under development)



UMIs designed by our patented algorithm provide state-of-the-art error correction

128 samples, across 4 tissue types were included. Matched FFPE tissue and plasma tested.

The fraction of ctDNA in plasma was calculated

bTMB is highly correlated with tissue TMB in samples with high ctDNA fraction ( $R=0.95$ ).  
Performance is equivalent with the predicted tumor fraction by plasma only ( $R=0.92$ )

Illumina data on file, 2018

illumina®

For Research Use Only. Not for use in diagnostic procedures.

# TruSight™ Oncology 500 | Partnerships

*Partnering with leading institutions to realize the vision of Precision Oncology*



**15 Early Access  
Partner Labs**

**...and  
expanding**

# TruSight™ Tumor 170 | TruSight™ Oncology 500

*Similar and automated workflows*

**Nucleic Acid to Variant Report in 3–4 days**



**FFPE  
Specimens**

**Sample Prep**

**Automated  
Library Prep  
and Enrichment**

**Sequencing**

**Analysis**

**TST170**

**TSO500**

- Multiple tissue types supported

- Commercial DNA/RNA extraction kits

- TruSight™ Oncology Kit
- DNA Probes – 152 genes
- RNA Probes -55 genes
- Automation with Hamilton or Beckman

- TruSight™ Oncology + UMI kit
- DNA Probes - 523 genes
- (optional: combine with TST170 RNA workflow)
- Automatable workflow

- NextSeq™
- NextSeq™ 550Dx\*
- NovaSeq™ - under development

\* For In vitro diagnostic use

3<sup>rd</sup> Party Examples:

- AMR: Pierian Dx
- EU: Sophia Dx
- S. Korea: NGene Bio

**For Research Use Only. Not for use in diagnostic procedures.**

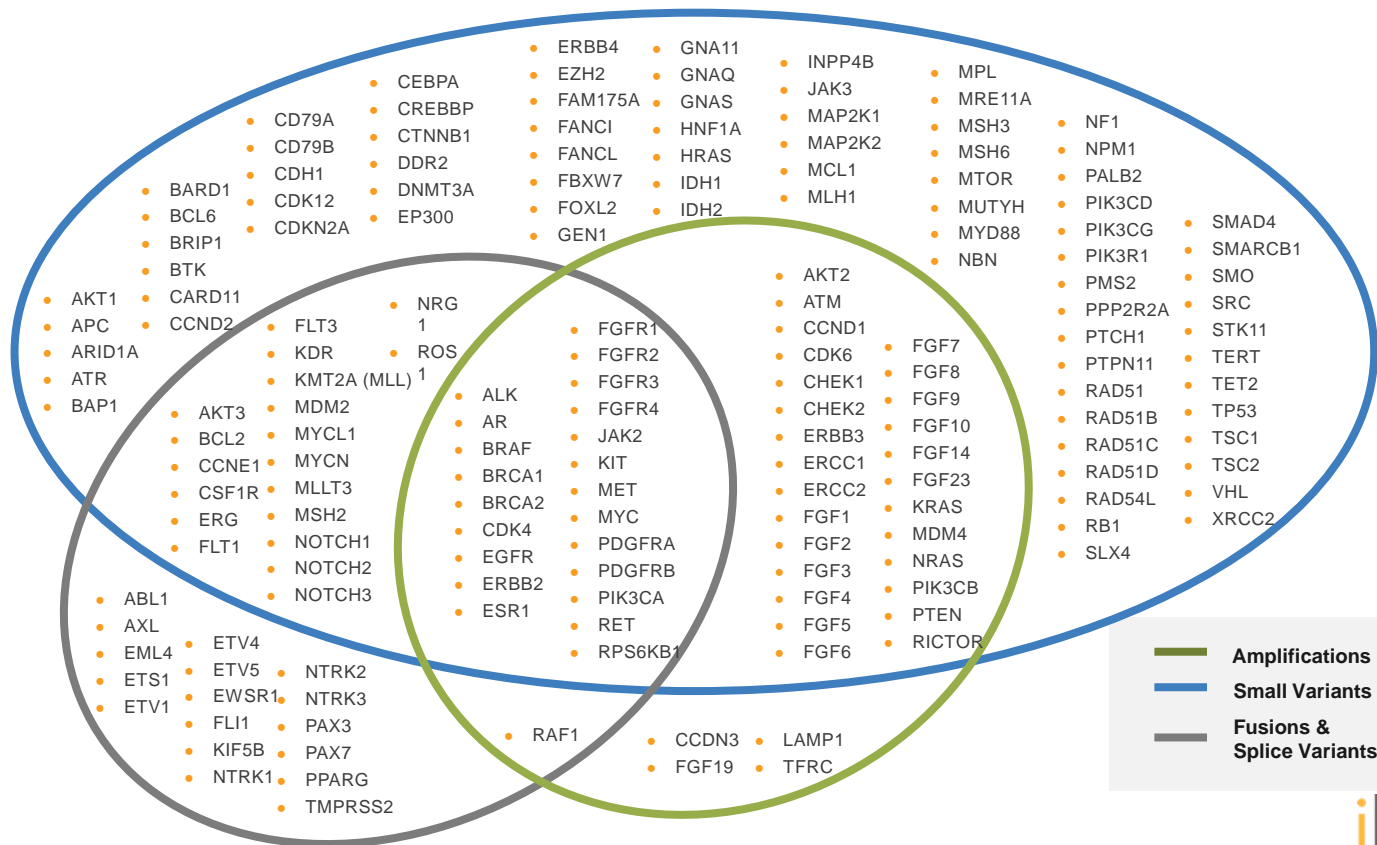
**illumina®**

## APPENDIX – ADDITIONAL PRODUCT INFORMATION



# TruSight Tumor 170 Content

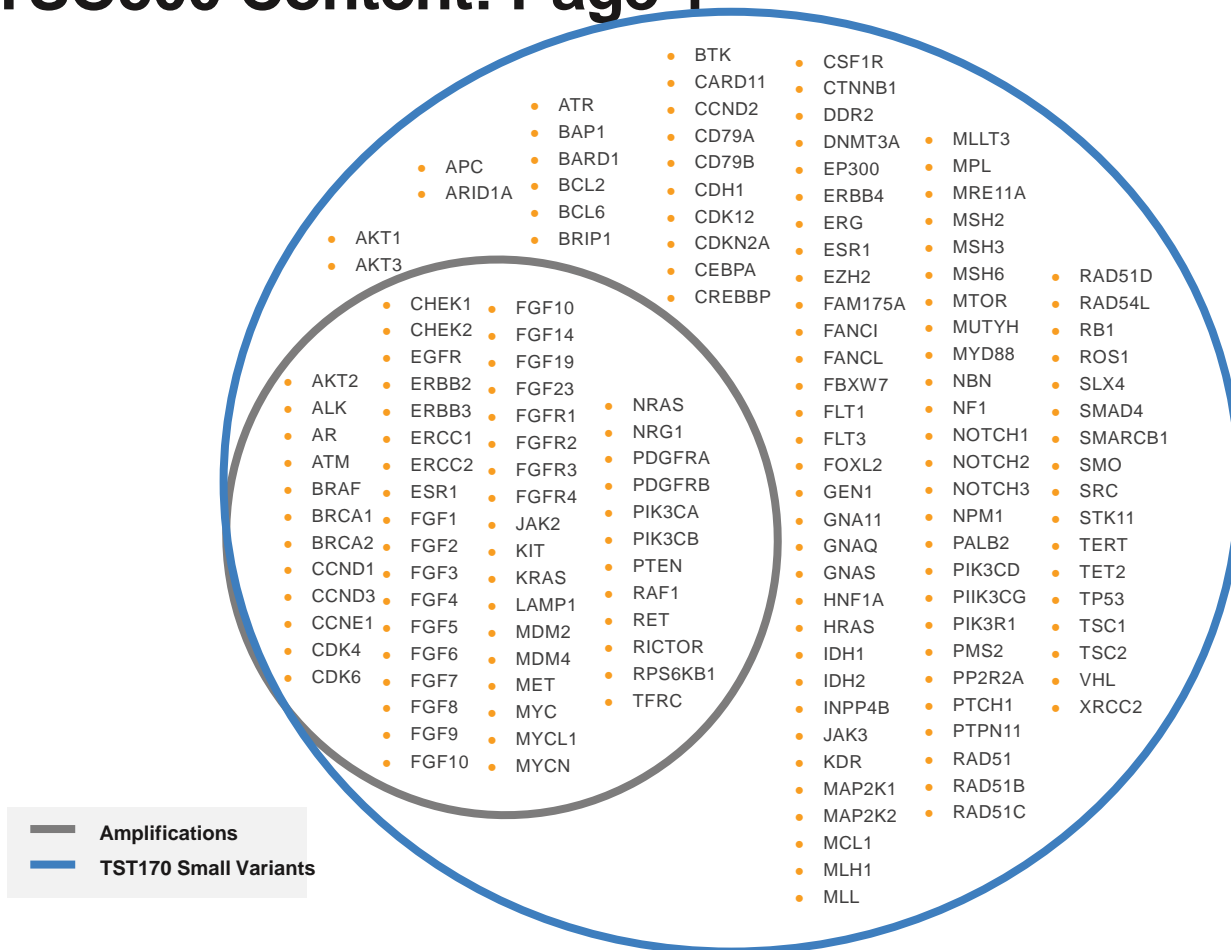
## Gene List and Variant Classification



For Research Use Only. Not for use in diagnostic procedures.



# TSO500 Content: Page 1



# TSO500 Content: Page 2

Small Variants  
Fusions

ABL1	CD74	EPHA7	GPR124	IL7R	MEF2B	PIK3C2G	RNF43	SUFU
ABL2	CDC73	EPHB1	GPS2	INHA	MEN1	PIK3C3	RPS6KA4	SUZ12
ACVR1	CDK8	ERCC3	GREM1	INHBA	MGA	PIK3R2	RPS6KB2	SYK
ACVR1B	CDKN1A	ERCC4	GRIN2A	INPP4A	MITF	PIK3R3	RPTOR	TAF1
ALOX12B	CDKN1B	ERCC5	GRM3	INSR	MST1	PIM1	RUNX1	TBX3
ANKRD11	CDKN2B	ERRF1	GSK3B	IRF2	MST1R	PLCG2	RUNX1T1	TCEB1
ANKRD26	CDKN2C	ETS1	H3F3A	IRF4	MYB	PLK2	RYBP	TCF3
ARAF	CENPA	ETV1	H3F3B	IRS1	MYOD1	PMAIP1	SDHA	TCF7L2
ARFRP1	CHD2	ETV4	H3F3C	IRS2	NAB2	PMS1	SDHAF2	TERC
ARID1B	CHD4	ETV5	HGF	JAK1	NCOA3	PNRC1	SDHB	TET1
ARID2	CIC	ETV6	HIST1H1C	JUN	NCOR1	POLD1	SDHC	TFE3
ARID5B	CRKL	EWSR1	HIST1H2BD	KAT6A	NEGR1	POLE	SDHD	TFRC
ASXL1	CRLF2	FAM123B	HIST1H3A	KDM5A	NF2	PPARG	SETBP1	TGFBR1
ASXL2	CSF3R	FAM46C	HIST1H3B	KDM5C	NFE2L2	PPM1D	SETD2	TGFBR2
ATRX	CSNK1A1	FANCA	HIST1H3C	KDM6A	NFKBIA	PPP2R1A	SF3B1	TMEM127
AURKA	CTCF	FANCC	HIST1H3D	KEAP1	NKX2-1	PPP6C	SH2B3	TMPRSS2
AURKB	CTLA4	FANCD2	HIST1H3E	KEL	NKX3-1	PRDM1	SH2D1A	TNFAIP3
AXIN1	CTNNA1	FANCE	HIST1H3F	KIF5B	NOTCH4	PREX2	SHQ1	TNFRSF14
AXIN2	CUL3	FANCF	HIST1H3G	KLIF4	NSD1	PRKAR1A	SLIT2	TOP1
AXL	CUX1	FANCG	HIST1H3H	KLHL6	NTRK1	PRKCI	SMAD2	TOP2A
B2M	CXCR4	FAS	HIST1H3I	KMT2B	NTRK2	PRKDC	SMAD3	TP63
BBC3	CYLD	FAT1	HIST1H3J	KMT2C	NTRK3	PRSS8	SMARCA4	TRAF2
BCL10	DAXX	FGF19	HIST2H3A	KMT2D	NUP93	PTPRD	SMARCD1	TRAF7
BCL2L1	DCUN1D1	FH	HIST2H3C	LAMP1	NUTM1	PTPRS	SMC1A	TSHR
BCL2L11	DDX41	FLCN	HIST2H3D	LATS1	PAK1	PTPRT	SMC3	U2AF1
BCL2L2	DHX15	FLI1	HIST3H3	LATS2	PAK3	QKI	SNCAIP	VEGFA
BCOR	DICER1	FLT4	HLA-A	LMO1	PAK7	RAB35	SOC3	VTCN1
BCORL1	DIS3	FOXA1	HLA-B	LRP1B	PARK2	RAC1	SOX10	WISP3
BCR	DNAJB1	FOXO1	HLA-C	LYN	PARP1	RAD21	SOX17	WT1
BIRC3	DNMT1	FOXP1	HNRNPK	LZTR1	PAX3	RAD50	SOX2	XIAP
BLM	DNMT3B	FRS2	HOXB13	MAGI2	PAX5	RAD52	SOX9	XPO1
BMPR1A	DOT1L	FUBP1	HSD3B1	MALT1	PAX7	RAF1	SPEN	YAP1
BRD4	E2F3	FYN	HSP90AA1	MAP2K4	PAX8	RANBP2	SPOP	YES1
BTG1	EED	GABRA6	ICOSLG	MAP3K1	PBRM1	RARA	SPTA1	ZBTB2
C11orf30	EGFL7	GATA1	ID3	MAP3K13	PDCD1	RASA1	SRSF2	ZBTB7A
CALR	EIF1AX	GATA2	IFNGR1	MAP3K14	PDCD1LG2	RBM10	STAG1	ZFXH3
CASP8	EIF4A2	GATA3	IGF1	MAP3K4	PDK1	RECQL4	STAG2	ZNF217
CBFB	EIF4E	GATA4	IGF1R	MAPK1	PDPK1	REL	STAT3	ZNF703
CBL	EML4	GATA6	IGF2	MAPK3	PGR	RFWO2	STAT4	ZRSR2
CD274	EPCAM	GID4	IKBKE	MAX	PHF6	RHEB	STAT5A	
CD276	EPHA3	GLI1	IKZF1	MDC1	PHOX2B	RHOA	STAT5B	
	EPHA5	GNA13	IL10	MED12	PIK3C2B	RIT1	STK40	

## FUSIONS + SPLICE VARIANTS

ABL1	KDR
AKT3	KIF5B
ALK	KIT
AR	KMT2A
AXL	(MLL)
BCL2	MET
BRAF	MLLT3
BRCA1	MSH2
BRCA2	MYC
CDK4	NOTCH1
CSF1R	NOTCH2
EGFR	NOTCH3
EML4	NRG1
ERBB2	NTRK1
ERG	NTRK2
ESR1	NTRK3
ETS1	PAX3
ETV1	PAX7
ETV4	PDGFRA
ETV5	PDGFRB
EWSR1	PIK3CA
	PPARG
FGFR1	RAF1
FGFR2	RET
FGFR3	ROS1
FGFR4	RPS6KB1
FLI1	TMPRSS2
FLT1	
FLT3	
JAK2	