Clonal expansion of cancer driver mutants by CarcSeq: a biomarker of carcinogenicity

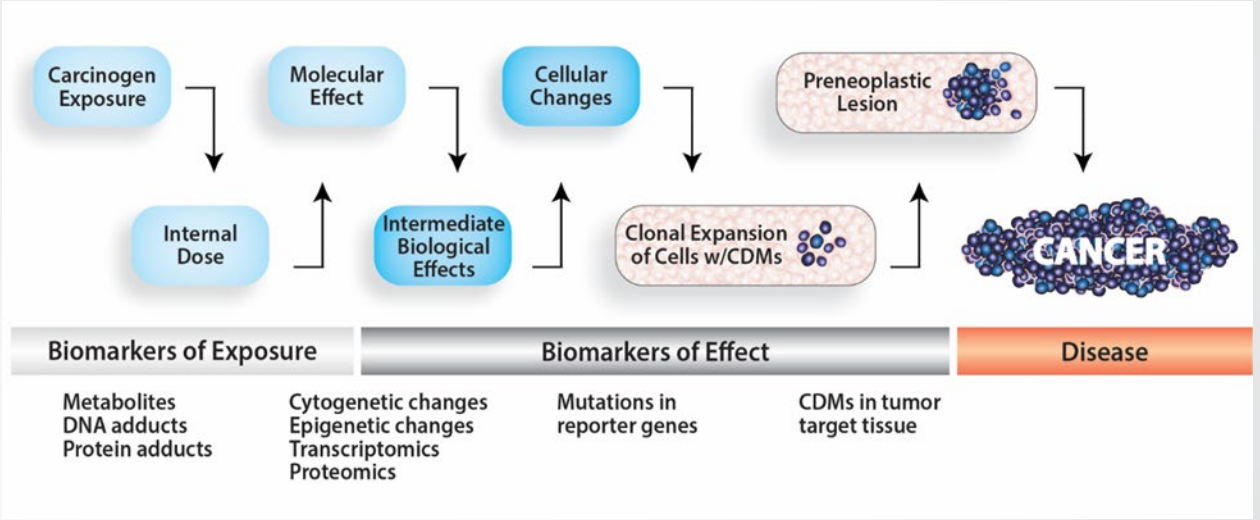
Barbara L. Parsons, US Food and Drug Administration
National Center for Toxicological Research

barbara.parsons@fda.hhs.gov

This presentation reflects the views of the author and does not necessarily reflect those of the US Food and Drug Administration

Section 1: Method Description

Rationale for developing clonal expansion of cancer driver gene mutants as a biomarker of future cancer risk



Parsons, B.L., M.B. Myers, F. Meng, Y. Wang, and P.B. McKinzie. 2010. Oncomutations as biomarkers of cancer risk. *Environmental and Molecular Mutagenesis* 51: 836-850.

- Cancer driver mutations (CDMs) are prevalent in normal human and rodent tissues
- Clonal expansion of cells carrying CDMs is an obligatory step in cancer development and clonal expansion of such mutants constitutes a biomarker of effect (an irreversible, carcinogenesis relevant change), temporally preceding the appearance of histologically observable lesions
- To strengthen the human relevance of the approach, the sequences selected for interrogation by CarcSeq are highly conserved across species and include the most prevalent hotspot CDMs in the 10 deadliest human cancers
  - lung/bronchus, colon/rectum, breast, prostate, non-Hodgkin lymphoma, bladder, kidney, melanoma, endometrium, and thyroid
  - *Apc* R1450, *Braf* V600, *Egfr* T790/L858, *Hras* G12/G13/Q61, *Kras* G12/G13/Q61, *Nfe2l2* E79/E82, *Pik3ca* E542/E545/H1047, *Setbp1* D868/G870/I871, *Stk11* F354, and *Tp53* R175/R248/R273)

Section 1: Method Description

(Overview of the approach)

Input is genomic DNA isolated from any human, rat or mouse tissue (frozen tissue samples)

CarcSeq

Individually amplify 13-15 regions of cancer driver gene sequence using PCR primers with random bases at their 5'-ends (unique molecular Identifier, UMI)

Products synthesized from the same sample are pooled, libraries are prepared and sequenced

For error correction by consensus, reads are sorted based on UMI. Recovery of ≥2 reads with the same UMI and ≥90% identify across reads are the criteria used for mutant or wild-type base calling, i.e., creating a single strand consensus sequence (SSCS)

Error corrected SSCS data are then used to calculate mutant fraction (MF) as the # mutant SSCSs/total # SSCSs, at every sequenced position

MF is measured with high throughput (≥ 300,000 SSCSs)

Only MFs ≥ 10^{-4} are considered, because: 1) it is fit for the purpose of quantifying clonally-expanded mutants, and 2) eliminates uncorrectable PCR errors that may occur in the first few cycles.

Clonal expansion is quantified as variability in MF, with Median Absolute Deviation in MF (MAD in MF) used as the metric of variability

Method Description: CarcSeq

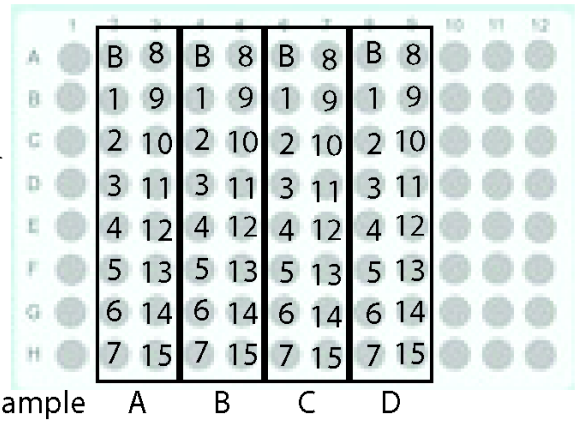
PCR Amplification and Product Characterization

Isolate genomic DNA

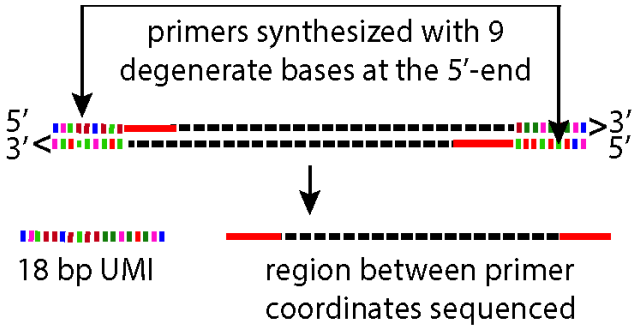
Agilent DNA Extraction Kit, (salting out method, high quality and yield from large tissue samples)

0.5 ug used in each individual PCR

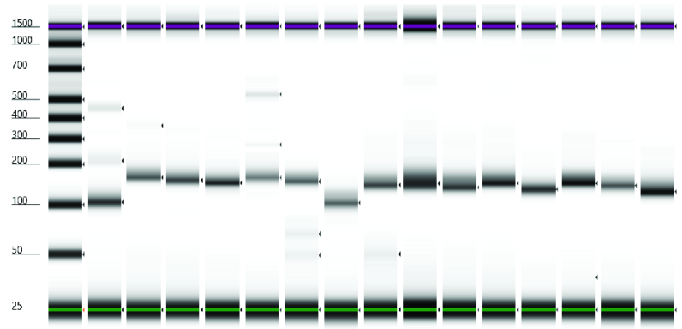
1.6×10^5 genome equivalents



13 - 15 regions of cancer driver genes (amplicons) amplified from each sample



Agilent TapeStation quantification of each product (pg/ul)

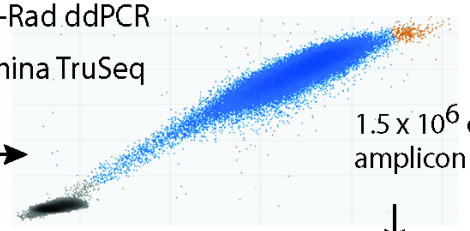


Library Preparation, Sequencing, and Bioinformatics

Pool aliquots of PCR products (empirically determined masses)

Illumina TruSeq ChIP Kit and Bio-Rad ddPCR Library Qunatification Kit for Illumina TruSeq

- End repair & A-tailing
- Ligate adapters, different indices
- Gel purity amplicon/adapters
- Quantify molecules by Illumina ddPCR



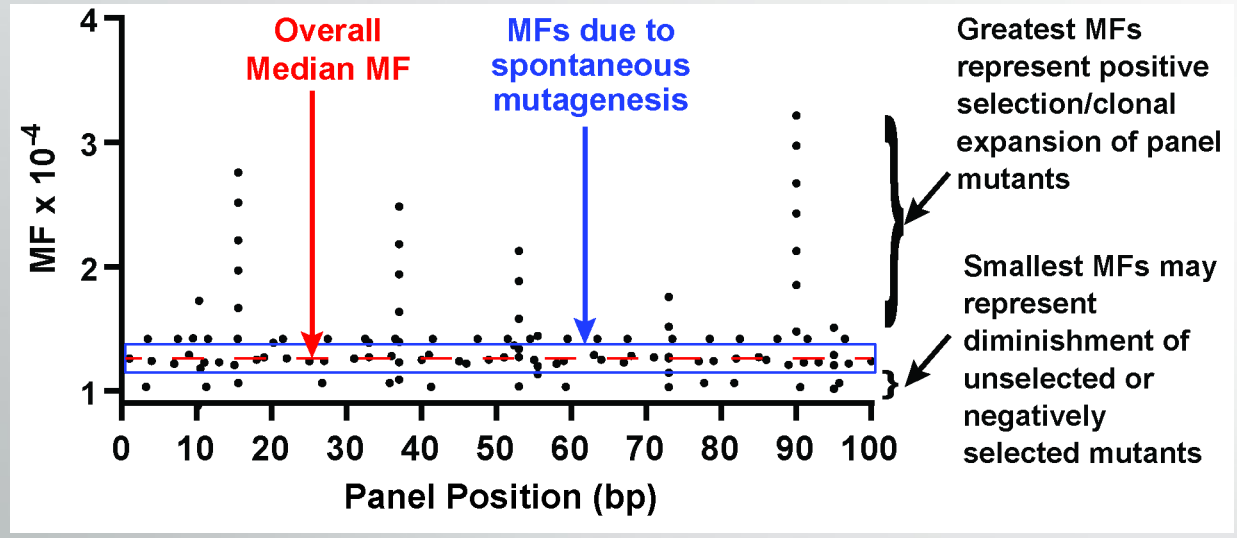
1.5×10^6 copies of each amplicon into final PCR

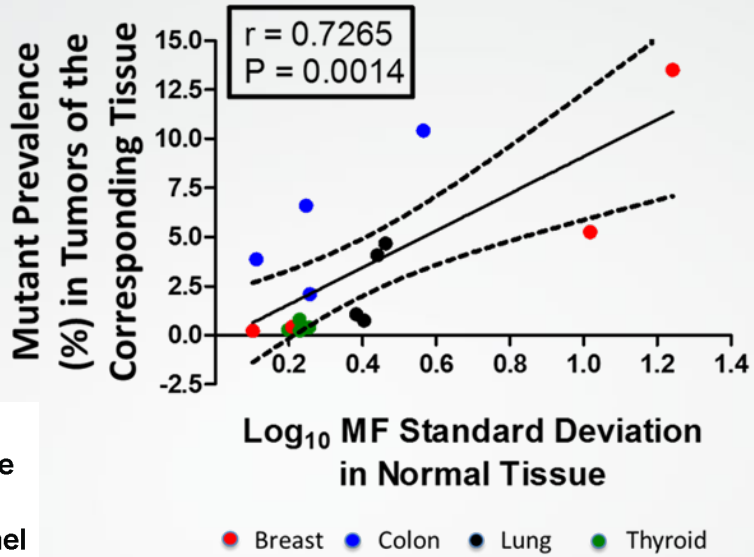
4 samples per mid-output flow cell sequenced on a NextSeq 500, which deconvolutes sample reads by index

- Reads sorted by length
- Reads sorted into UMI-based families
- Error-correction by consensus using Kennedy/Univeristy of Washington pipeline and unified consensus maker version 2017 or version 2.1.3 (https://github.com/Kennedy-Lab-UW/Duplex-Seq-Pipeline/releases)
- Construct single strand concensus sequences (SSCSs)
- A minimum of 2 members and 90% consensus required to create a single strand consensus sequence (SSCS)
- Analyze ≥ 300,000 SSCSs, resequence libraries if necessary
- Employ two filters:
  - Mutant fraction (#mutant SSCSs/# SSCS at any position) ≥ 10^{-4} (each mutation will be represented by 30 mutants
  - A coefficient of variation of ≥ 60% across samples

Method Description: MAD in MF

- Our ACB-PCR work showed variability in CDM MF correlated with CDM impact on tumor development
- A variety of metrics were analyzed for correlation with cancer risk factors (median MF, geomean MF, MF standard deviation) but MAD in MF exhibited the strongest correlations





Parsons *et al.* Variation in organ-specific *PIK3CA* and *KRAS* mutant levels in normal human tissues correlates with mutation prevalence in corresponding carcinomas. *Environ Mol Mutagen* (2017) 58: 466-476.

Parsons, B.L. (2018) Modern conception of carcinogenesis creates opportunities to advance cancer risk assessment. *Current Opinion in Toxicology* 11-12:1-9.

Measured MF			Median			Absolute Deviation		
Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
0.000125	0.000110	0.000132	0.000124	0.00013	0.000121	1.2220E-06	1.9926E-05	1.0253E-05
0.000141	0.000132	0.000164				1.6880E-05	2.0042E-06	4.2327E-05
0.000100	0.000141	0.000107				2.3887E-05	1.1545E-05	1.4431E-05
0.000104	0.000116	0.000133				2.0582E-05	1.3331E-05	1.1253E-05
0.000104	0.000116	0.000120				2.0538E-05	1.3397E-05	1.7900E-06
0.000138	0.000113	0.000114				1.4017E-05	1.6634E-05	7.5409E-06
0.000104	0.000135	0.000107				2.0563E-05	5.8541E-06	1.3905E-05
0.000129	0.000226	0.000133				4.5446E-06	9.6876E-05	1.1284E-05
0.000110	0.000167	0.000123				1.4007E-05	3.7071E-05	1.7900E-06
0.000100	0.000129	0.000104				2.3805E-05	7.6497E-07	1.7136E-05
0.000148	0.000135	0.000130				2.3993E-05	5.6973E-06	8.1934E-06
0.000110	0.000104	0.000126				1.3774E-05	2.5246E-05	4.9226E-06
0.000126	0.000130	0.000101				1.9427E-06	0.0000E+00	2.0350E-05
0.000123	0.000123	0.000126				1.2255E-06	6.2860E-06	4.9226E-06
0.000132	0.000101	0.000110				8.2277E-06	2.8345E-05	1.0988E-05
0.000123	0.000136	0.000104				1.2220E-06	6.0619E-06	1.7227E-05
0.000246	0.000104	0.000253				1.2234E-04	2.5310E-05	1.3132E-04
0.000128	0.000139	0.000101				4.2987E-06	9.3008E-06	2.0435E-05
	0.000126						3.2931E-06	
	0.000129						3.4084E-07	
	0.000139						9.1392E-06	
	0.000199						6.9167E-05	
	0.000138						8.6791E-06	
					Median Absolute Deviation	1.4012E-05	9.3008E-06	1.1269E-05

- Strengths of the MAD approach are that it utilizes all available data and robustly normalizes impacts of individual large MFs
- MAD in MF can be calculated for treatment groups or individual animals within treatment groups
- MAD in MF is calculated independently for tissue-specific and non-tissue specific driver gene MFs, as an approach for internal validation

Method Description: applicability domain

- Capable of detecting effects of non-genotoxic and genotoxic carcinogens
- No reason to expect there are classes of chemicals that would be inappropriate to assess using a biomarker of clonal expansion
- Preliminary evidence suggests exceeding the MTD as defined by a terminal body weight reduction of 10-12% relative to controls may reduce clonal expansion, consistent with the known effect of body weight reduction on delaying tumor development [van Berlo *et al.* (2022) *Regul Toxicol Pharmacol* 134: 105235]

Section 2: Context of Use

Context of Use

A. How is your method intended to be used?

It is envisioned that CarcSeq assessment of clonal expansion could be employed as part of early drug development, using samples collected from repeat dose, pre-clinical safety studies 3 – 6 months in duration

Context of Use

B. What regulatory testing need does your method address?

- CarcSeq is intended to provide a new information stream for building a weight of evidence carcinogenicity assessment as per ICH S1B(R1)
- Direct evidence of a potential carcinogenic liability that does not rely on histopathological evaluation could be derived from routinely conducted general toxicology studies of 3 to 6 months in duration
- These novel genetic analyses could be used by regulatory health authorities to determine the need for 2 year rat cancer bioassays
- Should CE of CDMs be observed, CarcSeq would provide exposure-response information that could contribute to the quantitative assessment of carcinogenic risk

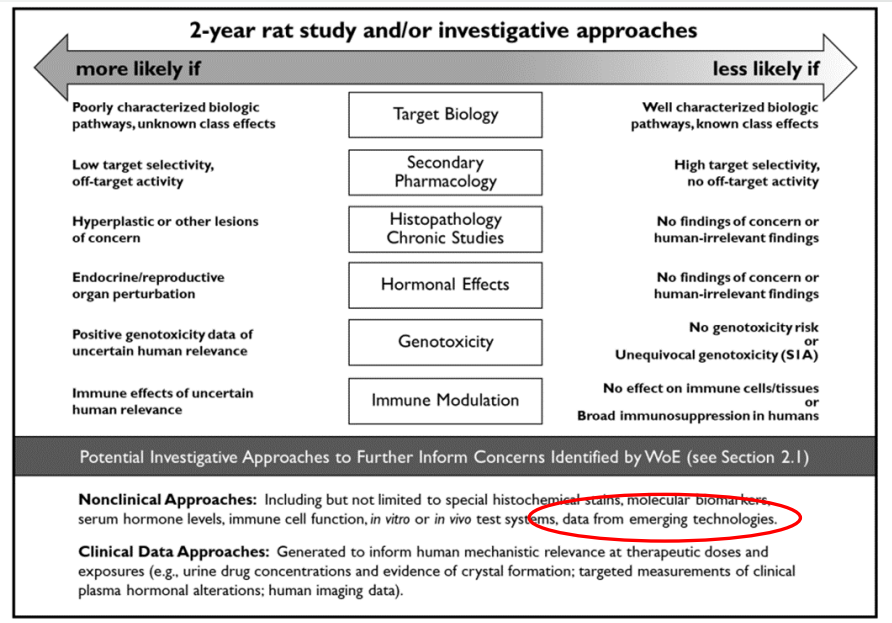


Figure from ICH S1B(R1) describing weight of evidence (WoE) carcinogenicity assessment

- Theoretically, information on cancer driver mutant clonal expansion could have value for chemical screening, hazard identification, potency evaluation, and developing adverse outcome pathways (AOPs).
- Identification of clonal expansion of specific driver gene mutations has the potential to inform mode of action
- Practically, how could CarcSeq be applied to general toxicology studies of 3 to 6 months in duration?
  - Tissues could be prospectively collected and frozen from repeat-dose studies
  - The tissues to be investigated by CarcSeq could be selected based on histopathology, consideration of the pharmacologic target, or applied to known tumor sensitive organs; alternatively, CarcSeq might be performed on DNA pooled from multiple organs

Context of Use

C. What regulatory space does your method address (e.g., agrochemicals, pharmaceuticals, medical devices, cosmetics, food/food additives, industrial chemicals)?

- The assay was developed primarily for use in drug development and regulatory review of investigational new drugs
- However, clonal expansion of CDMs as an early event could also be useful to derive an early indication of cancer risk in the development of agrochemicals
- The ability to perform the same DNA-based test in rodents and humans suggests that eventually the approach could be useful in the assessment of environmental and occupational exposures

Context of Use

D. Has data generated by your method been used for regulatory submissions?

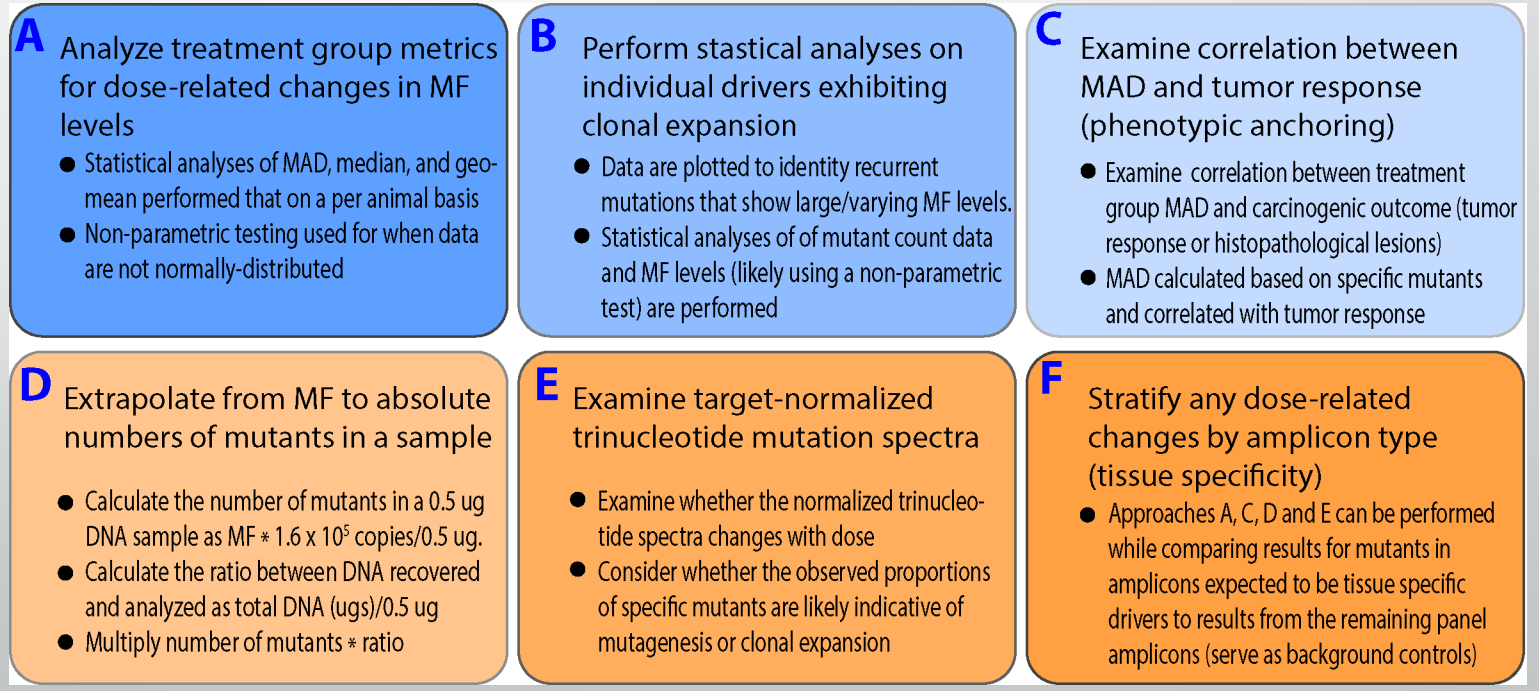
- CarcSeq data has not been used in a regulatory submission

Section 3: Biological Relevance

Biological Relevance

A. Mechanistic understanding: How does the information provided by your method support known mechanistic knowledge of the carcinogenesis process (e.g., an AOP or toxicologically relevant biological process)?

- Clonal expansion of cells carrying CDMs is an obligatory step in carcinogenesis
- The biomarker provides a direct read-out of clonal expansion of human-relevant driver mutations. Biological relevance is a strength of the CarcSeq clonal expansion biomarker approach
- The nature of the driver mutations observed can inform mode of action or identification of an AOP
- Statistical approaches have been developed to distinguish clonal expansion from mutagenesis

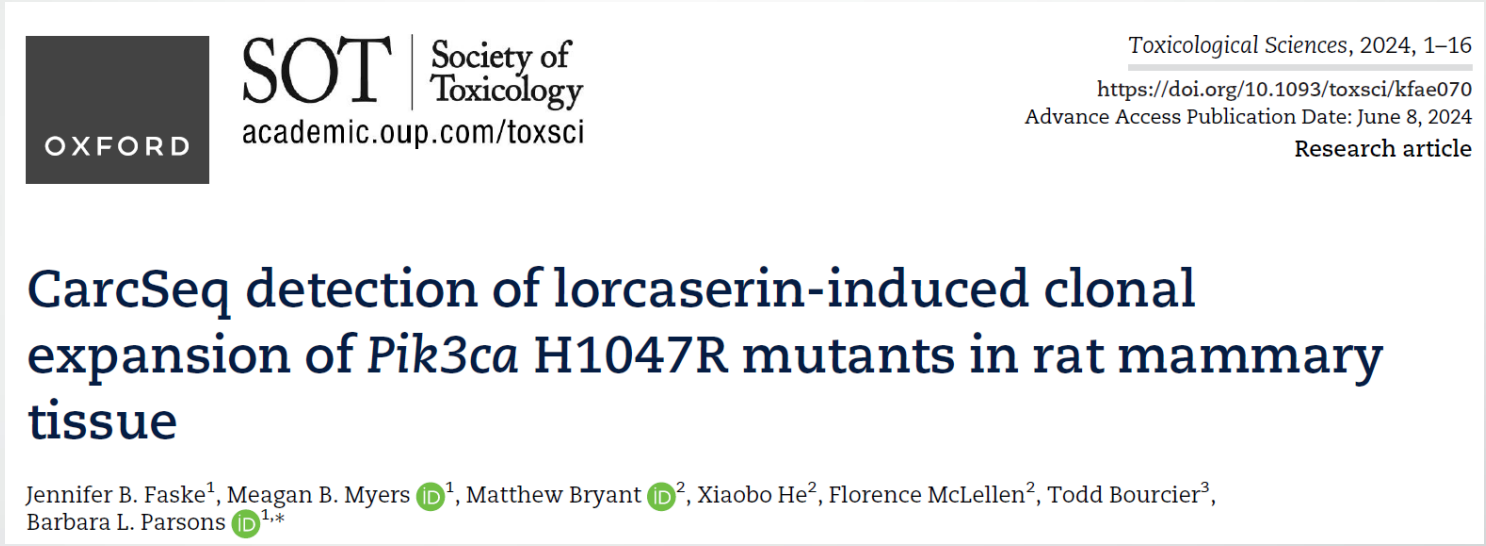


Biological Relevance

B. Reference compounds: What are well-characterized and understood compounds were used to assess the scientific validity or transferability of your method?

Two treatment studies have been conducted: lorcaserin, a non-genotoxic carcinogen and benzo[b]fluoranthene, a genotoxic carcinogen

Lorcaserin



OXFORD

SOT | Society of Toxicology

academic.oup.com/toxsci

Toxicological Sciences, 2024, 1–16
https://doi.org/10.1093/toxsci/kfae070
Advance Access Publication Date: June 8, 2024
Research article

CarcSeq detection of lorcaserin-induced clonal expansion of *Pik3ca* H1047R mutants in rat mammary tissue

Jennifer B. Faske[1], Meagan B. Myers[1], Matthew Bryant[2], Xiaobo He[2], Florence McLellen[2], Todd Bourcier[3], Barbara L. Parsons[1,*]

- This project was initiated by Todd Bourcier, Division of Pharmacology and Toxicology, Office of Cardiology, Hematology, Endocrinology, and Nephrology, Center for Drug Evaluation and Research, US FDA this project and supported by CDER, Office of New Drugs Grant, OND-22-M-1720 (Bourcier, Project Lead).
- Lorcaserin is a non-genotoxic, selective serotonin 2c receptor (5HT2c) agonist
- Approved in 2012 by FDA as an adjunct treatment for chronic weight management and withdrawn from US market in 2020 due to excess cancer risk identified in a cardiovascular outcome trial

Lorcaserin caused an increased incidence of multiple tumor types with increased multiplicity, lethality, and metastatic potential in rats and humans (although tumor types differ)

2-year rat bioassay findings

Males	Females
• Brain astrocyotoma • Mammary adenocarcinoma/fibroadenoma (!) • Skin/subcutis fibroma, squamous carcinoma • Schwannoma, systemic • Liver adenoma/carcinoma • Thyroid follicular cell adenoma	• Mammary adenocarcinoma/fibroadenoma • trend: brain astrocytoma

CAMELLIA-TIMI 61 Trial findings

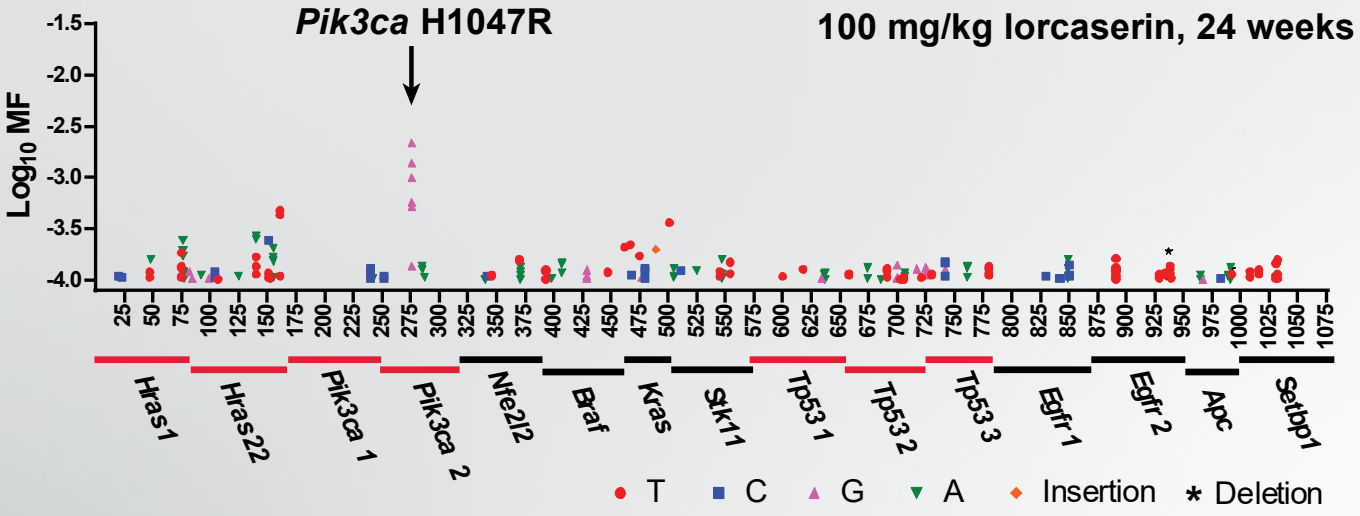
Table 1. Number of Cancers by Treatment Group in the CAMELLIA-TIMI 61 Trial, Full Analysis Set.*

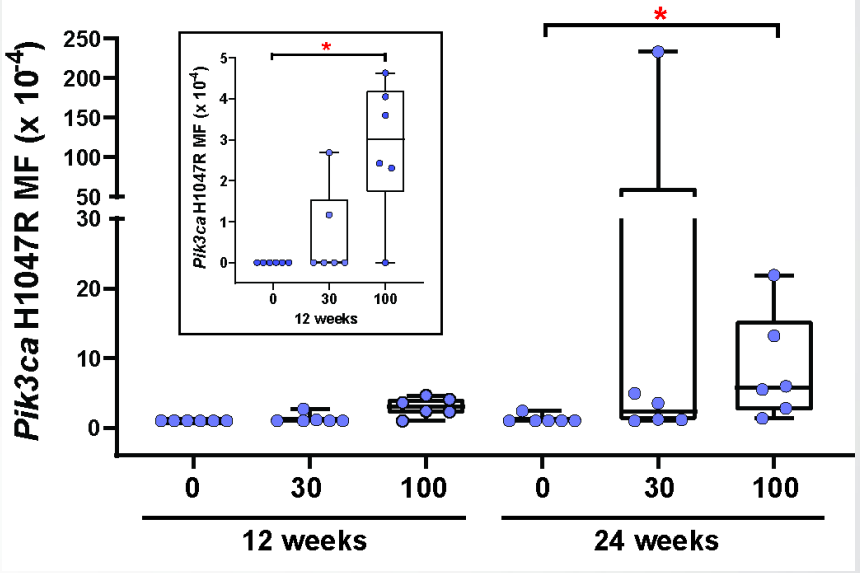
Variable	Lorcaserin (N = 6000)	Placebo (N = 6000)	Total (N = 12,000)
Cancers	520	470	990
Patients with cancers	462	423	885
Deaths from cancer	52	33	85
Patients with multiple primary lesions	20	8	28
Metastases	34	19	53

* Some patients had multiple cancers or multiple metastases.

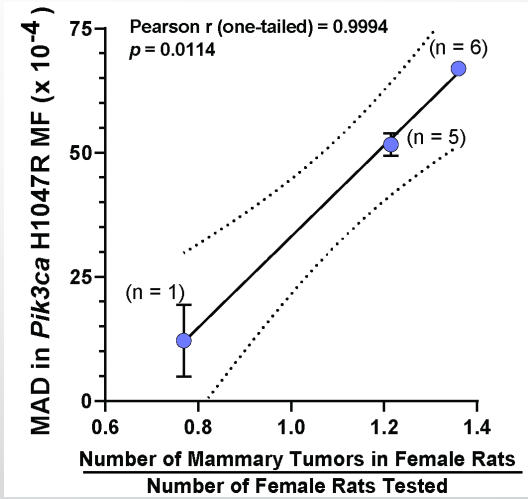
Analyzed DNA isolated from ¼ of the mammary tissue of female Sprague Dawley rats treated with 0, 30, or 100 mg/kg lorcaserin (mid and high bioassay doses) daily for 12 or 24 weeks (n = 6)

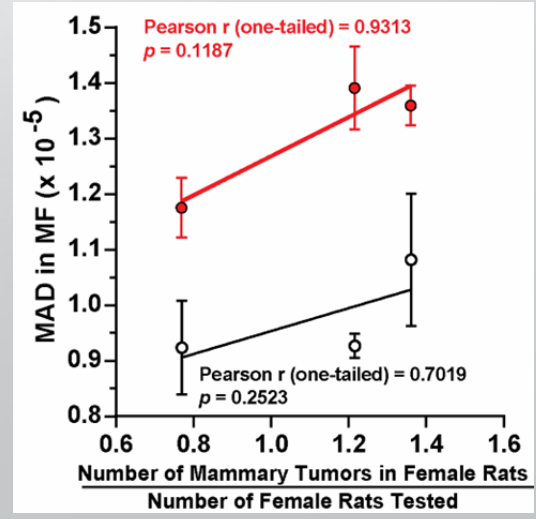
CarcSeq analysis performed by Jennifer Faske, US FDA, NCTR

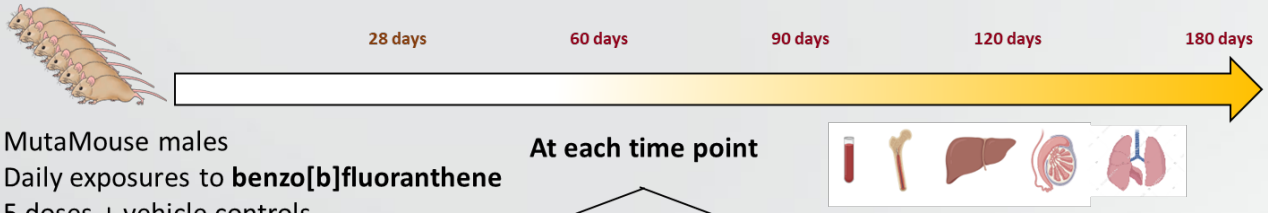




- Significant dose-related increases in *Pik3ca* H1047R MF and significant changes in the numbers of rats carrying the mutation, were observed after 12 and 24 weeks of treatment
- MAD in MF calculated based on *Pik3ca* H1047R MFs within dose groups treated for 24 weeks correlated with mammary tumor response in the rat bioassay
- Strong but non-significant correlation between MAD in MF and tumor response when MAD in MF was calculated using all mammary specific mutations for each dose group (*Hras*, *Pik3ca*, and *Tp53*)
- A weaker correlation was observed when MAD in MF was calculated from non-mammary-specific mutations (*Apc*, *Braf*, *Egfr*, *Kras*, *Nfe2l2*, *Setbp1*, and *Stk11*)







- MutaMouse males
- Daily exposures to **benzo[b]fluoranthene**
- 5 doses + vehicle controls
- N = 8 per dose group (4 for DS)

At each time point

Conventional endpoints
Established protocols
LacZ mutation (liver, bone marrow, germ cells)
Pig-a mutation (blood)
Micronucleus frequency (blood)
Histopathology

Genomics endpoints
TwinStrand MMP + *Pig-a* + *lacZ*
CarcSeq – CDM panel

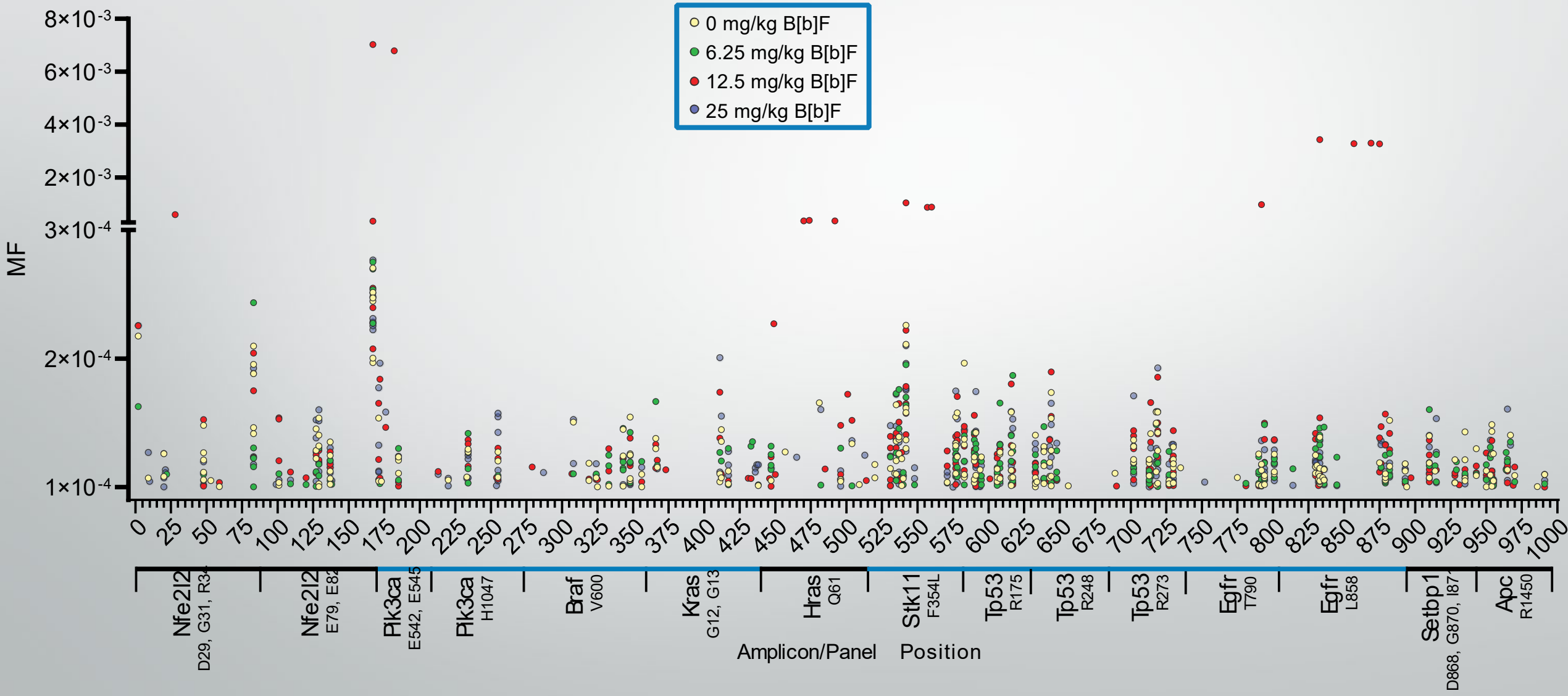
Concordance, correlation, potency analysis

Mutation frequency (genome and locus-specific)
Dose-response analysis, comparisons with conventional endpoints
Mutation spectrum analysis
Clonal expansion over time

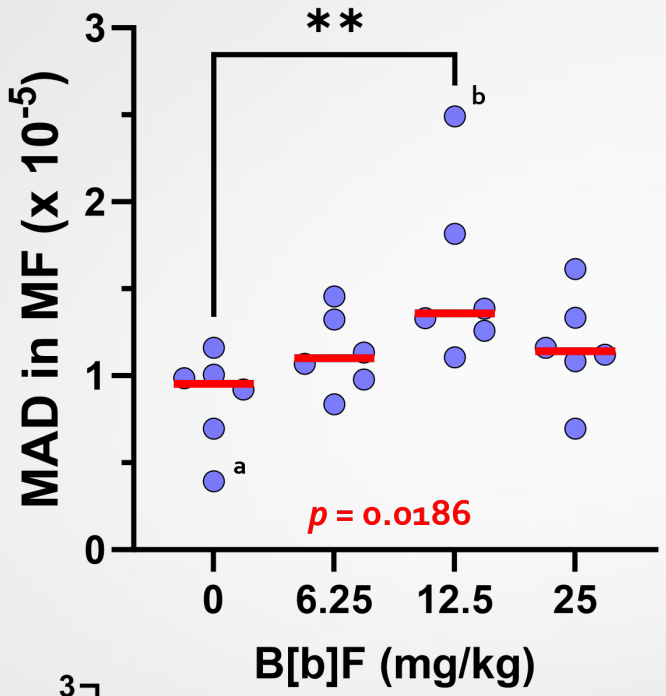
Benzo[b]fluoranthene

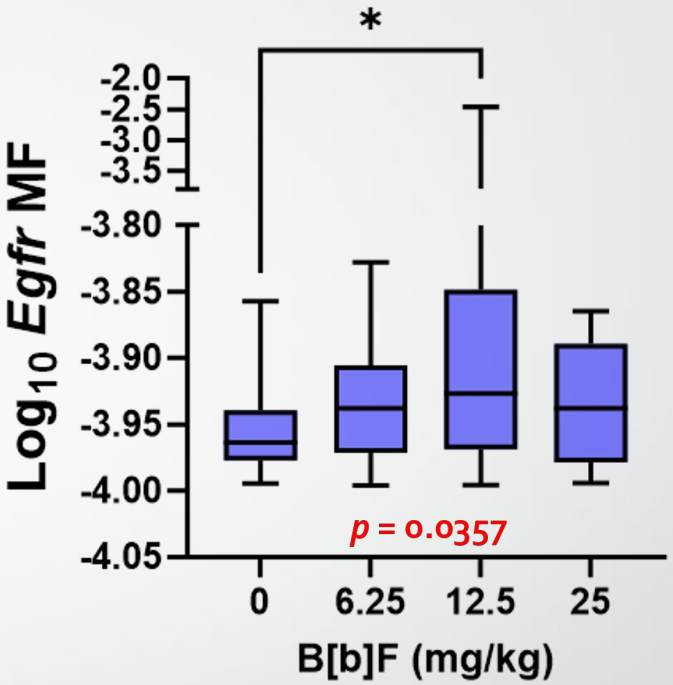
- Benzo[b]fluoranthene (B[b]F) is a genotoxic mouse lung carcinogen [a single 50 mg/kg i.p. injection at 6-8 weeks of age increased significantly the frequency of lung adenomas in A/J mice 8 months after treatment from 55% in controls to 80% , Mass *et al.* (1996) Carcinogenesis: 1701-1704.]
- Carole Yauk and Francesco Marchetti are leading a multi-endpoint international collaboration for which Dr. Yauk received a Burroughs Wellcome Innovations in Regulatory Sciences Award
- CarcSeq performed on DNA isolated from the superior lobe of MutaMouse males treated for 180 days with 0, 6.25, 12.5, or 25 mg/kg/day, using 6 mice per group
- CarcSeq work primarily conducted by Jennifer Faske, with bioinformatic support from Binsheng Gong, other contributors include: Danielle Leblanc, Andreas Zeller, Juergen Funk, Sabrina Kehm, Gu Zhou, Paul White, and Timothy Robison

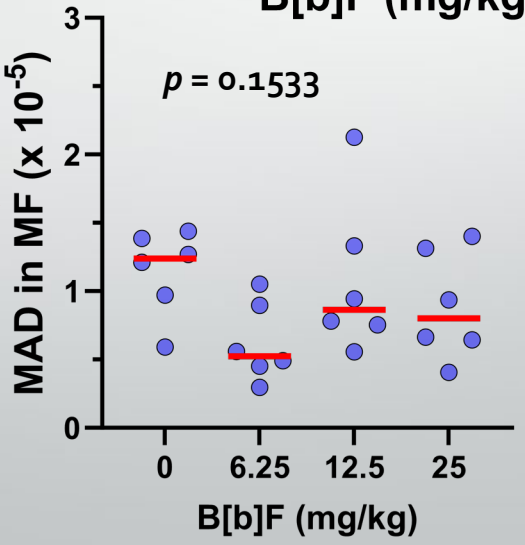
MF levels and distribution of mutants across the CarcSeq amplicon panel



- MAD in MF based on lung-specific drivers (*Braf*, *Egfr*, *Kras*, *Pik3ca*, *Stk11*, and *Tp53*) shows significant induction of clonal expansion based on ANOVA , with 0 and 12.5 mg/kg significantly different in a pairwise comparison
- Significant induction of *Egfr* MF also observed
- MAD in MF for non-lung-specific drivers (*Apc*, *Hras*, *Nfe2l2*, *and Setbp1*) was not significant



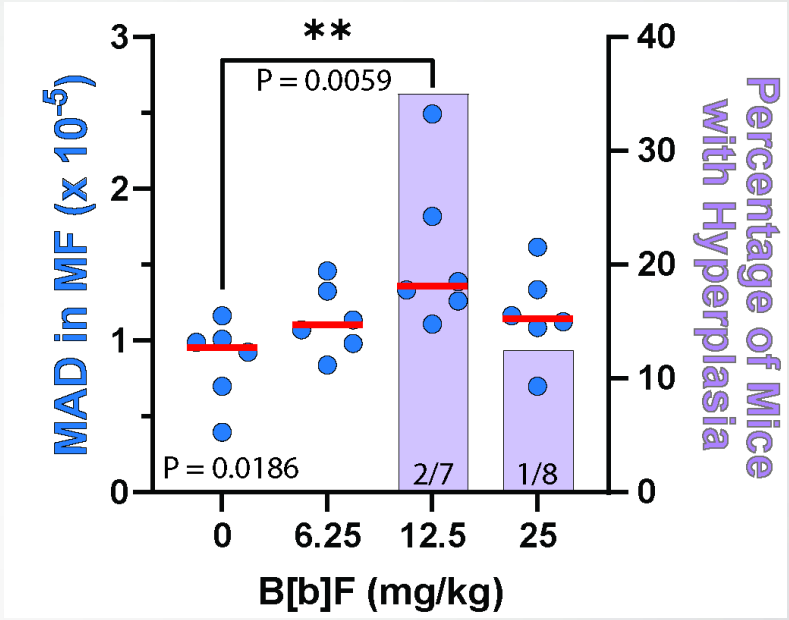




= treatment group median

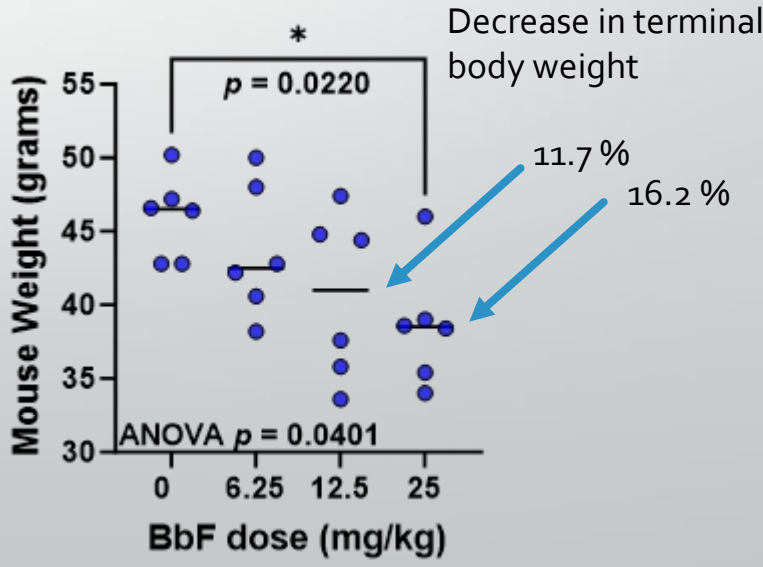
Mutation in Mammary-Specific Amplicons	Mouse with Smallest MAD[a]	Mouse with Greatest MAD[b]
Average MF (Median MF)	1.1693×10^{-4} (1.0700×10^{-4})	8.4454×10^{-4} (1.2909×10^{-4})
Unique Mutations	24	41
Total Mutants Recovered in 0.5 ug Lung DNA (Average)	896 (37)	10,917 (266)
Estimated Mutants in Lung Tissue Sample (~10 mg)	132,603	1,004,428

Dose mg/kg/day	0	1.56	3.125	6.25	12.5	25	ENU
No examined	8	8	7	8	7	8	3
Hyperplasia bronchioloalveolar	0/8	0/8	1/7	0/8	2/7	1/8	0/3
Adenoma	0/8	0/8	0/7	0/8	0/7	0/8	1/3



Histopathological analyses performed on a different lung lobe of the same mice

- Performed by Sabrina Kehn, Juergen Funk, and Andreas Zeller at Hoffmann–La Roche, Basel, Switzerland
- Very subtle bronchioloalveolar hyperplasia in a few animals that would most likely not have been considered test item-related if mode of action of the compound were not known
- No tumors observed with B[b]F treatment
- 1/3 (33%) positive control mice treated with a single 50 mg/kg dose of ENU developed a lung adenoma
- Decreases in MAD in MF and hyperplasia were observed at 25 mg/kg as compared to 12.5 mg/kg B[b]F, potentially due to exceeding the MTD



Biological Relevance

C. Comparison to existing laboratory animal methods: How does your method provide information that is equivalent or better than that from existing methods used for regulatory purposes? How does your method contribute to the reduction, refinement, or replacement of animal assays?

- No test for clonal expansion currently exists; but such a test is needed acutely for identification of non-genotoxic carcinogens
- Assessment of clonal expansion early in drug/agrochemical development pipelines or in chemical evaluations (EPA, CFSAN, NIOSH) using tissues from the ≤6-month repeat dose rodent studies, is intended to generate information to predict rodent tumor response in the absence of bioassay data, and without the use of additional animals (using tissues from required preclinical studies)
- Negative findings could be included as part of a carcinogenicity assessment that ascertains a rat bioassay is not needed, thereby reducing animal use
- Early positive findings could reduce late-stage development failures due to unexpected carcinogenicity, thereby reducing animal use

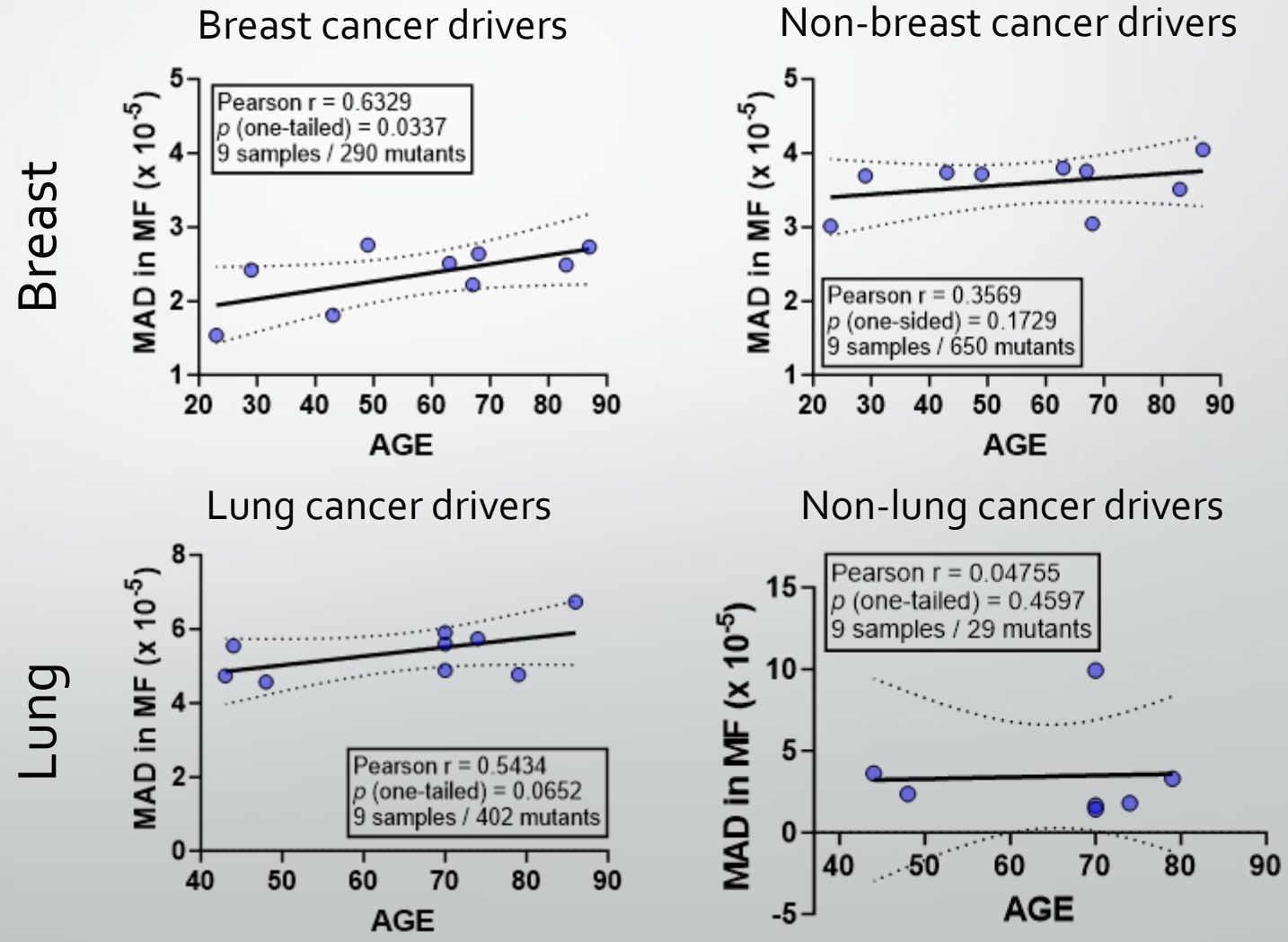
Section 4: Technical Characterization

Technical Characterization

B. How has <u>robustness</u> been evaluated?

MAD in MF was correlated with human age, a major cancer risk factor

- Measured MF in 9 normal human breast samples and 9 normal human lung samples
- Demonstrated that MAD in MF correlated with human age, a major cancer risk factor when based on the appropriate set of driver genes
- MAD calculated from non-tissue-specific drivers genes did not correlate with age



Technical Characterization

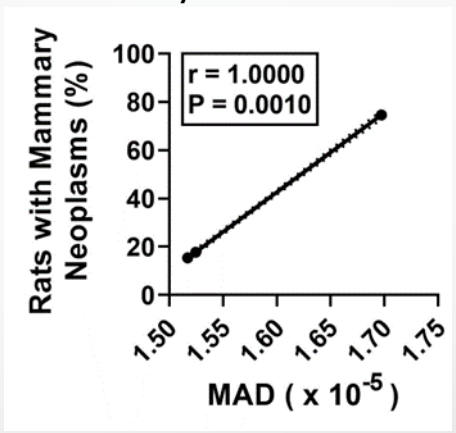
B. How has <u>robustness</u> (i.e., the ability of the method to be reproduced under different conditions or circumstances, without the occurrence of unexpected differences in the obtained results) been evaluated?

MAD in MF was correlated with bioassay control tumor response data

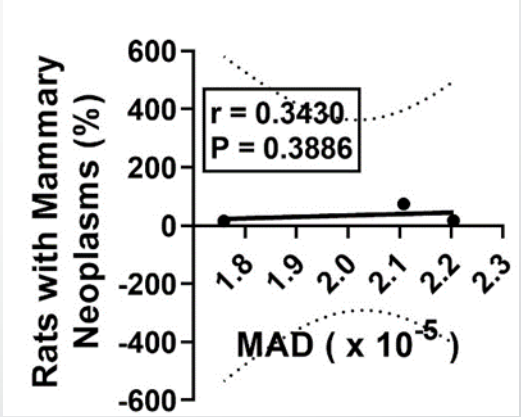
- MAD in MF calculated using appropriate driver genes correlated with the tumor response as impacted by rodent strain and sex
- MAD calculated from non-tissue-specific drivers genes showed no concordance with tumor response as impacted by rodent strain and sex

MAD in MF for F_{344}, Wistar Han, and SD rats correlated with historical control bioassay data (n = 10)

Mammary tumor drivers

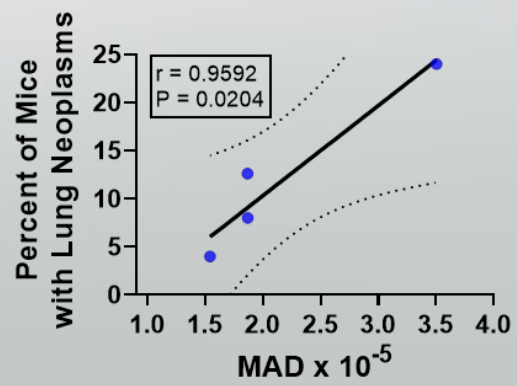


Non-mammary tumor drivers

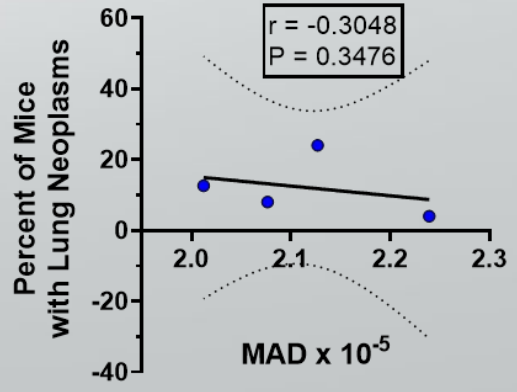


MAD in MF for male and female CD1 and $B6C_3F_1$ mice correlated with historical control bioassay data (n = 10)

Lung tumor drivers



Non-lung tumor drivers



Technical Characterization

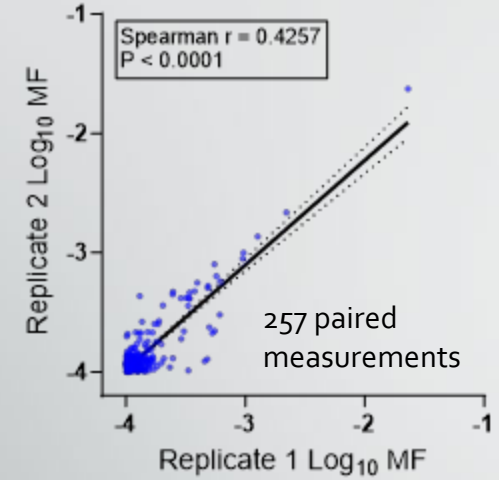
A. How have the sources of variability (e.g., interference, culture conditions, technique, contaminants) been evaluated?

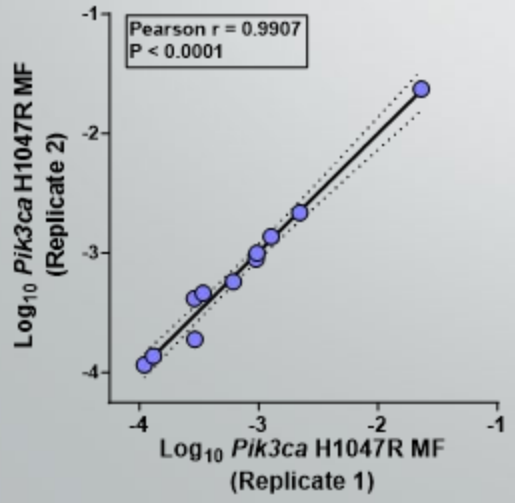
- Independent CarcSeq analyses of rat mammary samples treated with lorcaserin for 24 weeks
- Examined concordance between CarcSeq measured MF and measurement of the same MFs using orthogonal methods, allele-specific competitive blocker PCR (ACB-PCR) and Duplex Sequencing (DS)

Technical Characterization

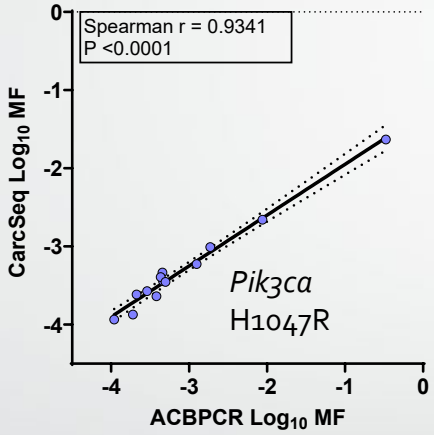
C. How has intra-laboratory reproducibility (i.e., the consistency of individual test results obtained within a laboratory using the same test protocol and test samples) been evaluated?

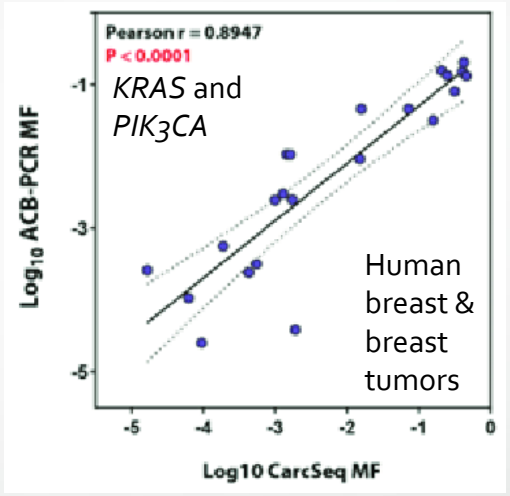
Replication of CarcSeq MF measurement

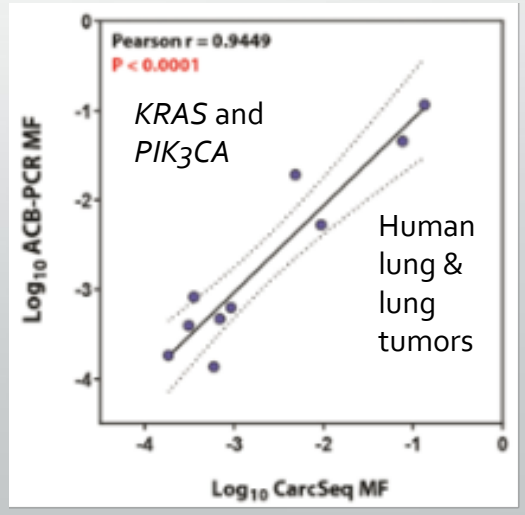


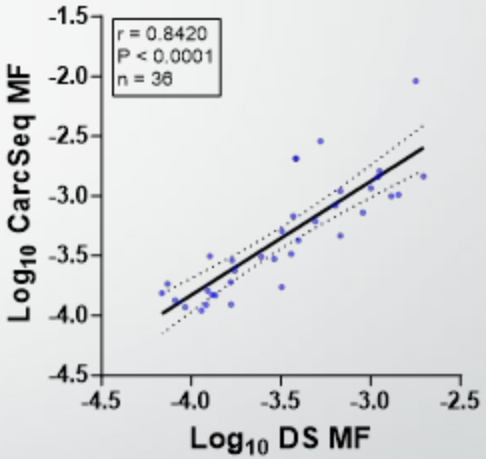


Concordance between CarcSeq and ACB-PCR or Duplex Sequencing









MF in normal human breast measured by DS in collaboration with TwinStrand (Jake Higgins, Fang Yin Lo, Mike Hipp, Lindsay Williams, Clint Valentine, and Jesse Salk

Technical Characterization

D. How has transferability (i.e., the ability of the method to be accurately and reliably performed in different, competent laboratories) been evaluated (if relevant)?

- The CarcSeq method has not been reproduced in another laboratory
- Open to collaborations in the analysis of samples or performing CarcSeq elsewhere
- Consulting on a project led by Drs. Eva Turley and Kathleen Hill (University of Western Ontario) to use CarcSeq to measure *Trp53* mutations in mouse skin
- A manuscript including detailed lab protocols is planned
- Access to the CarcSeq bioinformatic pipeline as a Docker container on GitHub may be provided upon request

Closing/Contact

Acknowledgements

Jennifer Faske

For her exceptional skill and had work in producing the lorcaserin and B[b]F data

Binsheng Gong

For sharing his bioinformatic expertise

Todd Bourcier

For consultation on lorcaserin study design and context of use

Thank you for your attention

barbara.parsons@fda.hhs.gov

Method Description: CarcSeq details

- CarcSeq amplicon panels have been developed for human, rat, and mouse. In each case, ~1,000 cancer driver gene bases are sequenced
- Interrogated bases encompass human hotspots for mutation or their homologues
  - *Apc* R1450, *Braf* V600, *Egfr* T790/L858, *Hras* G12/G13/Q61, *Kras* G12/G13/Q61, *Nfe2l2* E79/E82, *Pik3ca* E542/E545/H1047, *Setbp1* D868/G870/I871, *Stk11* F354, and *Tp53* R175/R248/R273
- The input genomic DNA isolated from otherwise unprocessed frozen tissue samples (human, rat, or mouse). Each PCR amplification uses 0.5 µg of genomic DNA, so 7 µg would be needed if 14 amplicons were interrogated.
- The output of CarcSeq is mutant fraction (MF), the number of mutants (of a particular type, e.g. mutant base A) at a given position divided by the total number of molecules characterized at that position (including all mutant and wild-type bases)
- These hotpots generally represent only 1-2% of all the mutations in a driver gene containing a hotspot. However, mutation in the surrounding bases occurs frequently
- Roughly, 75% of mutants recovered by CarcSeq are non-synonymous; ~25% are synonymous, which is in good agreement with 76.6% of human CDMs in the COSMIC database being non-synonymous and 23.4% synonymous [Sharma et al. (2019) *Nat Commun* 10:2569]
- Some synonymous mutations confer a selectable phenotype
  - Human synonymous CDMs are reported in SynMICdb, the Synonymous Mutations in Cancer database, https://synmicdb.dkfz.de/rsynmicdb/
  - Rodent homologues of human mutations in the database have been recovered using CarcSeq