SUPPLEMENTAL METHODS

Patient recruitment

All adult and pediatric cancer patients presenting to our institutions are invited to participate in the Cancer Research Study. Upon receipt of a test requisition and signed informed consent from the patient, samples are selected and analyzed in a CLIA-certified laboratory- the Center for Advanced Molecular Diagnostics (CAMD) at BWH. There were no restrictions imposed on the cancer type (both solid tumors and hematologic malignancies) to be tested, the stage or grade of disease (however in situ lesions were excluded), course of disease, or whether a primary or a metastatic sample was profiled.

OncoPanel

Eligible samples included formalin-fixed paraffin-embedded (FFPE) tissue sections or cytologic cell blocks, freshly frozen tissue or cell pellets, fresh peripheral blood or bone marrow, and slides containing smear preparations. Tumor content was estimated by an anatomic pathologist from corresponding stained slides, when available, or by cell count/flow cytometry for fresh blood or bone marrow samples. Samples with at least 20% malignant cells were analyzed. For tissue sections, including from surgical resections, biopsies, and cytology cell blocks, regions of adequate cancer measuring at least 3 mm in greatest linear dimension were either manually dissected off corresponding unstained sections or cored directly from the paraffin block. Paired germline samples were not analyzed. DNA was isolated with a commercial kit (Qiagen, Valencia, CA), following the manufacturer's instructions. DNA was quantified (PicoGreen, ThermoFisher Scientific, Waltham, MA), and 200 ng of DNA was used for library preparation (with a low input threshold of 50ng).

Hybrid-capture libraries were prepared as described previously (14, 55). DNA libraries were hybridized to a set of custom-designed capture probes targeting the full coding regions of 275 genes plus selected intronic regions of 30 genes, for a total of 282 unique genes/regions (**Supplemental Table 1**). This

report is focused on testing from August 2013 to July 2014; a new version of the test was implemented at the end of this period. Sequencing was performed using an Illumina HiSeq 2500 with 2x100 paired-end reads to a mean target coverage of 187X unique, high quality, mapped reads per sample (range 50 to 844X; 50X minimum required to pass).

In order to confirm patient identity and eliminate sample mix-up and cross-contamination, specimens were genotyped across 48 single nucleotide polymorphisms (SNPs) selected for 45 - 55% heterozygosity. Briefly, stock DNA was PCR amplified across each of the 48 SNPs follow by removal of primers and unincorporated nucleotides by SAP treatment. SNP identity at each of the targeted sites was resolved by single base extension and analysis by MALDI-TOF mass spectrometry using the MassArray System (Agena Bioscience). Genotyping results were correlated with results obtained during OncoPanel sequencing. Passing criteria required >80% concordance across genotyped loci between the two technologies. For specimens failing to meet 80% concordance with the fingerprinting assay, additional investigation is performed to identify similarly contaminated specimens and/or similar genomic profiles in a batch. Barring evidence of laboratory contamination, which will typically trigger specimen failure, additional investigation into the patient's history may be performed to determine if there is a clinical basis for chimerism (i.e. prior hematopoietic stem cell transplantation).

Data analysis

Data was analyzed by an internally-developed bioinformatics pipeline composed of reconfigured publically-available tools and internally-developed algorithms (VisCap Cancer, Phaser, BreaKmer(55)). Pooled sample reads were demultiplexed using Picard (http://broadinstitute.github.io/picard/command-line-overview.html), aligned to Human Genome Reference Consortium reference sequence GRCh37p13 (BWA(56)) and duplicate reads removed. GATK(57) was used to refine alignments around insertion/deletion (indel) sites and to perform quality

score recalibration. Single nucleotide variants (SNVs) were called using MuTect(58) and indels using Indelocator (http://www.broadinstitute.org/cancer/cga/indelocator). Annotation was performed using Oncotator(59). Multiple SNVs identified in the same codon were further analyzed using the GATK ReadBackedPhasing tool to determine if they represent a di- or tri-nucleotide change.

Because tumor tissues were tested without a paired normal from individual patients, additional informatics steps were taken to identify common single nucleotide polymorphisms (SNPS): any SNP present at >0.1% in Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/ accessed May 30, 2013) or present in dbSNP was filtered, however variants also present in at least twice in COSMIC were rescued for manual review. Variants that appeared two or more times in a panel of 150 normal samples sequenced in-house and are not present in COSMIC are also filtered. Following computational filtering, manual interrogation by a molecular pathologist was applied to consider tumor purity, ploidy and estimated allele fraction at each site, along with contextual information such as tumor type.

VisCap Cancer calls copy number changes based on log2 ratios that are calculated using a normalized depth of coverage against a median from a panel of normal (non-cancer) samples. Circular binary segmentation (CBS) is used to segment the data; segments are called via strict thresholding; the algorithm is configurable for different tumor fractions. Estimated (approximate) copy number was calculated using the ABSOLUTE method(60) and the formula CR = (TC/2)*P + (2/2)(1-P), where CR = median copy ratio for all intervals in a gene; TC = number of copies in the tumor sample; P = Purity. The OncoPanel assay does not cover a sufficient number of polymorphic sites in the genome to employ a computational algorithm (such as ABSOLUTE) to estimate tumor purity, therefore purity was calculated based on pathologist review. Pathology-based determination of tumor purity has been shown to have

limited precision between laboratories but reasonable accuracy; this approach tends to overestimate the tumor content(61), suggesting that our approximate ploidy calculation may underestimate the true copy numbers.

Unique, aligned (hg19) sequence reads with PHRED>30 were reviewed, annotated, and interpreted using Integrated Genome Viewer(62) and a suite of internally-developed Web-based tools. (58) A lower threshold of 50X mean target coverage was established for mutation calling- as previously reported, at loci with at least 50 high quality (Q20) reads, the power to detect a 15% allele fraction event is 99%(58).Samples with a mean target coverage of <50X were failed and excluded from further analysis. Individual variants present at <10% allele fraction (AF) or in regions with <50X coverage were flagged for manual review and interpreted by the reviewing laboratory scientists and molecular pathologists based on overall tumor percentage, read depth, complexity of alteration, and evidence for associated copy number alterations. These thresholds were established during assay validation to differentiate between absolute and partial precision in replicate experiments (data not shown). Hypermutated tumors were defined as those that, following filtration for germline variants, fell into the 95th percentile for median number of SNVs and/or indels per sample (>18).

Comparison to existing diagnostic techniques

For many samples, clinical analysis of selected genes was performed during the course of the patient's management. Single gene assays are performed by CLIA-validated methods using orthogonal technologies such as Sanger sequencing (*KIT, PDGFRA, ABL1*), pyrosequencing (*KRAS, BRAF, NRAS, NOTCH1*), allele-specific PCR or RT-PCR (*EGFR* codon 858, *KIT* codon 816, *JAK2* codon 617, *MYD88* codon 265, BCR-ABL1 b2a2 and b3a2), electrophoresis (*EGFR* exon 19, BCR-ABL1, PML-RARA, *IGH-BCL2*), as well

as FISH, array comparative genomic hybridization (OncoCopy(28)) and karyotype analysis for a broad array of structural and copy number variants. Individual methods are available upon request. Where available, these traditional analyses were compared with OncoPanel results.

Technical validation

Technical validation experiments were performed to determine precision, accuracy, sensitivity, specificity and limit of detection, using blood, fresh frozen, and FFPE samples that had existing genomic characterization using an orthogonal clinical test (e.g., OncoMap, pyrosequencing, Sanger sequencing, PCR/electrophoresis, real-time PCR, aCGH array, karyotype or FISH). 120 samples with known somatic mutations, indels, copy number alterations and/or rearrangements were selected. Normal (non-cancerous) FFPE liver and fresh blood was included to verify detection of wild type loci and to identify batch related sequencing artifacts. In order to obtain sufficient quantities of DNA, several isolations were performed from each sample, pooled to ensure sample homogeneity, and then aliquoted for use in validation.

Lower limit of detection was based on analysis of precision for 12 samples that were diluted with non-neoplastic DNA at 100% (no addition of diluent), 50%, and 20% sample. Each sample was run in triplicate at each dilution. For variants with coverage $\geq$50X and allele fraction $\geq$10%, alterations were detected across all three triplicate runs for 1309 of 1330 variants (98.4% concordance). With $\geq$50X coverage and $\geq$10% allele frequency, we saw 100% reproducibility (1014/1015 variants; **Supplemental Table 8**) across 160 genes identified in 10 samples. Clinical sensitivity and specificity across the 120 sample cohort was determined. For SNVs, OncoPanel was 97.8% sensitive (95% CI: 86.5 - 99.9%) and 100% specific (95% CI: 99.9 - 100%; **Supplemental Table 9**); for indels 97.7% sensitive (95% CI: 86.5 - 99.9%) and 100% specific (95% CI: 99.9 - 100%). For copy number changes (combining single copy gains, amplifications and deletions), the assay was 93% sensitive (95% CI: 87.1 - 96.4%) and 97.6% specific

(95% CI: 96.1 - 98.5%). For rearrangement detection, OncoPanel was 74% sensitive (95% CI: 53.4% - 88.1%) and 100% specific (95% CI: 69.9% - 100%). Six of 7 non-concordant rearrangement results mapped to IGH rearrangements. This is not unexpected as our assay covers only a fraction of the immunoglobulin loci that have been implicated in rearrangements.

Statistics

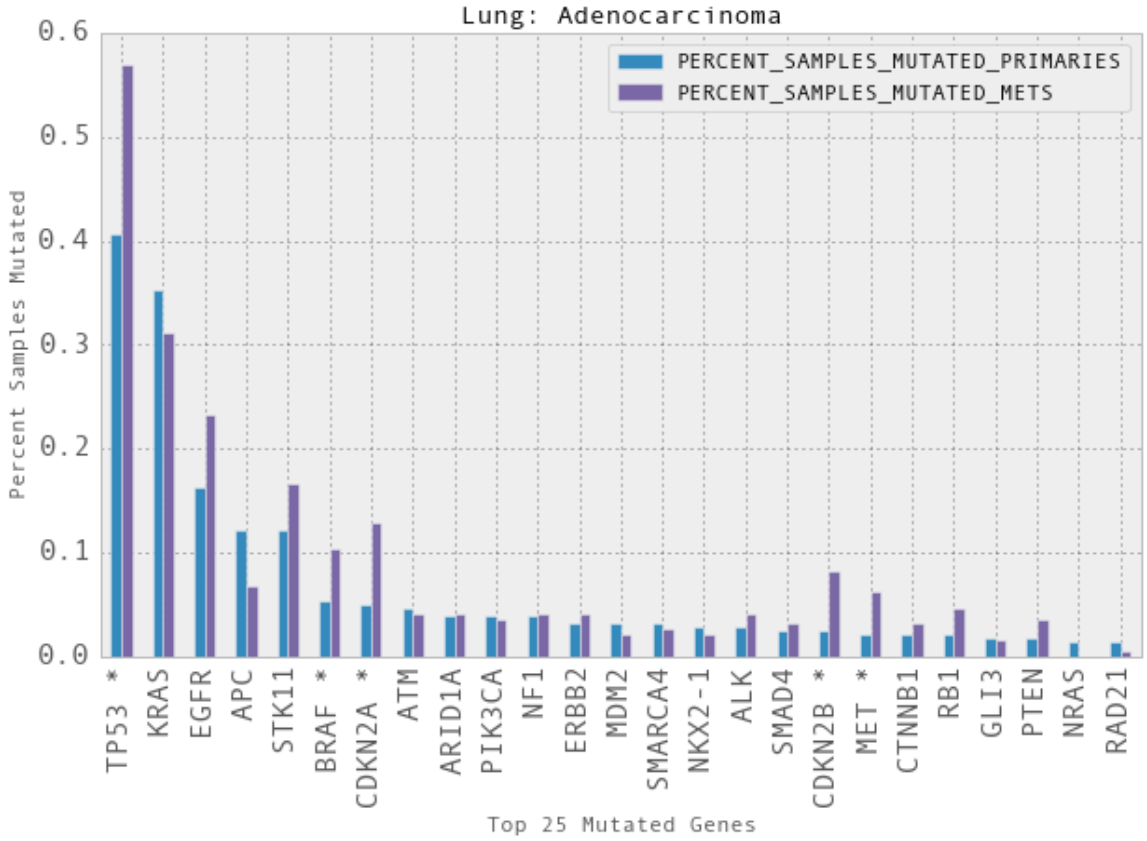
Categorical comparisons were performed using Fisher's exact or Chi square tests with Bonferroni correction for multiple comparisons. Sample means were compared using Student's T test assuming equal variance. P values of <0.05 were considered significant. Sensitivities and specificities with 95% confidence intervals were calculated using a publicly available clinical calculator (vassarstats.net). Confidence intervals were calculated using the Hmisc library in R using the binconf() function set to the "Wilson" calculation method.

Clinical use

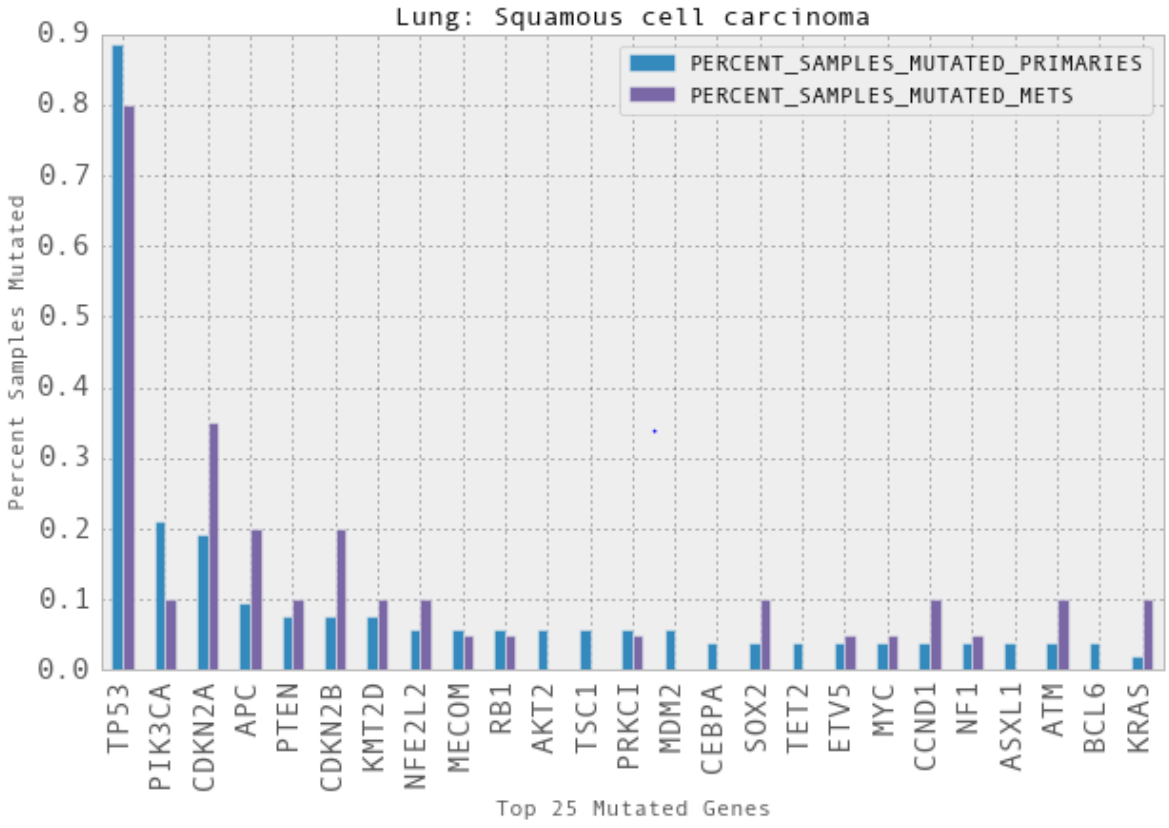
The resulting sequence data are analyzed and classified into 5 tiers of clinical relevance by a team of pathologists with molecular diagnostics expertise, incorporating information from each patient's electronic health record, and then provided to the patient's treating physician(s) as part of this study. Genomic, pathologic, and clinical data is deposited in a central knowledgebase which is linked to a database containing full clinical annotation. The knowledgebase can be queried by oncologists and laboratory scientists to facilitate development and enrollment of basket trials, inform tumor board discussions, and hone variant interpretations.

Supplemental Figure 1. Frequency of alterations in primary versus metastatic tumors among the top 25 most commonly altered tier 1-3 genes in selected diseases. Starred comparisons are significant by uncorrected Chi square test. Comparisons that remain significant after correction for multiple comparisons are noted in the results text.

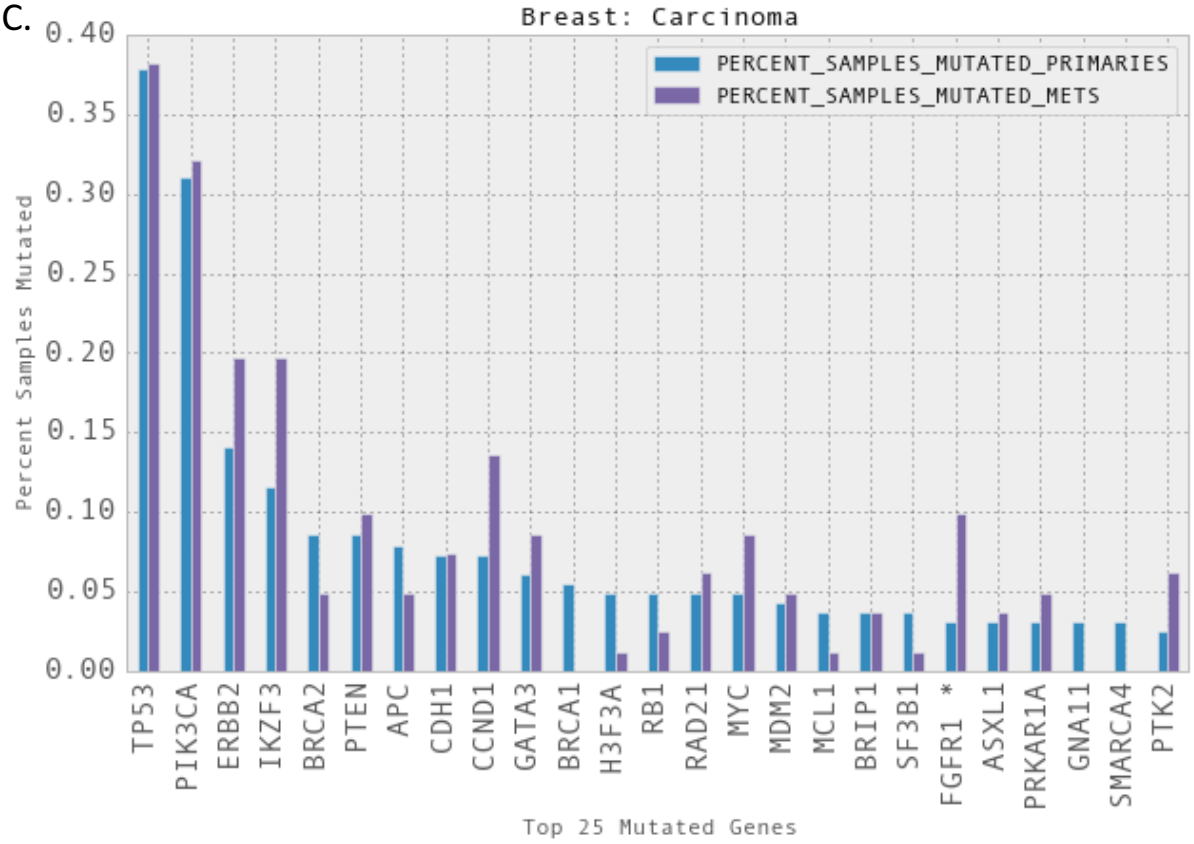
A.



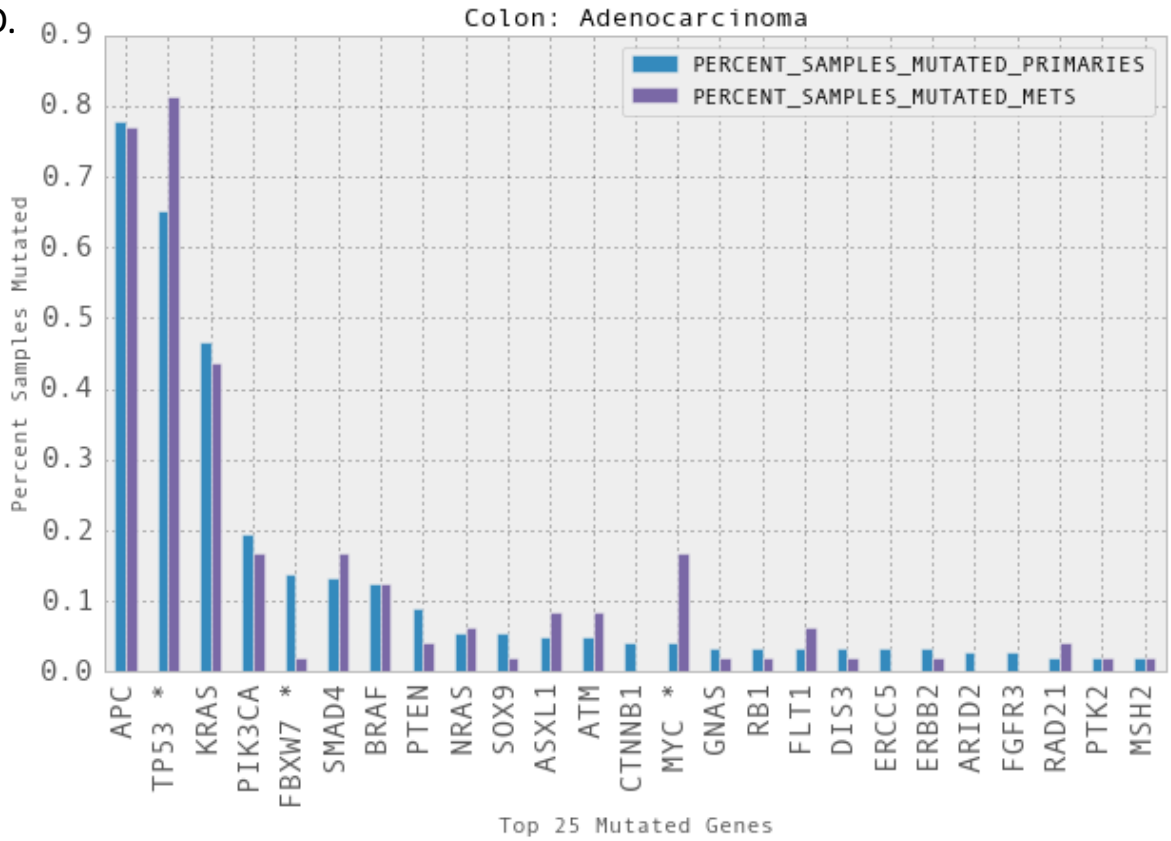
B.

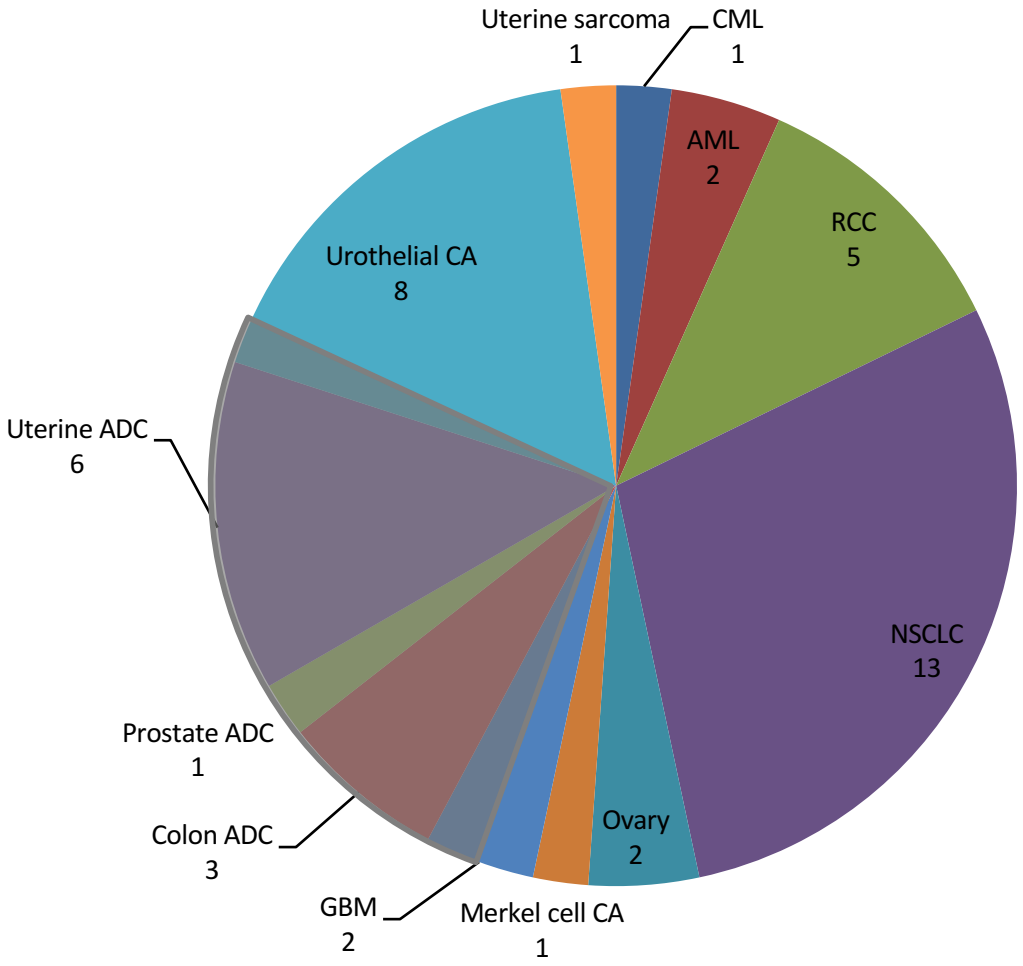


C.


Breast: Carcinoma
PERCENT_SAMPLES_MUTATED_PRIMARIES
PERCENT_SAMPLES_MUTATED_METS
Percent Samples Mutated
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
TP53
PIK3CA
ERBB2
IKZF3
BRCA2
PTEN
APC
CDH1
CCND1
GATA3
BRCA1
H3F3A
RB1
RAD21
MYC
MDM2
MCL1
BRIP1
SF3B1
FGFR1 *
ASXL1
PRKAR1A
GNA11
SMARCA4
PTK2
Top 25 Mutated Genes

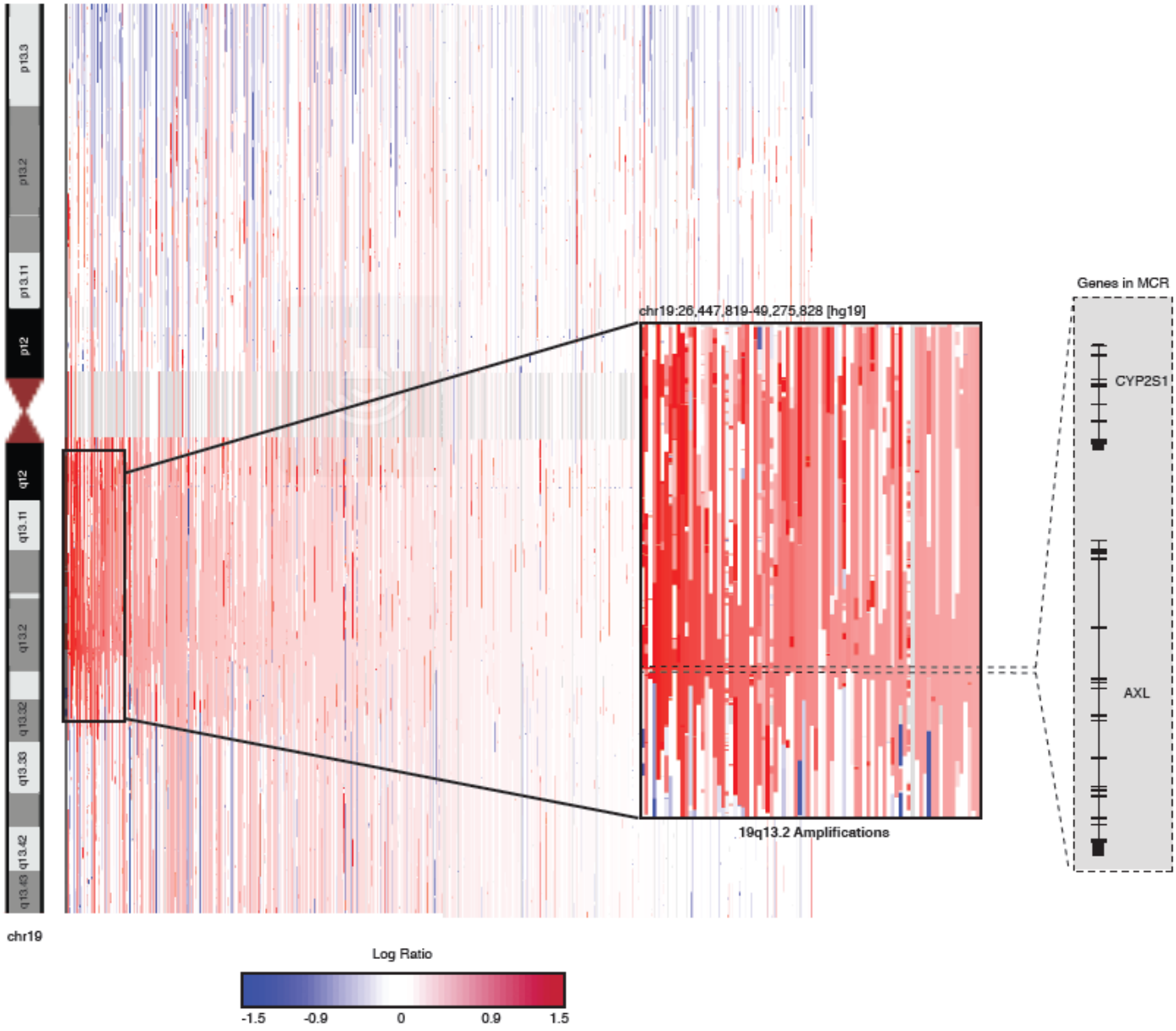
D.


Colon: Adenocarcinoma
PERCENT_SAMPLES_MUTATED_PRIMARIES
PERCENT_SAMPLES_MUTATED_METS
Percent Samples Mutated
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
APC
TP53 *
KRAS
PIK3CA
FBXW7 *
SMAD4
BRAF
PTEN
NRAS
SOX9
ASXL1
ATM
CTNNB1
MYC *
GNAS
RB1
FLT1
DIS3
ERCC5
ERBB2
ARID2
FGFR3
RAD21
PTK2
MSH2
Top 25 Mutated Genes



Tumor types	N (% of Oncopanel cohort)
All	43 (1.1)
Non-hypermutated	33 (0.9)

Supplemental Figure 2. Tumors with loss of function TSC1 or 2 variants. The case types containing at least one loss of function TSC1 or TSC2 variant are shown in the pie chart. The grey wedge overlay demarcates the tumors with evidence for hypermutation.



Supplemental Figure 3. The patterns of chromosome 19 copy number alterations across all tumor types tested by The Cancer Genome Atlas are shown. Individual tumors are arranged in columns, and sorted according to degree of gain at the *AXL* locus. In most cases containing *AXL* gain, the amplicon extends to include other genes located 19q. The minimally amplified region is shown in the inset.