Cis-Compound Mutations are Prevalent in Triple Negative Breast Cancer and Can Drive Tumor Progression

Nao Hiranuma[1], Jie Liu[2], Chaozhong Song[3,4,5], Jacob Goldsmith[5], Michael O. Dorschner[3,6], Colin C. Pritchard[3,7], Kimberly A. Burton[3,4,5], Elisabeth M. Mahen[3,4,5], Sibel Blau[3,8], Francis M. Senecal[3,8], Wayne L. Monsky[9], Stephanie Parker[8], Stephen C. Schmechel[3,6], Stephen K. Allison[9], Vijayakrishna K. Gadi[10,11], Sophie R. Salama[12], Amie J. Radenbaugh[12], Mary Goldman[12], Jill M. Johnsen[13], Shelly Heimfeld[14], Vitalina Komashko[15], Marissa LaMadrid-Hermannsfeldt[5], Zhijun Duan[3,4,5], Steven C. Benz[16], Patrick Soon-Shiong[16], David Haussler[12], Jingchun Zhu[3,12], Walter L. Ruzzo[1,2,3], William S. Noble[1,2,3], and C. Anthony Blau[2,3,4,5]

[1]Computer Science and Engineering, University of Washington, Seattle, WA
[2]Department of Genome Sciences, University of Washington, Seattle, WA
[3]Center for Cancer Innovation, University of Washington, Seattle, WA
[4]Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA
[5]Department of Medicine/Hematology, University of Washington, Seattle, WA
[6]Department of Pathology University of Washington, Seattle, WA
[7]Department of Laboratory Medicine, University of Washington, Seattle, WA
[8]Northwest Medical Specialties, Puyallup and Tacoma, WA
[9]Department of Radiology, University of Washington, Seattle, WA
[10]Department of Medicine/Oncology, University of Washington, Seattle, WA
[11]Seattle Cancer Care Alliance, Seattle, WA
[12]University of California at Santa Cruz, Santa Cruz, CA
[13]Bloodworks Northwest, Seattle, WA
[14]Fred Hutchinson Cancer Research Center Seattle, WA
[15]Trialomics LLC Seattle, WA
[16]Nantomics Inc., Culver City, CA

Correspondence to: Dr C. Anthony Blau, 850 Republican St., Box 358056, University of Washington, Seattle, WA 98109, USA tblau@uw.edu

About 16% of breast cancers fall into a clinically aggressive category designated "triple negative" (TNBC) due to a lack of ERBB2, estrogen receptor and progesterone receptor expression[1-3]. The mutational spectrum of TNBC has been characterized as part of The Cancer Genome Atlas (TCGA)[4]; however, snapshots of primary tumors cannot reveal the mechanisms by which TNBCs progress and spread. To address this limitation we initiated the Intensive Trial of OMics in Cancer (ITOMIC)-001, in which patients with metastatic TNBC undergo multiple biopsies over space and time[5]. Whole exome sequencing (WES) of 67 samples from 11 patients identified 426 genes containing multiple distinct single nucleotide variants (SNVs) within the same sample, instances we term "Multiple SNVs affecting the Same Gene and Sample" (MSSGS). We find that >90% of MSSGS result from cis-compound mutations (in which both SNVs affect the same allele), that MSSGS comprised of SNVs affecting adjacent nucleotides arise from single mutational events, and that most other MSSGS result from the sequential acquisition of SNVs. Some MSSGS drive cancer progression, as exemplified by a TNBC driven by FGFR2(S252W;Y375C). MSSGS are more prevalent in TNBC than other breast cancer subtypes and occur at higher-than-expected frequencies across TNBC samples within TCGA. MSSGS may denote genes that play as yet unrecognized roles in cancer progression.

The Intensive Trial of OMics in Cancer (ITOMIC)-001 (ClinicalTrials.gov Identifier: NCT01957514) enrolls patients with metastatic TNBC who have not received platinum based therapies and are scheduled to receive cisplatin[5]. Up to seven different metastatic sites are biopsied prior to cisplatin, following discontinuation of cisplatin, and following subsequent therapies. Additional samples are accessed as archival tissues, as leftovers following clinically indicated procedures, and from tissues taken at autopsy. Samples are selected for sequencing based on specimen size and tumor content (**Fig. 1a**). Results from 11 of the first 12 subjects are presented here because low tumor purities in Subject 08 precluded analysis. We performed WES of germline DNA and 67 tumor samples (**Extended Data Table 1**), including transiently cultured cells derived from a malignant pleural effusion in Subject 01 and two highly enriched samples of circulating tumor cells (CTCs) obtained following leukapheresis in Subject 02[5]. Whole genome sequencing (WGS) was also performed in 39 samples. The number of assessed samples per patient ranged from 2 to 16 (median 7), and the number of assessed time points ranged from 1 to 6 (median 3) (**Table 1**).

Our analysis focused on somatic single nucleotide variants (SNVs), which comprise the majority of mutations in breast cancer[4,6,7]. WES identified a total of 8449 SNVs, of which 7067 occurred within 5136 protein coding genes, including 43 breast cancer driver genes (0.84%)[7] (data not shown). Across all 67 samples, 1403 genes were affected by >1 SNV. Remarkably, 426 genes were found to contain >1 SNV within the same tumor sample (**Extended Data Table 1**), and we designated these instances Multiple SNVs affecting the Same Gene and Sample (MSSGS) (**Fig. 1b**).

MSSGS were observed in 65 of 67 samples, with a median of 8 and a range of 0 to 133 MSSGS per sample **(Extended Data Table 1)**. The distribution of median transcript sizes was no larger for MSSGS than for genes affected by single SNVs[8] (**Extended Data Fig. 1**). Concomitant WGS in 39 of 65 samples independently confirmed 48 of

150 evaluable MSSGS (32%), with most of the unconfirmed cases being attributable to the relatively lower read counts associated with WGS (**Extended Data Table 2 and Supplementary Methods**), although a fraction of MSSGS (11% - 15%) may represent artifacts associated with WES (**Supplementary Methods**). While somatic variants are unevenly distributed across cancer genomes[8], we found similar mutation densities in regions surrounding MSSGS compared to isolated SNVs (**Fig. 2a**) with the exception of an MSSGS involving *MUC4*, which fell within a region of kataegis[9] (**Fig. 2b**).

MSSGS can arise when two or more SNVs affect: i) the same gene in different tumor cells, ii) different alleles in the same cell, or iii) the same allele in the same cell (cis-compound mutation) (**Fig. 1c**). To assess the frequency with which SNVs contributing to MSSGS co-localize to form a cis-compound mutation we cloned and sequenced tumor DNA from eight patients, revealing a very high frequency of cis-compound mutations (12 of 13 evaluable MSSGS - 92%) (**Table 2**). Haplotype phasing[10] of 407 MSSGS similarly indicated that >90% were attributable to cis-compound mutations (**Extended Data Table 3**). This tendency for SNVs contributing to MSSGS to affect the same allele may result from a DNA strand bias in susceptibility to point mutations[11], or from a selection of genes bearing cis-compound mutations.

Assessing SNVs across different tumor samples from the same patient provides insight into the spatial and temporal origins of MSSGS. 130 of 426 MSSGS (~30.5%) were detected in at least two tumor samples from the same patient. Among these repeatedly detected MSSGS, three contiguous SNVs affecting *ZZZ3* were all detected in the same 4 of 7 samples from Subject 02 at nearly identical frequencies, suggesting their presence in the same cells (**Extended Data Table 1)**, and a similar pattern was observed for *ADAM33*, which was detected in the same 3 of 16 samples from Subject 01 (**Fig. 3a, Extended Data Table 1**). These patterns align with results from cloning and sequencing these MSSGS fragments (**Table 2**), which showed that either all SNVs contributing to the cis-compound mutation were observed or none were observed, consistent with their emergence from single mutational events.

Other MSSGS result from the sequential acquisition of SNVs, as seen for *FGFR2*(S252W;Y375C), in which S252W appeared after neoadjuvant chemotherapy and both S252W and Y375C were found in a metastatic lymph node 2 years later, persisting in all nine subsequent samples (**Fig. 4b, Extended Data Table 1**). In yet other cases contributing SNVs made separate and sporadic appearances prior to detection of the complete MSSGS, as exemplified by an MSSGS involving *F8* in Subject 01 (**Fig. 3a, Extended Data Table 1**).

Assessing allelic frequencies for all 426 MSSGS across 65 tumor samples revealed two distinct patterns. For the 71 MSSGS containing SNVs that affect consecutive nucleotides (including the above-referenced *ADAM33* and *ZZZ3*) we observed near perfect concordance in allelic frequencies between participating SNVs across tumor samples (**Fig. 3b, Extended Data Table 1**), consistent with their emergence from single mutational events (r=0.98). In contrast, concordance in allelic frequencies between SNVs separated by 1 or more nucleotides was significantly lower (r=0.27) (**Fig. 3c, Extended Data Table 1**). These findings indicate that the large majority of MSSGS containing immediately adjacent SNVs arose as single mutational events, whereas most other MSSGS were built from SNVs that arose separately.

Prior knowledge supports the functional significance of some MSSGS. For example, both missense mutations involving FGFR2(S252W;Y375C) (**Fig. 4a**) modify ligand binding[12,13]. FGFR2 S252W and Y375C have been separately associated with the congenital disorders Alpert Syndrome[14] and Beare-Stevenson Syndrome[15]*,* respectively, and while each mutation has also been separately associated with endometrial cancer and FGFR2(S252W;Y375C) has been observed in a single case of endometrial cancer, neither mutation has been reported previously in breast cancer. To further evaluate *FGFR2*(S252W;Y375C) we extracted RNA from transiently cultured breast cancer cells from Subject 01, performed reverse transcriptase - PCR, and sequenced the cloned PCR fragments. Among 13 clones examined, 6 contained both mutations, 5 contained S252W only, and 2 were wildtype (data not shown). These results are consistent with the temporal acquisition of these SNVs, in which the S252W mutation arose first and Y375C was acquired later (**Fig. 4b, Extended Data Table 1**). A drug susceptibility screen using the same transiently cultured breast cancer cells demonstrated a much greater susceptibility to the FGFR2 inhibitor, ponatinib, compared to 19 TNBC cell lines (**Fig. 4c**), and ponatinib treatment in the patient produced a regression of breast cancer infiltrates in the skin (**Fig. 4d**). Possibly indicative of a growth advantage conferred by other MSSGS are results from Table 2, which show that for MSSGS involving *ABHD16Ac*, *TMPRSS13* and *ROBO4*, sequences bearing the entire cis-compound mutation are more abundant than sequences containing only one of the component SNVs.

We reasoned that if cis-compound mutations can confer a selective advantage in TNBC, MSSGS might be more prevalent than would be predicted from a random co-localization of SNVs. To address this question we evaluated 106 TNBC samples from the TCGA. **Fig. 5s** shows the number of SNVs contributing to MSSGS against the total SNV count for each of 106 TNBC samples from TCGA[4]. We also plotted a theoretical curve under the null hypothesis that MSSGS arise from the random co-localization of SNVs (see **Supplementary Methods**). The assumptions inherent in generating this theoretical curve exert a greater impact at high SNV counts; therefore, in **Fig. 5b** we exclude the 5 samples containing >600 SNVs, and focus on the remaining 101 samples carrying fewer than 200 SNVs. As expected, the number of SNVs contributing to MSSGS increases in accordance with the total number of SNVs. However, **Fig. 5b** also shows that the number of observed SNVs contributing to MSSGS tends to exceed levels predicted by the theoretical curve. These results indicate that for many samples, the number of SNVs contributing to MSSGS is higher than predicted had they arisen independently. To further examine this phenomenon, we performed a permutation test for each of the 106 patients to test the null hypothesis that SNVs co-localize to form MSSGS randomly (details are provided in Methods). With the 106 resulting p-values, we accept six patients when we use the Benjamini-Hochberg (BH) procedure[16] to control the false discovery rate at 0.05 (**Fig. 5b**).

Finally, we assessed the frequency of MSSGS in 993 breast cancer samples from TCGA. While there were no significant differences in the mean or median numbers of MSSGS per sample (data not shown), nor in the fraction of MSSGS per SNV, a higher percentage of TNBCs contained at least 1 MSSGS (74.5%) compared to HER2+ breast cancers (61.5%), or ER+ breast cancers (49.7%) (**Fig. 5c-f**).

Evidence presented here supports the involvement of MSSGS in tumor growth based on our findings that: i) The large majority of MSSGS are attributable to cis-compound mutations; ii) FGFR2(S252W;Y375C) has unambiguous functional significance; iii) sequences bearing the entire cis-compound mutation are frequently more abundant than sequences bearing only one of the component SNVs; and iv) MSSGS are detected at significantly higher frequencies than would be predicted from the random co-localization of SNVs in subsets of patients with TNBC. MSSGS lack many features of conventional driver mutations because they are not widely detected across tumors from different individuals, are typically present at sub-clonal levels, and are frequently not associated with cancer initiation but rather cancer progression. Some of the MSSGS described here are unlikely drivers of cancer progression, exemplified by *MUC4*, which resides in a region of increased mutation density (**Extended Data Fig. 2**).

Epistastic interactions between missense mutations can enable proteins to evolve new conformations and functions[17,18]. Cis-compound mutations associated with gain-of-function[19,20] and loss-of-function[21] underpin inherited disorders such as Multiple Endocrine Neoplasia Type 2B[19], long QT syndrome[21], and familial Alzheimer's Disease[20]. Cis-compound mutations have also been described in cancer[22-32], provide an important mechanism of resistance to BCR-ABL targeted tyrosine kinase inhibitors[29-31], and can promote the re-acquisition of sensitivity in previously crizotinib resistant, ALK-rearranged lung cancer[32]. Driver mutations arising at metastatic sites of disease face a different set of constraints compared to driver mutations resident in primary tumors. Since cancers typically require years to become clinically evident, driver mutations conferring even a slight selective advantage can gain predominance due to the compounding effect of time[33]. In contrast, driver mutations acquired at metastatic sites of disease have a much shorter timeframe over which to gain a foothold, and the cells against which they compete have demonstrated a high degree of fitness. Therefore, driver mutations acquired at metastatic sites would be expected to be frequently present at subclonal levels. In conclusion, MSSGS appear to provide examples of "intra-molecular epistasis" in the promotion of cancer progression[34].

METHODS

Clinical Trial

ITOMIC-001 was approved by the Solid Tumor Scientific Review Committee and the Institutional Review Board of the Fred Hutchinson Cancer Research Center (FHCRC) and is registered at ClinicalTrials.gov (NCT01957514). Patients enroll from two sites: the Seattle Cancer Care Alliance (SCCA) and Northwest Medical Specialties (NWMS), a private practice with offices in Tacoma and Puyallup, Washington. Eligible subjects have metastatic TNBC, are platinum naive, and are scheduled to receive Cisplatin. Informed consent (Appendix 2) is conducted through detailed in person discussions involving, minimally, the principal investigator, the subject, and a subject family member, and last 1 - 2 hours. Study oversight is provided by a Data Safety Monitoring Board.

Biopsies

Biopsies of metastases involving lymph nodes, subcutaneous tissues, and liver were performed under ultrasound guidance using an 18 gauge BioPince Full-Core Biopsy instrument. Up to 5 disease sites were biopsied in a single setting, and multiple biopsies were performed per disease site. Bone marrow biopsies were taken from the iliac crest using a Jamshidi T-Handle needle. Skin biopsies were performed using 3 - 4 mm punch biopsy instrument. In Subject 2 circulating tumor cells were collected by leukapheresis. All biopsies were performed using local anesthesia and most were performed using conscious sedation. Biopsy specimens were processed in accordance with a standardized set of operating procedures and subjects were contacted one day and one week following the procedure to assess for complications.

Rapid On-Site Evaluation

Core biopsy samples of metastases involving lymph nodes, soft tissues, or liver were immediately photographed and divided orthogonally. One half was formalin fixed (for formalin fixation and paraffin embedding [FFPE]) and the other half was gently pressed against an RNAse treated slide to generate a touch prep, then immediately snap frozen. Cytological evaluation of the touch prep was performed by a pathologist in real time and, if necessary, additional biopsies were procured to optimize sample size and tumor content. The time interval between sample acquisition and the completion of processing (known as the cold ischemia time) was recorded and was, except for the bone marrow biopsies and leukaphereses, less than 5 minutes in all cases (Table 2).

Selection of Biopsy Specimens for Further Testing

FFPE specimens were sectioned, stained with hematoxylin and eosin (H&E), and evaluated by a pathologist for tumor cell content and necrosis. FFPE samples judged to have sufficient tumor content from different metastatic sites were stained for HER2 and ER. To select samples for sequencing we examined H&E sections of the FFPE portion and touch preps of the corresponding snap frozen portion of each biopsy core, with the goal of maximizing tumor cell content. An uneven distribution of tumor cells across biopsy cores sometimes caused discordance in estimates of tumor cell content between the FFPE portion of a sample (assessed by morphology) and the corresponding snap frozen portion (estimated from the touch prep and variant allele frequencies from UW-

OncoPlex).

Residual Clinical Materials

When feasible, leftover blood samples and pleural fluid specimens were processed and stored for analysis. The estimated time between sample collection and processing was recorded.

Autopsy Samples

At the time of this submission, Subjects 1, 2, 4, 5, 6, 7, 8, 11, and 12 had died and Subject 3 was lost to followup. Samples from Subjects 1, 2, 6, 11 and 12 were collected at the time of autopsy and processed as FFPE tissues.

DNA and RNA extraction

DNA and RNA were extracted from samples using AllPrep DNA/RNA/Protein Mini Kit from Qiagen following the manufacturer's protocol. The quality of the RNA was analyzed using the Agilent 2100 Bioanalyzer System.

Whole Exome Sequencing

WES is performed by the UW Northwest Clinical Genomics Laboratory, a CLIA--certified facility at the UW. DNA fragment libraries were constructed from tumor biopsies and blood (normal) using the Hyper Prep Library Preparation Kit (KAPA Biosystems, Wilmington, MA). Each library was enriched for protein and RNA coding portions of the human genome using the SeqCap EZ Exome v3.0 (Roche NimbleGen, Madison, WI) capture system. The target includes all coding content from the CCDS, RefSeq and miRBase databases. Paired-end sequencing (100 bp) of enriched libraries was performed using TruSeq Rapid Sequencing-by-Synthesis chemistry on a HiSeq 2500 (Illumina, San Diego, CA) sequencer according to the manufacturer's recommended protocol. Base calls were generated in real-time on the HiSeq instrument (RTA 1.17.21.3 software). After sequencing was complete, resulting reads were demultiplexed and sample-specific FASTQs produced. The average depth of coverage was 188 and the range was 166 to 208. The resulting reads are aligned to the genome human reference (hg19) using BWA (Burrows-Wheeler Aligner) , duplicate reads removed with Picard and variants called with GATK (Genome Analysis Toolkit) . BAM files are uploaded to DNAnexus (Mountainview, California) for subsequent somatic variant detection by researchers at UC Santa Cruz, UW and Data4Cure (using a proprietary algorithm).

Whole Genome Sequencing

Confirmation of a subset of MSSGs was done by NantOmics (Culver City, CA), which performed whole genome sequencing (WGS) in 39 of the 67 samples described here, using the same DNA that had been tested by WES. WGS Sequencing was performed on the Illumina HiSeq X sequencing platform using libraries prepared via the KAPA Hyper prep kit. Tumor genomes were sequenced to an average depth of 60x, and Normal genomes were sequenced to an average depth of 30x. Mutations were identified using the CLIA-validated NantOmics Contraster pipeline as previously described[35].

Germline Genome

Germline variant detection is performed to identify genes associated with inherited cancer syndromes, such as mutations involving *BRCA1* or *BRCA2*.

Data Management

Clinical data are entered into REDCap, a web--based application for Electronic Data Capture. The Institute of Translational Health Science's installation of REDCap is hosted on a secure server at UWMC. Large datasets that are not well suited for storage on REDCap are stored on the DNANexus platform (https://platform.dnanexus.com).

Clonal Analysis of MSSGS

MSSGS-containing DNA fragments were PCR-amplified using primers flanking all participating SNVs. The amplified PCR fragments were cloned into pBluescript II KS+ vector and transformed into *E. coli*. For each MSSGS, DNA from about 100 colonies was individually prepared and sequenced, and the number of clones containing germline sequence or one or multiple participating SNVs was determined.

Comparing sizes of genes affected by MSSGS versus isolated SNVs

To assess whether the formation of MSSGS is related to gene size, we compared the distributions of the median transcript sizes between genes affected by MSSGs versus genes affected by isolated SNVs. We performed the Wilcoxon rank-sum test to compare the means of the two distributions and found no significant difference (Extended Data Fig. 1). Gene transcript sizes were obtained from Biomart (Ensembl).

Haplotype phasing of SNVs associated with MSSGS

To assess the orientation of neighbouring SNVs within MSSGs, we used ReadBackedPhasing from Genome Analysis Toolkit version 3.3.0. Phasing information with at least 20.0 quality score was used to assess whether a pair of SNVs are cis or trans.

Evaluating the local mutation rates in regions surrounding MSSGS vs isolated SNVs.

To assess whether the formation of MSSGS is driven by elevated local mutation rates, we compared local mutation rates surrounding MSSGS with mutation rates genome wide. In order to calculate the local mutation rate for each MSSGS, we counted the number of SNVs identified by WGS in a 1 megabase window centered on each MSSGS (obtained by subtracting transcriptional stop site by the transcriptional start site, using RefSeq transcripts as a reference). We then calculated the local mutational rate across the genome by windowing each sample into 1MB windows and counting the number of SNVs. A plot of the local mutation rates surrounding MSSGS is depicted by the red circles shown in Figure 2a. Additionally, the distance between each neighboring SNV identified in the WGS samples was calculated to identify regions of kataegis[9], which we defined as at least 10 SNVs in a 100 kilobase window. Only one MSSGS, involving *MUC4*, landed in such a region (Figure 2b).

The theoretical distribution of SNVs contributing to MSSGS

Suppose that a patient carries m SNVs in the tumor cells. Let N (N=17,672) be the total number of genes subject to possible mutation in TNBC, and W be a vector of length N. We assume that each gene can be mutated with a probability proportional to W, and W

= $0.8W_1$ + 0.2W2, where W_1 is the normalized gene mutation frequency in TNBC samples in TCGA and W_2 is the normalized gene mutation frequency in non-TNBC BC samples in TCGA. We include the term $0.2W_2$ because there are only 106 TNBC samples (in which 7,933 genes are mutated), and we want to use W_2 to mimic the unseen genes (which can be potentially mutated in TNBC). In each round of permutation, we randomly assign each of the *m* SNVs to one of the 17,672 genes with a probability proportional to W, and then count SNVs in MSSGS. We repeat this procedure 10,000 times, and each time we obtain a count of the number of SNVs contributing to MSSGS. We then use the median of these 10,000 counts as the expected number of SNVs contributing to MSSGS. The entire simulation procedure is repeated for m from 1 up to 1500, where 1500 is the maximal number of SNVs observed in the TCGA cohort.

To assign a patient-specific P-value, we repeat the above simulation procedure using the observed value of m. We then compare the observed count of SNVs contributing to MSSGS with the 10,000 count values from the simulation. We define the P-value as (t + 1)/10001, where t is the number of times that the observed count of SNVs contributing to MSSGS is less than or equal to the count from simulations.

High Throughput Drug Screen

The drug sensitivity profiles of transiently cultured cells from a pleural effusion from Subject 1 and 19 TNBC lines were compared in a high throughput screen of 160 approved and investigational oncology drugs. Cells were seeded into non-tissue culture-treated 384-well plates. Twelve hours later, compounds were added and after 72 hours viability was assessed. Resulting dose curves were fitted using idbs' XLFit and to a 4

Parameter Logistic Dose Response Model. The TNBC lines tested were BT-20, BT549, HCC1143, HCC1187, HCC1395, HCC1599, HCC1806, HCC1937, HCC2157, HCC38, HCC70, Hs 578T, MDA-MB-157, MDA-MB-436, MDA-MB-231, MDA-MB-453, MDA-MB-468, MFM-223 and SW-527.

ACKNOWLEDGEMENTS

The patients and their families. We also gratefully acknowledge the assistance of Michelle Flores, Melodee Smith, Sherry Littlefield, Laura Connelly-Smith, James Riddle, Ednus Warren, Stan Riddell, Julie Gralow, Bernie McLaughlin, Jennifer Specht, Brent Wood, Katie Dougherty, Anju Thomas, Paul Martin, Corinne Fligner, David Yadock, Jennifer Spokely, Krystal Malhotra, Stephanie Parker, Daniel Campton, Josh Nordberg, Nick Drovetto, Nicole Heying, Leila Ritter and NW BioTrust. This publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000423. Funding for this work was provided by South Sound CARE, the Gary E. Milgard Family Foundation, the Pano Koumantaros Cancer Research Fund, the Washington Research Foundation, the University of Washington School of Medicine, and private donors.

AUTHOR CONTRIBUTIONS

N.H., J.L., S.R.S., A.J.R., M.G., J.M.J., V.K., M.L-M, S.C.B., D.H., J.Z., W.L.R., W.S.N. and C.A.B. contributed to data analysis and interpretation. C.Z., J.G., M.O.D., C.C.P., S.C.S., S.R.S., A.J.R., S.H., and C.A.B. contributed experimental results, K.A.B., E.M.M., S.B., F.M.S., W.L.M., S.P., S.K.A., V.K.G., and P.S.-S. contributed to the clinical trial, and N.H., J.L., Z.D., W.L.R., W.S.N. and C.A.B. wrote the manuscript.

REFERENCES

1. Engebraaten, O., Moen Vollan, H.K., Børresen-Dale, A.-L., Triple-negative breast cancer and the need for new therapeutic targets. *Am. J. Pathol.***183,** 1064-1074 (2013).

2. Turner, N.C. & Reis-Filho, J.S. Tackling the diversity of triple-negative breast cancer. *Clin. Cancer Res.* **19,** 6380-6388 (2013).

3. Mayer, I.A., Abramson, V.G., Lehmann, B.D. & Pietenpol, J.A. New strategies for triple-negative breast cancer--deciphering the heterogeneity. *Clin. Cancer Res.* **20,** 782-790 (2014).

4. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature.* 490,61-70 (2012).

5. Blau, C.A. *et al.* A Distributed Network for Intensive Longitudinal Monitoring in Metastatic Triple-Negative Breast Cancer. *J. Natl. Compr. Canc. Netw.* **14,** 8-17 (2016).

6. Shah, S.P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature.* **486,** 395-399 (2012).

7. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature.* **534,** 47-54 (2016).

8. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* **499,** 214-218 (2013).

9. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell.* **149,** 979-993 (2012).

10. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. **20**, 1297–1303 (2010).

11. Haradhvala, N.J. *et al.* Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell.* **164**, 538-549 (2016).

12. Dutt, A. *et al.* Drug-sensitive FGFR2 mutations in endometrial carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 8313-8317 (2008).

13. Yu, K., Herr, A.B., Waksman,G.& Ornitz, D.M. Loss of fibroblast growth factor receptor 2 ligand-binding specificity in Apert syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 14536-14541 (2000).

14. Wilkie, A.O.M. *et al.* Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome. *Nature Genet.,* **9,** 165–172 (1995).

15. McGaughran, J. *et al.* A case of Beare-Stevenson syndrome with a broad spectrum of features and a review of the FGFR2 Y375C mutation phenotype. *Clin. Dysmorphol.* **15,** 89-93 (2006).

16. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc., Series B* **57,** 289–300 (1995).

17. Ortlund, E.A., Bridgham, J.T., Redinbo, M.R. & Thornton, J.W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317,** 1544-1548 (2007).

18. McKeown, A.N. *et al.* Evolution of DNA specificity in a transcription factor family. *Cell.* **159,** 58-68 (2014).

19. Cranston, A.N. *et al.* RET is constitutively activated by novel tandem mutations that alter the active site resulting in multiple endocrine neoplasia type 2B. *Cancer Res.* **66,** 10179-10187 (2006).

20. Citron, M. *et al.* Mutation of the beta-amyloid precursor protein in familial Alzheimer's disease increases beta-protein production. *Nature.* **360,** 672-674 (1992).

21. Westenskow, P., Splawski, I, Timothy, K.W., Keating, M.T. & Sanguinetti, M.C. Compound mutations: a common cause of severe long-QT syndrome. *Circulation.* **109,** 1834-1841 (2004).

22. Jiang, L. *et al.* Low frequency KRAS mutations in colorectal cancer patients and the presence of multiple mutations in oncogenic drivers in non-small cell lung cancer patients. *Cancer Genet.* **206,** 330-339 (2013).

23. Stone, E. *et al.* High proportion of rare and compound epidermal growth factor receptor mutations in an Australian population of non-squamous non-small-cell lung cancer. *Intern. Med. J.* 44, 1188-1192 (2014).

24. Peng, L., Song, Z., & Jiao, S. Comparison of uncommon EGFR exon 21 L858R compound mutations with single mutation. *Onco. Targets Ther.* 8, 905-10 (2015).

25. Kim, E.Y. *et al.* Compound EGFR mutation is frequently detected with co-mutations of actionable genes and associated with poor clinical outcome in lung adenocarcinoma. *Cancer Biol. Ther.* **19,** 1-9 (2016).

26. Tiong, I.S. *et al.* Apparent 'JAK2-negative' polycythaemia vera due to compound mutations in exon 14. *Br. J. Haematol.* [Epub ahead of print] (2016).

27. Gao, J. *et al*. Heritable GATA2 mutations associated with familial AML-MDS: a case report and review of literature. *J. Hematol. Oncol.* **7**, 36 (2014).

28. Hahn, C.N. *et al.* Characterisation of a compound in-cis GATA2 germline mutation in a pedigree presenting with myelodysplastic syndrome/acute myeloid leukemia with concurrent thrombocytopenia. *Leukemia.* **29,**1795-1797 (2015).

29. Shah, N.P. *et al.* Sequential ABL kinase inhibitor therapy selects for compound drug-resistant BCR-ABL mutations with altered oncogenic potency. *J. Clin. Invest.* **117,** 2562-2569 (2007).

30. Khorashad, J.S. *et al.* BCR-ABL1 compound mutations in tyrosine kinase inhibitor-resistant CML: frequency and clonal relationships. *Blood.* **121,** 489-498 (2013).

31. Zabriskie, M.S. *et al.* BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in Ph chromosome-positive leukemia. *Cancer Cell.* **26,** 428-442 (2014).

32. Shaw, A.T. *et al.* Resensitization to Crizotinib by the Lorlatinib ALK Resistance Mutation L1198F. N. Engl. J. Med. 374, 54-61 (2016).

33. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 18545-18550 (2010).

34. Lehner, B. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27,** 323-331 (2011).

35. Sanborn JZ, *et al.* Double minute chromosomes in glioblastoma multiforme are revealed by precise reconstruction of oncogenic amplicons. *Cancer Res.* 73, 6036-6045 (2013).

LEGENDS FOR FIGURES AND TABLES

Figure 1. Schematic depiction of tumor samples collected and analyzed as part of ITOMIC-001 and description of MSSGS. (**a**) Schematic depiction of biopsies performed (black symbols) and sequenced (red symbols) in 11 of the first 12 subjects enrolled in ITOMIC-001. All treatments during participation in the study are as described. Subject 8 is not depicted because low tumor purities precluded analysis. (**b**) MSSGS are defined by the presence of two or more somatic SNVs occurring within the same gene and tumor sample. (**c**) MSSGS may arise from: SNVs occurring in different populations of tumor cells (left), SNVs affecting different alleles within the same cells (trans-compound mutation) (middle), SNVs affecting the same allele (cis-compound mutation) (right).

Figure 2. MSSGS are not associated with an increase in regional mutation density. (**a**) Y axis indicates numbers of SNVs in 1MB windows across the genome (gray circles) and surrounding MSSGs (red circles) for each sample examined by WGS (X axis). (**b**) Mutational density plot for sample 11-2-B3. The Y axis indicates the distance (in log scale) between two neighboring SNVs, and the X axis indicates the chromosomal position of each SNV-pair. SNVs in regions of kataegis are colored with red circles and the position of *MUC4* is labeled in blue.

Figure 3. MSSGS comprised of SNVs that affect consecutive nucleotides appear abruptly, whereas most other MSSGS develop over time. (**a**) Appearance of SNVs contributing to MSSGs in samples collected over time from Subject 1. Genomic

positions of SNVs are indicated. Colors depict VAFs as indicated in the key. Note that MSSGS comprised of SNVs affecting adjacent nucleotides (arrows) consistently appear together and at very similar VAFs. (**b**) VAFs of SNV pairs involving consecutive nucleotides are highly correlated (r=0.98). Each point denotes an SNV-pair in one tumor sample. (**c**) VAFs of SNV pairs involving non-consecutive nucleotides exhibit a much lower correlation (r=0.27).

Figure 4. A cis-compound mutation drives TNBC progression. (a) Schematic depiction of two missense mutations affecting the FGFR2 extracellular domain. S252W and Y375C are well-characterized germline mutations that cause Alpert and Beare Stevenson Cutis Gyrata Syndrome, respectively. Both mutations have been identified previously in endometrial cancer (S252W - 49 cases; Y375C - 6 cases) whereas neither mutation has been reported previously in breast cancer. The same cis-compound mutation has been reported in a single case of endometrial cancer. (**b**) Appearance of point mutations encoding S252W and Y375C over the disease course of Subject 1. Note that S252W was acquired first, followed by Y375C. (**c**) Cell viability curves in response to a 3.5 log range of ponatinib concentrations. Closed circles show mean and standard deviations of 19 TNBC lines. Open circles indicate the response curve for transiently cultured breast cancer cells from Subject 1. (**d**) Clinical response to ponatinib in Subject 1, with resolution of breast cancer infiltrates in the skin.

Figure 5. MSSGS arise at higher than predicted frequencies in TNBC. (a) Observed numbers of SNVs contributing to MSSGS (black dots) and theoretical

numbers of SNVs contributing to MSSGS (red curve) against the observed numbers of SNVs in TCGA TNBC samples. **(b)** Same as **(a)** but magnified to show results for 101 TNBC patients with <200 SNVs. Green-circled dots are accepted by BH procedure at a FDR level of 0.05. **(c-f)** Ratios of MSSGS/SNVs for each of 993 breast cancer samples contained within TCGA, ordered from high to low. Vertical dashed line divides samples containing 1 or more MSSGS from samples that contain no MSSGS, and the percentages of samples containing at least one MSSGS is indicated. **(c)** all 993 breast cancer samples; **(d)** 722 ER+ breast cancer samples; **(e)** 148 HER2+ breast cancer samples; **(f)** 106 TNBC samples. Note that a higher proportion of TNBC samples contain at least one MSSGS.

Table 1. Numbers of assessed patients, samples and timepoints.

Table 2. Results from cloning and sequencing 15 MSSGS-containing fragments. 12 of 13 evaluable MSSGS were associated with cis-compound mutations. Clonal analysis of 15 MSSGS. Estimated tumor content is based on tumor content in formalin fixed half of each biopsy specimen. CTCs: Circulating Tumor Cells; REF: Number of sequenced DNA fragments containing the reference (non-mutated) sequence; nt: Nucleotide; chr: Chromosome; aa: Amino Acid; * *ZZZ3* contained 3 SNVs, *F8* contained four MSSGS but only two were interrogated. All other MSSGS shown here contained two SNVs; ** *APOBR* could not be evaluated due to sequencing artifacts; ¶

Estimate following column enrichment and FACs sorting of CTCs; § Touch prep of the snap frozen tissue used for sequencing demonstrated numerous tumor cells.

LEGENDS FOR EXTENDED DATA FIGURES AND TABLES

Extended Data Figure 1. CDF plots depicting distributions of median transcript sizes for genes affected by MSSGs versus genes affected by isolated SNVs. Results indicate that there is no significant relationship between gene size and MSSG formation.

Extended Data Table 1. Summary of genes bearing multiple SNVs across 67 tumor samples. Rows indicate gene names, nucleotides at reference and alternate alleles, consequence, position and amino acid substitutions of missense mutations, codon changes, predicted functional consequences using Sift and Polyphen, and chromosomal positions. Columns indicate sample description, number of days post diagnosis, study day, whether tissue was snap frozen or stored as formalin fixed paraffin embedded tissue (FFPE) or both, cold ischemia time, and estimated tumor purity. 3 columns describe each sample (designated 01-0-B1, etc). Ref = number of reads for reference allele; Alt = number of reads for alternate allele; VAF = variant allele frequency. Blank cells denote a value of zero.

Extended Data Table 2. Confirmation of MSSGS by WGS. To confirm MSSGS identified by WES, 39 samples evaluated by WES were also examined using WGS. Results for each sample and MSSGS-associated SNV are shown. WGS_Ref and WGS_Alt indicate read counts for MSSGS-associated SNVs identified independently by WGS. WGS_RAW_Depth and WGS_RAW_ALT indicate read depths for MSSGS-associated SNVs identified by WES but not independently identified by WGS. P values were calculated using Fisher's exact test.

Extended Data Table 3. Haplotype phasing[10] reveals that 394 of 407 MSSGS accessible for analysis (97%) co-localized to the same allele.

ACKNOWLEDGEMENTS

We acknowledge foremost the patients and their families. We also gratefully acknowledge the assistance of Michelle Flores, Melodee Smith, Sherry Littlefield, Laura Connelly -Smith, James Riddle, Ednus Warren, Stan Riddell, Julie Gralow, Bernie McLaughlin, Jennifer Specht, Paul Martin, Brent Wood, Katy Dougherty, Anju Thomas, Corinne Fligner, David Yadock, Jennifer Spokely, Krystal Malhotra, Zhijun Duan, and NW BioTrust. We thank Kate Sweeney for assistance with figure preparation. This publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000423.

Figures and Tables

Figure 1

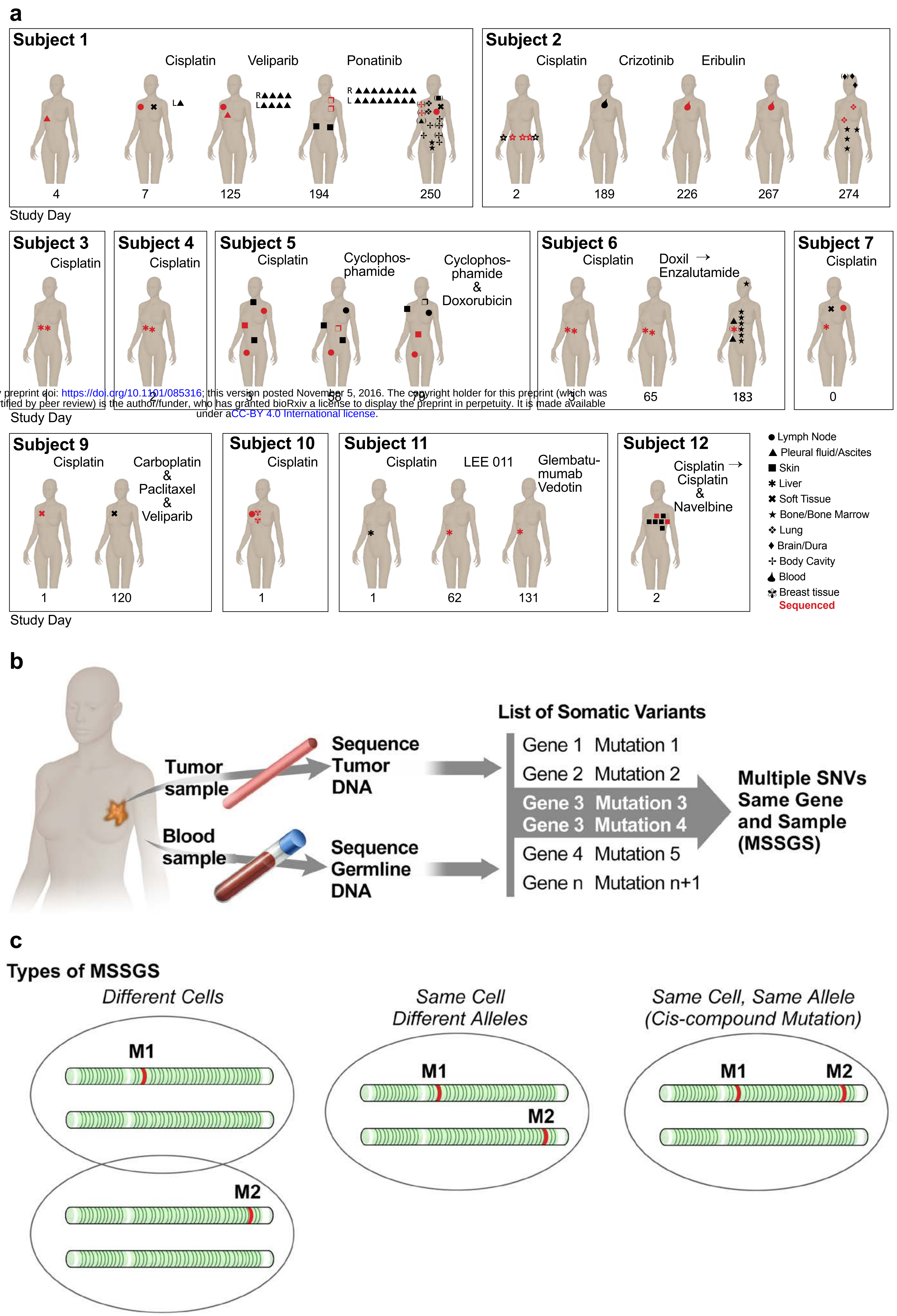

a
Subject 1
Cisplatin
Veliparib
Ponatinib
R
L
4
7
125
194
250
Subject 2
Cisplatin
Crizotinib
Eribulin
2
189
226
267
274
Study Day
Subject 3
Cisplatin
Subject 4
Cisplatin
2
Subject 5
Cisplatin
Cyclophos-phamide
Cyclophos-phamide & Doxorubicin
3
58
79
Subject 6
Cisplatin
Doxil → Enzalutamide
3
65
183
Subject 7
Cisplatin
0
Study Day
Subject 9
Cisplatin
Carboplatin & Paclitaxel & Veliparib
1
120
Subject 10
Cisplatin
1
Subject 11
Cisplatin
LEE 011
Glembatu-mumab Vedotin
1
62
131
Subject 12
Cisplatin → Cisplatin & Navelbine
2
Study Day
● Lymph Node
▲ Pleural fluid/Ascites
■ Skin
✱ Liver
✖ Soft Tissue
★ Bone/Bone Marrow
❖ Lung
♦ Brain/Dura
✢ Body Cavity
Blood
Breast tissue
Sequenced
b
Tumor sample
Blood sample
Sequence Tumor DNA
Sequence Germline DNA
List of Somatic Variants
Gene 1 Mutation 1
Gene 2 Mutation 2
Gene 3 Mutation 3
Gene 3 Mutation 4
Gene 4 Mutation 5
Gene n Mutation n+1
Multiple SNVs Same Gene and Sample (MSSGS)
c
Types of MSSGS
Different Cells
M1
M2
Same Cell Different Alleles
M1
M2
Same Cell, Same Allele (Cis-compound Mutation)
M1
M2

Figure 2

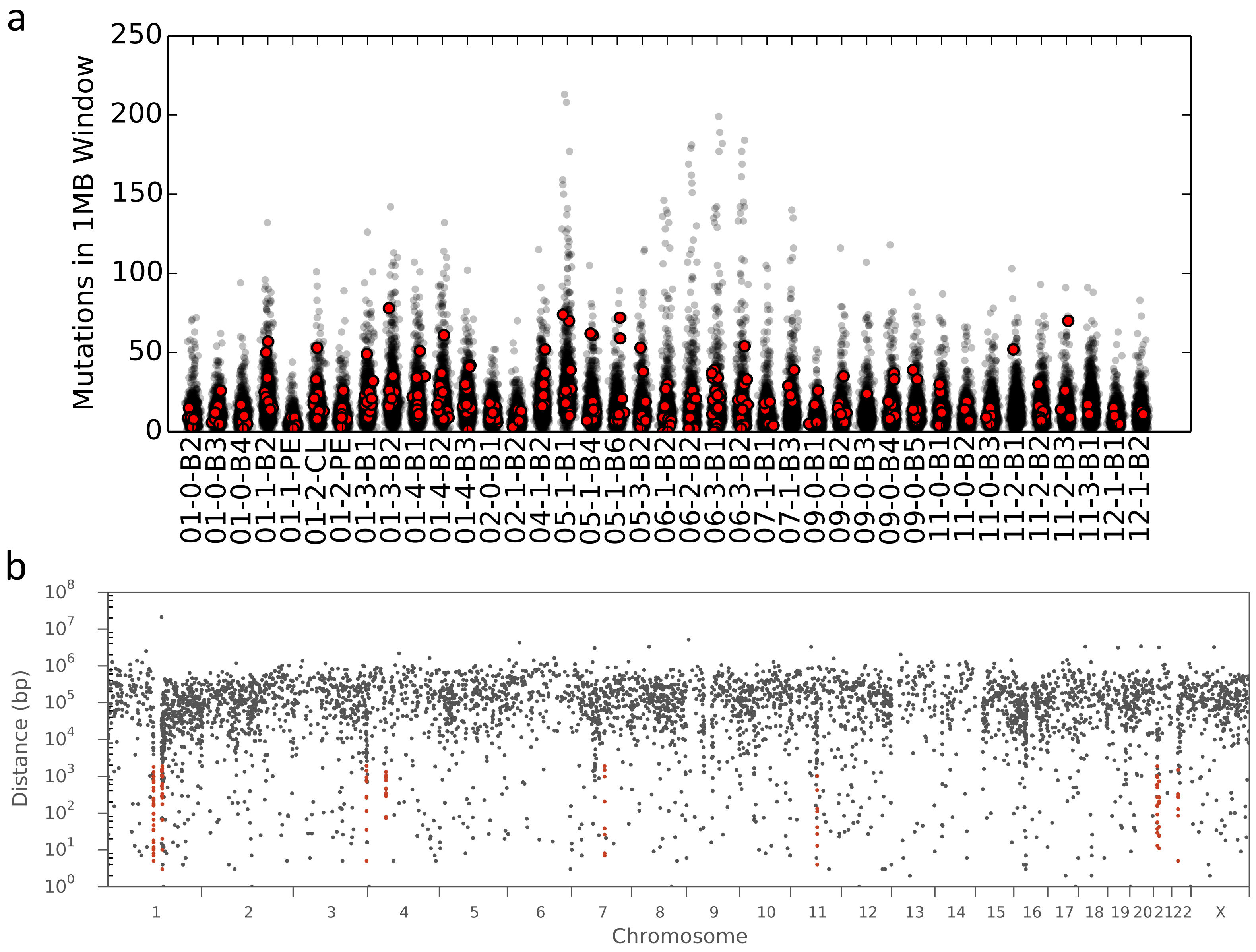

a
Mutations in 1MB Window
0
50
100
150
200
250
01-0-B2
01-0-B3
01-0-B4
01-1-B2
01-1-PE
01-2-CL
01-2-PE
01-3-B1
01-3-B2
01-4-B1
01-4-B2
01-4-B3
02-0-B1
02-1-B2
04-1-B2
05-1-B1
05-1-B4
05-1-B6
05-3-B2
06-1-B2
06-2-B2
06-3-B1
06-3-B2
07-1-B1
07-1-B3
09-0-B1
09-0-B2
09-0-B3
09-0-B4
09-0-B5
11-0-B1
11-0-B2
11-0-B3
11-2-B1
11-2-B2
11-2-B3
11-3-B1
12-1-B1
12-1-B2
b
Distance (bp)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

Figure 3

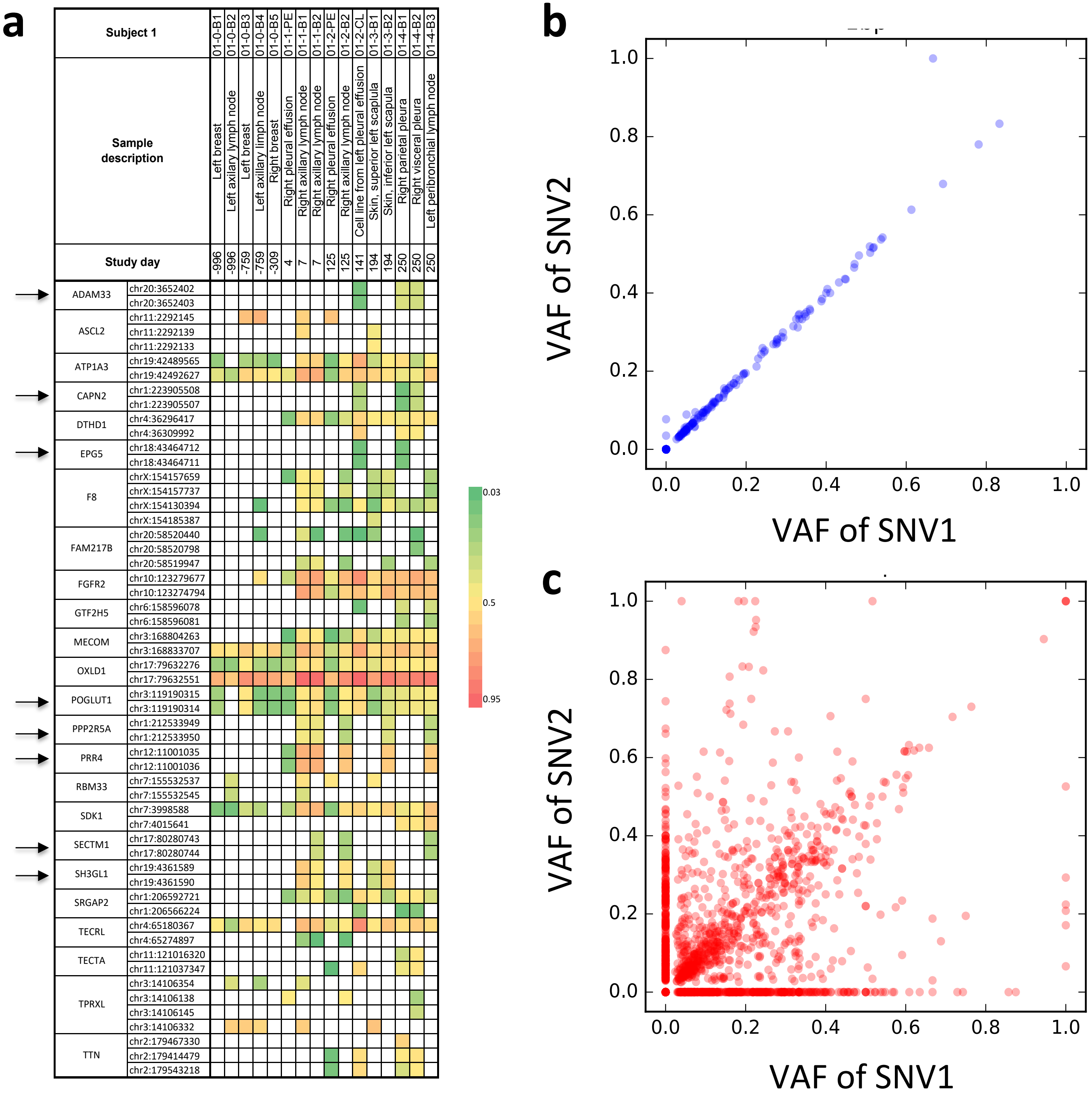
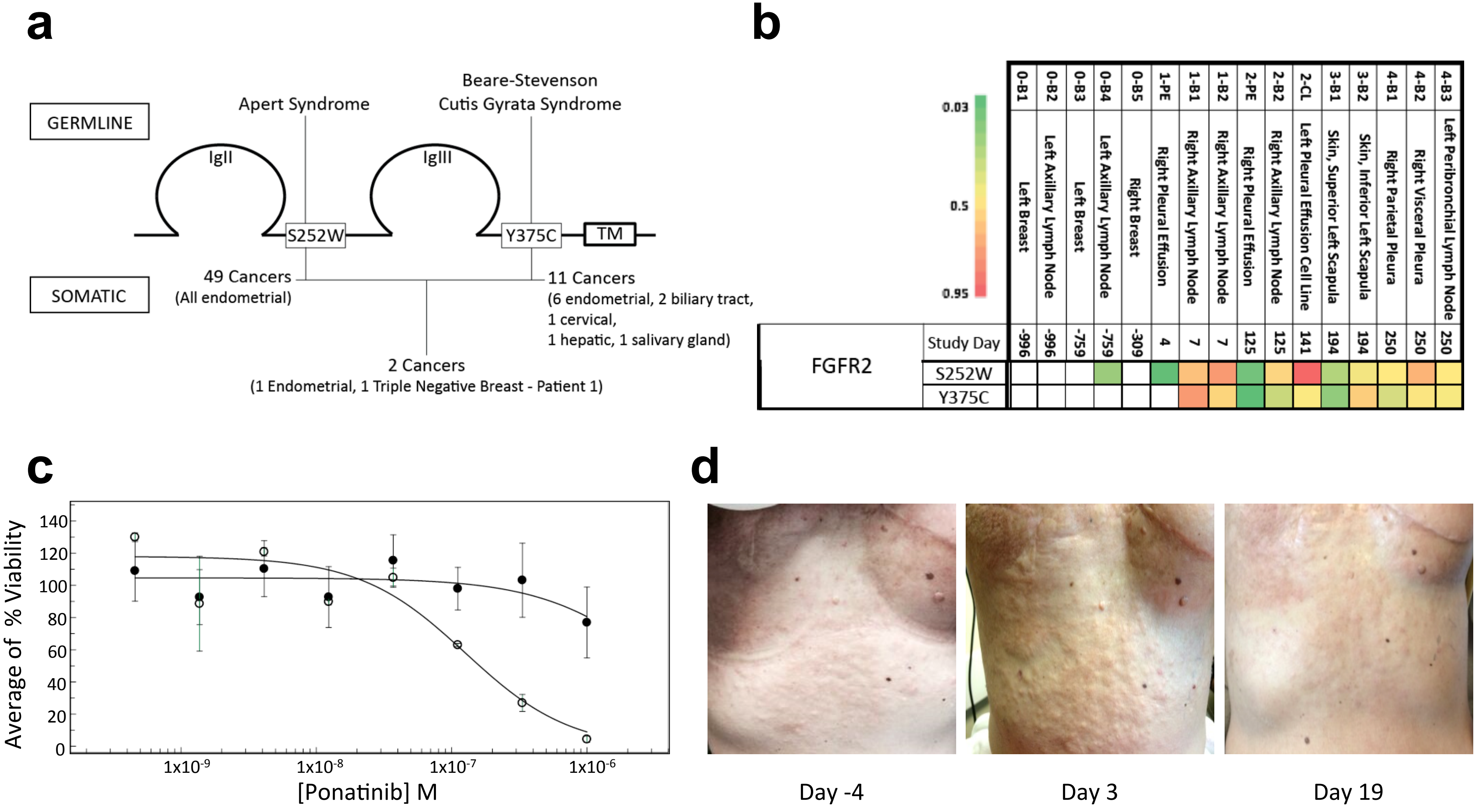

a
Subject 1
01-0-B1
01-0-B2
01-0-B3
01-0-B4
01-0-B5
01-1-PE
01-1-B1
01-1-B2
01-2-PE
01-2-B2
01-2-CL
01-3-B1
01-3-B2
01-4-B1
01-4-B2
01-4-B3
Sample description
Left breast
Left axilary lymph node
Left breast
Left axillary limph node
Right breast
Right pleural effusion
Right axillary lymph node
Right axillary lymph node
Right pleural effusion
Right axillary lymph node
Cell line from left pleural effusion
Skin, superior left scaplula
Skin, inferior left scapula
Right parietal pleura
Right visceral pleura
Left peribronchial lymph node
Study day
-996
-996
-759
-759
-309
4
7
7
125
125
141
194
194
250
250
250
ADAM33
chr20:3652402
chr20:3652403
ASCL2
chr11:2292145
chr11:2292139
chr11:2292133
ATP1A3
chr19:42489565
chr19:42492627
CAPN2
chr1:223905508
chr1:223905507
DTHD1
chr4:36296417
chr4:36309992
EPG5
chr18:43464712
chr18:43464711
F8
chrX:154157659
chrX:154157737
chrX:154130394
chrX:154185387
FAM217B
chr20:58520440
chr20:58520798
chr20:58519947
FGFR2
chr10:123279677
chr10:123274794
GTF2H5
chr6:158596078
chr6:158596081
MECOM
chr3:168804263
chr3:168833707
OXLD1
chr17:79632276
chr17:79632551
POGLUT1
chr3:119190315
chr3:119190314
PPP2R5A
chr1:212533949
chr1:212533950
PRR4
chr12:11001035
chr12:11001036
RBM33
chr7:155532537
chr7:155532545
SDK1
chr7:3998588
chr7:4015641
SECTM1
chr17:80280743
chr17:80280744
SH3GL1
chr19:4361589
chr19:4361590
SRGAP2
chr1:206592721
chr1:206566224
TECRL
chr4:65180367
chr4:65274897
TECTA
chr11:121016320
chr11:121037347
TPRXL
chr3:14106354
chr3:14106138
chr3:14106145
chr3:14106332
TTN
chr2:179467330
chr2:179414479
chr2:179543218
0.03
0.5
0.95
b
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
VAF of SNV2
VAF of SNV1
c
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
VAF of SNV2
VAF of SNV1

Figure 4



		0-B1	0-B2	0-B3	0-B4	0-B5	1-PE	1-B1	1-B2	2-PE	2-B2	2-CL	3-B1	3-B2	4-B1	4-B2	4-B3
		Left Breast	Left Axillary Lymph Node	Left Breast	Left Axillary Lymph Node	Right Breast	Right Pleural Effusion	Right Axillary Lymph Node	Right Axillary Lymph Node	Right Pleural Effusion	Right Axillary Lymph Node	Left Pleural Effusion Cell Line	Skin, Superior Left Scapula	Skin, Inferior Left Scapula	Right Parietal Pleura	Right Visceral Pleura	Left Peribronchial Lymph Node
FGFR2	Study Day	-996	-996	-759	-759	-309	4	7	7	125	125	141	194	194	250	250	250
	S252W				🟩		🟩	🟧	🟧	🟩	🟧	🟥	🟩	🟨	🟨	🟧	🟨
	Y375C							🟧	🟨	🟩	🟨	🟨	🟩	🟧	🟨	🟨	🟨

Figure 5

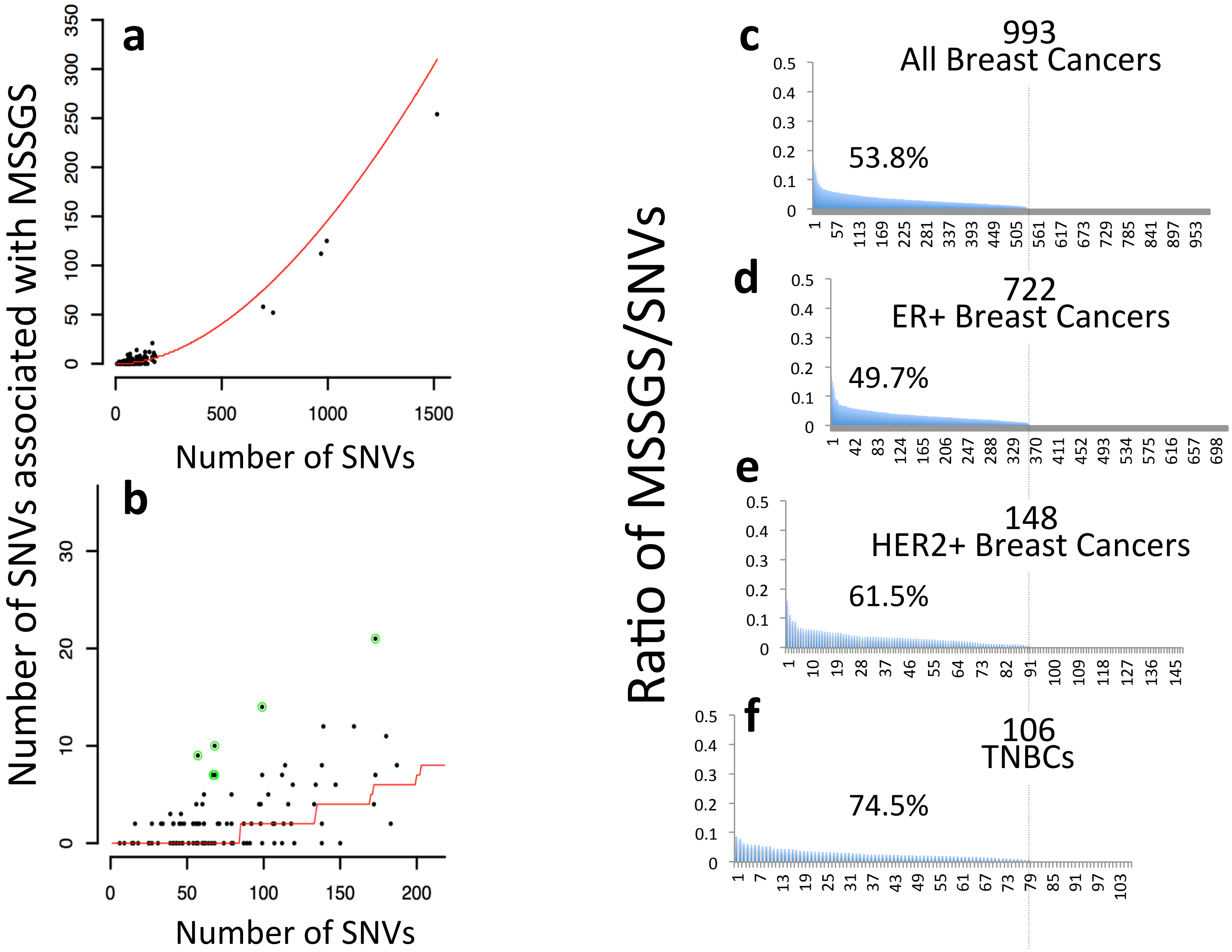

a
Number of SNVs associated with MSSGS
0
50
100
150
200
250
300
350
0
500
1000
1500
Number of SNVs
b
0
10
20
30
0
50
100
150
200
Number of SNVs
Ratio of MSSGS/SNVs
c
993
All Breast Cancers
0.5
0.4
0.3
0.2
0.1
0
53.8%
1 57 113 169 225 281 337 393 449 505 561 617 673 729 785 841 897 953
d
722
ER+ Breast Cancers
0.5
0.4
0.3
0.2
0.1
0
49.7%
1 42 83 124 165 206 247 288 329 370 411 452 493 534 575 616 657 698
e
148
HER2+ Breast Cancers
0.5
0.4
0.3
0.2
0.1
0
61.5%
1 10 19 28 37 46 55 64 73 82 91 100 109 118 127 136 145
f
106
TNBCs
0.5
0.4
0.3
0.2
0.1
0
74.5%
1 7 13 19 25 31 37 43 49 55 61 67 73 79 85 91 97 103

Table 1

Subject	**number of assessed samples**	**number of assessed timepoints**
1	16	11
2	7	5
3	2	1
4	2	1
5	7	3
6	7	4
7	2	1
9	8	4
10	3	1
11	7	3
12	6	3
Total	67	37
Mean	6.090909091	3.363636364
Median	7	3
Range	2 to 16	1 to 11

Table 2

				WES		DNA fragment sequencing				
#	Subject	Estimated Tumor Content (Disease Site)	Gene	SNV1 nt change chr position aa change VAF	SNV2 nt change chr position aa change VAF	REF %	SNV1 %	SNV2 %	SNV1+ SNV2 %	Clones Tested (Assessable)
1	06	85% (Liver)	*ABHD16Ac*	C-T 6:31659632 Q230H 0.473	C-G 6:31659583 E247K 0.424	64%	4%	2.7%	28%	75 (94)
2	01	75% (Pleura)	*ADAM33*	C-A 20:3652402 Q577H 0.131	T-A 20:3652403 Q577L 0.130	96.8%	0%	0%	3.2%	94 (94)
3	06	70% (Liver)	*APOBR***	G-C 16:28507445 E361D 0.161	G-T 16:28507452 G364W 0.107	100%	-	-	-	4 (94)
4	10	60% (Breast)	*C20orf26*	G-C 20:20037448 NA 0.152	G-C 20:20037264 NA 0.11	81%	6.6%	7.7%	4.4%	91 (94)
5	01	0% § (Skin)	*CBR4*	C-A 4:169923348 V137F 0.228	T-A 4:169923351 I136F 0.145	100%	0%	0%	0%	92 (92)
6	04	70% (Liver)	*CCDC171*	T-G 9:15744665 K815T 0.11	A-C 9:15744263 NA 0.106	94.5%	2.2%	1.1%	2.2%	92 (94)
7	01	0% § (Skin)	*F8**	A-T X:154157659 L1469* 0.107	T-A X:154157737 K1443I 0.101	96%	1%	1%	3%	93 (94)
8	04	65% (Liver)	*FLG2*	A-T 1:152329531 L244* 0.259	C-G 1:152329951 R104P 0.209	72%	14%	3%	1%	93 (94)
9	06	70% (Liver)	*MGA*	G-C 15:42042226 E2141Q 0.374	G-C 15:42041736 Q1977H 0.385	50 %	23%	17%	10%	87 (94)
10	05	30% (Skin)	*TCTE1*	C-G 6:44254146 C134S 0.122	A-G 6:44254159 Y130H 0.128	95%	2%	3%	0%	93 (94)
11	01	75% (Pleura)	*TMPRSS13*	T-C 11:117789342 Q78R 0.165	G-C 11:117789345 A77G 0.120	48%	2%	0%	49%	91 (94)
12	06	95% (Liver)	*DMPK*	C-T 19:46283223 R32K 0.32	C-T 19:46283028 NA 0.331	70%	7%	9%	14%	100 (101)
13	02	100%¶ (CTCs)	*ROBO4*	G-C 11:124754988 Q984E 0.606	G-C 11:124755134 S935* 0.555	28%	9%	3%	60%	102 (103)
14	02	100%¶ (CTCs)	*ZZZ3**	TCC-AAT 1:78045223 1:78045224 1:78045225 R690N, T691S 0.41, 0.4, 0.4		57%	0%	0%	43%	100 (101)
15	07	90% (Liver)	*FANCC*	G-C 9:98011451 F41L 0.295	G-C 9:98011537 Q13E 0.22	62%	13%	6%	19%	102 (103)

Estimated tumor content is based on tumor content in formalin fixed half of each biopsy specimen.
CTCs: Circulating Tumor Cells
REF: Number of sequenced DNA fragments containing the reference (non-mutated) sequence.
nt: Nucleotide
chr: Chromosome
aa: Amino Acid
* *ZZZ3* contained 3 SNVs, *F8* contained four MSSGs but only two were interrogated. All other MSSGs shown here contained two SNVs.
** *APOBR* could not be evaluated due to sequencing artifacts.
¶ Estimate following column enrichment and FACs sorting of CTCs.
§ Touch prep of the snap frozen tissue used for sequencing demonstrated numerous tumor cells.