bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Evolving neoantigen profiles in colorectal cancers with DNA repair defects

Giuseppe Rospo*[1], Annalisa Lorenzato*[1,2], Nabil Amirouchene-Angelozzi[3], Alessandro Magrì [1,2], Carlotta Cancelliere[1], Giorgio Corti[1], Carola Negrino[1,2], Vito Amodio[1,2], Monica Montone[1], Alice Bartolini[1], Ludovic Barault[1,2], Luca Novara[1], Claudio Isella[1], Enzo Medico[1,2], Andrea Bertotti[1,2], Livio Trusolino[1,2], Giovanni Germano[1,2], Federica Di Nicolantonio[1,2] and Alberto Bardelli[1,2]

[1]Candiolo Cancer Institute, FPO-IRCCS, 10060 Candiolo (TO), Italy;
[2]University of Turin, Department of Oncology, 10060 Candiolo (TO), Italy;
[3]FIRC Institute of Molecular Oncology (IFOM), 20139 Milan, Italy;

*These authors contributed equally

Corresponding author
Alberto Bardelli (alberto.bardelli@unito.it)

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Background Neoantigens that arise as a consequence of tumour-specific mutations can be recognized by T lymphocytes leading to effective immune surveillance. In colorectal cancer (CRC) and other tumour types, a high number of neoantigens is associated with patient response to immune therapies. The molecular processes governing the generation of neoantigens and their turnover in cancer cells are poorly understood. We exploited CRC as a model system to understand how alterations in DNA repair pathways modulate neoantigen profiles over time.

Methods We performed Whole Exome Sequencing (WES) and RNA sequencing (RNAseq) in CRC cell lines, *in vitro* and *vivo*, and in CRC patient-derived xenografts (PDXs) to track longitudinally genomic profiles, clonal evolution, mutational signatures and predicted neoantigens.

Results The majority of CRC models showed remarkably stable mutational and neoantigen profiles, however those carrying defects in DNA repair genes continuously diversified. Rapidly evolving and evolutionary stable CRCs displayed characteristic genomic signatures, and transcriptional profiles. Downregulation of molecules implicated in antigen presentation occurred selectively in highly mutated and rapidly-evolving CRC.

Conclusions These results indicate that CRC carrying alterations in DNA repair pathways display dynamic neoantigen patterns that fluctuate over time. We define CRC subsets characterized by slow and fast evolvability and link this phenotype to downregulation of antigen-presenting cellular mechanisms. Longitudinal monitoring of the neoantigen landscape could be relevant in the context of precision medicine.

Keywords: Colorectal cancer, neoantigen, DNA repair, immune response, mutational signature

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Background

Anticancer therapies based on immune-checkpoint blockade are often remarkably effective but benefit only a minor fraction of cancer patients (1). Several biomarkers of response and resistance to immune modulators have been proposed (2) (3). Among these, the overall mutational burden (number of somatic variants per megabase (Mb)) and the number of predicted neoantigens were highlighted in multiple studies (4) (5) (6). The predictive values of mutational and antigen burdens are still being evaluated in the clinical settings. Both parameters are presently assessed on DNA extracted from individual tissue samples and are typically measured only once in the clinical history of each patient. Alterations in DNA repair pathways, including mutations or promoter hypermethylation of mismatch repair (MMR) effectors (*MLH1*, *MSH2*, etc) or DNA polymerases (polymerase ε and δ) (7) are known to increase the mutational burden and the neoantigen profiles of cancers (8). Whether, and to what extent, neoantigen profiles evolve over time as a result of the inherent genomic instability of individual tumours is largely unknown. We recently reported that in mouse models, inactivation of DNA mismatch repair increases the mutational burden and leads to dynamic mutational profiles resulting in effective cancer immune response (9). Here we exploit CRCs as a model system to understand whether mutational burden and neoantigen profile of human tumours evolve over time as a result of their distinctive genomic landscapes.

Methods

CRC cell culture conditions

All cell lines were maintained in their original culturing conditions according to supplier guidelines. Cells were ordinarily supplemented with FBS 10%, 2mM L-glutamine, antibiotics (100U/mL penicillin and 100 mg/mL streptomycin) and grown in a 37°C and 5% CO2 air incubator. To study evolution of cell populations, cell lines were not cloned prior to the experiment or at any subsequent timepoint. Cell lines were thawed in a 10 cm dish. After thaw-recovery, each cell line was screened for the absence of Mycoplasma contamination and checked for its identity, referred below as Quality Control (QC)). To preserve heterogeneity, upon thawing individual lines were

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expanded to at least 10^8 cells. At this point for each model, cells were counted, and the percentage of alive/dead cells was calculated. At the beginning of the experiment (T0), 4 x10^7 live cells were distributed as follows: (A) 2 x10^6 cells were re-plated in a 10cm dish for *in vitro* propagation, (B) 3x10^7 cells were used for *in vivo* experiments, (C) 2 x10^6 cells were frozen, (D) 3 pellets (2x10^6 cells each) were frozen for DNA, RNA and protein extraction. Cells plated as in A were kept in culture changing medium twice a week and dividing them at constant splitting rate, determined before initiating the experiment. In details, splitting was performed before full confluency was achieved. The number of cells that were split and the number of passages and days of culture were recorded for each cell model to calculate the doubling time. During *in vitro* culture, cell populations were collected at the following pre-determined time points: 30 days (T30), 60 days (T60) and 90 days (T90) days from T0. At each time point a fraction of the cells were put aside (note that this did not affect the rate of passaging described below) and pellets (2x10^6 each) were collected for DNA, RNA and protein extraction. QC was repeated at each time point.

Cell quality control (QC)

Cells were screened for absence of Mycoplasma contamination using the Venor®GeM Classic kiy (Minerva biolabs). The identity of each cell line was checked before starting each experiment and after every genomic DNA extraction by PowerPlex® 16 HS System (Promega), through Short Tandem Repeats (STR) at 16 different loci (D5S818, D13S317, D7S820, D16S539, D21S11, vWA, TH01, TPOX, CSF1PO, D18S51, D3S1358, D8S1179, FGA, Penta D, Penta E, and amelogenin). Amplicons from multiplex PCRs were separated by capillary electrophoresis (3730 DNA Analyzer, Applied Biosystems) and analysed using GeneMapper v 3.7 software (Life Technologies).

Microsatellite instability (MSI) status

The MSI status was assessed with the MSI Analysis System kit (Promega). The analysis requires a multiplex amplification of seven markers including five mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24 and MONO-27) and two pentanucleotide repeat markers (Penta C and Penta D). The products were analysed by capillary electophoresis in a single injection (3730 DNA Analyzer, ABI

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

capillary electrophoresis system (Applied Biosystems). Then the results were analysed using GeneMapper V5.0 software.

DNA extraction and exome sequencing

Genomic DNA (gDNA) was extracted from CRC cell lines, xenografts, and PDXs using Maxwell® RSC Blood DNA kit (AS1400, Promega). DNA was sent to IntegraGen SA (Evry, France) that performed library preparation, exome capture, sequencing, and data demultiplexing. Final DNA libraries were pair-end sequenced on Illumina HiSeq4000 as paired end 100b reads.

Mutational analysis in cell lines

When cell lines were passaged in mice or when analysing patient derived xenografts, Fastq files were first processed with Xenome (10) to remove reads of mouse origin. Reads files were aligned to the human reference hg38 using BWA-mem algorithm (11) and then the “rmdup” samtools command was used to remove PCR duplicates (12). On the resulting aligned files, we observed a median depth of 138x with 98% of targeted-region covered by at least one read. Bioinformatic modules previously developed (13) (9) by our laboratory were used to identify single nucleotide variants (SNVs) and indels. The mutational characterization of the 64 cell lines at time point 0 was assessed by calling the alterations against the hg38 reference annotation. Then a series of filters were used to remove germline variants and artifacts: alleles supported by only reads with the same strand, excluding start and end read positions from the count, were discarded; variants called with allelic frequency lower than 10% as well a p-value greater than 0.05 (binomial test calculated on allele count and depth of each sample) were excluded; common dbSNP version 147 and a panel of normal (40 samples) from previous sequencing were used to annotate and filter germline variants and sequencing artifacts. The variant calls of 45 cell lines at time point 90 and the 18 cell lines explanted from mice were performed using the allele comparison strategy between the same cell line at time 0 and time point 90 and xenograft respectively. Only variants present at time point 90 (or in xenograft) were kept. Artifact removal was employed as described above. To calculate the tumour mutational burden (number of variants/Mb), only coding variants were considered. Those variants were used to predict neoantigens using previously published methods (14) (9). Briefly, RNAseq data were used as input of “OptitypePipeline” (15) to assess

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the HLA status of each samples at time point 0, then NetMHC 4.0 software (16) was employed to analyse mutated peptides derived from variants calls using kmer of 8-11 length. Next, for each SNV we modified the corresponding cDNA in the selected position and we examined the 5' and 3' context. The latter was set taking into account the length (in terms of aminoacids) with which the putative antigen could bind HLA. We translated the cDNA and feed mutant peptide to netMHC with the proper HLA(s). For frameshifts, we applied the same approach considering every possible peptide generated by the new frame. Finally, RNAseq data were used to annotate and then filter according to expression values (Fragments Per Kilobase Million (FPKM) > 10). Only predicted neoantigens with a strong binding affinity (Rank < 0.5) were considered for further analysis.

Mutational analysis of patient-derived xenograft

WES of patient-derived xenografts was performed at IntegraGen SA (Evry, France). Sequenced samples included a microsatellite stable (MSS), a microsatellite unstable (MSI), and a *POLE* mutant case (5, 7, and 6 respectively). Samples were analysed with the same bioinformatic pipeline applied to cell lines and murine reads were first removed using Xenome (10). A median depth of 130x and with 98% of targeted-region covered by at least one read was observed. All 18 PDXs samples were characterized by calling alterations against the hg38 reference annotation. For each generation, with the exception of first one, the mutational evolution was inferred by subtracting the mutations of the previous generation. Second generation samples were compared to the first-generation samples, samples from the 3rd generation were compared to the 2nd generation samples, and so on.

Ploidy Estimation

Gene copy-number (GCN) was calculated in a two-steps approach: initially we treated the cell line as diploid and considered the median read depth of all coding regions as the level for 2N ploidy. We also calculated the median read depth for every gene. The ratio between the two median values was then considered as the relative GCN. In the second step, to estimate the overall ploidy we segmented all chromosomes using a custom script that implements circular binary segmentation. Finally, we exploited the distribution of allelic frequencies for individual segments to assess the

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

absolute GCN. This was necessary since distinct ploidy levels have different expected distributions. For example, a 2N ploidy status has a bell-shaped curve with a peak on 50%, a 3N ploidy is expected to have two peaks on 33% and 66%, etc.

Mutational Signature

Mutational signatures were calculated using the web application 'Mutational Signatures in Cancer' (MuSiCa) (17). The profile of each signature is calculated using the six substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G (all substitutions are referred to by the pyrimidine of the mutated Watson–Crick base pair). Information on nucleotides 5' and 3' to each mutated base are incorporated to generate 96 possible mutation types. For each sample, a tab-separated values file was created with chromosome, position, reference, and alternate alleles. Only samples with at least 10 mutations were included. The output file of MuSiCa, that includes the contribution values of 30 signatures (18), was used to create a clustermap with *seaborn,* a *Python* data visualization library, setting Euclidean metric and the average linkage method.

Doubling time

Cell lines were passaged *in vitro* for a minimum of 85 to a maximum of 103 days. Each passage was performed before full confluency was achieved and the total number of doublings was annotated for each cell model. Two parameters, number of passages (n), and days of culture (t), were used to estimate the growth rate (GR) and the doubling time (DT) assuming that: every division is an independent random event; probability distribution of division is equal for all cells and it is an exponential distribution; the number of cells in each plate before confluence is fixed (K). The growth rate is defined as $GR=\log_n(2)\div DT$ (19). The estimated number of cells at time t is defined as $N(t)=N(0)\times e^{(GR\times t)}$ where N(0) is the number of cells at time 0. Therefore $GR=\log_n(N(t)\div N(0))\div t$ where $N(t)\div N(0)=(K\times 2^n)\div(K\times 2^0)=2^n$ and so $GR=\log_n(2^n)\div t$. Finally, $DT=t\times\log_n(2)\div \log_n(2^n)$.

RNA extraction and RNAseq analysis

Total RNA was extracted from a pellet of CRC cells (2×10^6 cells) using Maxwell® RSC miRNA Tissue Kit (AS1460, Promega), according to the manufacturer's

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

protocol. The quantification of RNA was performed by Thermo Scientific Nanodrop 1000 (Agilent) and Qubit 3.0 Fluorometer (LifeTechnologies). RNA integrity was evaluated with the Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano Kit. Total RNA (800 ng) with RNA integrity number (RIN) score between 9 and 10 was used as input to the Illumina TruSeq RNA Sample Prep Kit v2-Set B (48Rxn), according to the manufacturer's protocol. The standard RNA fragmentation profile was used (94 °C for 8 min for the TruSeq RNA Sample Prep Kit). PCR-amplified RNA-seq library quality was assessed using the Agilent DNA 1000 kit on the Agilent 2100 BioAnalyzer and quantified using Qubit 3.0 Fluorometer (LifeTechnologies). Libraries were diluted to 10 nM using Tris-HCl (10 mM pH 8.5) and then pooled together. Diluted pools were denatured according to standard Illumina protocol and 1.8 pM were run on NextSeq500 using high output Reagent cartridge V2 for 150 cycles. A single-read 150-cycle run was performed. FastQ files produced by Illumina NextSeq500 were aligned using MapSplice2 (20) transcriptome-aware aligner using hg38 assembly as reference genome. The resulting BAM files were post-processed to translate genomic coordinates to transcriptomic ones and to filter out alignments carrying insertions or deletions (which RSEM does not support) or falling outside the transcriptome regions. The post-processed BAM alignment was given as input to RSEM (21) for gene expression quantification using GENCODE v22 as gene annotation.

Differential expression analysis

The abundance quantification generated with RSEM provides the FPKM and the expected counts for each gene. The latter was used to perform genes differential expression analysis with DESeq2 R package (library Bioconductor) (22) given two distinct groups of interest, one of which considered as the reference. Genes were considered as differentially expressed if the adjusted p-value was less than 0.05, and the log2 fold change was less or equal to -1 (if median FPKM value of the reference group was greater or equal to 10) or the log2 fold change was greater or equal to 1 (if median FPKM of the target group was greater or equal to 10). The analyses were performed between the following groups: MSI *vs* MSS (reference); hypermutated *vs* non-hypermutated (reference); "EVOLVING-CRC" *vs* "STABLE-CRC" (reference). Hypermutated group included MSI and MSS POLE mutated cell lines (18 samples). "EVOLVING-CRC" group included all samples with at least 10 alterations acquired

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

per day. A multi-factor configuration of the expression analysis was designed including extra variables of interest such as growth rates or the number of mutations normalized to doubling time.

Pathway analysis

Genes differentially expressed were then analysed with g:Profiler (23), an online pathway analysis tool that takes a list of genes and assigns them to different families of biological functions. We set the query options to select significant biological processes only and we retained (for further analysis) only the top most families of the hierarchy (depth 1).

Xenograft mouse model

Each CRC cell line (5×10^6 cells) was injected subcutaneously into both flanks of two 6-week-old female NOD (nonobese diabetic)/SCID (severe combined immunodeficient) mice (Charles River Laboratory). Tumour size was measured twice a week and calculated using the formula: $V = ((d)^2 \times (D)) \div 2$ (d = minor tumour axis; D = major tumour axis). Tumours were explanted when they reached a volume of $1000mm^3$. The investigators were not blinded, and measurements were acquired before the identification of the cages.

Patient-derived mouse model

Tissue from hepatic metastasectomy of CRC patients was collected at surgery and implanted in NOD-SCID mice as described previously (24). When reaching a volume of 1500-$2000mm^3$, the tumours were explanted, fragmented, and serially passaged in new mice. At each passage, part of the material was frozen for molecular analyses. Samples' genetic identity was determined by Sequenom-based analysis of 24 highly variable SNPs of germline DNA (Table 5), confirmed by analysing pre-implantation tumour material, and then validated every second passage in mice. The study population consisted of matched tumour and normal samples from 3 CRC patients that underwent surgical resection of liver metastases at the Candiolo Cancer Institute (Candiolo, Torino, Italy) and at the Mauriziano Umberto I Hospital (Torino) between

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2009 and 2013. Patients signed informed consent, and the study was approved by the relevant institutional Ethics Committees.

Western blotting analysis

Proteins were extracted by solubilizing the cells in boiling SDS buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, and 1% SDS). Samples were boiled for 5 minutes at 95°C and sonicated for 10 seconds. Extracts were clarified by centrifugation, normalized with the BCA Protein Assay Reagent kit (Thermo). Equal amounts of proteins (20µg) were loaded in each lane. Proteins were separated by PAGE and transferred to nitrocellulose sheets. Western blot detection was performed with enhanced chemiluminescence system (GE Healthcare) and peroxidase conjugated secondary antibodies (Amersham). The following primary antibodies were used for western blotting: anti-beta2 Microglobulin [EP2978Y] (ab75853, Abcam), anti-MLH1 (ab92312, Abcam), anti-MSH2 (ab70270), Abcam), anti-MSH6 [EPR3945],(ab92471, Abcam), anti-MSH3 PA527864, Invitrogen, anti-PMS2 EPR3947 (Cell Marque Corporation, USA), anti-actin (I-19) (sc1616, Santa Cruz), anti-HSP 90α/β (H-114, sc-7947, Santa Cruz). Images were acquired with Chemidoc (Biorad) and western blot band intensity was analysed using Image Lab software (Biorad).

Results

We selected from our database 64 CRC cell lines designed to recapitulate clinically relevant characteristics of CRC patients (Table 1 and Suppl. Fig. 1a). Whole exome sequencing and RNAseq were performed on all models. Using previously developed computational tools and bioinformatic algorithms (13) (25) (14) (26), we measured mutational burden (alterations per Mb) assessing both SNVs and frameshifts (Fig. 1a,b) Scrutiny of genomic alterations highlighted that MSI cell lines and those carrying known *POLE* hotspot mutations had higher number of mutations per Mb as compared to MSS cell lines (Fig. 1a). The type of DNA repair alterations occurring in each model affected the nature of mutations: MSI cells displayed higher number of frameshifts and indels than *POLE* mutant cell lines, the opposite was true for SNVs (Fig. 1c, d).

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Alterations in MMR and *POLE* genes are listed in Table 2 and Suppl. Fig. 1b. The cell line with the highest number of variants (SNU1040) carried inactivating alterations in both *MLH1* and *POLE* (Suppl. Fig. 1b). Altogether, these results are consistent with what has been reported in CRC patients carrying alterations in the MMR DNA repair pathway, indicating that the cell models included in this study broadly recapitulate what is observed in clinical specimens (27).

To assess whether, and to what extent the basal mutational profiles (Time 0: T0) evolved over time, we passaged 45 cell lines for 90 days and collected a second set of samples (Time 90: T90) (Suppl. Fig. 2). These were subjected to WES and analysed using the computational pipeline described above. Across all cell lines globally, the total mutational burden was similar between T0 and T90 (Suppl. Fig. 3). However, when the T0 and T90 mutational profiles were compared, prominent differences were detected among models sharing specific DNA repair defects (Fig. 2a). Specifically, the mutational landscapes of most MSI and *POLE* mutant cells evolved very rapidly through the generation of novel SNVs and frameshifts (Fig. 2a). On the contrary, the majority of MSS models showed more stable profiles (Fig. 2a). We sought to minimize confounding effects due to differences in cell-intrinsic doubling times (Table 1), we therefore calculated the doubling time of all cell models (Table 1, Suppl. Fig. 4). Notably, evolvability trends remained apparent after normalization for doubling time (Suppl. Fig. 5). We designated rapidly evolving CRC cells as "EVOLVING-CRC" and evolutionary stable CRC cell as "STABLE-CRC" (Table 1).

We empirically define EVOLVING-CRCs as those cells that acquire 10 alterations (or more) per day after normalizing mutation data to the doubling time of cell lines (Table 1). Moreover, EVOLVING-CRCs often carried alterations in multiple genes involved in distinct DNA repair functions, suggesting that defects in several DNA damage response pathways might be co-selected (Suppl. Fig. 1b). The expression of MMR genes was assessed by western blot at T0 and T90 and no differences were observed (Suppl. Fig. 6).

The genome of four CRC lines classified as MSS (SNU1235, COCM1, HDC142, and SNU1411) exhibited dynamic mutational profiles (Fig. 2). In an attempt to decipher the molecular basis of these findings, whole exome data of the outliers were carefully

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

examined, focusing on genes previously implicated in DNA repair pathways that are not routinely subjected to scrutiny in CRC patients. We found that SNU1235 and HDC142 models carried biallelic alterations in the *EXO1* (S510*) and *MUTYH* (S179C) genes, respectively. The exonuclease EXO1 is implicated in both MMR (it binds MLH1) and base excision repair (28), while *MUTYH* encodes a DNA glycosylase that is involved in oxidative DNA damage repair and is part of the base excision repair pathway (29). Germline mutations in *MUTYH* cause MUTYH Associated Polyposis (MAP) (30). Scrutiny of the COCM1 exome revealed a *POLE* variant (A629D). A629 is localized in a region of *POLE* highly conserved during evolution (Suppl. Fig. 7). The A629D change is potentially damaging according to the SIFT (31) and Polyphen (32) algorithms, which predict the putative impact of amino acid substitutions on human proteins using structural and comparative evolutionary considerations.

We next addressed how longitudinal evolution of CRC cell genomes affected their predicted neoantigen profile. To this end, WES, RNAseq, and HLA prediction data were combined as previously described (9). In detail, we identified genomic variants that satisfied three criteria: (i) emerged over time, (ii) occurred in transcribed genes, and (iii) scored positively when HLA I matching algorithms were applied. The variants that emerged after deploying the above computational pipeline were classified as putative neoantigens (Fig. 2b). Hypermutated and “EVOLVING-CRC” cells displayed higher levels of putative neoantigens compared to slowly evolving CRC cells (Fig. 2b). Moreover, and consistent with their predicted effects on antigenicity, a high prevalence of indels and associated frameshifts, which occur in MSI CRCs, translated into higher numbers of predicted neoantigens in this subset (Fig. 2b).

Next, we studied whether in parallel to mutation’ gains we could also detect loss of variants over time. For this reason, we tracked lost and gained alteration in ‘evolving’ cell lines over time. As expected, variants that did not change over time showed high allelic frequency, likely reflecting their clonal (trunk) status. Mutations that emerged or were lost showed lower allelic frequency (Fig. 3).

Mutational signatures are characteristic combinations of mutation types arising from mutagenesis processes such as alterations in DNA replication, exposure to DNA

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

damaging agents, tissue culture conditions, and DNA enzymatic editing (18). In human tumours, over 30 mutational signatures have been identified, a subset of which are linked to defective DNA repair pathways. For example, signatures 6, 15, 20, and 26 are associated with MMR defects, signature 10 is linked to inactivating mutation in the proofreading domain of DNA polymerases, while signature 18 appears to explain the rise of 8-oxoG:A mismatches due to *MUTYH* biallelic alteration (33).

We reasoned that the remarkable evolvability observed in a subset of CRC cells might be reflected in their mutational signatures. To test this, we first identified mutational signatures at T0. As expected, MSI cells displayed signatures 6, 15, 20, and 26, while *POLE* mutant cells showed primarily mutational signature 10 (Suppl. Fig. 8).

We next assessed which signatures were acquired (remained active) during replication of the cells *in vitro* by comparing samples collected at T0 and T90. We found that in most instances, DNA alterations linked to MMR and *POLE* defects continued to occur over time, indicating that the corresponding DNA repair capabilities were permanently disabled (Fig. 4a).

Replication of cancer cell populations in 2D is thought to encounter little or no selective pressure as the cells are cultured in the same conditions for many generations before the experiment is started. To monitor mutational and neoantigen evolution under more stressful (selective) conditions, CRC cells including MSS, MSI, and *POLE* models were transplanted in immunodeficient (NOD SCID) mice and allowed to grow until they reached approximately 1000mm^3 in size, after which tumours were excised. Although NOD SCID mice have no adaptive immunity, the mouse stromal microenvironment and elements of cellular innate immunity are known to affect the growth of human cancer cells *in vivo* (34). DNA samples were obtained before implantation and at the end of the experiment. WES was performed, and the data were analysed with the same bioinformatic pipeline applied to cells grown *in vitro*. The mutational profiles revealed higher evolutionary rates *in vivo* than *in vitro* (Suppl. Fig. 9a,b). This translated in increased levels of predicted neoantigens *in vivo*, (Suppl Fig 9c). Notably, mutational signatures linked to MSI status and *POLE* mutations were more marked *in vivo* than *in vitro* (Fig. 4b, Suppl Fig 10).

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Next, we asked whether the evolutionary trajectories observed in CRC cells with alterations in DNA repair pathways also occurred in human CRC with analogous molecular profiles. To this end, we selected MMR proficient, MMR deficient, and POLE mutant cases (Table 3) from our extensive patient-derived CRC xenograft biobank (35). Each model was serially transplanted for at least four generations in immunodeficient mice as described in the phylogenetic tree (Fig. 5a). Samples collected at each transplantation were subjected to WES. In some instances, simultaneous transplantation of the same tumour in two animals allowed acquisition of independent measurements for each generation. NGS data were analysed with the bioinformatic pipeline applied to cells grown *in vitro*. These experiments revealed remarkable differences in the evolvability of MSS, MSI, and *POLE* CRC models *in vivo*, and indicated that these characteristics also occurred in patient-derived CRC samples (Fig. 5b, c). As expected, high frequency (clonal-trunk) variants were conserved across generations. Interestingly, the *in vivo* results differ from those obtained in cell models *in vitro*. We find that in PDX models, not only sub-clonal but also clonal populations can emerge in subsequent generation of colorectal cancers with DNA repair defects (Fig 6).

Furthermore, in MSI and *POLE* patient-derived xenografts, the mutational signatures were continuously (re)generated and could be clearly recognized (Suppl. Fig. 11 and 12). In non-mutator (slow evolving) cell lines very few mutations emerged over time, and thus the possibility to assess mutational signatures was limited. Because of this, in the slowly-evolving models we were unable to reliably generate mutational signatures.

Distinct subsets of CRCs can be recognized based on histological characteristics, as well as their genomic, epigenetic, and transcriptional profiles. As a result, CRC can be classified into specific subsets, which are often correlated with divergent clinical outcomes (36) (37). The rate of genomic evolution and the dynamics of neoantigen profile have not yet been systematically explored as a method to classify CRC. We therefore asked whether any molecular traits (beyond alterations in DNA repair genes) could distinguish “EVOLVING-CRC” and “STABLE-CRC”. To address this question, we performed unbiased gene copy and transcriptional comparative

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analyses of CRC cell lines. As previously reported, MSI CRC cells typically carried a close to diploid chromosomal status, while MSS showed elevated aneuploidy (Fig. 7) (38). Interestingly, the most rapidly evolving *POLE* mutant lines, SNU81 and HDC114, also displayed a diploid prevalent phenotype. Nonetheless, copy number and ploidy status could not distinguish “EVOLVING” and “STABLE” CRC models.

Next, we performed RNAseq on the entire dataset to explore whether transcriptional profiles could classify rapidly evolving CRC lines. Differential analysis of RNAseq data was initially performed comparing the MSS and MSI sample groups. The list of differentially expressed genes was consistent to results previously reported in this setting, and 168 genes were differentially expressed between these two groups (Table 4) (39). Next, we evaluated genes differentially expressed in hypermutated versus non-hypermutated cells, grouping together MSI and *POLE* mutated cell lines and comparing them to the MSS lines (Fig. 8a). Notably, proteins associated with immune response and predominantly with antigen-presenting and antigen recognition functions were consistently downregulated in cell lines with high mutational burden (Fig. 8b). Next, we compared “EVOLVING” and “STABLE” CRC models. The number of genes differentially expressed with significant p-value was smaller due to the reduced number of available samples (Fig. 9a). Beta-2 microglobulin (B2M) was downregulated in most “EVOLVING” as compared to “STABLE” CRCs (Fig. 9b, c). Downregulation of B2M was confirmed at the protein level (Fig. 9c) and was frequently associated with premature stop codons in the *B2M* gene (Fig. 9d). Interestingly, the four MSS models (COCM1, SNU1235, SNU1411, and HDC142) with low mutational burden but dynamic mutational profile also displayed low levels of B2M (Fig.9b, c). Comparison of “EVOLVING” and “STABLE” CRC models pinpointed other genes differentially expressed including CPNE1, IRF1, and PMSB10. These genes are also involved in immune-related processes and their downregulation might similarly reduce immune surveillance of “EVOLVING” CRCs (Fig. 9a and Suppl. Fig. 13). We next performed the analysis showed in Fig 9a in a multivariate fashion taking into account the growth rates of the cells or the number of mutations normalized to the doubling time. The number of statistically significant genes in the multivariate analyses (Suppl Fig 14) was lower but consistent with the findings of Fig 9a. In the future it would be interesting to assess whether the differential expression of genes in

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fast evolving CRC models has a functional impact. This aspect cannot be causally predicted at this stage.

Discussion

In the past decade it has become clear that most human tumours are highly molecularly heterogeneous, and this affects prognosis and the emergence of therapeutic resistance (40). How tumour-specific somatic variations can lead to distinct neoantigen profiles and ultimately to immune surveillance has also been partially elucidated. The number of neoantigens depends on several factors. For example lung cancers associated with smoking habits have high levels of mutations (41) (42), whereas the development of skin melanomas is correlated with UV light mediated mutagenicity (43). Both smoking and UV exposure occur during defined periods and their mutagenicity is transient, leading to high -but relatively stable- mutational profiles (44) (45). Another class of tumours with high mutational burden is characterized not by exposure to external carcinogens, but rather by the intrinsic inability of tumour cells to efficiently repair DNA. The latter is due to epigenetic or genetic alterations in key effectors of DNA repair pathways, rather than acute or chronic carcinogen exposure. In this work, we used CRC as a model system to understand whether, and to what, extent alterations of DNA repair pathway components modulate neoantigen profiles over time *in vitro* and *in vivo*. Tumours carrying alterations affecting DNA repair genes maintained their molecular characteristics over time and, in most instances, the functional consequence of those alterations is continuous and propagated at every generation. An exception was represented by two *POLE* mutant CRC cell lines (HROC69 and HCC2998) which despite having high mutational burden did not appreciably evolve over time. The reason(s) for this phenotype are presently unclear. Interestingly these two *POLE* mutant cells that evolved poorly over time had less marked mutational signatures, possibly suggesting that, in these models, polymerase defects may undergo some form of functional compensation.

The longitudinal analysis of cell and PDX models highlighted several aspects. For example, MSI and POLE mutated tumours tended to acquire SNV or short

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

insertions/deletions over time. These alterations can lead to novel putative neoantigens potentially trigger the host immune system. In addition to well-known DDR genes (*MLH1, MSH2, MSH6, PMS2, POLE*), our study, indicate that other genes involved in DNA repair pathway may lead to accumulations of mutations possibly translating in novel epitopes. *EXO1* and *MUTYH* are two of such examples. Profiling of these genes in the clinical setting may help to intercept tumours not classified as unstable or with hypermutator phenotype but nevertheless continuously evolving and accumulating mutations.

Our analysis suggests that in parallel to mutations gains, loss of variants also occurs during cell propagation. Our data indicate that in hypermutated CRCs, including MSI and *POLE* mutated models expanded *in vitro*, these events are mainly confined to subclones. A limitation of this study is that longitudinal characterization of lost and gained mutations *in vitro* could be influenced by sampling of cell populations during cell passaging. We also report that in the propagation of PDX, possibly due to selection imposed by the microenvironment, not only subclonal but also clonal variants emerge *de novo* over time. Based on these results, we speculate that in CRC patients with DNA repair defects metastatic seeding or therapeutic debulking can lead to the emergence of new subsets of clonal neoantigens. This could have implications for the development of therapies relying on the presence of clonal neo antigens, such as ICP, CAR-T and vaccines.

Both cell lines and PDX have been widely employed to test anticancer compounds (46-48), however experimental reproducibility has occasionally been questioned (49, 50). The molecular evolvability that we find to occur during serial passaging of cells and PDX may partly account for the discrepant results obtained with these models (51-53).

A limitation of the present study is that it examined the evolution of cell lines and xenografts but cannot address the impact of the immune system in the evolutionary dynamics due to intrinsic limitations of the models we used.

Our data indicate that alterations in DNA repair genes facilitate acquisition of neoantigens. These novel putative epitopes can be recognized by the immune

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

system. Accordingly, we confirm that CRCs with high number of mutations (hypermutated CRCs) selectively downregulate components of the neoantigen presentation process, such as *B2M*, thus restricting the ability of the host immune system to detect them. Our results further suggest that non hypermutated CRCs, that display fast evolving mutational and antigen profiles, also show downregulation of components implicated in neoantigen presentation. The differences in expression of molecules involved in immune functions we observed in the CRC models could have originated from adaption previously experienced in the patient as a mechanism of escape from negative pressure of the immune system related to the elevated neoantigens' production rate.

Conclusions

In summary, we identified, and functionally highlighted CRC subsets characterized by slow and fast genome evolvability. CRCs carrying alterations in genes involved in DNA repair (including *MLH1*, *MSH2*, *MSH6*, *MUTYH*, *EXO1*, and *POLE*) display dynamic neoantigen patterns that fluctuate over time. Furthermore, we find that in CRC cells and patient-derived tumour xenografts, DNA repair defects leading to high mutational burden and neoantigen evolvability are associated with inactivation or downregulation of antigen-presentation functions. Longitudinal monitoring the neoantigen landscape of CRC and other tumour types may have clinical implications. While tracking time-dependent neoantigen evolution in the tissue of cancer patients might be difficult or impossible to achieve, monitoring predicted neoantigens in circulating tumour DNA is already within reach. Accordingly, longitudinal liquid biopsies could be deployed to assess whether and how time and/or therapeutic regimens affect the mutational burden and the neoantigen profiles in individual patients. Neoantigen clonality profiles could be valuable to develop specific vaccines and deploy immunomodulatory molecules in the context of precision oncology.

Declarations

Ethics approval and consent to participate

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

All animal procedures were approved by the Ethical Commission of Candiolo Cancer Institute IRCCS, of the University of Turin and by the Italian Ministry of Health. They were performed in accordance with institutional guidelines and international law and policies. The number of mice included in the experiments, the inclusion/exclusion criteria and the observed tumour size limits were based on institutional guidelines. Guidelines limited us to using 6–8-week-old female and male NOD–SCID mice. Mice were obtained from Charles River. Informed consent for research use was obtained from all patients at the enrolling institution prior to tissue banking and study approval was obtained from the different centres.

Consent for publication

Not applicable

Availability of data and material

The datasets used in the current study are available from the corresponding author upon reasonable request.

Authors' contributions

A.B., G.R., A.L., and N.A. conceived and designed the study. A.L., N.A., A.M., C.C., V.A., M.M., A.Bart., and G.G. performed the experiments. G.R., G.C., C.N., L.N., and C.I. performed bioinformatics analysis. A.B., G.R., A.L., and G.G. interpreted the data. L.T. and A.Ber. provided samples and interpreted the data. A.B., G.R., and A.L. wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge Francesco Galimi for providing gDNA and RNA from PDXs. We thank Daniela Cantarella for performing RNAseq. We thank all members of our

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

laboratory for valuable suggestions, critical review of the data, and the manuscript, as well as all laboratory technicians for their help.

Competing interests

G. Germano has ownership interest in Neophore. A. Bardelli has ownership interest in Phoremost and Neophore and is a consultant/advisory board member for Phoremost and Neophore. No potential conflicts of interest were disclosed by the other authors.

Funding

The research leading to these results has also received funding from AIRC IG program: ID 20697 (AB) 21407 (FDN), 20697 (ABer), 18532 (LT). From the AIRC Special Program 5 per mille metastases Project n 21091 (AB, FDN, EM, LT, ABer). By Fondazione Piemontese per la Ricerca sul Cancro-ONLUS, 5 × 1,000 Ministero della Salute 2011, 2014 and 2015 (AB, LT, FDN, EM). RC 2017 Ministero della Salute (AB, LT, FDN and EM). Partly funded also by Fondo per la Ricerca Locale (ex 60%), Università di Torino, 2017 (AL and FDN). This work was also supported by the European Community's H2020 grant agreement no. 635342-2 MoTriColor (AB);

Figures and Table legends

Fig. 1 Analysis of mutational burden in a panel of 64 CRC cell lines

Mutational characterization and comparison of SNVs and frameshifts among MSS (46 samples), MSI (12 samples) and POLE mutated (6 samples) of CRC models. a) The distribution of SNVs per Mb of coding DNA at time 0 is shown for each cell line. b) The number of frameshift mutations at time 0 is shown for each cell line. c) The number of SNVs per each group is shown ('MSS' refers to MSS cells without POLE mutations; 'MSI' includes MSI cells, as well as the SNU1040 cell line which is both MSI and POLE mutated; 'POLE' includes only MSS cell lines carrying a POLE mutation). d) The number of frameshifts per group is shown. The center line of each box plot indicates the median. $P < 0.0001$.

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 2 *In vitro* evolution of mutational landscape in 45 CRC cell lines

Mutational characterization of CRC cells after 90 days of culture (T90*) in vitro*. a) Bar charts show the number of novel alterations (SNVs and frameshifts) acquired at T90 (not present at T0) for each cell line. b) The number of predicted neoantigens (see Methods) is shown. Each bar represents putative neoepitopes derived from SNVs and frameshifts.

Fig. 3 Lost and gained mutations across 'evolving' CRC cell lines

For each CRC model the allelic frequency of SNVs at T0 and T90 are shown. Mutations were called against the reference genome (hg38) with allelic frequency > 1. The y-axis reports all the mutations found in each cell line, whereas the time points data are reported on x-axis.

Fig. 4 Mutational signatures associated with alterations emerging during *in vitro* or *in vivo* CRC propagation

Analysis of 30 validated cancer associated mutational signatures in hypermutated/rapidly evolving CRC cell lines. Signatures associated to MMR-deficient (6, 15, 20 and 26), *POLE*-dependent (10) and *MUTYH*-associated polyposis (18) are highlighted. Analysis and clustering were performed as reported in Methods. a) Heatmap of signature contributions during replication of CRC cells *in vitro* by analysing alterations acquired at T90. b) Heatmap of signature contributions during replication of the CRC cells *in vivo* by comparing xenograft tumours to the corresponding cells at T0 (See Methods for detailed information).

Fig. 5 Genomic evolution in patient-derived xenografts

Phylogeny of the indicated patient-derived xenograft and their molecular characterization. a) MSS, MSI and *POLE* mutant samples were serially transplanted for at least four generations (F1-F4) in NOD/SCID mice as shown. Samples collected at each passage were subjected to WES. b) WES data of each generation were compared with those obtained from the previous generation. Bar graphs show *de novo* acquired SNVs and frameshifts at each generation. c) The number of predicted neoantigens in each PDX is shown. Each bar represents putative neoepitopes derived from SNVs and frameshifts. (See Methods for detailed information).

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig. 6 Lost and gained mutations across the indicated PDX generations

The color code defines allelic frequencies of acquired SNVs at each generation (with allelic frequency > 1). The y-axis lists all SNV identified in each branch, the mouse generation (genealogy) are reported on the x-axis.

Fig. 7 Analysis of cell ploidy in a panel of 64 CRC cell lines

Heatmap showing distribution of ploidy for every segmented region in each cell line. Samples are sorted from most to less mutated as reported in Fig. 1. The percentage (ploidy) is calculated as described in detail in the Methods.

Fig. 8 Transcriptional analysis of CRC cell lines

Differential expression analysis between hypermutated and non-hypermutated cells. a) 183 unique genes differentially expressed between hypermutated (MSI/*POLE*) versus non-hypermutated CRC cells (MSS). Log2 expression values, along with the mean change in expression are shown. b) Pathway analysis of genes differentially expressed between hypermutated versus non-hypermutated CRC cells using g:Profiler application (see Methods).

Fig. 9 Beta2 microglobulin (B2M) expression is downregulated in EVOLVING-CRC

Transcriptional and protein levels of the B2M gene. a) Genes differentially expressed in EVOLVING-CRC relative to STABLE-CRC with a significant p-value ($P < 0.05$). b) Waterfall chart showing B2M expression at RNA level across a panel of 45 CRC cell lines. c) Western blot analysis of B2M expression. In gray are highlighted samples for which T90 sequencing were not available. Blots were reprobed with anti-HSP90 antibody to confirm equal loading. d) *B2M* gene alterations on 64 CRC cell lines at T0 (upper panel) and codon affected (lower panel).

Supplementary Fig. 1 Genomic features of 64 CRC cell lines

Molecular characterization of the indicated CRC models at T0 using cbioportal oncoprint graphic representation. a) Schematic diagram showing *BRAF*, *KRAS*, *NRAS* and Microsatellite/*POLE* status of CRC cell lines. b) Schematic diagram

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

showing genetic alterations in MMR genes and DNA proofreading polymerase *POLE* in CRC cell lines.

Supplementary Fig. 2 Outline of the experimental workflow to assess evolution of CRC cell lines *in vitro* and *in vivo*

CRC cell lines were thawed and kept in culture for 90 days. WES and RNAseq were performed at the beginning of the experiment (T0) for 64 cell lines. Forty-five samples collected at T90 were also subjected to WES. In a few instances at T0 an equal number of cells was injected in two immunodeficient mice. When tumours reached approximately a volume of $1000m^3$ in size, they were excised and sequenced.

Supplementary Fig. 3 Genetic alterations in CRC cell lines at T0 and T90

Bar chart showing the number of alterations at the beginning of experiment (T0) and after 90 days of culture (T90). All the variants are called against the hg38 reference.

Supplementary Fig. 4 Doubling time of MSS, MSI and POLE mutated cell lines

Cell doubling time in MSS, MSI and MSS *POLE* mutant cells. The number of days per group is shown. The centre line of each box plot indicates the median.

Supplementary Fig. 5 *In vitro* evolution of mutational landscape in CRC cell line normalized to the doubling time

Mutational characterization of CRC cells after 90 days of culture (T90) normalized to the doubling time. a) The bar chart shows the number of new alterations acquired at T90 (absent at T0) normalized to the cell doubling time. b) The number of predicted neoantigens per cell doubling (see Methods) is shown. Each bar represents putative neoepitopes derived from SNVs and frameshifts. c) The bar chart shows the number of new alterations per day acquired at T90 (absent at T0). d) The number of predicted neoantigens acquired per day (see Methods) is listed. Each bar represents putative neoepitopes derived from SNVs and frameshifts.

Supplementary Fig. 6 MMR proteins expression in cell models

Western blot analysis on hypermutated samples (MSI and/or POLE mutant cell lines) to assess the status of the MMR proteins at Time 0 (T0) and Time 90 (T90).

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Fig. 7 POLE protein sequence alignment across species

The yellow box indicates alanine (A) 629, which is highly conserved across species but is mutated to aspartic acid (D) in the COCM1 cell line.

Supplementary Fig. 8 Analysis of mutational signatures in CRC cell lines

Heatmap and clustering of 30 cancer associated signatures across 64 CRC cell lines at T0. Signatures associated with MMR-deficiency (6, 15, 20 and 26), POLE mutations (10) and MUTYH-associated polyposis (18) are highlighted. Analysis and clustering were performed as reported in Methods.

Supplementary Fig. 9 Comparison of mutational profiles in cell lines grown *in vitro* and *in vivo*

Alterations acquired by the indicated cell lines after 90 days in cell culture and upon transplantation in mice. a) Number of SNVs/Mb acquired *in vitro* (blue) and *in vivo* (yellow). b) Number of frameshifts acquired *in vitro* (blue) and *in vivo* (yellow) c) Number of predicted neoantigens in CRC cell lines after injection in mice (See Method for details).

Supplementary Fig 10 Comparison of signatures contribution between CRCs models propagated *in vitro* and *in vivo*.

The contributions of the associated signatures based on the molecular status detected *in vitro* and *in vivo* were compared in each model. For the MSI samples we compared the maximum contribution provided by signatures 6, 15, 20 or 26 (linked to MMRd). For the MSS POLE mutant cells, contribution was from signature number 10. For the cell line carrying MUTYH biallelic alterations contribution was from signature 18.

Supplementary Fig. 11 Analysis of mutational signatures in patient-derived xenograft

Contributions of 30 cancer associated signatures in patient-derived CRC xenografts. (a) Clustered heatmap according to signatures contribution in the indicated PDXs. (b)

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Signature profiles in PDXs using the six substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G. Analysis and clustering were performed as reported in Methods.

Supplementary Fig. 12 Mutational signatures acquired during propagation of patient-derived xenografts

Contributions of 30 cancer-associated signatures and signature profiles across evolving PDXs. Genomic variants of the individual PDXs were compared to the corresponding previous generation to infer signatures contributions. a) Clustered heatmap of signature contributions during evolution of the indicated PDX generations. b) Signature profiles acquired in each PDX generation using the six substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G. Alterations of each sample were inferred comparing two consecutive generations (see Method for detailed information).

Supplementary Fig. 13 Gene differentially expressed in EVOLVING-CRC

Waterfall charts show the z-score expression values of IRF1, CPNE1 and PSMB10 across a panel of 45 CRC cell lines.

Supplementary Fig. 14 Gene differentially expressed in EVOLVING-CRC in univariate and multivariate analyses. Log2 fold-change of the genes listed in Fig 9a according to univariate and multivariate analyses considering the mutator status or the growth rates of the cells.

Table 1 Molecular and functional characteristics of the indicated cell lines

Table 2 POLE mutations in CRC cells.

Table 3 Molecular characterization of patient-derived xenografts

Table 4 List of genes differentially expressed in the indicated cell lines

Table 5 List of SNPs used to identify patient-derived xenografts

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Overman MJ, McDermott R, Leach JL, Lonardi S, Lenz HJ, Morse MA, et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. Lancet Oncol. 2017;18(9):1182-91.
2. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. Lancet Oncol. 2016;17(12):e542-e51.
3. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018;359(6382):1350-5.
4. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science. 2018;362(6411).
5. Lee CH, Yelensky R, Jooss K, Chan TA. Update on Tumor Neoantigens and Their Utility: Why It Is Good to Be Different. Trends Immunol. 2018;39(7):536-48.
6. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019;51(2):202-6.
7. Shlien A, Campbell BB, de Borja R, Alexandrov LB, Merico D, Wedge D, et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. Nat Genet. 2015;47(3):257-62.
8. Turajlic S, Litchfield K, Xu H, Rosenthal R, McGranahan N, Reading JL, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. Lancet Oncol. 2017;18(8):1009-21.
9. Germano G, Lamba S, Rospo G, Barault L, Magrì A, Maione F, et al. Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. Nature. 2017;552(7683):116-20.
10. Conway T, Wazny J, Bromage A, Tymms M, Sooraj D, Williams ED, et al. Xenome--a tool for classifying reads from xenograft samples. Bioinformatics. 2012;28(12):i172-8.
11. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.
12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
13. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med. 2015.
14. Germano G, Amirouchene-Angelozzi N, Rospo G, Bardelli A. The Clinical Impact of the Genomic Landscape of Mismatch Repair-Deficient Cancers. Cancer Discov. 2018.
15. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics. 2014;30(23):3310-6.

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

16. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2016;32(4):511-7.
17. Díaz-Gay M, Vila-Casadesús M, Franch-Expósito S, Hernández-Illán E, Lozano JJ, Castellví-Bel S. Mutational Signatures in Cancer (MuSiCa): a web application to implement mutational signatures analysis in cancer samples. BMC Bioinformatics. 2018;19(1):224.
18. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.
19. Hirsch HR, Engelberg J. Determination of the cell doubling-time distribution from culture growth-rate data. J Theor Biol. 1965;9(2):297-302.
20. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 2010;38(18):e178.
21. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
23. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016;44(W1):W83-9.
24. Galimi F, Torti D, Sassi F, Isella C, Corà D, Gastaldi S, et al. Genetic and expression analysis of MET, MACC1, and HGF in metastatic colorectal cancer: response to met inhibition in patient xenografts and pathologic correlations. Clin Cancer Res. 2011;17(10):3146-56.
25. Siravegna G, Lazzari L, Crisafulli G, Sartore-Bianchi A, Mussolin B, Cassingena A, et al. Radiologic and Genomic Evolution of Individual Metastases during HER2 Blockade in Colorectal Cancer. Cancer Cell. 2018;34(1):148-62.e7.
26. Corti G. A genomic analysis workflow for colorectal cancer precision oncology. In press ed. Clinical Colorectal Cancer.
27. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.
28. Goellner EM, Putnam CD, Kolodner RD. Exonuclease 1-dependent and independent mismatch repair. DNA Repair (Amst). 2015;32:24-32.
29. Boiteux S, Coste F, Castaing B. Repair of 8-oxo-7,8-dihydroguanine in prokaryotic and eukaryotic cells: Properties and biological roles of the Fpg and OGG1 DNA N-glycosylases. Free Radic Biol Med. 2017;107:179-201.
30. Dallosso AR, Dolwani S, Jones N, Jones S, Colley J, Maynard J, et al. Inherited predisposition to colorectal adenomas caused by multiple rare alleles of MUTYH but not OGG1, NUDT1, NTH1 or NEIL 1, 2 or 3. Gut. 2008;57(9):1252-5.
31. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(Web Server issue):W452-7.
32. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.
33. Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, et al. A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. EBioMedicine. 2017;20:39-49.

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol. 2013;14(10):1014-22.
35. Bertotti A, Papp E, Jones S, Adleff V, Anagnostou V, Lupo B, et al. The genomic landscape of response to EGFR blockade in colorectal cancer. Nature. 2015;526(7572):263-7.
36. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350-6.
37. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, et al. Stromal contribution to the colorectal cancer transcriptome. Nat Genet. 2015;47(4):312-9.
38. Sinicrope FA, Rego RL, Halling KC, Foster N, Sargent DJ, La Plant B, et al. Prognostic impact of microsatellite instability and DNA ploidy in human colon carcinoma patients. Gastroenterology. 2006;131(3):729-37.
39. Tian S, Roepman P, Popovici V, Michaut M, Majewski I, Salazar R, et al. A robust genomic signature for the detection of colorectal cancer patients with microsatellite instability phenotype and high mutation frequency. J Pathol. 2012;228(4):586-95.
40. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15(2):81-94.
41. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012;150(6):1121-34.
42. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. Mutational signatures associated with tobacco smoking in human cancer. Science. 2016;354(6312):618-22.
43. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, et al. A landscape of driver mutations in melanoma. Cell. 2012;150(2):251-63.
44. Network CGAR. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543-50.
45. Network CGA. Genomic Classification of Cutaneous Melanoma. Cell. 2015;161(7):1681-96.
46. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013;41(Database issue):D955-61.
47. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012;483(7391):570-5.
48. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483(7391):603-7.
49. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A Landscape of Pharmacogenomic Interactions in Cancer. Cell. 2016;166(3):740-54.
50. Vanden Heuvel JP, Bullenkamp J, Biology RPC. Registered report: Systematic identification of genomic markers of drug sensitivity in cancer cells. Elife. 2016;5.
51. Liu Y, Mi Y, Mueller T, Kreibich S, Williams EG, Van Drogen A, et al. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. Nat Biotechnol. 2019;37(3):314-22.
52. Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, et al. Genetic and transcriptional evolution alters cancer cell line drug response. Nature. 2018;560(7718):325-30.

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

53. Ben-David U, Ha G, Tseng YY, Greenwald NF, Oh C, Shih J, et al. Patient-derived xenografts undergo mouse-specific tumor evolution. Nat Genet. 2017;49(11):1567-75.

Table 1

SAMPLE	Microsatellites Status	Altered MSI Markers	Genome Evolvability	Split ratio	*In vitro* Doubling Time	Growth rate
C10	Stable	-	Stable	0.34	2.33	0.30
C106	Stable	-	Stable	0.35	2.50	0.28
C125PM	Stable	-	Stable	0.34	2.37	0.29
C32	Stable	-	Stable	0.18	1.54	0.45
C70	Stable	-	Stable	0.4	2.76	0.25
C75	Stable	-	Stable	0.36	2.46	0.28
C99	Stable	-	NA	NA	NA	NA
CACO2	Stable	-	Stable	0.26	1.83	0.38
CAR1	Stable	-	Stable	0.36	2.47	0.28
CCK81	Unstable	bat26-nr21-bat25-mono27-nr24	Evolving	0.34	2.53	0.27

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CL14	Stable	-	NA	NA	NA	NA
COCM1	Stable	-	Evolving	0.3	2.34	0.30
COGA1	Unstable	bat26-mono27-nr24	NA	NA	NA	NA
COGA2	Stable	-	Stable	0.32	2.22	0.31
COGA5	Stable	-	Stable	0.20	1.69	0.41
COGA8	Stable	-	Stable	0.22	1.67	0.41
COLO201	Stable	-	NA	NA	NA	NA
COLO94H	Stable	-	Stable	0.45	2.81	0.25
DIFI	Stable	-	NA	NA	NA	NA
DLD1	Unstable	bat26-nr21-bat25-mono27-nr24	Evolving	0.07	0.98	0.71
HCA24	Stable	-	Evolving	0.33	2.22	0.31
HCA46	Stable	-	Stable	0.4	2.41	0.29
HCC2998	Stable	-	Stable	0.34	2.33	0.30
HDC114	Stable	-	Evolving	0.23	1.69	0.41
HDC142	Stable	-	Evolving	0.35	2.42	0.29
HDC82	Stable	-	NA	NA	NA	NA
HRA16	Stable	-	NA	NA	NA	NA
HROC24	Unstable	bat26-nr21-bat25-mono27-nr24	Evolving	0.16	1.23	0.56
HROC32	Stable	-	Stable	0.62	6.07	0.11
HROC334	Stable	-	Stable	0.45	2.83	0.24
HROC39	Stable	-	Stable	0.5	3.92	0.18
HROC69	Stable	-	Stable	0.33	2.26	0.31
HT115	Stable	-	Evolving	0.24	1.78	0.39
HT29	Stable	-	Stable	0.16	1.38	0.50
HT55	Stable	-	Stable	0.33	2.21	0.31
LIM1215	Unstable	bat26-nr21-bat25-mono27-nr24	Evolving	0.15	1.24	0.56
LIM2099	Stable	-	Stable	0.33	2.26	0.31
LOVO	Unstable	nr21-bat25-mono27-nr24	NA	NA	NA	NA
LS180	Unstable	bat26-nr21-bat25-mono27-nr24	Evolving	0.25	1.90	0.37
LS411N	Unstable	bat26-nr21-bat25-mono27-nr24	Evolving	0.32	2.29	0.30
MDST8	Stable	-	Stable	0.15	1.31	0.53

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

NCIH716	Stable	-	NA	NA	NA	NA
OUMS23	Stable	-	Stable	0.26	1.70	0.41
OXCO3	Stable	-	Stable	0.23	1.76	0.39
RW7213	Stable	-	Stable	0.4	2.60	0.27
SNU1040	Unstable	bat26-nr21-bat25-mono27-nr24	Evolving	0.54	4.14	0.17
SNU1181	Stable	-	Stable	0.62	4.65	0.15
SNU1235	Stable	-	Evolving	0.35	2.33	0.30
SNU1411	Stable	-	Evolving	0.44	2.76	0.25
SNU1460	Stable	-	NA	NA	NA	NA
SNU1684	Unstable	bat26-nr21-bat25-mono27-nr24	NA	NA	NA	NA
SNU175	Unstable	bat26-nr21-bat25-mono27	NA	NA	NA	NA
SNU283	Stable	-	NA	NA	NA	NA
SNU479	Stable	-	NA	NA	NA	NA
SNU81	Stable	-	Evolving	0.42	2.33	0.30
SNU977	Stable	-	Stable	0.42	2.71	0.26
SNUC1	Stable	-	NA	NA	NA	NA
SW1417	Stable	-	NA	NA	NA	NA
SW1463	Stable	-	NA	NA	NA	NA
SW480	Stable	-	Stable	0.18	1.64	0.42
SW837	Stable	-	Stable	0.27	2.07	0.34
V411	Stable	-	Stable	0.22	1.91	0.36
V481	Unstable	bat26-nr21-bat25-mono27-nr24	NA	NA	NA	NA
WIDR	Stable	-	NA	NA	NA	NA

Table 2

Time point 0								
			DNA			RNA		
SAMPLE	Coord (HG38)	AA Change	Normal allele counts	Altered allele counts	Allelic Frequency	Normal allele counts	Altered allele counts	Allelic Frequency
HCA24	chr12:132673261	p.S459F	33	26	44,07%	9	9	50,00%
HCC2998	chr12:132676598	p.P286R	54	44	44,90%	27	25	48,08%
HDC114	chr12:132673271	p.A456P	33	36	52,17%	17	28	62,22%
HROC69	chr12:132673703	p.V411L	88	64	42,11%	31	16	34,04%
HT115	chr12:132673703	p.V411L	100	62	38,27%	24	10	29,41%
SNU1040	chr12:132675740	p.F367L	34	16	32,00%	25	37	59,68%
SNU81	chr12:132676598	p.P286R	40	30	42,86%	22	26	54,17%
Time point 90								
SAMPLE	Coord (HG38)	AA	DNA			RNA		

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

		Change	Normal allele counts	Altered allele counts	Allelic Frequency	Normal allele counts	Altered allele counts	Allelic Frequency
HCA24	chr12:132673261	p.S459F	33	27	45,00%	17	20	54,05%
HCC2998	chr12:132676598	p.P286R	37	32	46,38%	11	17	60,71%
HDC114	chr12:132673271	p.A456P	37	20	35,09%	30	37	55,22%
HROC69	chr12:132673703	p.V411L	83	44	34,65%	49	23	31,94%
HT115	chr12:132673703	p.V411L	146	38	20,65%	18	4	18,18%
SNU1040	chr12:132675740	p.F367L	24	17	41,46%	11	33	75,00%
SNU81	chr12:132676598	p.P286R	65	56	46,28%	15	22	59,46%

MSS
MSI

Table 3

Sample	MSI Status	BRAF	KRAS	NRAS	MLH1	MSH2	MSH3	MSH6	PMS2	POLE
CRC106	MSI	V600E	WT	WT	WT	WT	WT	R577H / p.T1085Tfs7*	WT	WT
CRC542	MSS	WT	WT	WT	WT	WT	WT	WT	WT	WT
CRC371	MSS	WT	WT	E132K	R265C	WT	WT	E946*	WT	V411L

*Only non-synonymous variants present in COSMIC database are reported

Table 4 (MSS vs MSI)

gene_id	log2FoldChange	padj
ABCB6	-1,12	0,01
ACAD11	1,18	0,00
ACOT8	-1,04	0,00
AHCYL2	1,12	0,02
ALDH6A1	1,34	0,00
AMACR	-1,54	0,01
AOC1	-1,57	0,04
APBB3	1,04	0,00
ARHGAP18	-1,01	0,02
ATOX1	-1,00	0,01
ATP5E	-1,22	0,00
ATP9A	-1,77	0,00
B2M	-1,58	0,00

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B4GALT5	-1,14	0,00
BCL2L15	-1,73	0,01
BCL9L	-1,09	0,02
BPTFP1	1,09	0,03
C15orf52	-1,28	0,05
C20orf24	-1,14	0,00
C6orf48	1,05	0,01
CAPG	-1,30	0,02
CCDC71L	-1,08	0,02
CCND3	-1,26	0,00
CCPG1	1,16	0,02
CD59	-1,01	0,03
CD82	-1,37	0,03
CDKN2A	-3,20	0,00
CEACAM1	-1,89	0,00
CEACAM6	-1,71	0,05
CGN	-1,06	0,02
CHKA	-1,13	0,00
CHMP4B	-1,16	0,00
CLDN1	-1,37	0,05
CLDN3	-1,58	0,01
COL17A1	-1,61	0,03
COTL1	-1,61	0,00
CPNE1	-1,05	0,00
CPT1B	1,13	0,00
CTSA	-1,10	0,00
CTSD	-1,02	0,03
CTSH	-1,22	0,03
CTSS	-2,46	0,00
CTSV	-1,81	0,00
CTSZ	-1,08	0,01
CUEDC1	-1,33	0,02
CXXC5	-1,24	0,01
CYB5D1	1,24	0,00
CYFIP2	1,37	0,03
DAB2IP	-1,14	0,00
DCBLD2	-2,43	0,00
DDX27	-1,02	0,00
DNTTIP1	-1,25	0,00
DPEP1	-1,88	0,04
DYNLRB1	-1,04	0,00
EDN1	-1,79	0,01
EMP1	-1,47	0,03
EPB41L1	-1,09	0,01
EPB41L4A-AS1	1,04	0,01
EPS8L1	-1,13	0,03
FCGRT	-1,69	0,00
FECH	1,07	0,00
FKBP1A	-1,02	0,01
FZD7	-1,02	0,03
GABBR1	1,70	0,01
GABRE	-1,83	0,00

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

GCAT	1,39	0,00
GLS2	1,38	0,00
GNE	-1,13	0,04
GSN	-1,30	0,03
HELZ2	-1,14	0,01
HIST1H2AC	-1,36	0,01
HIST1H2BD	-1,92	0,00
HIST1H2BK	-1,48	0,00
HIST2H2AA3	-1,13	0,03
HIST2H4A	-2,11	0,00
HSPA1A	-1,83	0,02
HSPB1	-1,41	0,03
HSPH1	-1,43	0,00
IDH2	-1,01	0,04
IDS	-1,30	0,02
IFI6	-1,76	0,03
IRF1	-1,02	0,01
KRT20	-1,73	0,04
KRT23	-4,08	0,00
LAMC2	-1,77	0,00
LFNG	-1,19	0,02
LGALS1	-2,88	0,00
LINC01089	-1,05	0,01
LIPG	1,38	0,00
LPCAT2	-1,11	0,01
LRRC75A-AS1	1,01	0,04
LRRC8A	-1,23	0,00
LTBP3	-1,37	0,01
LY6E	-1,41	0,02
LY75	-1,60	0,01
MACF1	-1,23	0,01
MALL	-1,20	0,03
MAP7D1	-1,03	0,02
MAPRE3	-1,11	0,02
MDM2	1,37	0,00
MGLL	-1,35	0,01
MIR4435-2HG	-1,09	0,05
MMP14	-1,53	0,02
MOCOS	1,18	0,01
MORC4	1,15	0,00
MUC20	-1,64	0,02
MYBL2	-1,13	0,01
MYL5	1,04	0,01
NABP1	-1,35	0,01
NDUFC2	-1,22	0,00
OXR1	-1,12	0,01
PDP1	-1,01	0,03
PEA15	-1,29	0,00
PFDN4	-1,04	0,00
PIGT	-1,00	0,00
PLA2G6	1,17	0,00
PLS3	-1,21	0,03

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PMEPA1	-2,81	0,00
PML	-1,04	0,00
POLE4	-1,01	0,01
PPM1M	1,46	0,00
PPP1R14D	-1,98	0,01
PPP1R18	-1,27	0,01
PRADC1	-1,19	0,01
PRAP1	-1,79	0,04
QPCT	-1,89	0,02
QPRT	-2,16	0,00
REG4	2,81	0,00
ROMO1	-1,14	0,01
RPL22L1	1,62	0,00
RPL32P29	1,31	0,00
S100A11	-1,62	0,00
S100A2	-2,35	0,00
S100A4	-2,70	0,00
SDC4	-1,30	0,00
SESN2	1,23	0,01
SGK2	-1,40	0,04
SLC20A2	-1,07	0,01
SLC2A1	-1,22	0,04
SLC39A5	-1,93	0,02
SLC6A6	-1,40	0,00
SNHG8	1,01	0,01
SNORA73B	1,12	0,02
SPINK1	-2,25	0,01
STAT1	-1,06	0,01
SULT2B1	-1,78	0,01
SUPT4H1	1,22	0,00
SYT7	-1,82	0,00
TCF7	-1,03	0,04
TFF1	2,17	0,02
TGFBI	-2,43	0,00
TIMP2	-2,16	0,01
TM4SF1	-1,70	0,01
TMEM52	1,08	0,03
TMPRSS4	-1,35	0,04
TNFSF9	2,01	0,00
TRIM7	1,24	0,05
TSPAN6	-1,49	0,00
TUBA4A	-1,11	0,00
TUBE1	1,12	0,01
TXNDC9	-1,04	0,00
UCA1	-1,79	0,03
UCP2	-1,38	0,03
UNC13D	-1,81	0,01
VAMP8	-1,09	0,00
VOPP1	-1,13	0,01
ZMYND8	-1,11	0,00
ZNFX1	-1,08	0,00

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4 (Hypermutated vs Non-hypermutated)

gene_id	log2FC	padj
ABHD12	-1,10	0,00
ACOT8	-1,09	0,00
AHCYL2	1,00	0,01
AKR1C3	-1,52	0,02
ALDH6A1	1,30	0,00
AMN	1,06	0,04
ANXA6	-1,94	0,00
AOC1	-1,96	0,00
ARHGEF10	1,15	0,04
ARL4C	-1,56	0,01
ATOX1	-1,12	0,00
ATP5E	-1,31	0,00
ATP8B1	1,09	0,00
ATP9A	-1,77	0,00
B2M	-1,82	0,00
BCAS4	-1,06	0,00
BCL2L1	-1,02	0,00
BNIP3	1,31	0,02
C15orf52	-1,10	0,04
C20orf24	-1,21	0,00
C2orf54	-2,00	0,00
C7orf50	-1,20	0,00
CAPG	-1,02	0,03
CCDC71L	-1,00	0,01
CD59	-1,13	0,00
CD82	-1,31	0,01
CDKN2A	-1,43	0,03
CEACAM1	-1,42	0,00
CFD	1,65	0,00
CHKA	-1,17	0,00
CHMP4B	-1,25	0,00
CLDN3	-1,45	0,00
COL17A1	-1,33	0,03
COTL1	-1,67	0,00
CPNE1	-1,09	0,00
CST3	-1,29	0,00
CTSA	-1,01	0,00
CTSD	-1,27	0,00
CTSH	-1,31	0,00
CTSS	-2,26	0,00
CTSV	-2,11	0,00
CTSZ	-1,05	0,00
CUEDC1	-1,59	0,00
CXXC5	-1,01	0,01
CYB5D1	1,17	0,00
CYTOR	-1,42	0,00
DBNDD2	-1,06	0,00
DCBLD2	-2,20	0,00
DGAT2	-1,04	0,01

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DHRS3	-1,09	0,01
DNTTIP1	-1,23	0,00
DYNLRB1	-1,25	0,00
EDN1	-1,27	0,03
EHD1	-1,01	0,00
EMP1	-1,60	0,00
EPB41L1	-1,06	0,00
EPS8L1	-1,20	0,00
FAH	-1,28	0,00
FAM84B	-1,10	0,02
FCGRT	-1,60	0,00
FECH	1,27	0,00
FKBP10	-1,41	0,04
FKBP1A	-1,16	0,00
FUNDC2	-1,04	0,00
FUNDC2P1	-1,08	0,00
FZD7	-1,19	0,00
GABARAPL1	-1,35	0,01
GABRE	-1,53	0,00
GCAT	1,01	0,01
GNE	-1,03	0,02
GPC1	-1,01	0,05
GSN	-1,75	0,00
HELZ2	-1,02	0,01
HIST1H2BD	-1,04	0,01
HIST2H4A	-1,38	0,00
HSD11B2	-1,11	0,03
HSPB1	-1,20	0,02
IDS	-1,60	0,00
IFI27	-1,48	0,03
IFI27L2	-1,03	0,03
IFI6	-1,86	0,00
IGFBP4	-1,15	0,02
IL33	1,58	0,04
IRF1	-1,28	0,00
ISG15	-1,13	0,04
ITGA3	-1,23	0,01
ITGB5	-1,09	0,00
KIFC3	-1,20	0,04
KLK6	-2,00	0,00
KRT20	-1,56	0,02
KRT23	-3,88	0,00
KRT80	-1,58	0,00
LAMC2	-1,57	0,00
LFNG	-1,23	0,00
LGALS1	-2,44	0,00
LIPG	1,38	0,00
LITAF	-1,20	0,01
LTBP3	-1,63	0,00
LTBP4	1,17	0,00
LY6E	-1,74	0,00
LY6G6D	-2,03	0,01

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

LY75	-1,60	0,00
MAPRE3	-1,23	0,00
MBOAT2	-1,03	0,02
MCRIP1	-1,02	0,00
MDK	-1,36	0,01
MDM2	1,09	0,00
MELTF	-1,18	0,01
MGLL	-1,20	0,00
MIR4435-2HG	-1,47	0,00
MMP14	-1,64	0,00
MOCOS	1,13	0,00
MOSPD1	-1,01	0,00
MUC20	-1,58	0,00
MYBL2	-1,06	0,00
NABP1	-1,05	0,01
NDUFC2	-1,11	0,00
PDP1	-1,16	0,00
PEA15	-1,30	0,00
PFDN4	-1,20	0,00
PHLDB1	-1,05	0,05
PLAUR	-1,18	0,00
PLTP	-1,53	0,03
PMEPA1	-2,92	0,00
PML	-1,03	0,00
POLD4	-1,06	0,00
PPP1R14D	-2,05	0,00
PRDX5	-1,09	0,00
PRR15	-1,06	0,03
PRSS23	-1,09	0,02
PSMA7	-1,00	0,00
PSMB10	-1,27	0,00
QPCT	-2,23	0,00
QPRT	-1,32	0,05
RASL11A	-1,19	0,04
REG4	2,99	0,00
RGCC	-2,97	0,00
RGMB	1,49	0,00
RHOD	-1,22	0,01
ROMO1	-1,28	0,00
RPL22L1	1,12	0,01
RTFDC1	-1,04	0,00
S100A11	-1,65	0,00
S100A2	-2,87	0,00
S100A4	-1,55	0,02
SDC4	-1,30	0,00
SLC20A2	-1,01	0,00
SLC2A1	-1,42	0,00
SLC2A4RG	-1,04	0,00
SLC2A8	-1,04	0,00
SLC39A5	-1,70	0,01
SLCO1B3	1,46	0,05
SMAD3	-1,08	0,00

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SMIM22	-1,19	0,01
SNORA73B	1,00	0,01
SPINK1	-2,72	0,00
SULT2B1	-1,60	0,00
SYT7	-1,55	0,00
TACSTD2	-2,32	0,00
TCF7	-1,18	0,00
TGFBI	-2,28	0,00
TIMP2	-2,59	0,00
TM4SF1	-1,78	0,00
TMEM176A	-2,16	0,00
TMEM176B	-1,85	0,01
TMEM185A	-1,07	0,00
TMPRSS4	-1,35	0,01
TNFRSF14-AS1	1,19	0,02
TNFRSF1B	-1,44	0,00
TRIB1	-1,11	0,02
TRIM7	1,08	0,04
TSC22D1	-1,01	0,01
TUBA4A	-1,01	0,00
TUBB2A	-1,49	0,00
TUBE1	1,12	0,00
UCA1	-1,49	0,03
UCP2	-1,16	0,02
VAMP8	-1,10	0,00
VEGFB	-1,25	0,00
VOPP1	-1,27	0,00
VSIR	-1,68	0,00
ZFP90	1,13	0,00
ZMYND8	-1,07	0,00

Table 4 (EVOLVING-CRC vs STABLE-CRC)

gene_id	log2FoldChange	padj
ABCB6	-1,00	0,04
AHI1	1,04	0,02
ANXA6	-1,83	0,02
ATOX1	-1,07	0,01
B2M	-1,57	0,00
CPNE1	-1,12	0,00
FAT1	1,09	0,04
GTF3A	-1,09	0,03
IGFBP6	-2,01	0,02
IRF1	-1,05	0,02
LGALS1	-2,45	0,00
LY6E	-1,47	0,02
PAM	-1,27	0,03

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PEA15	-1,12	0,01
PLA2G2A	2,56	0,01
PSMB10	-1,05	0,02
S100A2	-1,96	0,00
SDC4	-1,04	0,05
SLC7A11	1,63	0,01
TGFBI	-1,50	0,04
TIMP2	-1,82	0,03
TUBA1A	-1,71	0,05
TUBB2A	-1,55	0,03
TUBE1	1,27	0,01
VEGFA	1,08	0,05

Table 5

Probe	SNP	GENE	Chr	Functional Consequence
AMG_mid100				
rs11017876	[A\G]	DOCK1	10:127402700	intron variant
rs1106334	[C\T]		8:70100576	
rs1155741	[C\T]	ITGA9	3:37585621	intron variant
rs11655512	[A\G]	LOC339260	17:20948422	intron variant
rs11940551	[G\T]		4:27160856	
rs1210110	[A\G]	PRDM2	1:13770326	intron variant
rs1364054	[C\T]	LINC00299	2:8038605	intron variant
rs1528601	[C\G]		16:51064516	
rs161792	[A\T]	LOC101928166	3:152181915	intron variant
rs17272796	[C\T]	PLCL2	3:17035776	intron variant

bioRxiv preprint doi: https://doi.org/10.1101/667097; this version posted June 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rs242076	[C\T]	SYN3 - TIMP3	22:32833844	intron variant
rs4775699	[C\T]	SEMA6D	15:47581352	intron variant
rs4793172	[A\T]	DCAKD	17:45054112	intron variant
rs4905366	[A\G]		14:95636762	
rs6603251	[C\T]	PPP2R3B	Y:359845	intron variant
rs6734275	[A\G]	LOC105374785	2:67014042	intron variant
rs685449	[A\T]	RGS17	6:153023396	intron variant
rs7555566	[A\G]	KAZN	1:14478378	intron variant
rs7584993	[A\C]		2:222981224	
rs7808249	[A\G]	CROT	7:87354399	intron variant
rs9293511	[C\T]	LOC105379072	5:89120537	intron variant
rs9352613	[A\G]		6:78714716	
rs9572094	[C\T]	LOC105370159	13:34678745	intron variant

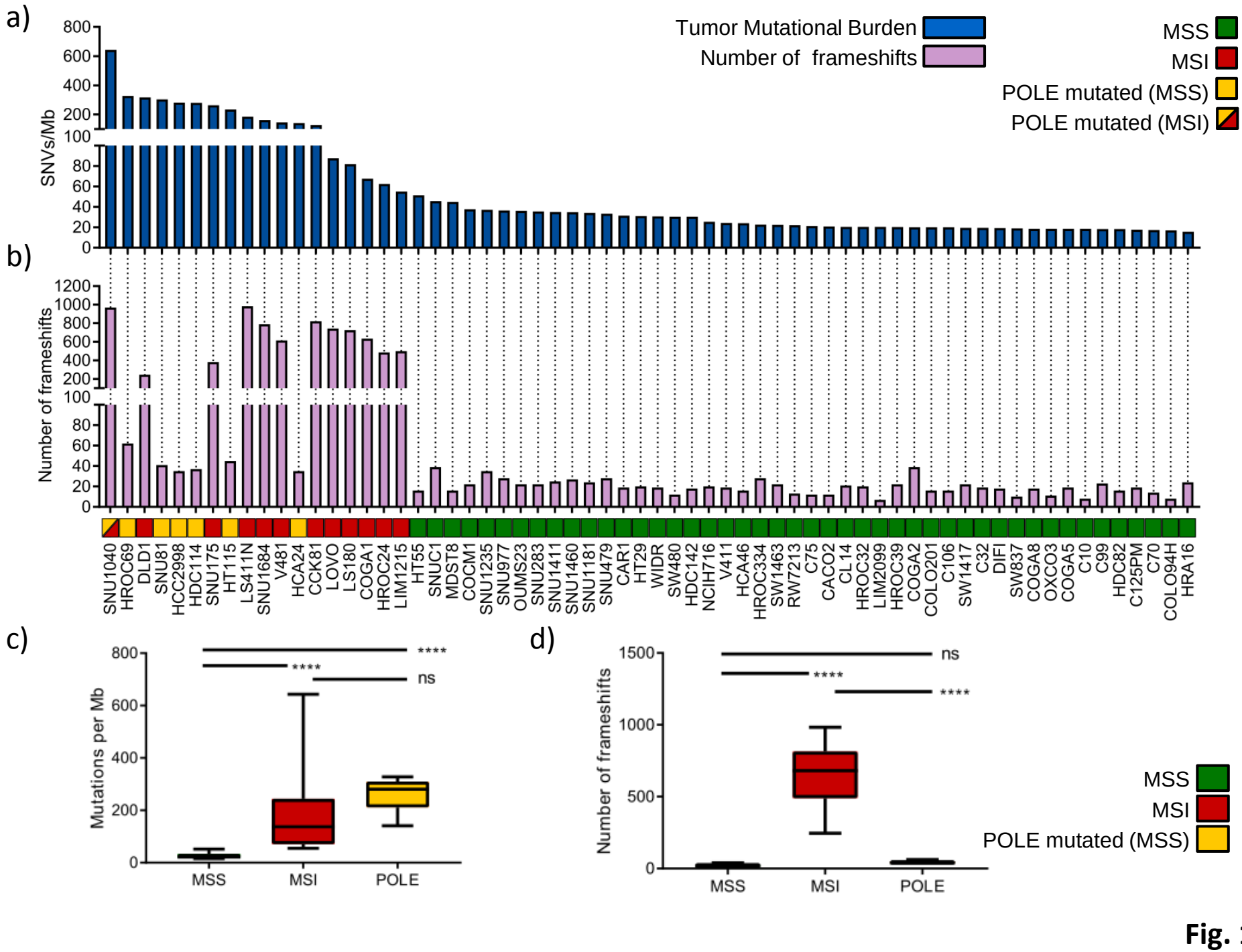

a)
Tumor Mutational Burden
Number of frameshifts
MSS
MSI
POLE mutated (MSS)
POLE mutated (MSI)
SNVs/Mb
b)
Number of frameshifts
SNU1040 HROC69 DLD1 SNU81 HCC2998 HDC114 SNU175 HT115 LS411N SNU1684 V481 HCA24 CCK81 LOVO LS180 COGA1 HROC24 LIM1215 HT55 SNUC1 MDST8 COCM1 SNU1235 SNU977 OUMS23 SNU283 SNU1411 SNU1460 SNU1181 SNU479 CAR1 HT29 WIDR SW480 HDC142 NCIH716 V411 HCA46 HROC334 SW1463 RW7213 C75 CACO2 CL14 HROC32 LIM2099 HROC39 COGA2 COLO201 C106 SW1417 C32 DIFI SW837 COGA8 OXCO3 COGA5 C10 C99 HDC82 C125PM C70 COLO94H HRA16
c)
Mutations per Mb
****
ns
MSS
MSI
POLE
d)
Number of frameshifts
ns
****
****
MSS
MSI
POLE
MSS
MSI
POLE mutated (MSS)

Fig. 1

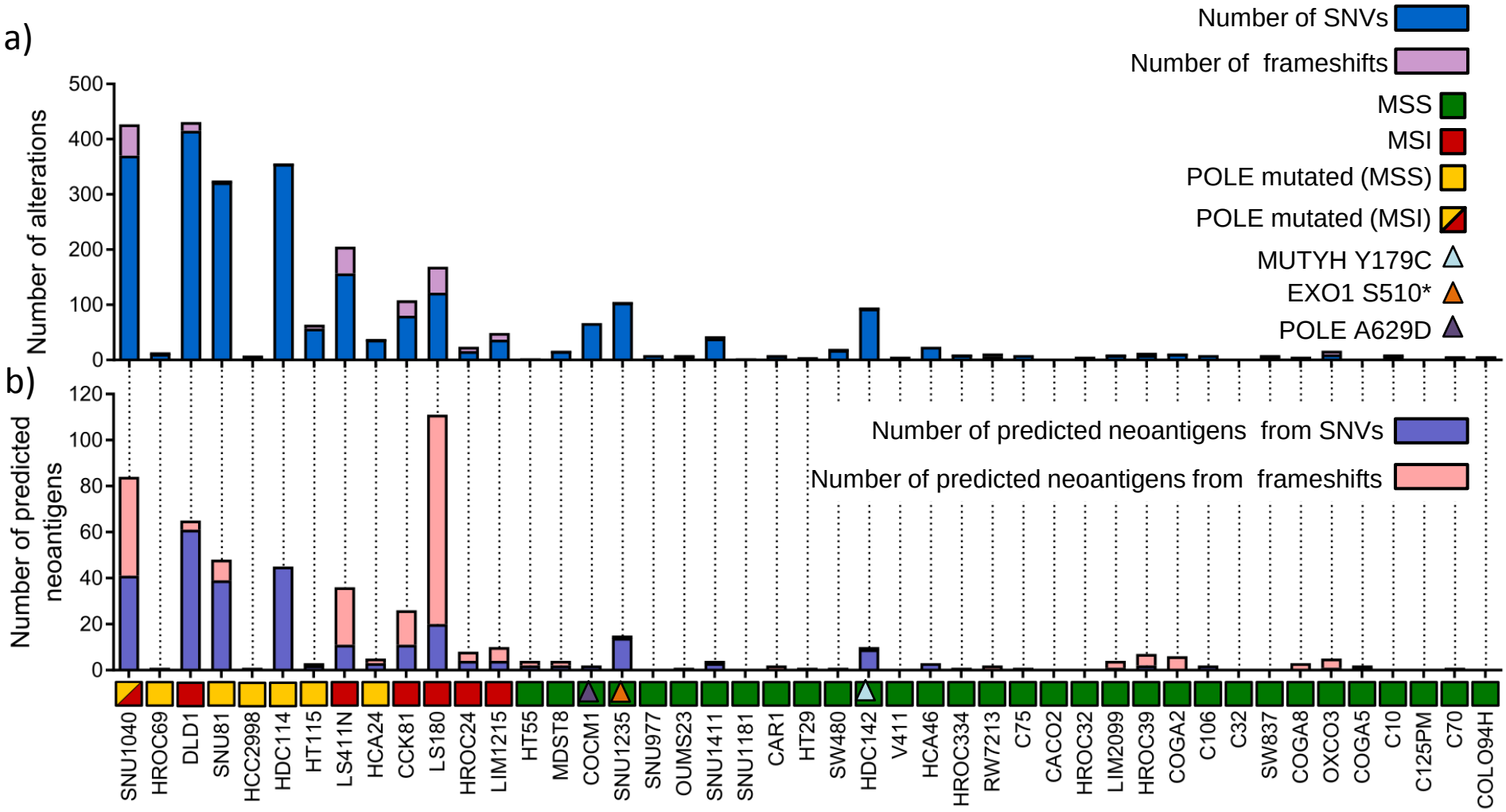

a)
Number of alterations
500
400
300
200
100
0
Number of SNVs
Number of frameshifts
MSS
MSI
POLE mutated (MSS)
POLE mutated (MSI)
MUTYH Y179C
EXO1 S510*
POLE A629D
b)
Number of predicted neoantigens
120
100
80
60
40
20
0
Number of predicted neoantigens from SNVs
Number of predicted neoantigens from frameshifts
SNU1040
HROC69
DLD1
SNU81
HCC2998
HDC114
HT115
LS411N
HCA24
CCK81
LS180
HROC24
LIM1215
HT55
MDST8
COCM1
SNU1235
SNU977
OUMS23
SNU1411
SNU1181
CAR1
HT29
SW480
HDC142
V411
HCA46
HROC334
RW7213
C75
CACO2
HROC32
LIM2099
HROC39
COGA2
C106
C32
SW837
COGA8
OXCO3
COGA5
C10
C125PM
C70
COLO94H

Fig. 2

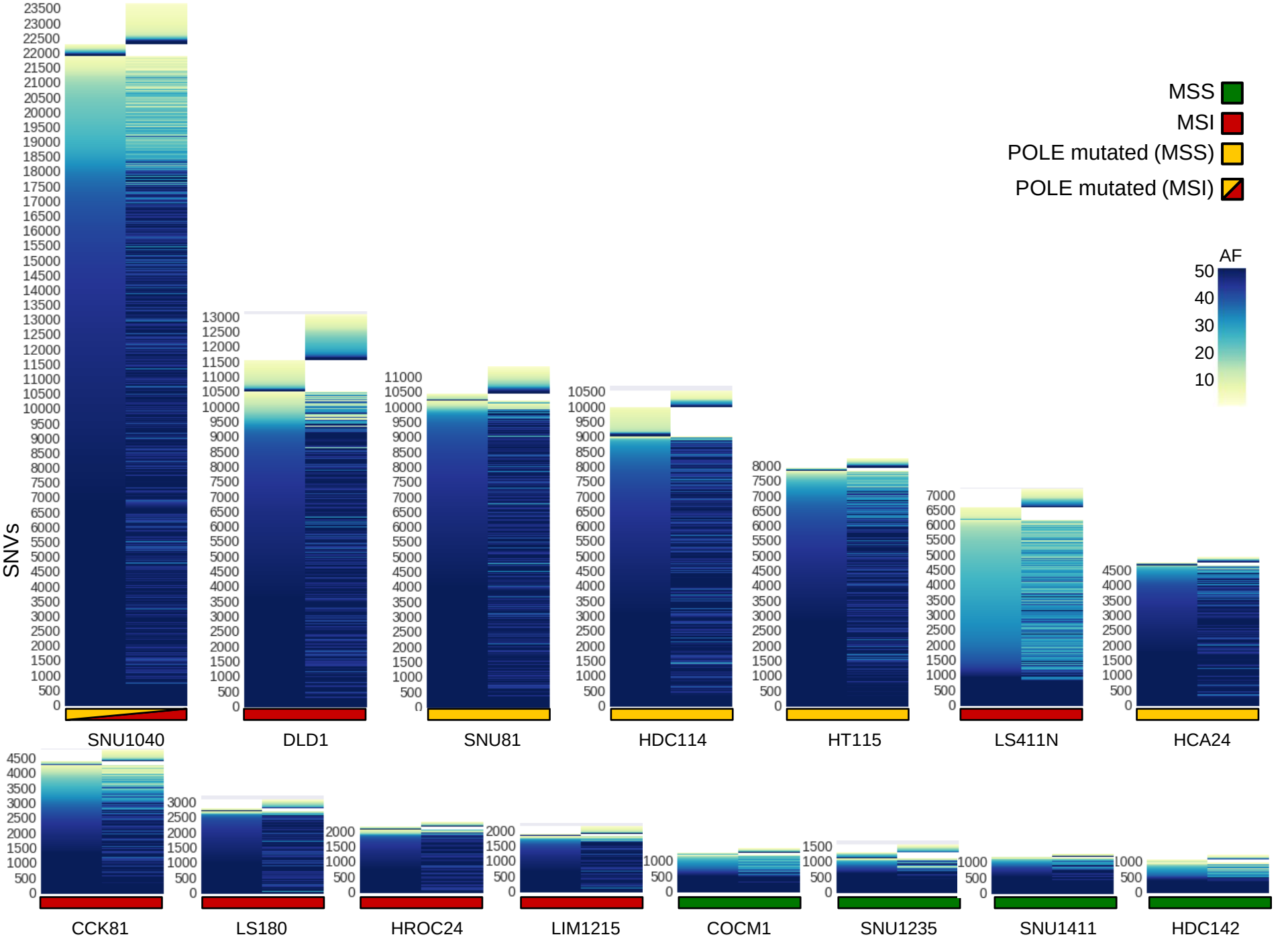

MSS
MSI
POLE mutated (MSS)
POLE mutated (MSI)
AF
50
40
30
20
10
SNVs
SNU1040
DLD1
SNU81
HDC114
HT115
LS411N
HCA24
CCK81
LS180
HROC24
LIM1215
COCM1
SNU1235
SNU1411
HDC142

Fig. 3

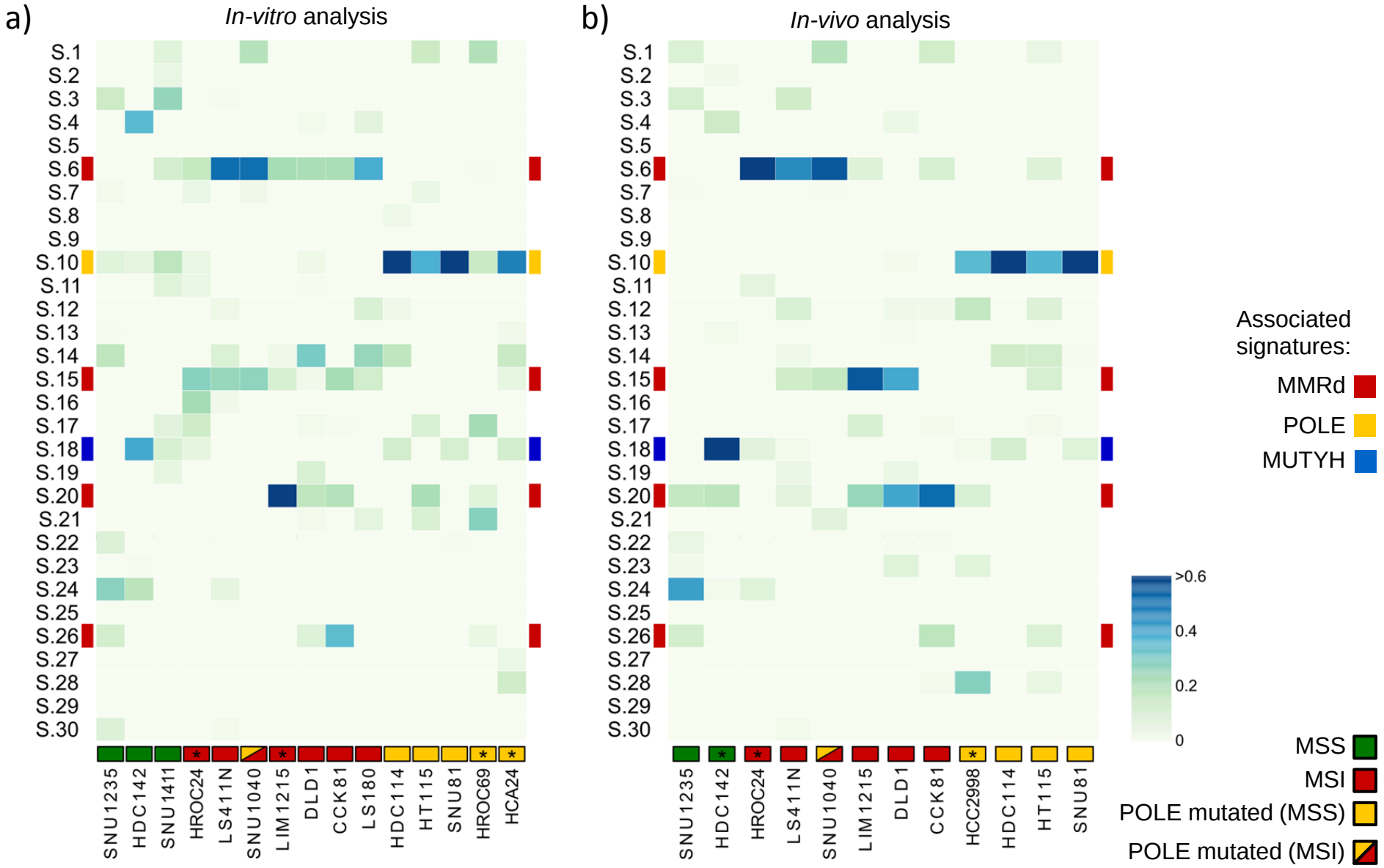

a)
In-vitro analysis
b)
In-vivo analysis
S.1
S.2
S.3
S.4
S.5
S.6
S.7
S.8
S.9
S.10
S.11
S.12
S.13
S.14
S.15
S.16
S.17
S.18
S.19
S.20
S.21
S.22
S.23
S.24
S.25
S.26
S.27
S.28
S.29
S.30
SNU1235
HDC142
SNU1411
HROC24
LS411N
SNU1040
LIM1215
DLD1
CCK81
LS180
HDC114
HT115
SNU81
HROC69
HCA24
SNU1235
HDC142
HROC24
LS411N
SNU1040
LIM1215
DLD1
CCK81
HCC2998
HDC114
HT115
SNU81
Associated signatures:
MMRd
POLE
MUTYH
>0.6
0.4
0.2
0
MSS
MSI
POLE mutated (MSS)
POLE mutated (MSI)
* Number of alterations <50

Fig. 4

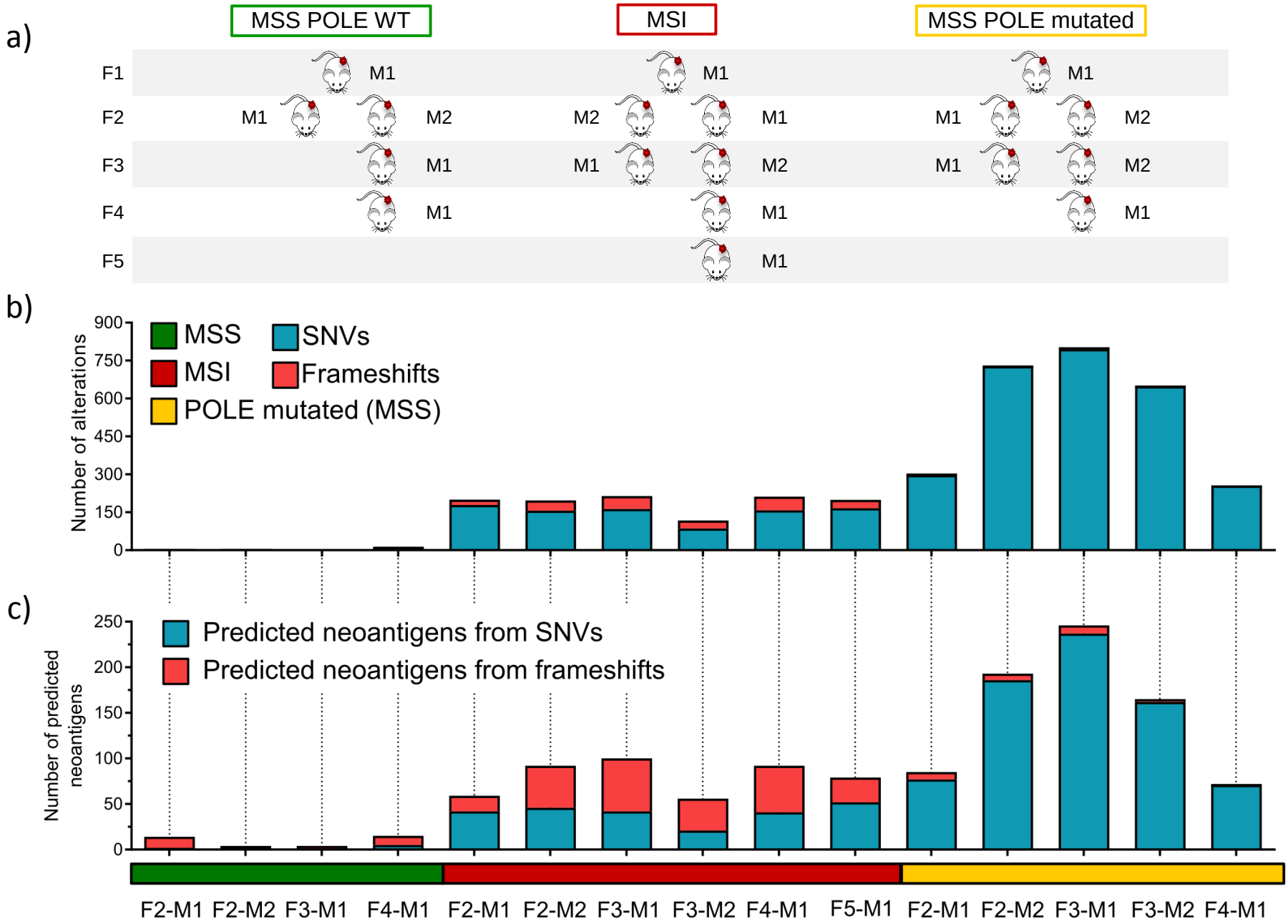

a)
MSS POLE WT
MSI
MSS POLE mutated
F1
F2
F3
F4
F5
M1
M2
b)
MSS
MSI
POLE mutated (MSS)
SNVs
Frameshifts
Number of alterations
0
150
300
450
600
750
900
c)
Predicted neoantigens from SNVs
Predicted neoantigens from frameshifts
Number of predicted neoantigens
0
50
100
150
200
250
F2-M1
F2-M2
F3-M1
F4-M1
F2-M1
F2-M2
F3-M1
F3-M2
F4-M1
F5-M1
F2-M1
F2-M2
F3-M1
F3-M2
F4-M1

Fig. 5

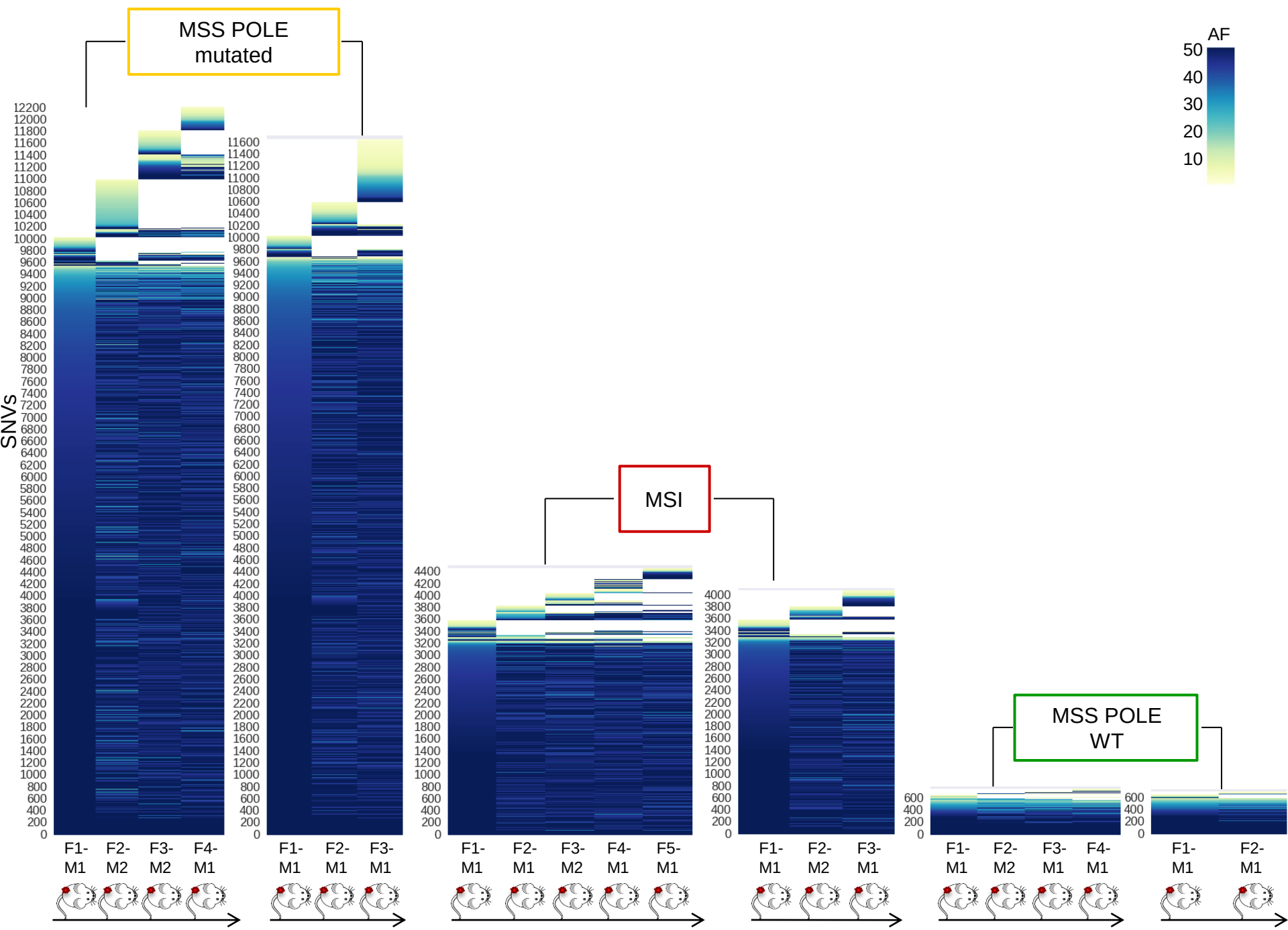

MSS POLE mutated
MSI
MSS POLE WT
AF
50
40
30
20
10
SNVs
F1-M1
F2-M2
F3-M2
F4-M1
F2-M1
F3-M1
F5-M1

Fig. 6

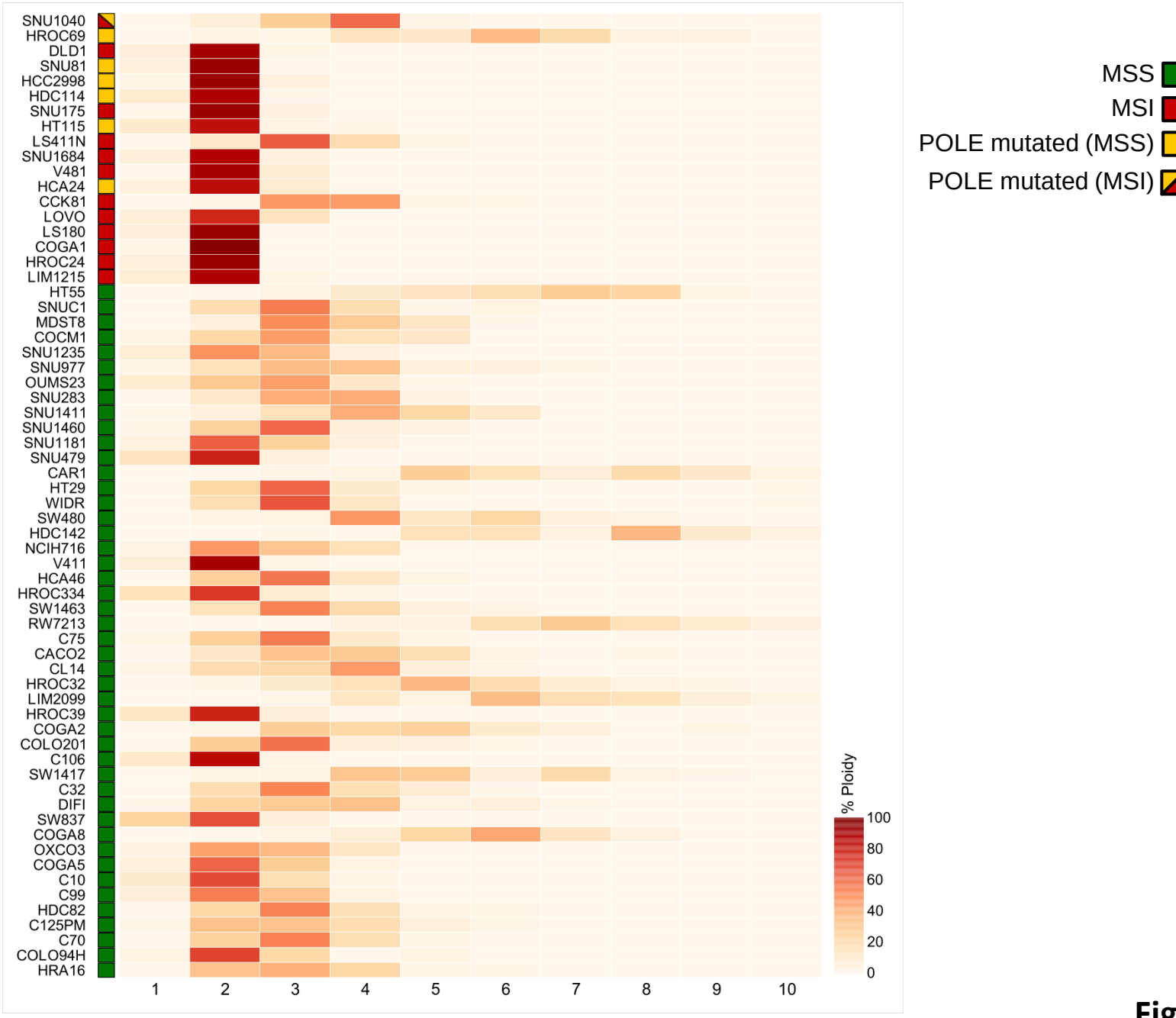

SNU1040
HROC69
DLD1
SNU81
HCC2998
HDC114
SNU175
HT115
LS411N
SNU1684
V481
HCA24
CCK81
LOVO
LS180
COGA1
HROC24
LIM1215
HT55
SNUC1
MDST8
COCM1
SNU1235
SNU977
OUMS23
SNU283
SNU1411
SNU1460
SNU1181
SNU479
CAR1
HT29
WIDR
SW480
HDC142
NCIH716
V411
HCA46
HROC334
SW1463
RW7213
C75
CACO2
CL14
HROC32
LIM2099
HROC39
COGA2
COLO201
C106
SW1417
C32
DIFI
SW837
COGA8
OXCO3
COGA5
C10
C99
HDC82
C125PM
C70
COLO94H
HRA16
1
2
3
4
5
6
7
8
9
10
% Ploidy
100
80
60
40
20
0
MSS
MSI
POLE mutated (MSS)
POLE mutated (MSI)

Fig. 7

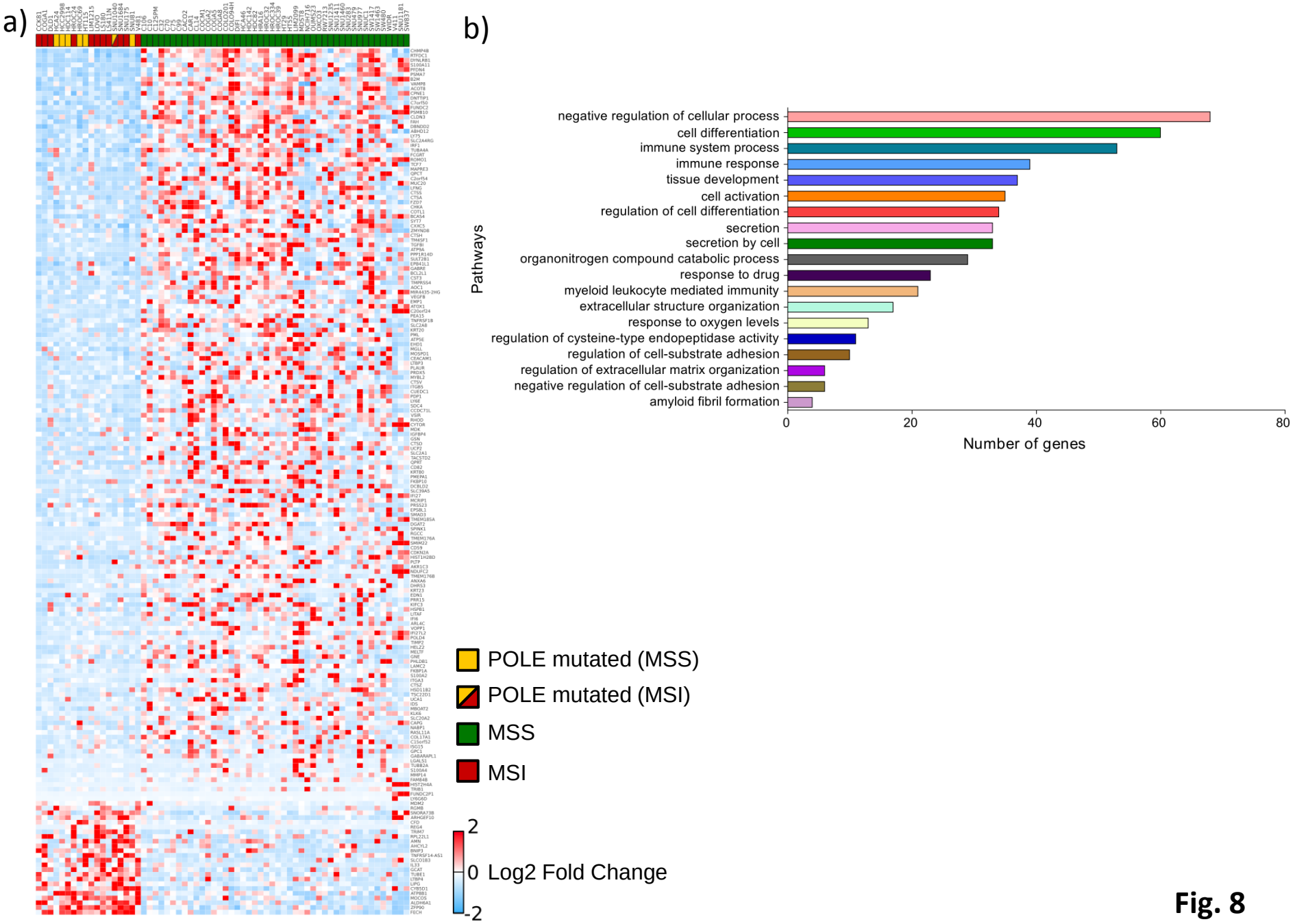

a)
b)
negative regulation of cellular process
cell differentiation
immune system process
immune response
tissue development
cell activation
regulation of cell differentiation
secretion
secretion by cell
organonitrogen compound catabolic process
response to drug
myeloid leukocyte mediated immunity
extracellular structure organization
response to oxygen levels
regulation of cysteine-type endopeptidase activity
regulation of cell-substrate adhesion
regulation of extracellular matrix organization
negative regulation of cell-substrate adhesion
amyloid fibril formation
Pathways
0
20
40
60
80
Number of genes
POLE mutated (MSS)
POLE mutated (MSI)
MSS
MSI
2
0
-2
Log2 Fold Change

Fig. 8

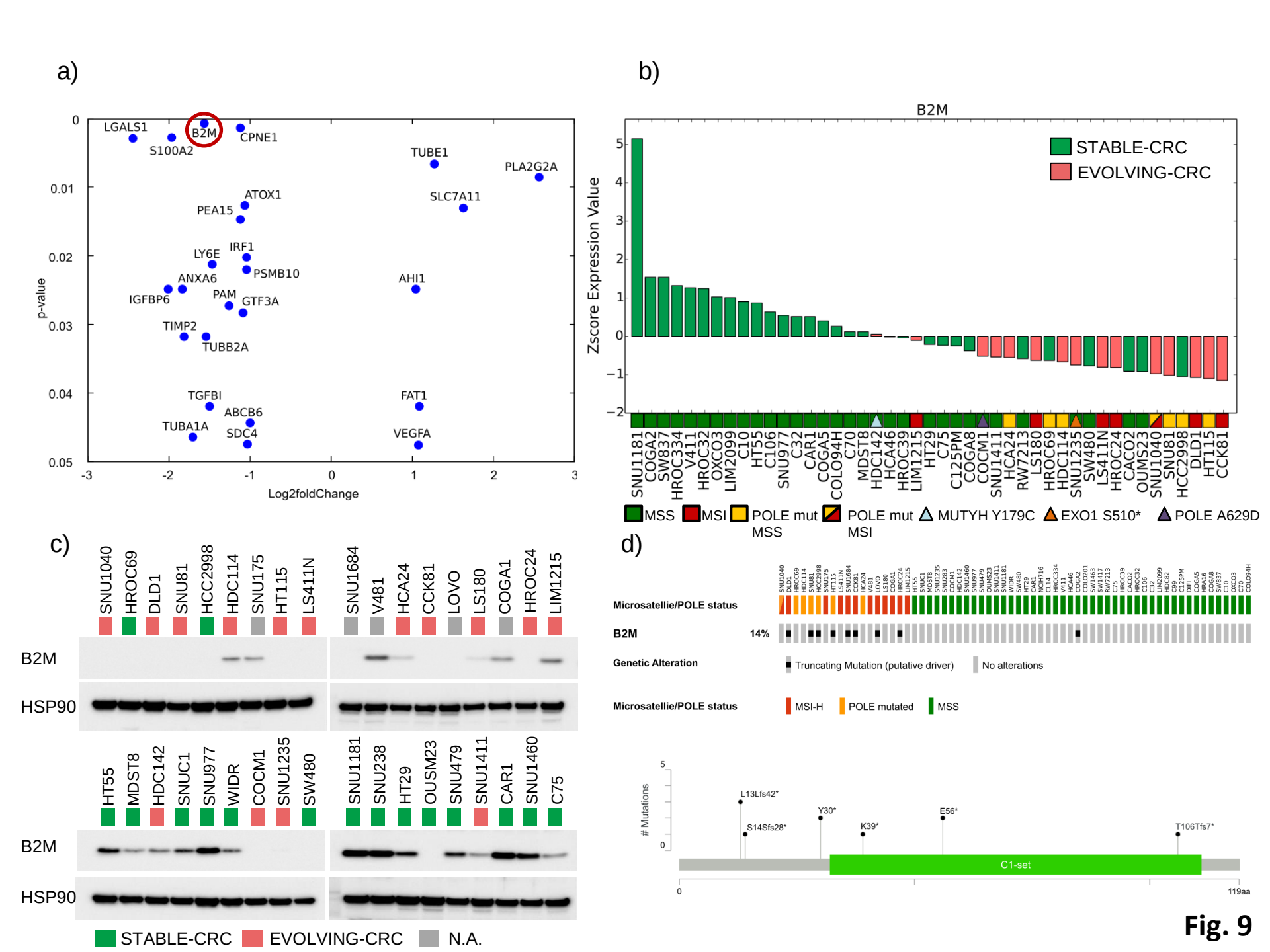
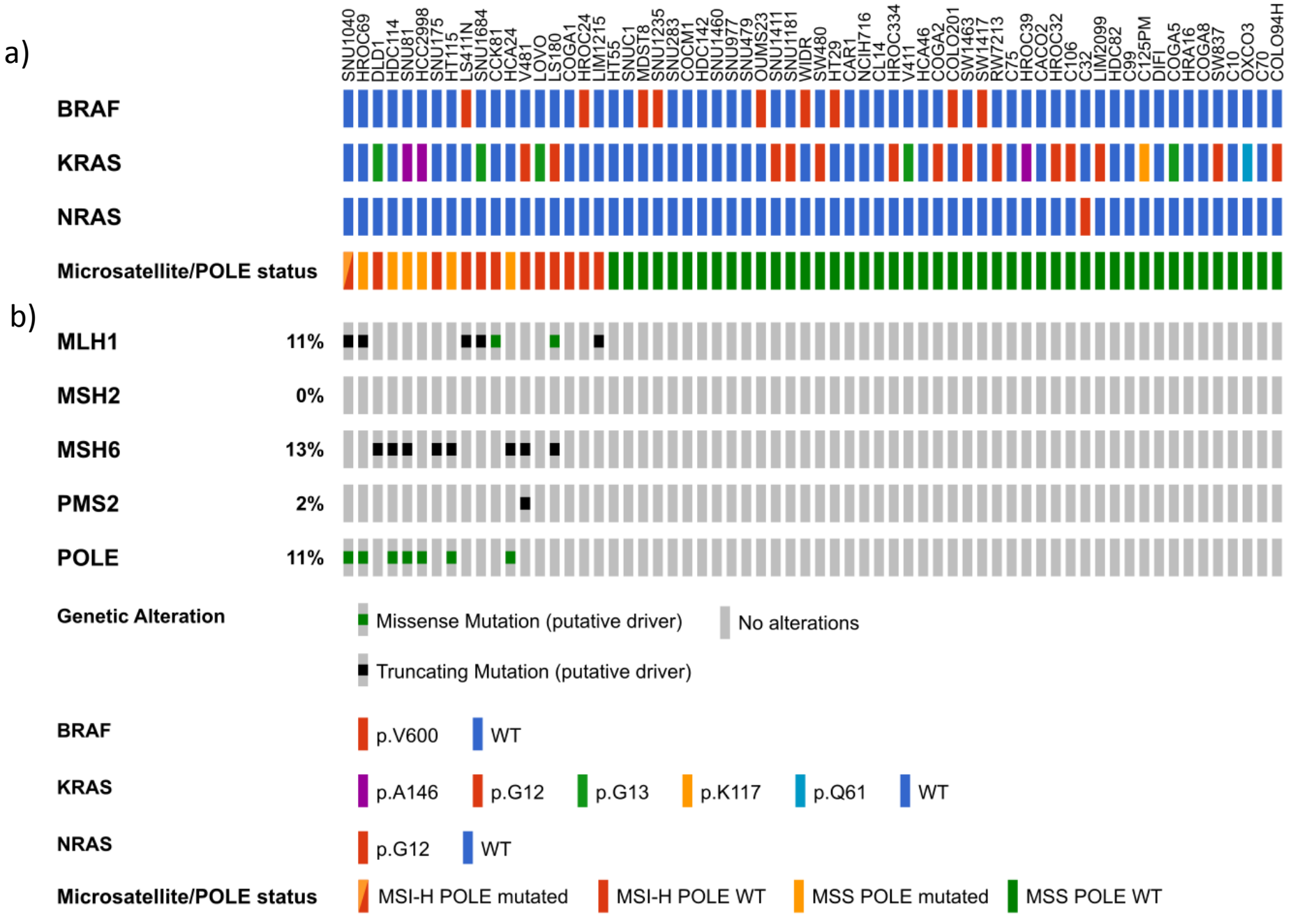
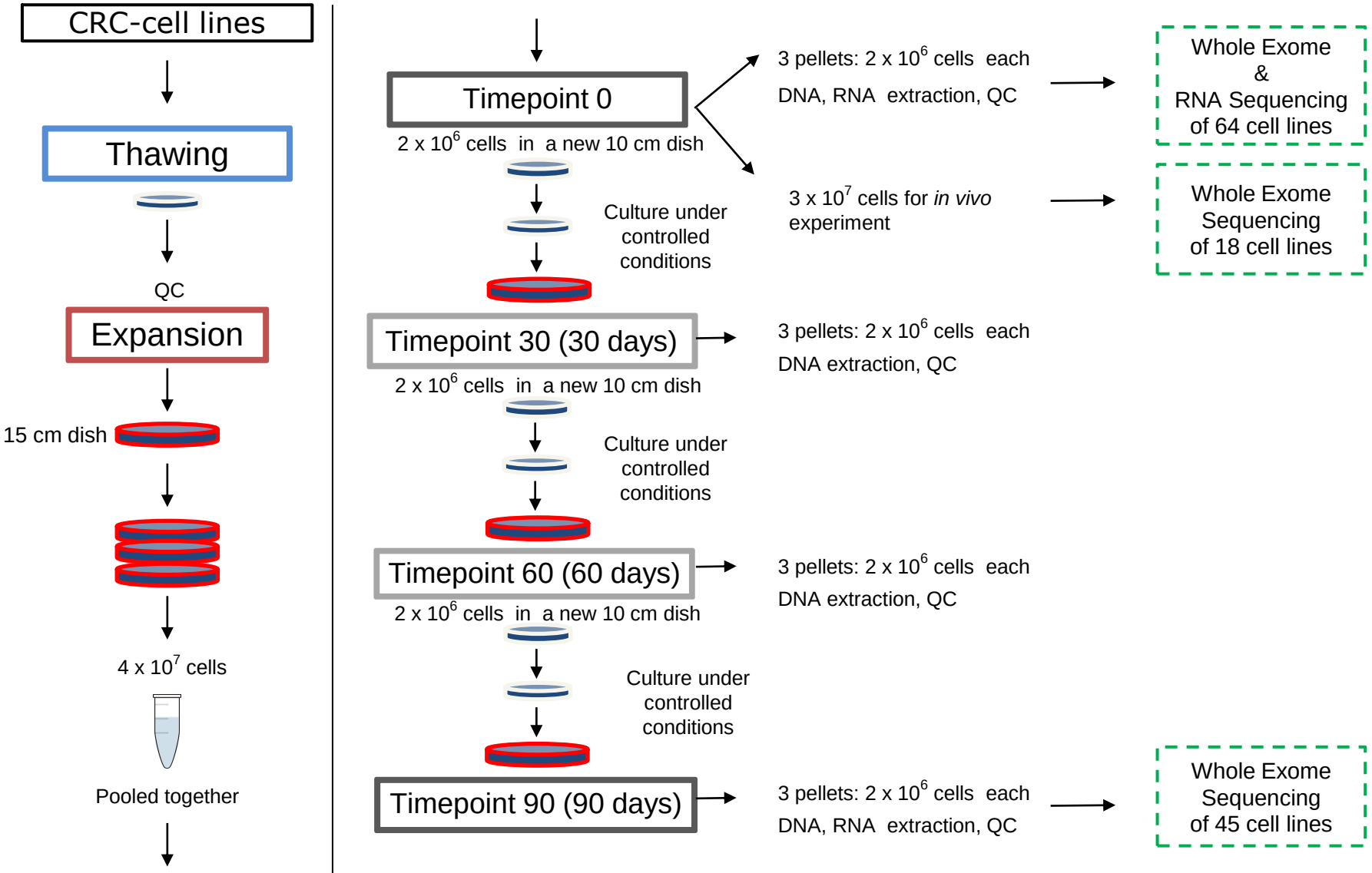

a)
b)
c)
d)
p-value
Log2foldChange
LGALS1
B2M
CPNE1
S100A2
TUBE1
PLA2G2A
ATOX1
SLC7A11
PEA15
IRF1
LY6E
PSMB10
ANXA6
AHI1
IGFBP6
PAM
GTF3A
TIMP2
TUBB2A
TGFBI
FAT1
ABCB6
TUBA1A
SDC4
VEGFA
Zscore Expression Value
STABLE-CRC
EVOLVING-CRC
MSS
MSI
POLE mut MSS
POLE mut MSI
MUTYH Y179C
EXO1 S510*
POLE A629D
SNU1040
HROC69
DLD1
SNU81
HCC2998
HDC114
SNU175
HT115
LS411N
SNU1684
V481
HCA24
CCK81
LOVO
LS180
COGA1
HROC24
LIM1215
B2M
HSP90
HT55
MDST8
HDC142
SNUC1
SNU977
WIDR
COCM1
SNU1235
SW480
SNU1181
SNU238
HT29
OUSM23
SNU479
SNU1411
CAR1
SNU1460
C75
STABLE-CRC
EVOLVING-CRC
N.A.
Microsatellie/POLE status
B2M
14%
Genetic Alteration
Truncating Mutation (putative driver)
No alterations
MSI-H
POLE mutated
MSS
Mutations
L13Lfs42*
S14Sfs28*
Y30*
K39*
E56*
T106Tfs7*
C1-set
119aa

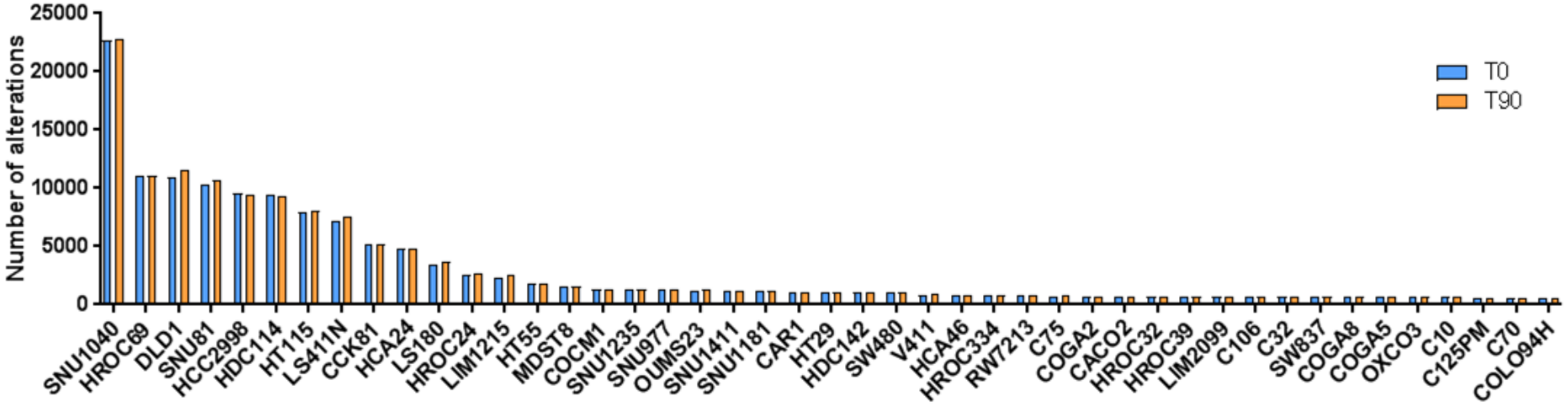
Fig. 9


a)
SNU1040
HROC69
DLD1
HDC114
SNU81
HCC2998
SNU175
HT115
LS411N
SNU1684
CCK81
HCA24
V481
LOVO
LS180
COGA1
HROC24
LIM1215
HT55
SNUC1
MDST8
SNU1235
SNU283
COCM1
HDC142
SNU1460
SNU977
SNU479
OUMS23
SNU1411
SNU1181
WIDR
SW480
HT29
CAR1
NCIH716
CL14
HROC334
V411
HCA46
COGA2
COLO201
SW1463
SW1417
RW7213
C75
HROC39
CACO2
HROC32
C106
C32
LIM2099
HDC82
C99
C125PM
DIFI
COGA5
HRA16
COGA8
SW837
C10
OXCO3
C70
COLO94H
BRAF
KRAS
NRAS
Microsatellite/POLE status
b)
MLH1 11%
MSH2 0%
MSH6 13%
PMS2 2%
POLE 11%
Genetic Alteration
Missense Mutation (putative driver)
No alterations
Truncating Mutation (putative driver)
BRAF
p.V600
WT
KRAS
p.A146
p.G12
p.G13
p.K117
p.Q61
WT
NRAS
p.G12
WT
Microsatellite/POLE status
MSI-H POLE mutated
MSI-H POLE WT
MSS POLE mutated
MSS POLE WT

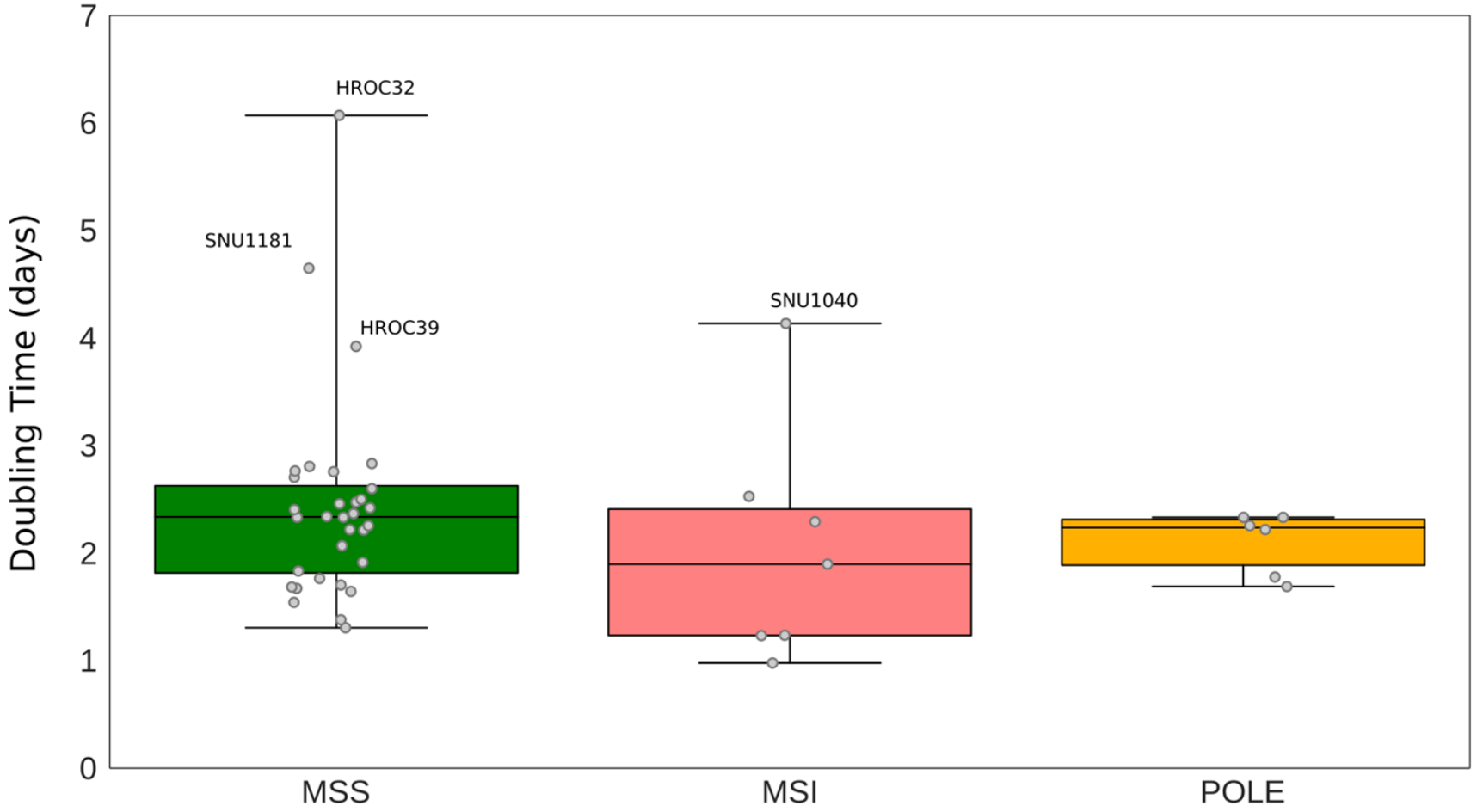
Suppl. Fig. 1


CRC-cell lines
Thawing
QC
Expansion
15 cm dish
4 x 10^7 cells
Pooled together
Timepoint 0
2 x 10^6 cells in a new 10 cm dish
Culture under controlled conditions
Timepoint 30 (30 days)
2 x 10^6 cells in a new 10 cm dish
Culture under controlled conditions
Timepoint 60 (60 days)
2 x 10^6 cells in a new 10 cm dish
Culture under controlled conditions
Timepoint 90 (90 days)
3 pellets: 2 x 10^6 cells each
DNA, RNA extraction, QC
3 x 10^7 cells for in vivo experiment
3 pellets: 2 x 10^6 cells each
DNA extraction, QC
3 pellets: 2 x 10^6 cells each
DNA extraction, QC
3 pellets: 2 x 10^6 cells each
DNA, RNA extraction, QC
Whole Exome & RNA Sequencing of 64 cell lines
Whole Exome Sequencing of 18 cell lines
Whole Exome Sequencing of 45 cell lines

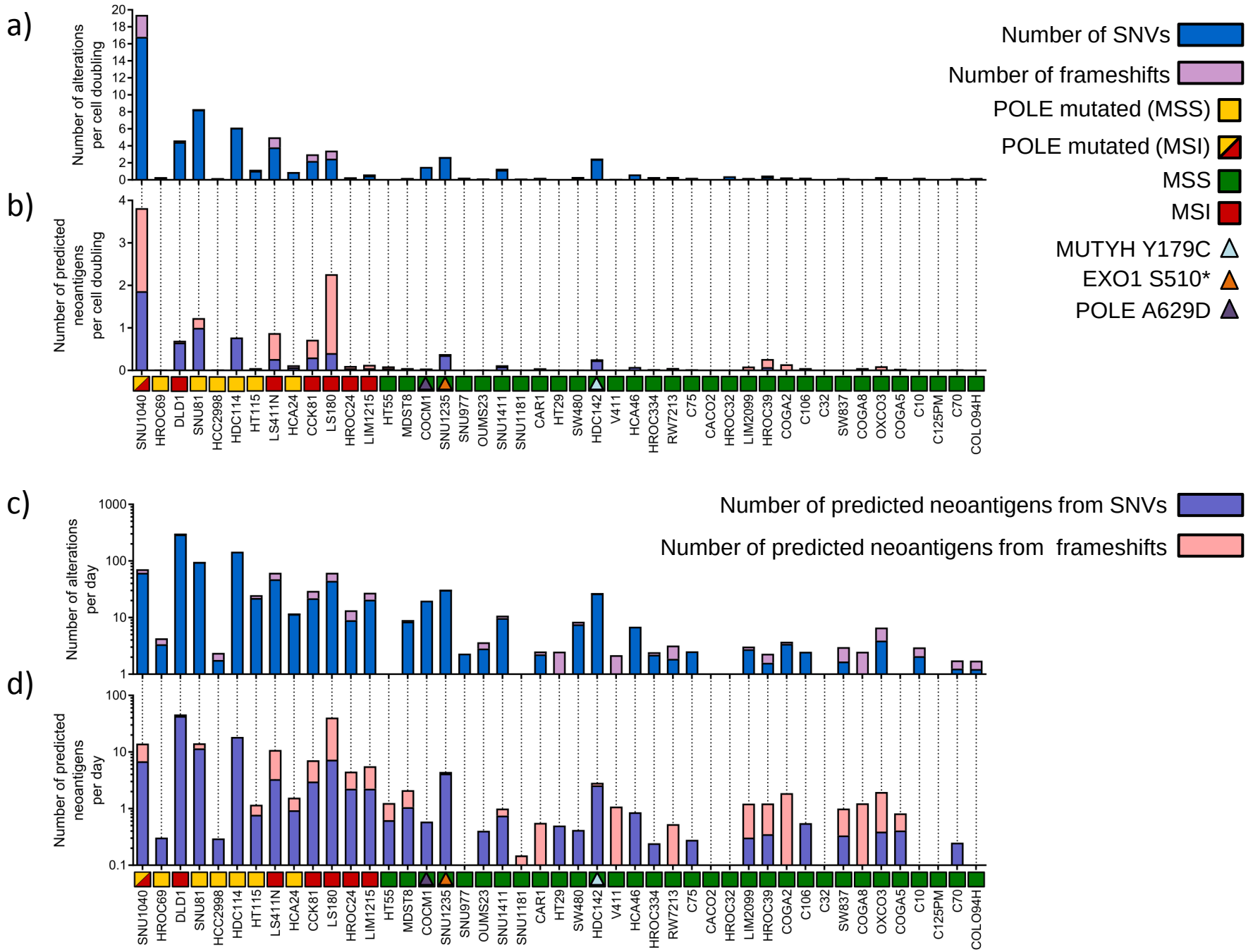
Suppl. Fig. 2


Number of alterations
25000
20000
15000
10000
5000
0
T0
T90
SNU1040
HROC69
DLD1
SNU81
HCC2998
HDC114
HT115
LS411N
CCK81
HCA24
LS180
HROC24
LIM1215
HT55
MDST8
COCM1
SNU1235
SNU977
OUMS23
SNU1411
SNU1181
CAR1
HT29
HDC142
SW480
V411
HCA46
HROC334
RW7213
C75
COGA2
CACO2
HROC32
HROC39
LIM2099
C106
C32
SW837
COGA8
COGA5
OXCO3
C10
C125PM
C70
COLO94H

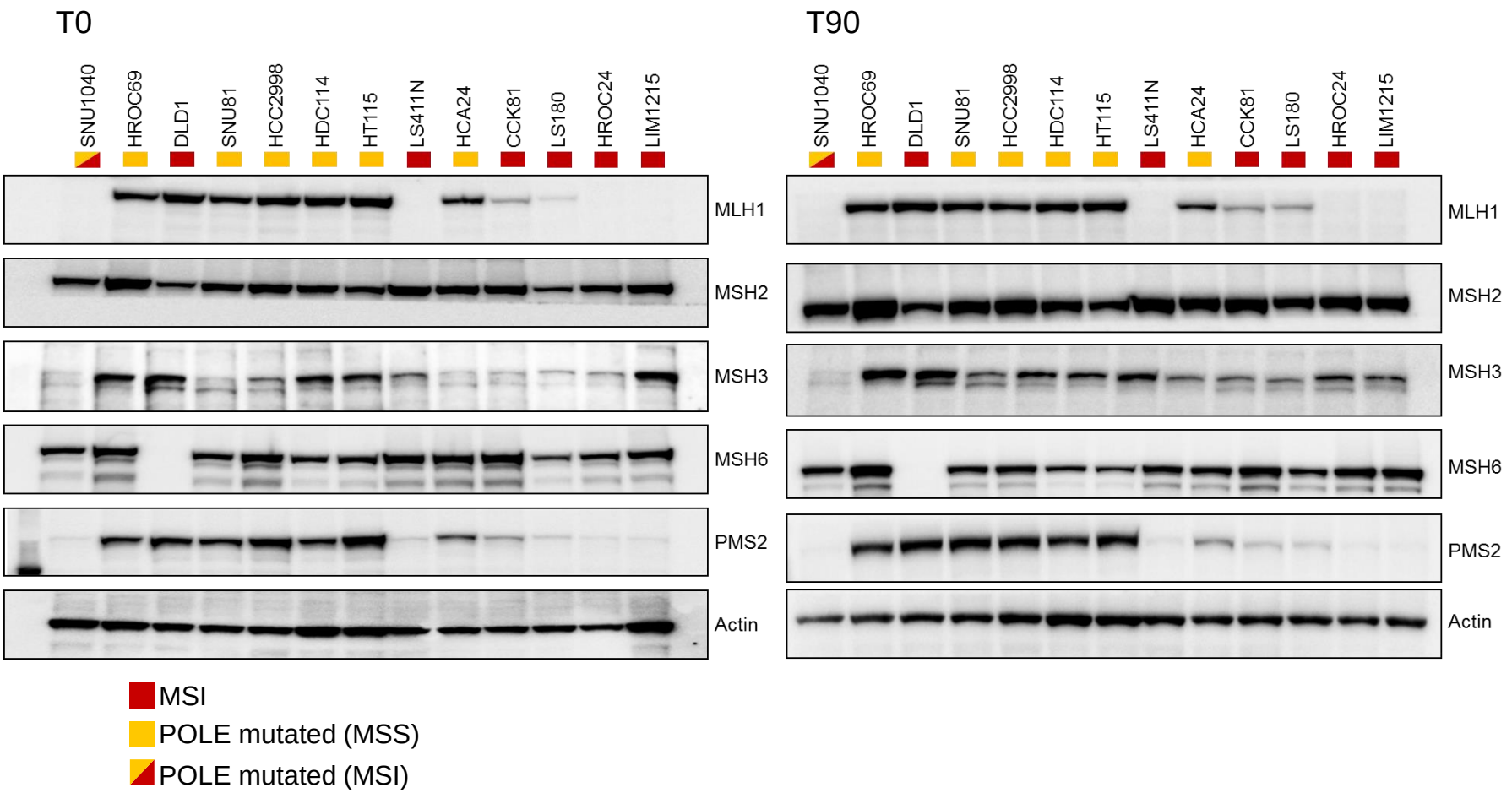
Suppl. Fig. 3


Doubling Time (days)
0
1
2
3
4
5
6
7
HROC32
SNU1181
HROC39
SNU1040
MSS
MSI
POLE

Suppl. Fig. 4


a)
Number of alterations per cell doubling
b)
Number of predicted neoantigens per cell doubling
SNU1040
HROC69
DLD1
SNU81
HCC2998
HDC114
HT115
LS411N
HCA24
CCK81
LS180
HROC24
LIM1215
HT55
MDST8
COCM1
SNU1235
SNU977
OUMS23
SNU1411
SNU1181
CAR1
HT29
SW480
HDC142
V411
HCA46
HROC334
RW7213
C75
CACO2
HROC32
LIM2099
HROC39
COGA2
C106
C32
SW837
COGA8
OXCO3
COGA5
C10
C125PM
C70
COL94H
Number of SNVs
Number of frameshifts
POLE mutated (MSS)
POLE mutated (MSI)
MSS
MSI
MUTYH Y179C
EXO1 S510*
POLE A629D
c)
Number of alterations per day
d)
Number of predicted neoantigens per day
Number of predicted neoantigens from SNVs
Number of predicted neoantigens from frameshifts

Suppl. Fig. 5


T0
T90
SNU1040
HROC69
DLD1
SNU81
HCC2998
HDC114
HT115
LS411N
HCA24
CCK81
LS180
HROC24
LIM1215
MLH1
MSH2
MSH3
MSH6
PMS2
Actin
MSI
POLE mutated (MSS)
POLE mutated (MSI)

Suppl. Fig. 6

POLE A629D

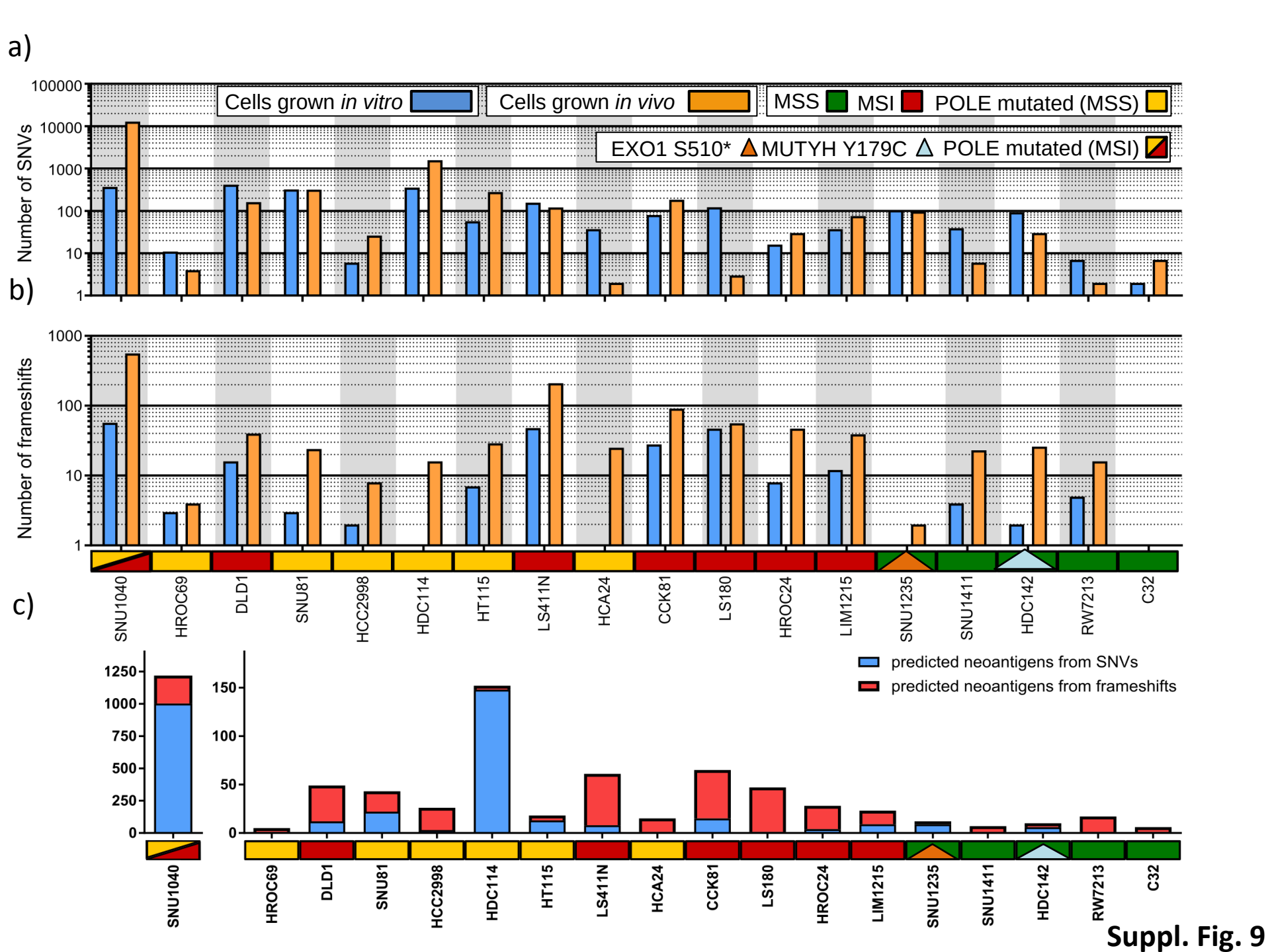
Species	Gene	Protein
H.sapiens	POLE	NP_006222.2
P.troglodytes	POLE	XP_522564.4
C.lupus	POLE	XP_543348.3
B.taurus	POLE	NP_001178358.1
M.musculus	Pole	NP_035262.2
R.norvegicus	Pole	NP_001100622.2
G.gallus	POLE	XP_004934482.1
D.rerio	pole	NP_001121995.1
D.melanogaster	DNApol-epsilon	NP_524462.2
A.gambiae	AgaP_AGAP004615	XP_315205.4
C.elegans	F33H2.5	NP_493616.1
S.cerevisiae	POL2	NP_014137.1
K.lactis	KLLA0C02585g	XP_452312.1
E.gossypii	AGOS_AFR657C	NP_986205.2
S.pombe	cdc20	NP_596354.1
318829	MGG_03850	XP_003720041.1
N.crassa	NCU04548	XP_955939.2
A.thaliana	TIL2	NP_180280.2
A.thaliana	TIL1	NP_172303.5
O.sativa	Os02g0511900	NP_001046939.2
X.tropicalis	pole	XP_004910692.1

```
NP_006222.2     603 IKSKLASLKDVPSRIECPLIYHLDVGAMYPNIILTNRLQPSAMVDEATCA 652
XP_522564.4     659 IKSKLASLKDVPSRIECPLIYHLDVGAMYPNIILTNRLQPSAMVDEATCA 708
XP_543348.3     560 IKTKLTSLKDVPNRIECPLIYHLDVGAMYPNIILTNRLQPSAVVDEATCA 609
NP_001178358.1  603 IKTKLTSLKDVPNRIECPLIYHLDVGAMYPNIILTNRLQPSAMVDEATCA 652
NP_035262.2     603 IKTKLTSLKDVPNRIECPLIYHLDVGAMYPNIILTNRLQPSAIVDEATCA 652
NP_001100622.2  603 IKTKLTSLKDVPNRIECPLIYHLDVGAMYPNIILTNRLQPSAIVDEATCA 652
XP_004934482.1  628 IKVKLNSLKDVPNRIECPLIYHLDVGAMYPNIILTNRLQPSAMVDEATCA 677
NP_001121995.1  602 IKRKLISLKEVPNRIECPLIYHLDVGAMYPNIILTNRLQPSAMVDEATCA 651
NP_524462.2     601 IVQGLQGLHDIPNRLEQPVIYHLDVGAMYPNIILTNRLQPSAMVSDLDCA 650
XP_315205.4     597 ILSALQQLYDIPTRLEQPVIYHLDVGAMYPNIILTNRLQPSSMVNDADCA 646
NP_493616.1     588 VQDAFDGMINVPTRLENPRIYHLDVGAMYPNIILTNRLQPCAMVTEEICM 637
NP_014137.1     617 ITQKLLELKENNIRNELPLIYHVDVASMYPNIMTTNRLQPDSIKAERDCA 666
XP_452312.1     594 ITAQLTDLKINNKRNELPLIYHVDVASMYPNIMTTNRLQPDSMKDEKDCA 643
NP_986205.2     591 ITEQLQELKMNNKRKELPLIYHVDVASMYPNIMTTNRLQPDSMKTERDCA 640
NP_596354.1     603 ILKKLSDLRDRPKRSEKPRIYHLDVASMYPNIMITNRLQPDSVKDESFCA 652
XP_003720041.1  621 IKAKLLALRDTPNRQERPLIYHLDVASMYPNIMTTNRLQPDSMISEADCA 670
XP_955939.2     621 ITARLMNLKETPNRLERPLIYHLDVASMYPNIMTTNRLQPDSMISESTCA 670
NP_180280.2     567 IREKLEKLRDDPIREEGPLIYHLDVAAMYPNIILTNRLQPPSIVTDEVCT 616
NP_172303.5     575 IKEKLEKLRDDPIREEGPLIYHLDVAAMYPNIILTNRLQPPSIVTDEICT 624
NP_001046939.2  606 IKQKLVSLRDHPIREECPLIYHLDVAAMYPNIILTNRLQPPSIVTDVDCT 655
XP_004910692.1  601 IKKKLNSLKEVPNRIECPLIYHLDVGAMYPNIILTNRLQPSAMVDEVTCA 650
```

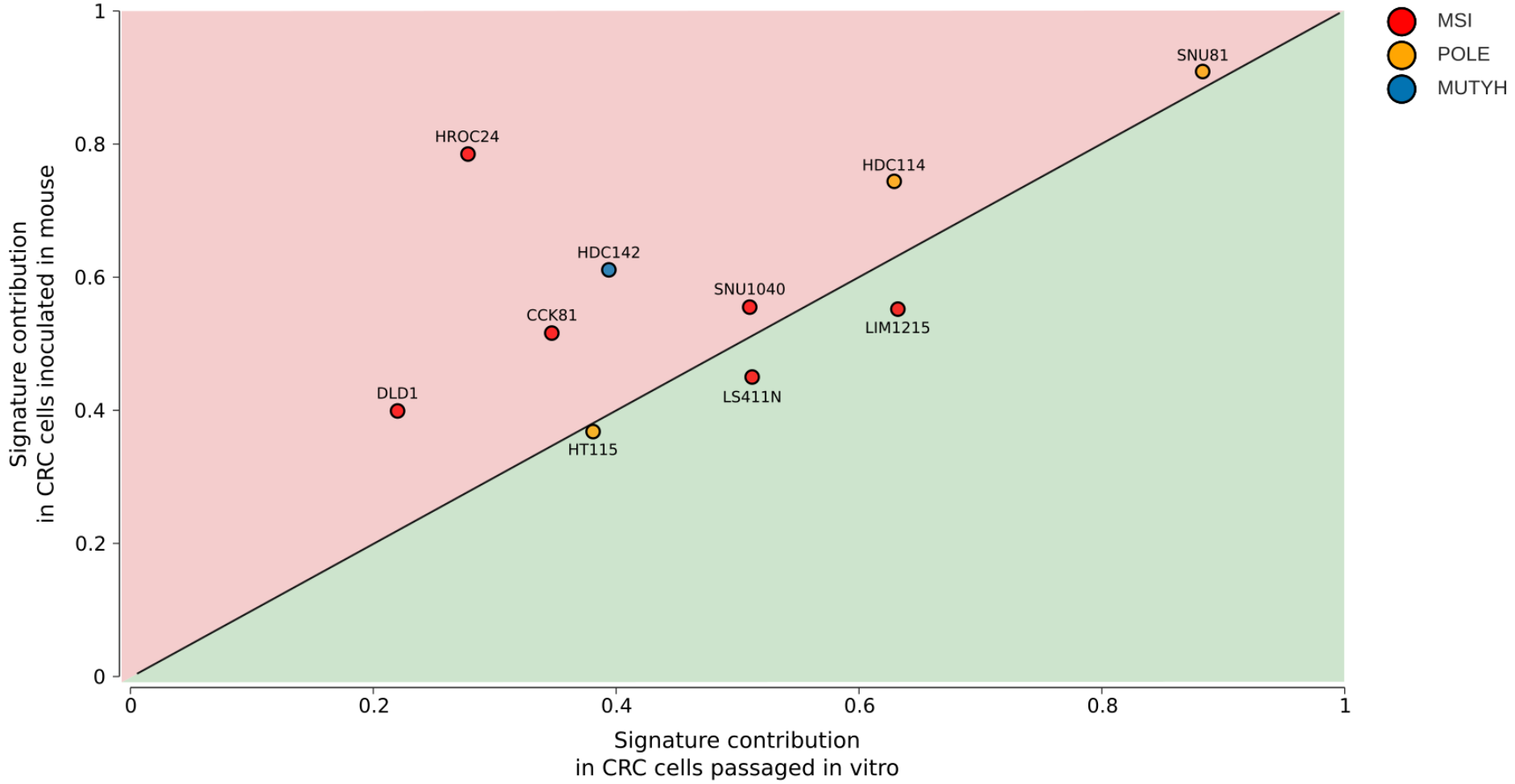
Suppl. Fig. 7

MSS
MSI
POLE mutated (MSS)
POLE mutated (MSI)
S.1
S.2
S.3
S.4
S.5
S.6
S.7
S.8
S.9
S.10
S.11
S.12
S.13
S.14
S.15
S.16
S.17
S.18
S.19
S.20
S.21
S.22
S.23
S.24
S.25
S.26
S.27
S.28
S.29
S.30
HROC69
HT115
HCC2998
HDC114
HCA24
SNU81
DLD1
HT55
MDST8
HDC142
SW480
V481
COGA2
C10
HRA16
V411
CL14
CACO2
SNUC1
COLO94H
HT29
WIDR
C70
NCIH716
COCM1
HCA46
OXCO3
SNU1460
HROC39
SNU977
OUMS23
SNU479
DIFI
LIM2099
COGA5
C125PM
COLO201
SNU1181
SW837
C32
C99
RW7213
HDC82
CAR1
HROC334
SNU283
C106
SW1463
SW1417
COGA8
HROC32
C75
SNU1235
SNU1411
SNU1040
SNU1684
LS180
HROC24
SNU175
LS411N
CCK81
LIM1215
COGA1
LOVO
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Associated signatures:
MMRd
POLE
MUTYH

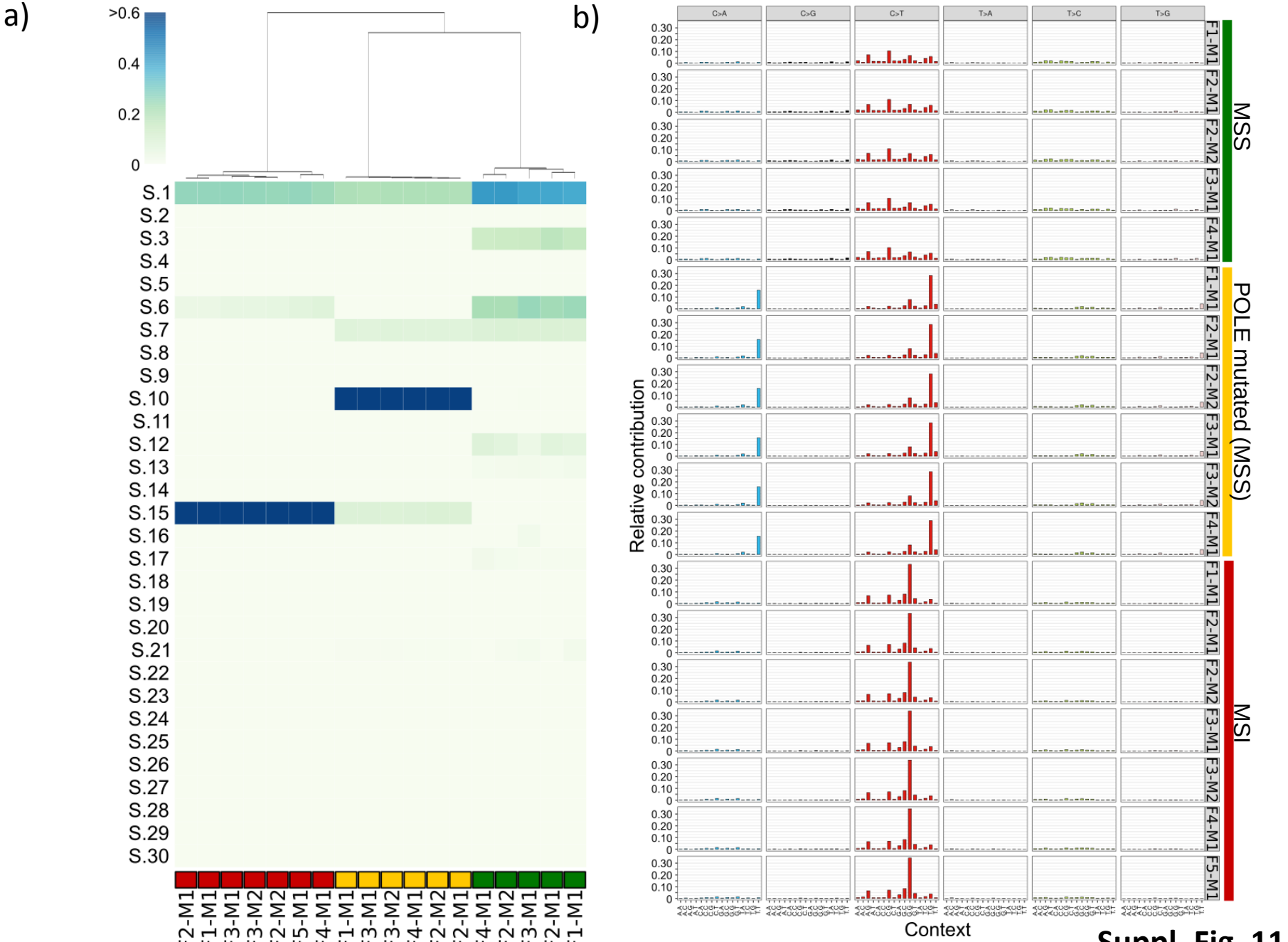
Suppl. Fig. 8


a)
b)
c)
Cells grown *in vitro*
Cells grown *in vivo*
MSS
MSI
POLE mutated (MSS)
EXO1 S510*
MUTYH Y179C
POLE mutated (MSI)
Number of SNVs
Number of frameshifts
100000
10000
1000
100
10
1
SNU1040
HROC69
DLD1
SNU81
HCC2998
HDC114
HT115
LS411N
HCA24
CCK81
LS180
HROC24
LIM1215
SNU1235
SNU1411
HDC142
RW7213
C32
predicted neoantigens from SNVs
predicted neoantigens from frameshifts
1250
1000
750
500
250
0
150
100
50
0

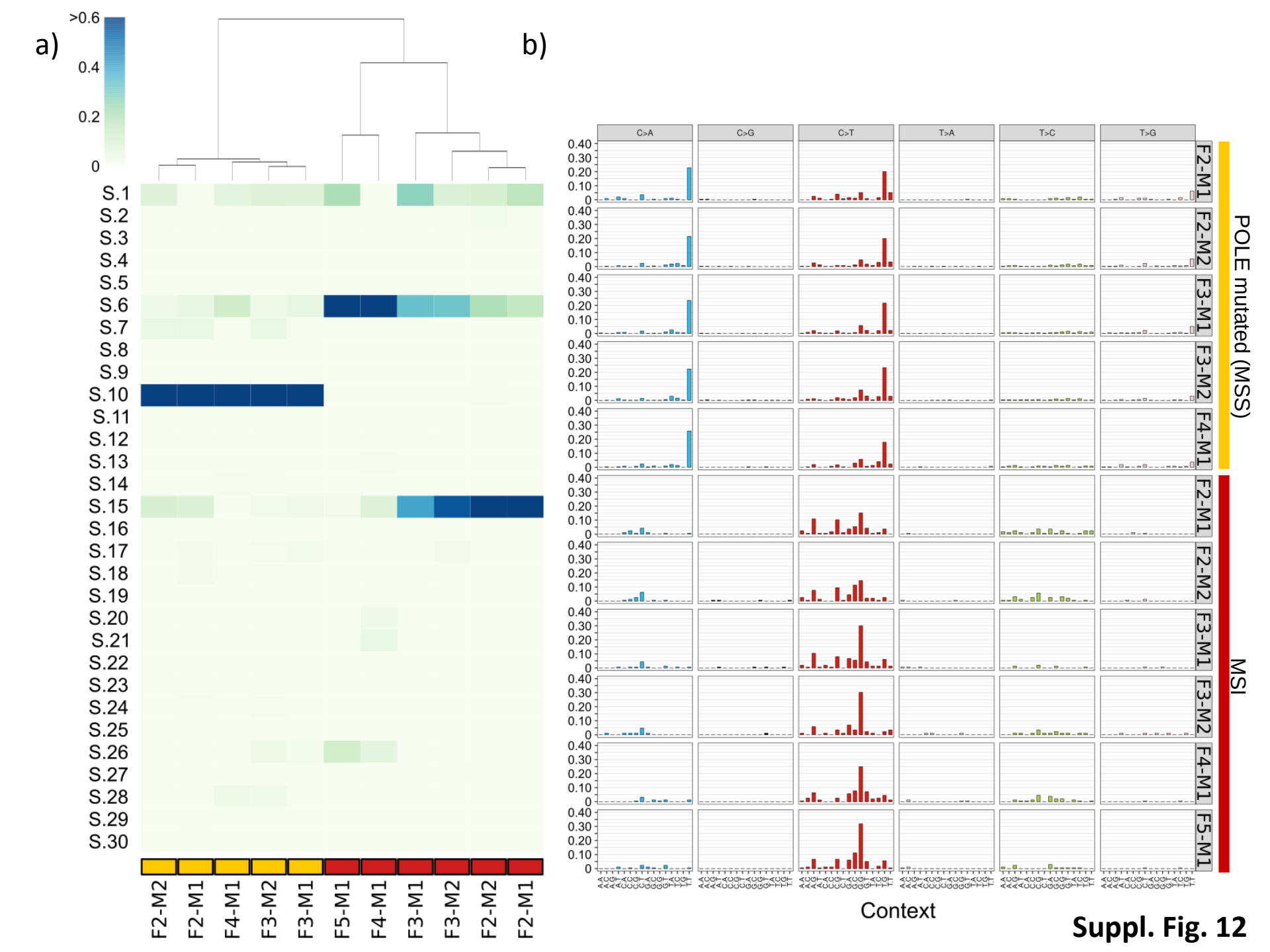
Suppl. Fig. 9


MSI
POLE
MUTYH
SNU81
HROC24
HDC114
HDC142
SNU1040
CCK81
LIM1215
DLD1
LS411N
HT115
Signature contribution
in CRC cells inoculated in mouse
Signature contribution
in CRC cells passaged in vitro
0
0.2
0.4
0.6
0.8
1

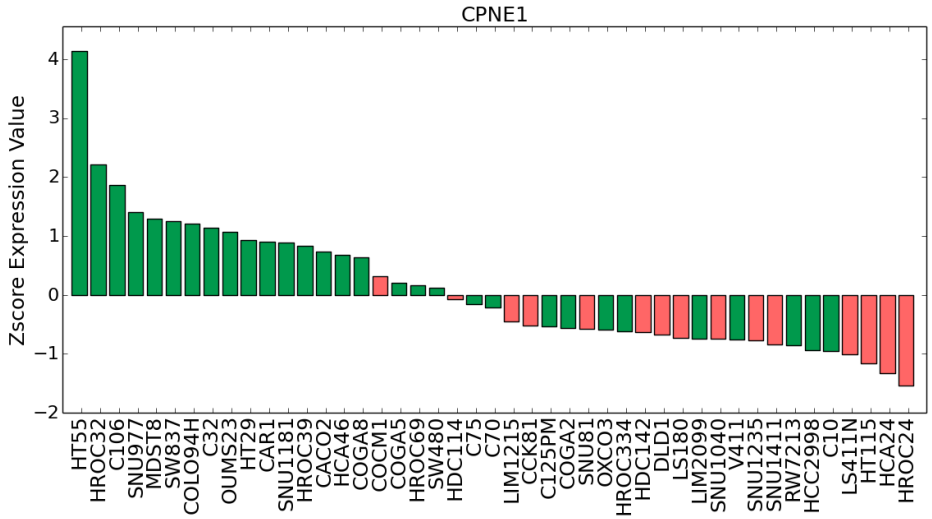
Suppl. Fig. 10

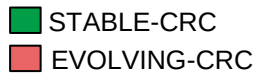

a)
b)
>0.6
0.4
0.2
0
S.1
S.2
S.3
S.4
S.5
S.6
S.7
S.8
S.9
S.10
S.11
S.12
S.13
S.14
S.15
S.16
S.17
S.18
S.19
S.20
S.21
S.22
S.23
S.24
S.25
S.26
S.27
S.28
S.29
S.30
F2-M1
F1-M1
F3-M1
F3-M2
F2-M2
F5-M1
F4-M1
F1-M1
F3-M1
F3-M2
F4-M1
F2-M2
F2-M1
F4-M1
F2-M2
F3-M1
F2-M1
F1-M1
C>A
C>G
C>T
T>A
T>C
T>G
F1-M1
F2-M1
F2-M2
F3-M1
F4-M1
F1-M1
F2-M1
F2-M2
F3-M1
F3-M2
F4-M1
F1-M1
F2-M1
F2-M2
F3-M1
F3-M2
F4-M1
F5-M1
MSS
POLE mutated (MSS)
MSI
Relative contribution
Context

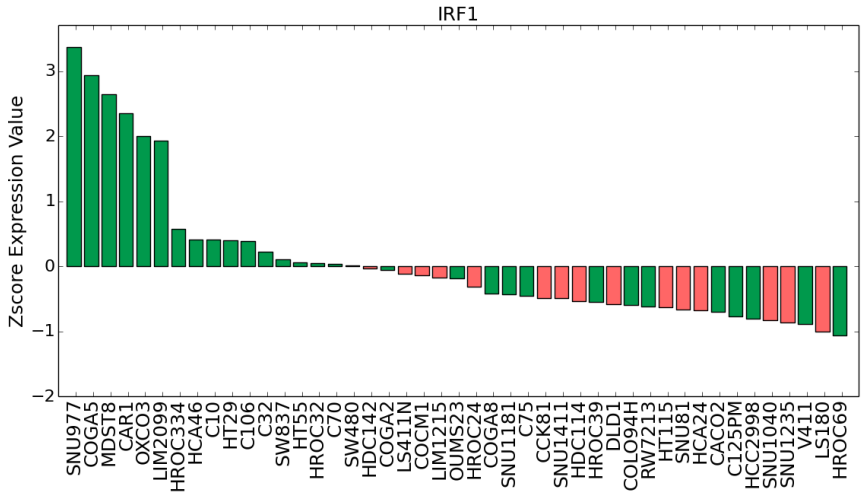
Suppl. Fig. 11

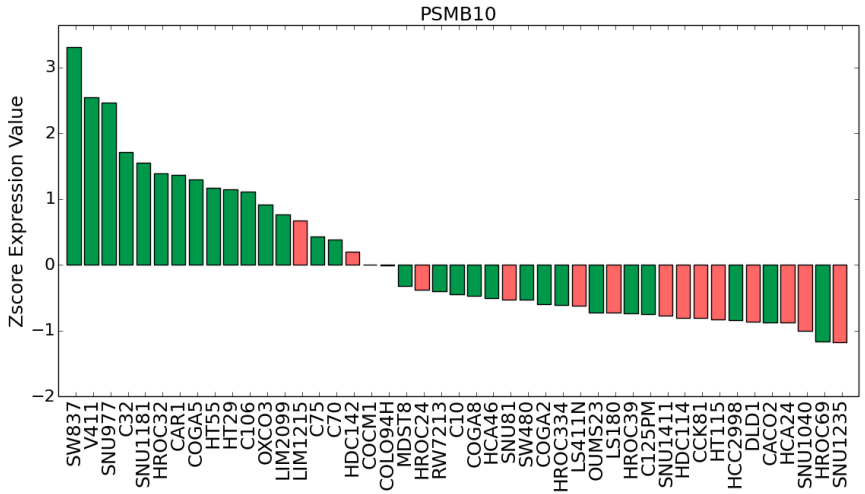

a)
b)
>0.6
0.4
0.2
0
S.1
S.2
S.3
S.4
S.5
S.6
S.7
S.8
S.9
S.10
S.11
S.12
S.13
S.14
S.15
S.16
S.17
S.18
S.19
S.20
S.21
S.22
S.23
S.24
S.25
S.26
S.27
S.28
S.29
S.30
F2-M2
F2-M1
F4-M1
F3-M2
F3-M1
F5-M1
F4-M1
F3-M1
F3-M2
F2-M2
F2-M1
C>A
C>G
C>T
T>A
T>C
T>G
F2-M1
F2-M2
F3-M1
F3-M2
F4-M1
F5-M1
POLE mutated (MSS)
MSI
Context

Suppl. Fig. 12

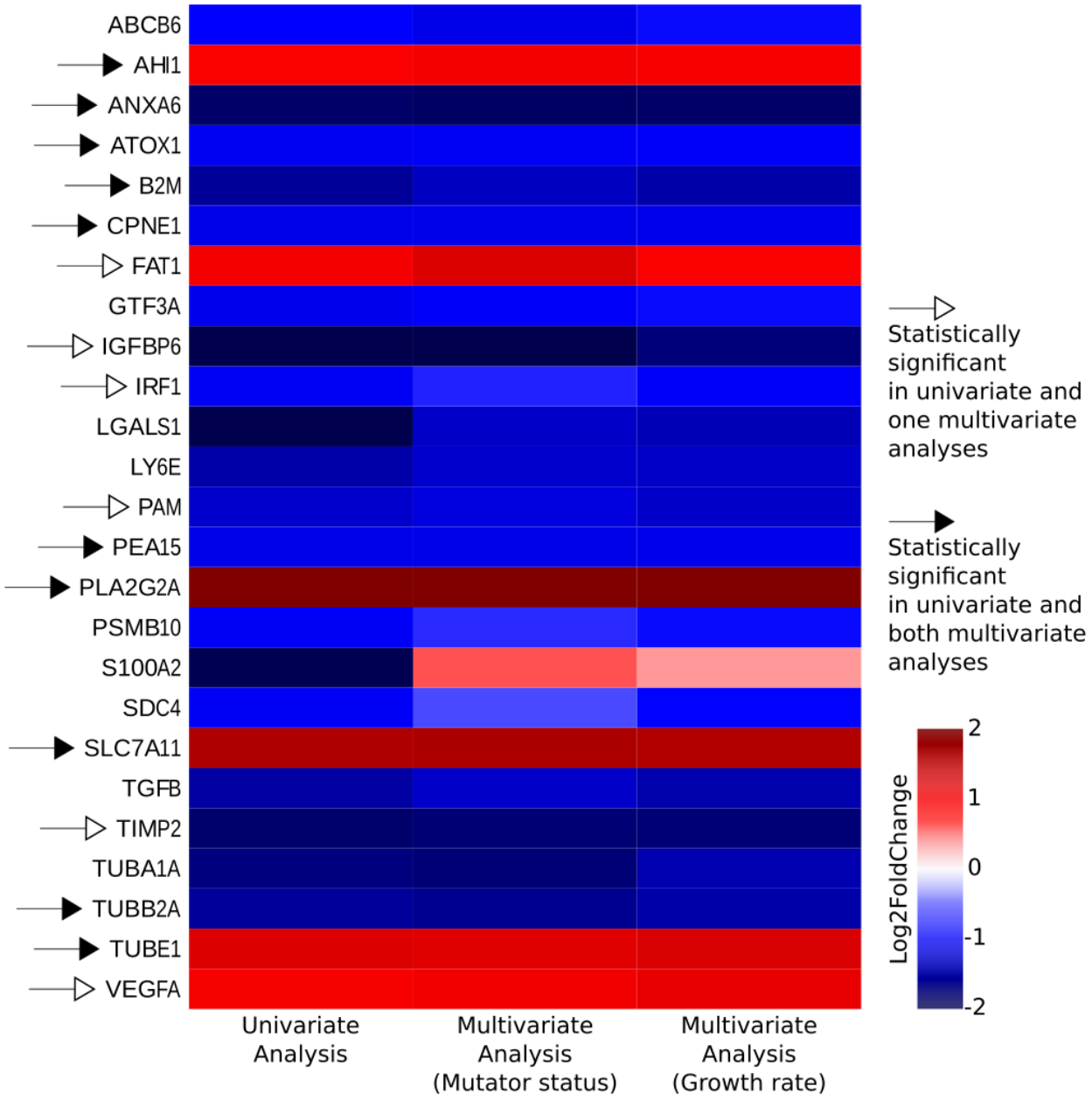

CPNE1
Zscore Expression Value
4
3
2
1
0
−1
−2
HT55
HROC32
C106
SNU977
MDST8
SW837
COLO94H
C32
OUMS23
HT29
CAR1
SNU1181
HROC39
CACO2
HCA46
COGA8
COCM1
COGA5
HROC69
SW480
HDC114
C75
C70
LIM1215
CCK81
C125PM
COGA2
SNU81
OXCO3
HROC334
HDC142
DLD1
LS180
LIM2099
SNU1040
V411
SNU1235
SNU1411
RW7213
HCC2998
C10
LS411N
HT115
HCA24
HROC24


STABLE-CRC
EVOLVING-CRC


IRF1
Zscore Expression Value
3
2
1
0
−1
−2
SNU977
COGA5
MDST8
CAR1
OXCO3
LIM2099
HROC334
HCA46
C10
HT29
C106
C32
SW837
HT55
HROC32
C70
SW480
HDC142
COGA2
LS411N
COCM1
LIM1215
OUMS23
HROC24
COGA8
SNU1181
C75
CCK81
SNU1411
HDC114
HROC39
DLD1
COLO94H
RW7213
HT115
SNU81
HCA24
CACO2
C125PM
HCC2998
SNU1040
SNU1235
V411
LS180
HROC69


PSMB10
Zscore Expression Value
3
2
1
0
−1
−2
SW837
V411
SNU977
C32
SNU1181
HROC32
CAR1
COGA5
HT55
HT29
C106
OXCO3
LIM2099
LIM1215
C75
C70
HDC142
COCM1
COLO94H
MDST8
HROC24
RW7213
C10
COGA8
HCA46
SNU81
SW480
COGA2
HROC334
LS411N
OUMS23
LS180
HROC39
C125PM
SNU1411
HDC114
CCK81
HT115
HCC2998
DLD1
CACO2
HCA24
SNU1040
HROC69
SNU1235

Suppl. Fig. 13


ABCB6
AHI1
ANXA6
ATOX1
B2M
CPNE1
FAT1
GTF3A
IGFBP6
IRF1
LGALS1
LY6E
PAM
PEA15
PLA2G2A
PSMB10
S100A2
SDC4
SLC7A11
TGFB
TIMP2
TUBA1A
TUBB2A
TUBE1
VEGFA
Univariate Analysis
Multivariate Analysis (Mutator status)
Multivariate Analysis (Growth rate)
Statistically significant in univariate and one multivariate analyses
Statistically significant in univariate and both multivariate analyses
Log2FoldChange
2
1
0
-1
-2

Suppl. Fig. 14