**Vignette Documentation**

[**Summary**](#_udkioui1l9gz) **3**

[Cancer Vignette Example](#_18v74meczzru) 4

[**Overall Schematic**](#_isxzbojdnfjk) **5**

[Summary](#_victaslz5pbw) 5

[Time and Date Stamp](#_31d23echm096) 5

[**Cohort Metadata**](#_77gx6hzfvvtc) **6**

[Summary](#_76citdlue1by) 6

[Figure Description](#_3zm8ij1cfka7) 6

[**Tumor Characteristics**](#_g8bbheildks1) **7**

[Summary](#_8mpxlj9efxyk) 7

[Figure Description](#_rhz0vbkeoo68) 7

[**Mutational Landscape**](#_wqopetz2dkvf) **8**

[Summary](#_rkbnkeoc50no) 8

[Figure Description](#_8jr7czr5eunk) 8

[**Processes Underlying Mutations**](#_a25kgz54ba0d) **10**

[Summary](#_e8imf6nr884l) 10

[Figure Description](#_xlcqa3gmberb) 10

[**Copy Number Alteration Profile**](#_q06dostvp8e7) **11**

[Summary](#_ihj5o8vs2czz) 11

[Figure Description](#_7pih7oalszr1) 11

[**Cancer Driver Landscape**](#_xwoff9fa6a75) **12**

[Summary](#_6dry6saf6woj) 12

[Figure Description](#_i28ldjk943ks) 12

[**Potentially Actionable Events**](#_u7kmplk0w09e) **14**

[Summary](#_e0tmip2bik4n) 14

[Figure Description](#_vimchnb8x2ce) 15

[**Germline Predisposition**](#_dbg3athvqybh) **16**

[Summary](#_dvngxoogdlrr) 16

[Figure Description](#_4d1mff67ermg) 16

[**WGS vs Panel Coverage**](#_n39ea47ffpr2) **17**

[Summary](#_o4tnca9i93l3) 17

[Figure Description](#_pmkqosnssy85) 18

[**Citations**](#_anqavsx3rjc4) **19**

## Summary

This document describes in detail the figures shown in the cancer vignettes. While we share many vignettes across different cancer cohorts, each uses the same underlying data from the Hartwig Medical Foundation database. Links to the Hartwig Medical Foundation database, tools, and data request guide are found here:

* [Hartwig database description](https://www.hartwigmedicalfoundation.nl/en/data/database/)
* [Overview of all Hartwig tools](https://github.com/hartwigmedical/hmftools?tab=readme-ov-file)
* [Data request access guide](https://www.hartwigmedicalfoundation.nl/en/data/data-access-request/)

It should be emphasized that the vignettes are not created to support any clinical diagnosis nor to draw solid scientific conclusions. The goal is to share the rich whole genome sequencing (WGS) data available in the Hartwig database and to inspire researchers and clinicians to take a deeper dive into potentially interesting aspects that become visible in these vignettes.

The vignettes will be updated periodically, and many cohorts contain a mixture of disease ontologies. The date of the creation and included disease ontology identifies (DOIDs) in each cohort are shared in the bottom middle of each vignette.

### 

### Cancer Vignette Example

Below shows an example of the vignette for Non-Small Cell Lung Carcinoma (NSCLC). The vignette is organized into 10 panels – Cohort Metadata, Tumor Characteristics, Mutational Landscape, Processes Underlying Mutations, Schematic figure, Copy Number Alteration Profile, Cancer Driver Landscape, Potential Actionable events, Germline Predisposition, and WGS vs Panel Coverage.



In the following sections we describe each figure present within these 10 panels.

## 

## 

## 

## Overall Schematic

### Summary

The schematic shows the primary tumor location of the metastatic tumor. The labels indicate the number of patients with WGS that passed the quality control filters in the specific cancer cohort and Pan-Cancer. Some patients had multiple WGS tumor samples in the database, for these patients we only included the sample with the highest tumor purity.

## 

### Time and Date Stamp

The creation date and the disease ontology identifiers (DOIDs) of patients included in the specific cancer cohort are included in the bottom middle of each vignette.



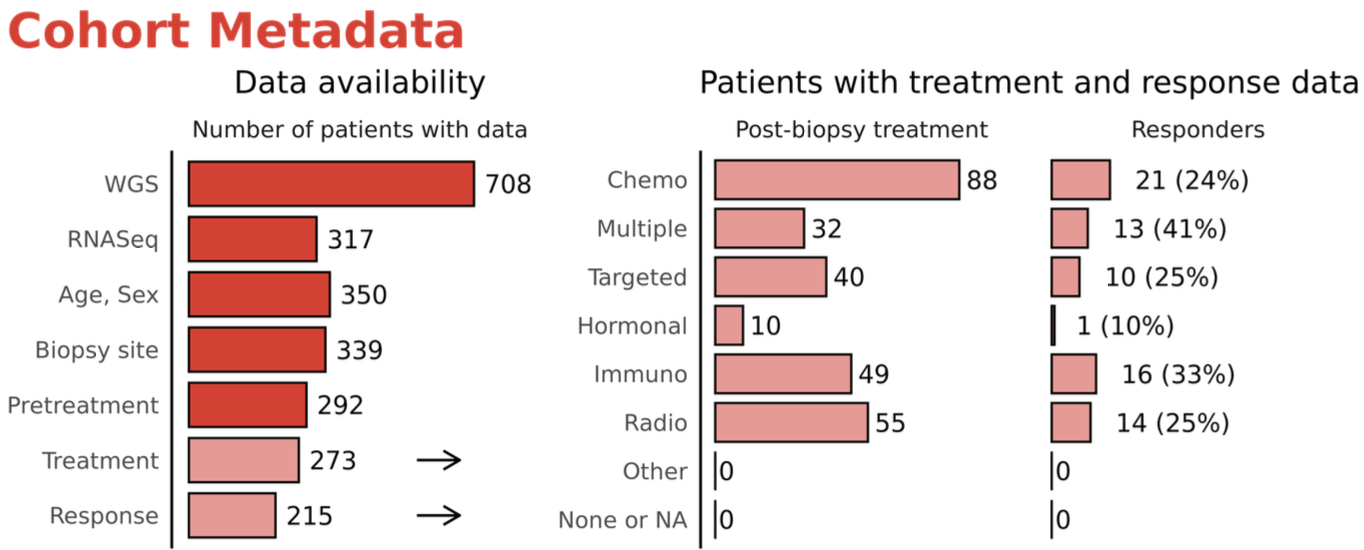
## 

## 

## Cohort Metadata

### Summary

The Cohort Metadata panel summarizes the available clinical metadata for each cohort. The data summarized is available upon data request. For each patient in the Hartwig database tumor-normal whole genome sequencing (WGS) [(Roepman et al. 2021)](https://paperpile.com/c/k7SKn3/vLUs) has been performed on the tumor samples, and the subsequent WGS output is used to create figures throughout the vignette. In addition to WGS, for a subset of patients in the Hartwig database there is clinical information and bulk RNA tumor expression data available (RNASeq).



### Figure Description

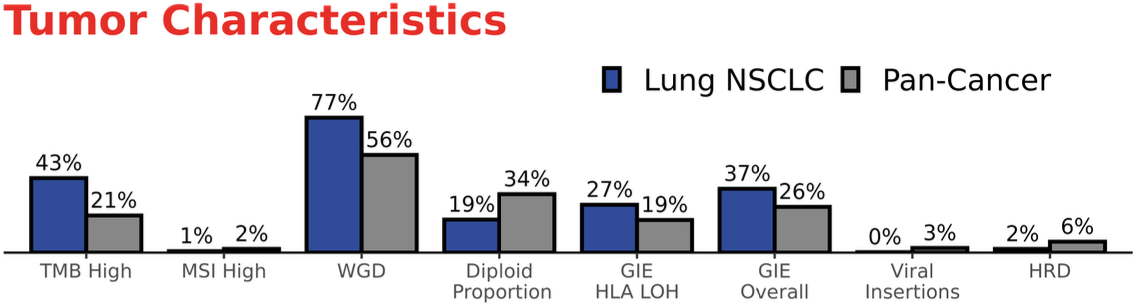
The left figure shows Data Availability. WGS represents the number of patients with processed tumor-normal WGS data in the Hartwig database. RNASeq displays the number of patients with matched tumor RNA-seq data (note that no RNA-seq data is available for patients for which WGS was not successful). Age, Sex, Biopsy site, Pretreatment, and Treatment are clinical fields that are available for a subset of the patients in the Hartwig database. The Pretreatment field refers to treatment received prior to the patient biopsy and the sequencing of their tumors, while Treatment refers to the post-biopsy anti-cancer therapies received by the patient. The Response field refers to patients who had their post-biopsy response measured with either the RECIST 1.1 or RANO criteria.

The right figure Patients with treatment and response data shows a more detailed breakdown of the available post biopsy treatment data which includes the type of treatment and the number of patients with a positive response. To be explicit, the treatment acronyms have the following meanings: Chemo – Chemotherapy, Targeted – Targeted Therapy, Hormonal – Hormonal Therapy, Immuno – Immunotherapy, Radio – Radiotherapy, Other – Other types of systemic therapy. Multiple refers to patients who received a treatment line that involved a combination of systemic therapies (e.g. Chemotherapy and Immunotherapy). The “Responders” counts the patients whose best response evaluation was either a partial (PR) or complete response (CR); the percentage of responders in added brackets. We note that patients who received multiple therapy lines were counted multiple times in this figure. Also, we emphasize this figure only includes patients with both their post-biopsy treatment and response measured.

## Tumor Characteristics

### Summary

This figure provides a high-level summary of fundamental tumor characteristics within specific cancer type cohorts compared to the pan-cancer reference. [PURPLE](https://github.com/hartwigmedical/hmftools/blob/master/purple/README.md) tools [(Priestley et al. 2019)](https://paperpile.com/c/k7SKn3/AkJj) are used to estimate the tumor mutational burden (TMB), microsatellite instability (MSI), whole genome doubling (WGD), and diploid proportion. To estimate genetic immune escape (GIE) overall we use the methodology described by [(Martínez-Jiménez et al. 2023)](https://paperpile.com/c/k7SKn3/nkuG) and estimate GIE loss of heterozygosity (LOH) of the human leukocyte antigen (HLA) with [LILAC](https://github.com/hartwigmedical/hmftools/blob/master/lilac/README.md). Viral insertions are predicted using the [Virus Interpreter](https://github.com/hartwigmedical/hmftools/tree/master/virus-interpreter) tool. Finally, homologous recombination deficiency (HRD) is estimated with the [CHORD](https://github.com/UMCUGenetics/CHORD) tool [(Nguyen et al. 2020)](https://paperpile.com/c/k7SKn3/J7wB).



### Figure Description

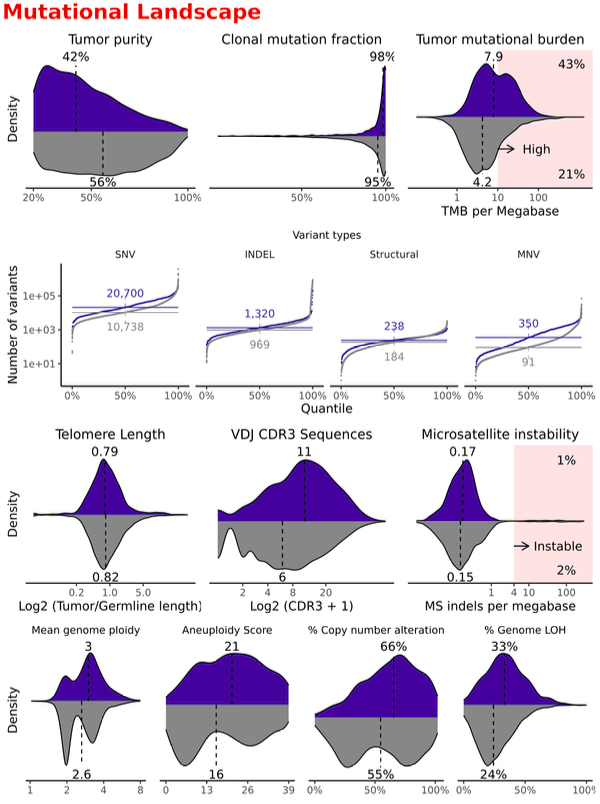
The fields in this figure are described as follows:

* TMB High: % of patients with > 10 somatic mutations per megabase. All mutations passing the quality control thresholds are included.
* MSI High: % of patients with > 4 microsatellite indels per megabase.
* WGD: % of patients with a whole genome doubling events across the cohort. This is set to true if more than 10 autosomes have major allele copy number > 1.5.
* Diploid proportion: the mean diploid proportion across all patients in the cohort.
* GIE HLA LOH: % of patients with a loss of heterozygosity (LOH) of one or more human leukocyte antigen (HLA) alleles.
* GIE Overall: % of patients with any genetic immune escape (GIE) event.
* Viral Insertions: % of patients with a viral insertion in their cancer genome.
* HRD: % of patients labeled to have homologous recombination deficiency with CHORD score > 0.5.

## Mutational Landscape

### Summary

The mutational landscape figure compares the distribution of various molecular feature values within the cohort to the pan-cancer reference. For most figures in this panel, the underlying data is provided by [PURPLE](https://github.com/hartwigmedical/hmftools/blob/master/purple/README.md) [(Priestley et al. 2019)](https://paperpile.com/c/k7SKn3/AkJj), while the telomere length is computed from the [TEAL](https://github.com/hartwigmedical/hmftools/blob/master/teal/README.md) tool, and VDJ CDR3 sequences are output from [CIDER](https://github.com/hartwigmedical/hmftools/tree/master/cider). For each figure, the dark blue represents the distribution of feature values across the highlighted cohort, while gray represents the pan-cancer reference values. The numeric labels corresponding to the vertical dashed lines annotate median values.



### Figure Description

The fields in the figure are described in more detail below:

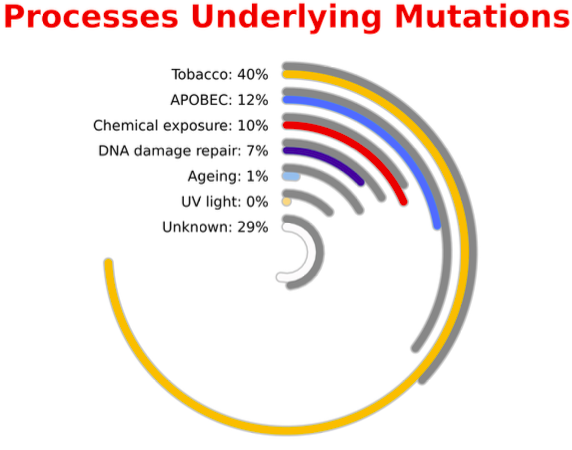
* Tumor purity: Estimated tumor purity from the PURPLE tool. Only tumor samples with an estimated tumor purity tumor purity greater than 20% are included in the data request.
* Clonal mutation fraction: For each patient, the numbers refer to the percentage of variants with clonal likelihood greater than .8.
* Tumor mutational burden: Number of somatic variants per megabase that pass the quality control thresholds (all variants). Patients with > 10 mutations per megabase are labeled TMB High, and the overall %’s are added in the right corners.
* Variant Types: Cumulative distribution frequency (CDF) plots of different variant types.
  + SNV - Single nucleotide variants
  + INDEL - Insertions or deletions
  + MNV - Multiple nucleotide variants
  + Structural - Structural variants
* Telomere Length: Log base 2 ratio of the estimated tumor versus germline telomere lengths.
* VDJ CDR3 Sequences: Number of captured CDR3 VDJ sequences (includes all loci TRG, TRA/TRG, TRB, IGK, IGH, IGL).
* % Genome LOH: For each patient, the proportion of the genome undergoing LOH is defined as regions with minor allele copy number < .3 and major allele copy number > .7 as determined by PURPLE.
* Microsatellite Instability: Average number of microsatellite inserts per megabase. Patients with > 4 microsatellite inserts per megabase are labeled MSI High, and the overall MSI High % 's are added in the right corners.
* Mean genome ploidy: Average ploidy of the tumor sample after adjusting for purity.
* Aneuploidy: The number of autosomal chromosome arm-level copy number events (gains or losses) in each patient sample. The total scores range from 0-39 reflecting the total number of autosomal arms (for chromosome 13-15 and 20-22 only the q arms are counted). The scores were computed following the logic described in [(Taylor et al. 2018)](https://paperpile.com/c/k7SKn3/aEk6).
* % Copy number alteration: The % of the genomes with somatic copy number alterations (defined by copy number < 1.5 or > 2.5 for autosomal chromosomes).

## 

## Processes Underlying Mutations

### Summary

This figure summarizes estimated mutational processes underlying the somatic mutations in the Hartwig database. The processes are approximated based on single base substitution (SBS) signatures developed by [(Alexandrov et al. 2013, 2023)](https://paperpile.com/c/k7SKn3/hS5r+q1ds). The somatic VCF files used to compute trinucleotide contexts are output by [PURPLE](https://github.com/hartwigmedical/hmftools/blob/master/purple/README.md) [(Priestley et al. 2019)](https://paperpile.com/c/k7SKn3/AkJj).



### Figure Description

The above percentages show the average contributions of each underlying process to the overall SBS mutational burden across the cohort. For each patient, the SBS mutational burden corresponding to each process is estimated as follows:

1. SBSs and trinucleotide contexts are extracted from the observed somatic variants.
2. Aggregated counts of SBS mutations and their tri-nucleotide contexts are computed.
3. SBS contributions were estimated using the R package deconstructSigs with version [COSMIC](https://cancer.sanger.ac.uk/signatures/) 3.4 reference.
4. SBS TMB was computed by multiplying the SBS contribution by the sample TMB.
5. SBS TMBs were grouped based on the following rules:

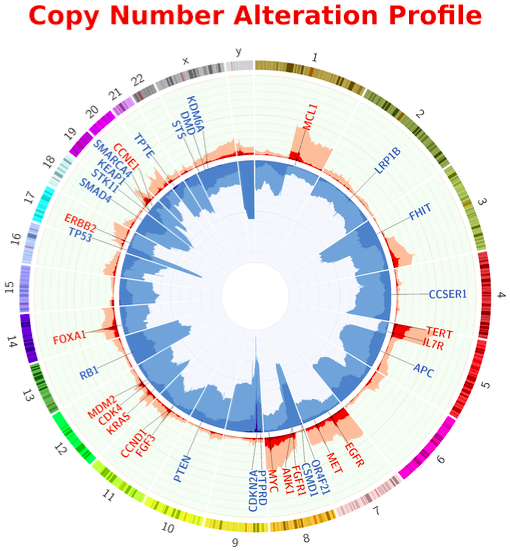
|  |  |
| --- | --- |
| Underlying Process | SBS COSMIC 3.4 Signatures |
| Age | 1, 5, 21 |
| APOBEC | 2, 13 |
| Chemical Exposure | 11, 22, 24, 25, 31, 32, 35, 42, 86, 87, 88, 90, 99 |
| DNA Damage Repair | 3, 9, 10,14, 15, 20, 26, 36, 44, 84, 85 |
| Tobacco | 4, 92 |
| UV Light | 7, 38 |
| Unknown | 8,12,16,17,19, 23, 33, 34, 37, 39, 40, 41, 43, 45-60, 89, 91, 93-98 |

Finally, the mutational burdens attributed to each Underlying Process were aggregated across the cohorts, and the mean contributions were shown in the figure.

## Copy Number Alteration Profile

### Summary

This CIRCOS plot [(Krzywinski et al. 2009)](https://paperpile.com/c/k7SKn3/9VrR) provides a high-level summary of recurrent copy number losses and gains across the cohort by genomic position. Driver genes with the frequent amplifications and homozygous deletions are labeled. Estimation of the copy number and driver events is done by [PURPLE](https://github.com/hartwigmedical/hmftools/blob/master/purple/README.md) [(Priestley et al. 2019)](https://paperpile.com/c/k7SKn3/AkJj). The copy number and driver event estimates from PURPLE are available in the data request.



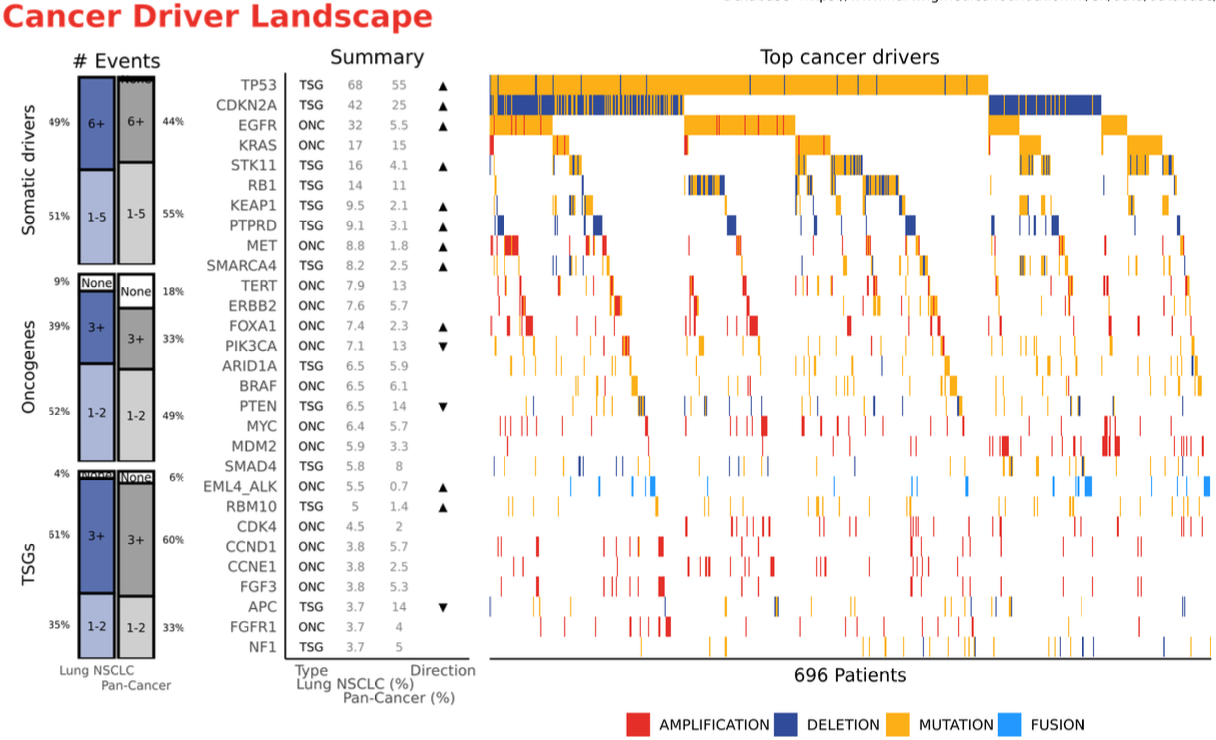
### Figure Description

The figure shows the proportion of the cohort patients with amplification and deletion events by genomic position. The inner ring shows the percentage of tumors with homozygous deletion (very dark blue), LOH or significant copy number loss (defined by copy number < .6 x sample ploidy; dark blue) and near copy neutral LOH (light blue). The outer ring shows the percentage of tumors with high-level amplification (> 3 x sample ploidy; very dark red), moderate amplification (> 2 x sample ploidy; dark red) and low-level amplification (> 1.4 x sample ploidy; light red). The scale on both rings is 0-100% and inverted for the inner ring. The most frequently observed high-level driver gene amplifications (red text) and homozygous deletions (blue text) are added.

## Cancer Driver Landscape

### Summary

The cancer drive landscape figure summarizes the inferred somatic driver events in the database. The amplification, deletion, and mutation based driver events are estimated by [PURPLE](https://github.com/hartwigmedical/hmftools/blob/master/purple/README.md) [(Priestley et al. 2019)](https://paperpile.com/c/k7SKn3/AkJj) while the fusion driver events are annotated by [LINX](https://github.com/hartwigmedical/hmftools/blob/master/linx/README.md) [(Shale et al. 2022)](https://paperpile.com/c/k7SKn3/wbUK). Within PURPLE, driver events are estimated from a panel of 474 genes (see [Hartwig Resource files](https://resources.hartwigmedicalfoundation.nl)) created initially from the union of Marticorena significantly mutated genes [(Martincorena et al. 2018)](https://paperpile.com/c/k7SKn3/QDzG), Cosmic curated genes [(Forbes et al. 2017)](https://paperpile.com/c/k7SKn3/Td1K), and significantly mutated genes within the Hartwig database. See the supplementary information of [(Priestley et al. 2019)](https://paperpile.com/c/k7SKn3/AkJj) for more detail on the algorithm developed to identify driver events. The output of the predicted driver events from both PURPLE and LINX are available in the data request. Only driver events with predictor driver likelihood greater than 0.8 were included in the vignette figure.



### Figure Description

In the # Events figure shown on the left, the distribution of the # driver events output per patient are summarized. The number of events is summarized within the bars and relative percentages within cohorts is shown outside the bars. Pan-cancer in gray is added as a reference for comparison. The specified size groupings of (None, 1-2, 3+) for Tumor Suppressors (TSGs) and Oncogenes and (None, 1-5, 6+) for Overall are arbitrary and simply chosen to highlight roughly the distribution of driver event counts.

In the middle Summary figure, driver genes are ordered from top to bottom according to their frequencies within the cohort. The Type column indicates whether the genes are labeled Tumor Suppressors (TSG) or Oncogenes (ONC). The cohort's name (%) and Pan-Cancer (%) fields show the prevalence of the events within the cohort. For the Direction field, the up and down arrows are added to highlight clearer differences as compared with the Pan-Cancer cohort. An up arrow indicates that either the absolute percentage difference was greater than 5%, or the relative difference was greater than 3 times that of Pan-Cancer; while a down arrow indicates either the absolute difference is less than 5% below, or the relative difference is less than 1/3 of Pan-Cancer.

In the right aligned Top cancer drivers figure, the order of the gene rows from top to bottom match those in the Summary figure. Only patients with at least one driver event were included in this figure, and the total number of patients included labeled on x-axis. The color coding for the event types is based on grouping of the output from the driver catalog. The following table shows the grouping for the driver event types:

|  |  |
| --- | --- |
| Event Type | Groupings |
| Amplification | Partial of Full Amplification |
| Deletion | Deletions or Homozygous Disruptions |
| Mutation | Somatic mutations |
| Fusion | Fusions |

## 

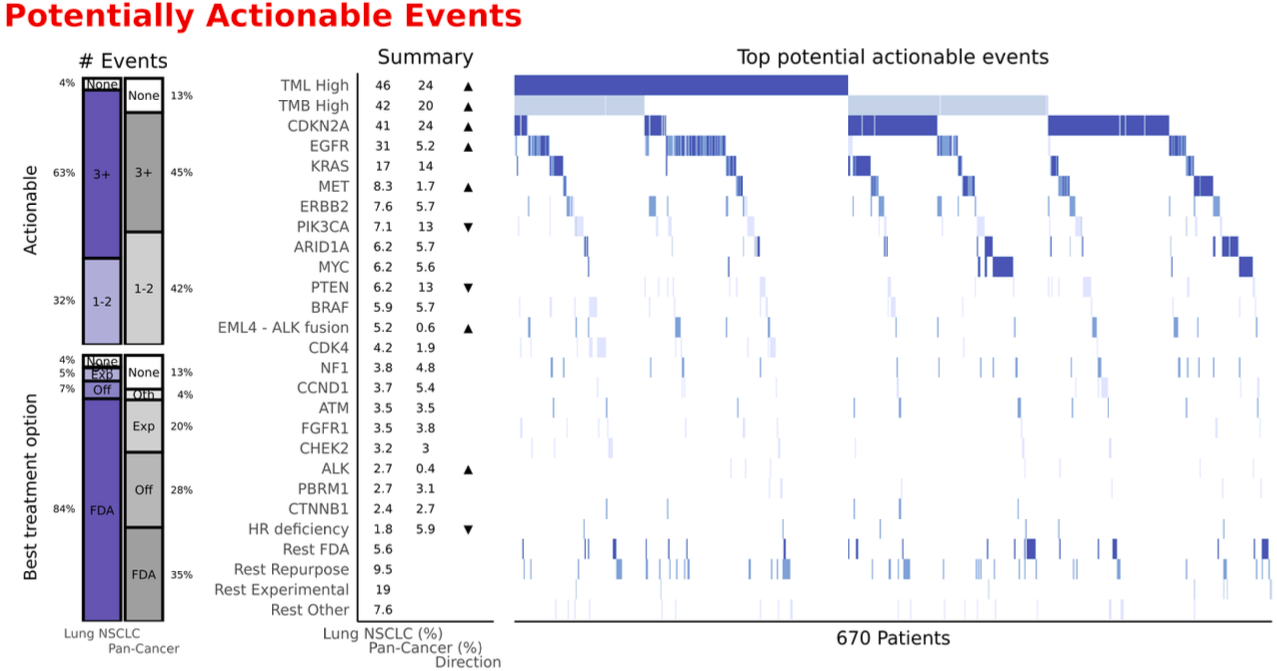
## Potentially Actionable Events

### Summary

This figure displays the genomic events labeled as potentially actionable using reference the Jackson Laboratory [Clinical Knowledgebase](https://www.jax.org/clinical-genomics/ckb) [(Patterson et al. 2016)](https://paperpile.com/c/k7SKn3/y2Pq) (JAX, [Genomenon](https://www.genomenon.com/)). Only reported events by the [PROTECT](https://github.com/hartwigmedical/oncoact/blob/master/protect/README.md) (Personalized Report on Treatment and Eligibility for Clinical Trials) tool with evidence levels A (On or Off-Label), B (On or Off-Label), or C On-label are included. These events correspond to the biomarkers that predict response reported in the Hartwig ORANGE report.

The biomarkers are categorized into best evidence level groups derived from CKB labels that follow guidelines described in [(Li et al. 2017)](https://paperpile.com/c/k7SKn3/A3De). The 4 groups are defined in the table below ordered from highest to lowest evidence level.

|  |  |  |
| --- | --- | --- |
| Evidence Group | Evidence in CKB | Description based on ([(Li et al. 2017)](https://paperpile.com/c/k7SKn3/A3De) |
| FDA Approved | A On-Label | FDA-approved therapy included in professional guidelines. |
| Drug Repurposing | A Off-Label | FDA Approved therapy in different tumor types. \*Vast majority of A Off-Label events also have B On-Label evidence in the CKB database. |
| Experimental/  Other Guidelines | B On-Label | Well-powered studies with consensus from experts in the field. |
| Other | B Off-Label,  C On-Label | FDA Approved therapies or well power study within a different tumor type (but no well powered study within tumor type), or multiple small published studies with some consensus. |



### Figure Description

In the # Events figure shown on the left, the top Actionable facet shows the distribution of the # of all potentially actionable events across patients. The number of events is summarized within the bars and relative percentages within cohorts is shown outside the bars. The bottom Best treatment option facet plot shows the distribution across patients of the best possible treatment option according to evidence level within the bars (ordered by FDA Approved, Drug Repurposing, Experimental/Other Guidelines, Other) and displays the relative percentages within cohorts outside the bars.

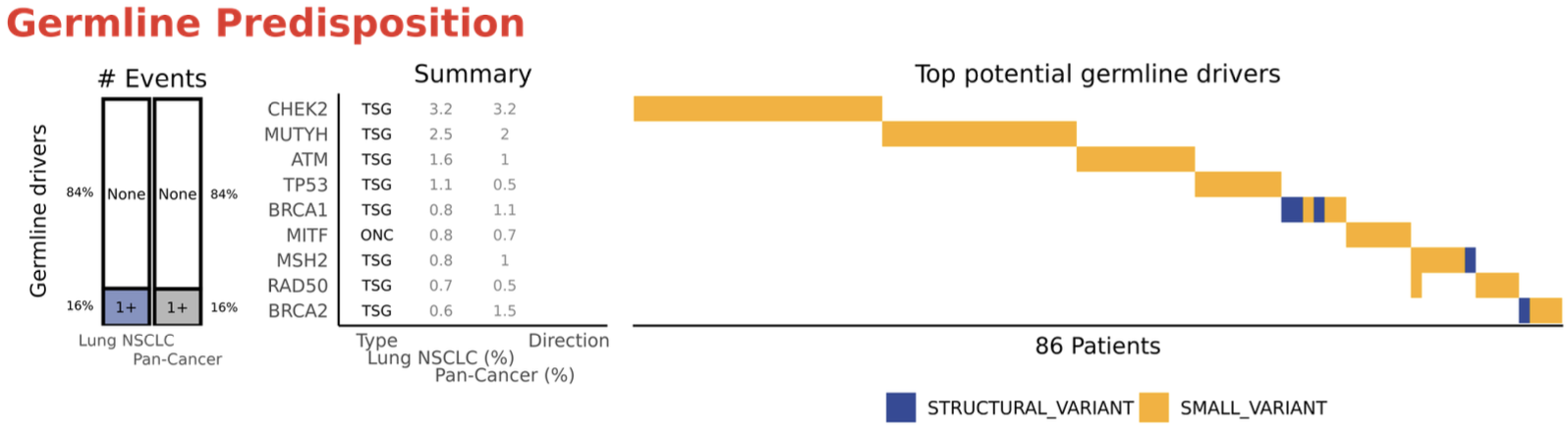
In the middle Summary figure, the potentially actionable events are ordered row-wise from high to low according to their frequencies within the cohort. The cohort's name (%) and Pan-Cancer (%) fields show the prevalence of the events within the cohort. For the Direction field, the up and down arrows are added to highlight clearer differences with the Pan-Cancer cohort. An up arrow indicates that either the absolute percentage difference was greater than 5%, or the relative difference was greater than 3 times that of Pan-Cancer; while a down arrow indicates either the absolute difference is less than 5% below, or the relative difference is less than 1/3 of Pan-Cancer.

In the right aligned Top potential actionable events figure, the order of the gene rows from top to bottom match those in the Summary figure. Only patients with at least one potentially actionable event were included in this figure, and the total number of patients included labeled on x-axis.

## Germline Predisposition

### Summary

The germline predisposition figure summarizes the predicted germline variants in the database. The germline disposition genes are selected from a list of 152 genes curated by [(Huang et al. 2018)](https://paperpile.com/c/k7SKn3/s1vB). For SNV and INDEL, in each of the 152 genes the germline variants were called using the GATK HaplotypeCaller [(Poplin et al. 2018)](https://paperpile.com/c/k7SKn3/issL), and predisposition events were labeled based on the ClinVar [(Landrum et al. 2016)](https://paperpile.com/c/k7SKn3/ypNS) annotation of their canonical transcripts. See the Germline predisposition variant callingsection in the supplementary information of [(Priestley et al. 2019)](https://paperpile.com/c/k7SKn3/AkJj) for complete details. The germline predisposition variants are found within the [PURPLE](https://github.com/hartwigmedical/hmftools/blob/master/purple/README.md) driver catalog output in the data request.



### Figure Description

In the # Events figure shown on the left, the distribution of the # driver events output per patient are summarized. The number of events is summarized within the bars and relative percentages within cohorts is shown outside the bars. Pan-cancer in gray is added as a reference for comparison.

In the middle Summary figure, predisposition genes are ordered from top to bottom according to their frequencies within the cohort. The Type column indicates whether the genes are labeled Tumor Suppressors (TSG) or Oncogenes (ONC). The cohort's name (%) and Pan-Cancer (%) fields show the prevalence of the events within the cohort. For the Direction field, the up and down arrows are added to highlight clearer differences with the Pan-Cancer cohort. An up arrow indicates that either the absolute percentage difference was greater than 5%, or the relative difference was greater than 3 times that of Pan-Cancer; while a down arrow indicates either the absolute difference is less than 5% below, or the relative difference is less than 1/3 of Pan-Cancer.

In the right aligned Top germline drivers figure, the order of the gene rows from top to bottom match those in the Summary figure. Only patients with at least one germline driver were included in this figure, and the total number of patients included labeled on x-axis. The color coding for the event types is based on grouping of the output from the driver catalog.

|  |  |
| --- | --- |
| Event Type | Groupings |
| Structural Variant | Germline Disruptions or Deletions |
| Small Variant | Germline Mutations |

## WGS vs Panel Coverage

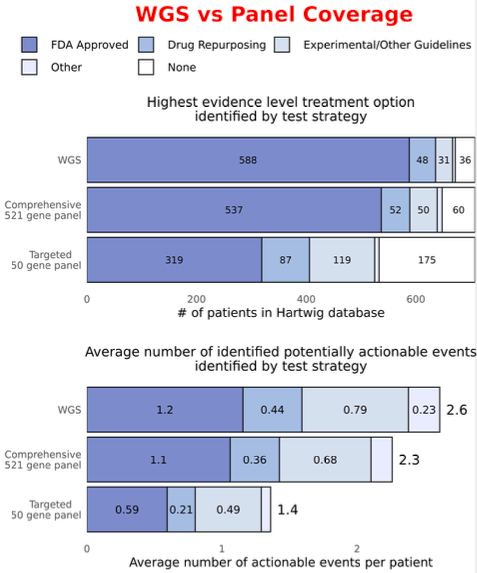
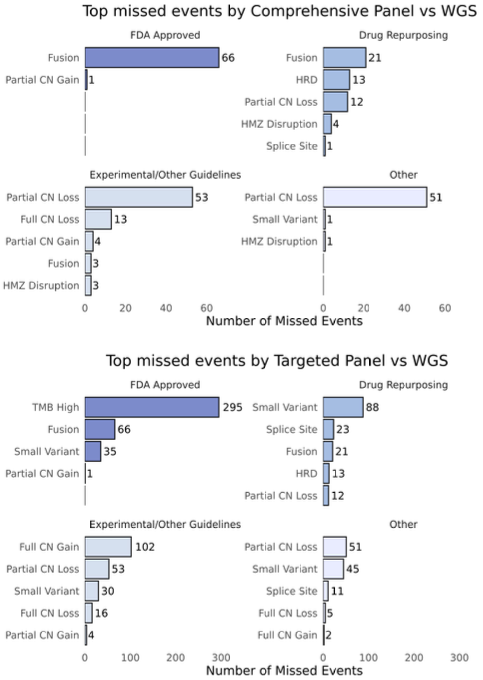
### Summary

The final panel summarizes results from an *in-silico* study where we performed a comparison of the actionability coverage of WGS compared to two targeted gene panels: a comprehensive gene panel and targeted gene panel. In this study, the comprehensive panel is represented by the [TruSight Oncology 500](https://emea.illumina.com/products/by-type/clinical-research-products/trusight-oncology-500.html) (TSO500) panel, including 521 genes; while the Targeted panel is based on the [AmpliSeq](https://www.thermofisher.com/es/es/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-select-targets/ampliseq-target-selection.html#panels) custom panel, including 50 genes.

The actionability coverage comparison study was conducted as follows. First, WGS biomarker events reported by [PROTECT](https://github.com/hartwigmedical/oncoact/blob/master/protect/README.md) were set as a reference of all actionable events. The study scope was restricted to reported genomic events (no RNA), with evidence for treatment response (i.e., excluding resistance markers), whose evidence source was the [Clinical Knowledgebase](https://www.jax.org/clinical-genomics/ckb) [(Patterson et al. 2016)](https://paperpile.com/c/k7SKn3/y2Pq). These events were then classified into four ordinal evidence levels described in the Potentially Actionable Events section of this document (FDA Approved, Drug Repurposing, Experimental/Other Guidelines, or Other). To measure panel coverage, we specified a set of rules based on descriptions of the panels gathered from the company websites ([TSO500](https://emea.illumina.com/products/by-brand/trusight-oncology/tso-500-portfolio.html) for comprehensive and [AmpliSeq](https://www.google.com/url?q=https://www.thermofisher.com/es/es/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-select-targets/ampliseq-target-selection.html%23panels&sa=D&source=docs&ust=1720188301272085&usg=AOvVaw2xefSmtY-yLLFtBNvvfxBx) for targeted) and the genomic coordinates of the commercial panel [BED](https://samtools.github.io/hts-specs/BEDv1.pdf) files [(Quinlan and Hall 2010)](https://paperpile.com/c/k7SKn3/I8WU). The comprehensive panel BED file was based on an experimental validation of the TSO500 panel, while the targeted panel used the AmpliSeq design BED file. We specify the follow rules to measure the actionability coverage:

|  |  |
| --- | --- |
| Event Type | Rules for panel coverage measurement |
| Small Variant  Splice Site Variants | * Only events observed in genomic locations captured within the panel BED files are covered. |
| Copy Number Variants  (Amplifications and Deletions) | * Only genes included in the panels can capture CN events. * Gene deletion detected if minimum copy number < .5. * Gene amplification captured if maximum copy number > 6. |
| Genomic Signatures  (TMB High,  MSI High, HRD) | * TMB, MSI High are captured by the comprehensive panel. * Only MSI High is captured by the targeted panel. * HRD is missed by both panels |
| Other Events  (Fusions,  Homozygous disruption,  Viral Inserts, HLA Type) | * All events are assumed to be missed by both the panels. |

Finally, we applied the above coverage rules across all events in the Hartwig database, and thereby computed coverage labels for the comprehensive and targeted panels. Aggregations of these coverage labels served as the basis for the presented figures.

### Figure Description

The topmost Highest evidence level treatment option identified by each test strategy figure compares the distributions of best evidence level treatment options measured by WGS, the comprehensive gene panel, and targeted gene panels. The evidence levels are ordered high to low from FDA Approved > Drug Repurposing > Experimental/Other Guidelines > Other > None (see description of evidence levels in Potentially Actionable Events section). The x-axis shows the total number of patients in the Hartwig database. The number of patients with each best evidence level treatment option is labeled within the bars.

The Average number of identified potentially actionable events identified by test strategy figure compares the average number of events within each evidence level measured by WGS, the comprehensive gene panel, and targeted gene panels - across all patients in the cohort. The y-axes represent the WGS, comprehensive panel, and targeted panel, and x-axes show the average number of events. The labels within the bars indicate the average event counts within each evidence level, while the outer labels show the overall average number of events.

The Top missed events by Comprehensive/Targeted Panels vs WGS figures summarize the most frequent missed event types by the respective comprehensive and targeted panels, faceted by evidence level. For each figure, the y-axes share the 5 most frequent missed event types within each evidence level, while the x-axes represent the number of missed events. The range of the x-axes differ between the comprehensive and targeted panel figures.

## 

## Citations

[Alexandrov, Ludmil B., Jaegil Kim, Nicholas J. Haradhvala, Mi Ni Huang, Alvin Wei Tian Ng, Yang Wu, Arnoud Boot, et al. 2023. “Author Correction: The Repertoire of Mutational Signatures in Human Cancer.” *Nature* 614 (7948): E41.](http://paperpile.com/b/k7SKn3/q1ds)

[Alexandrov, Ludmil B., Serena Nik-Zainal, David C. Wedge, Samuel A. J. R. Aparicio, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, et al. 2013. “Signatures of Mutational Processes in Human Cancer.” *Nature* 500 (7463): 415–21.](http://paperpile.com/b/k7SKn3/hS5r)

[Forbes, Simon A., David Beare, Harry Boutselakis, Sally Bamford, Nidhi Bindal, John Tate, Charlotte G. Cole, et al. 2017. “COSMIC: Somatic Cancer Genetics at High-Resolution.” *Nucleic Acids Research* 45 (D1): D777–83.](http://paperpile.com/b/k7SKn3/Td1K)

[Huang, Kuan-Lin, R. Jay Mashl, Yige Wu, Deborah I. Ritter, Jiayin Wang, Clara Oh, Marta Paczkowska, et al. 2018. “Pathogenic Germline Variants in 10,389 Adult Cancers.” *Cell* 173 (2): 355–70.e14.](http://paperpile.com/b/k7SKn3/s1vB)

[Krzywinski, Martin, Jacqueline Schein, Inanç Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. 2009. “Circos: An Information Aesthetic for Comparative Genomics.” *Genome Research* 19 (9): 1639–45.](http://paperpile.com/b/k7SKn3/9VrR)

[Landrum, Melissa J., Jennifer M. Lee, Mark Benson, Garth Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, et al. 2016. “ClinVar: Public Archive of Interpretations of Clinically Relevant Variants.” *Nucleic Acids Research* 44 (D1): D862–68.](http://paperpile.com/b/k7SKn3/ypNS)

[Li, Marilyn M., Michael Datto, Eric J. Duncavage, Shashikant Kulkarni, Neal I. Lindeman, Somak Roy, Apostolia M. Tsimberidou, et al. 2017. “Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists.” *The Journal of Molecular Diagnostics: JMD* 19 (1): 4–23.](http://paperpile.com/b/k7SKn3/A3De)

[Martincorena, Iñigo, Keiran M. Raine, Moritz Gerstung, Kevin J. Dawson, Kerstin Haase, Peter Van Loo, Helen Davies, Michael R. Stratton, and Peter J. Campbell. 2018. “Universal Patterns of Selection in Cancer and Somatic Tissues.” *Cell* 173 (7): 1823.](http://paperpile.com/b/k7SKn3/QDzG)

[Martínez-Jiménez, Francisco, Peter Priestley, Charles Shale, Jonathan Baber, Erik Rozemuller, and Edwin Cuppen. 2023. “Genetic Immune Escape Landscape in Primary and Metastatic Cancer.” *Nature Genetics* 55 (5): 820–31.](http://paperpile.com/b/k7SKn3/nkuG)

[Nguyen, Luan, John W M Martens, Arne Van Hoeck, and Edwin Cuppen. 2020. “Pan-Cancer Landscape of Homologous Recombination Deficiency.” *Nature Communications* 11 (1): 5584.](http://paperpile.com/b/k7SKn3/J7wB)

[Patterson, Sara E., Rangjiao Liu, Cara M. Statz, Daniel Durkin, Anuradha Lakshminarayana, and Susan M. Mockus. 2016. “The Clinical Trial Landscape in Oncology and Connectivity of Somatic Mutational Profiles to Targeted Therapies.” *Human Genomics* 10 (January): 4.](http://paperpile.com/b/k7SKn3/y2Pq)

[Poplin, Ryan, Valentin Ruano-Rubio, Mark A. DePristo, Tim J. Fennell, Mauricio O. Carneiro, Geraldine A. Van der Auwera, David E. Kling, et al. 2018. “Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples.” *bioRxiv*. https://doi.org/](http://paperpile.com/b/k7SKn3/issL)[10.1101/201178](http://dx.doi.org/10.1101/201178)[.](http://paperpile.com/b/k7SKn3/issL)

[Priestley, Peter, Jonathan Baber, Martijn P. Lolkema, Neeltje Steeghs, Ewart de Bruijn, Charles Shale, Korneel Duyvesteyn, et al. 2019. “Pan-Cancer Whole-Genome Analyses of Metastatic Solid Tumours.” *Nature* 575 (7781): 210–16.](http://paperpile.com/b/k7SKn3/AkJj)

[Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics*  26 (6): 841–42.](http://paperpile.com/b/k7SKn3/I8WU)

[Roepman, Paul, Ewart de Bruijn, Stef van Lieshout, Lieke Schoenmaker, Mirjam C. Boelens, Hendrikus J. Dubbink, Willemina R. R. Geurts-Giele, et al. 2021. “Clinical Validation of Whole Genome Sequencing for Cancer Diagnostics.” *The Journal of Molecular Diagnostics: JMD* 23 (7): 816–33.](http://paperpile.com/b/k7SKn3/vLUs)

[Shale, Charles, Daniel L. Cameron, Jonathan Baber, Marie Wong, Mark J. Cowley, Anthony T. Papenfuss, Edwin Cuppen, and Peter Priestley. 2022. “Unscrambling Cancer Genomes via Integrated Analysis of Structural Variation and Copy Number.” *Cell Genomics* 2 (4): 100112.](http://paperpile.com/b/k7SKn3/wbUK)

[Taylor, Alison M., Juliann Shih, Gavin Ha, Galen F. Gao, Xiaoyang Zhang, Ashton C. Berger, Steven E. Schumacher, et al. 2018. “Genomic and Functional Approaches to Understanding Cancer Aneuploidy.” *Cancer Cell* 33 (4): 676–89.e3.](http://paperpile.com/b/k7SKn3/aEk6)