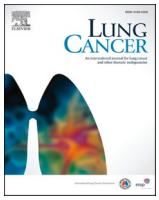
Lung Cancer 182 (2023) 107286

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan



Reliability of panel-based mutational signatures for immune-checkpoint-inhibition efficacy prediction in non-small cell lung cancer

H.C. Donker [a,b], K. Cuppens [c,d,e], G. Froyen [e,f], H.J.M. Groen [g], T.J.N. Hiltermann [g], B. Maes [e,f], E. Schuuring [h], P.-J. Volders [f], G.A. Lunter [i], B. van Es [j,k,*]

[a] Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
[b] Global Computational Biology & Digital Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riß, Germany
[c] Department of Pulmonology and Thoracic Oncology, Jessa Hospital, Hasselt, Belgium
[d] Department of Thoracic Oncology, The Netherlands Cancer Institute, Amsterdam, the Netherlands
[e] Faculty of Medicine and Life Sciences - LCRC, Hasselt University, Diepenbeek, Belgium
[f] Laboratory of Molecular Diagnostics, Dept Clinical Biology, Jessa Hospital, Hasselt, Belgium
[g] Department of Pulmonary Diseases, University of Groningen and University Medical Center Groningen, the Netherlands
[h] Department of Pathology, University of Groningen and University Medical Center Groningen, the Netherlands
[i] Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, Oxford University, Oxford, UK
[j] Central Diagnostic Laboratory, University Medical Centre Utrecht, Utrecht University, Heidelberglaan 100, 3508 GA Utrecht, the Netherlands
[k] MedxAI, Theophile de Bockstraat 77-1, 1058VA Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Immunotherapy
Mutational signatures
Non-small cell lung cancer
Panel sequencing

ABSTRACT

Objectives: Mutational signatures (MS) are gaining traction for deriving therapeutic insights for immune checkpoint inhibition (ICI). We asked if MS attributions from comprehensive targeted sequencing assays are reliable enough for predicting ICI efficacy in non-small cell lung cancer (NSCLC).

Methods: Somatic mutations of $m = 126$ patients were assayed using panel-based sequencing of 523 cancer-related genes. *In silico* simulations of MS attributions for various panels were performed on a separate dataset of $m = 101$ whole genome sequenced patients. Non-synonymous mutations were deconvoluted using COSMIC v3.3 signatures and used to test a previously published machine learning classifier.

Results: The ICI efficacy predictor performed poorly with an accuracy of $0.51^{+0.09}_{-0.09}$, average precision of $0.52^{+0.11}_{-0.11}$, and an area under the receiver operating characteristic curve of $0.50^{+0.10}_{-0.09}$. Theoretical arguments, experimental data, and *in silico* simulations pointed to false negative rates (FNR) related to panel size. A secondary effect was observed, where deconvolution of small ensembles of point mutations lead to reconstruction errors and misattributions.

Conclusion: MS attributions from current targeted panel sequencing are not reliable enough to predict ICI efficacy. We suggest that, for downstream classification tasks in NSCLC, signature attributions be based on whole exome or genome sequencing instead.

Abbreviations: αCTLA-4, cytotoxic T lymphocyte associated antigen-4 inhibitor; αPD-1, programmed death-1 inhibitor; αPD-L1, programmed death-ligand 1 inhibitor; αTNFRSF7, tumour necrosis factor receptor superfamily type 7 inhibitor; αVEGF-A, vascular endothelial growth factor inhibitor; AI, artificial intelligence; BH, Benjamini-Hochberg; DB, durable benefit; F1 CDx, FoundationOne CDx; FFPE, formalin-fixed paraffin-embedded; FN, false negative; FNR, false negative rate; ICI, immune checkpoint inhibition; IQR, interquartile range; Mb, megabase; MS, mutational signatures; MSK-IMPACT, Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets; NMF, non-negative matrix factorisation; NSCLC, non-small cell lung cancer; PDL1, programmed death ligand-1; Rad, radians; ROC AUC, area under the receiver operating characteristic curve; SBS, single base substitution; TSO500, TruSight Oncology 500; TMB, tumour mutational burden; WGS, whole genome sequencing.

* Corresponding author at: Central Diagnostic Laboratory, University Medical Centre Utrecht, Utrecht University, Heidelberglaan 100, 3508 GA Utrecht, the Netherlands.

E-mail addresses: h.c.donker@umcg.nl (H.C. Donker), Kristof.Cuppens@jessazh.be (K. Cuppens), h.j.m.groen@umcg.nl (H.J.M. Groen), t.j.n.hiltermann@umcg.nl (T.J.N. Hiltermann), brigitte.maes@jessazh.be (B. Maes), e.schuuring@umcg.nl (E. Schuuring), Pieter-Jan.Volders@jessazh.be (P.-J. Volders), g.a.lunter@umcg.nl (G.A. Lunter), bes3@umcutrecht.nl (B. van Es).

https://doi.org/10.1016/j.lungcan.2023.107286
Received 9 May 2023; Received in revised form 20 June 2023; Accepted 23 June 2023
Available online 3 July 2023

1. Introduction

Immune checkpoint inhibitor (ICI) therapy is the backbone of first line treatment for the majority of patients with advanced-stage non-small cell lung cancer (NSCLC) who have tested negative for targetable mutations in the tumour. Unfortunately, only a subset of these patients respond to ICI treatment [1]. Even though impressive durable responses can be seen, meta-analysis of 12 randomised phase-III studies found that median progression free survival typically improves by at most 2.4 months [1]. This heterogeneity in the population led to a surge of research looking for efficacy predictors [2] to single out patients that benefit from treatment. Two well-studied biomarkers are tumour mutational burden (TMB) [3] and programmed-death ligand-1 (PDL1) protein expression [4], but both are imperfect predictors. Currently, mutational signatures are becoming of increasing interest as predictive markers [5]. Specifically for ICI, evidence is mounting that particular signatures relate to treatment efficacy in lung cancer [6–12]. In particular, Ref. [11] used non-synonymous somatic mutations attributed to COSMIC single base substitution (SBS) signatures SBS4 [13,14] and SBS87 [5,15]. Earlier work suggested that these two signatures can be viewed as refinements of TMB [11]. Both mutational signatures were used as input for an artificial intelligence (AI) classifier that predicts durable benefit (DB) from ICI [11]. To facilitate clinical adoption of this AI-powered technology, it is desirable to base the input of the classifier (the mutational signature attributions) on a cheap and fast next-generation sequencing panel [16]. One such example is the Illumina TruSight Oncology 500 (TSO500) assay [17]. Earlier work described the panel-derived mutational signature attributions from samples with a high somatic mutation count [18–20]. However, it is not clear whether the signatures remain reliable in an unselected population. A related work in an unselected population studied SBS4-based clusters as a surrogate of smoking history [21]. Their *in silico* analysis indicated that most SBS mutational signatures could be called (but their calling criterion was unspecified) [21]. However, all previous work was based on both synonymous and non-synonymous mutations. The classifier of Donker et al. [11] requires non-synonymous SBS4 and SBS87 mutational signature attributions as input. Since high quality input data is crucial for the reliability of the classifier we revisit Ref. [22], scrutinise attributions to signatures and validate their downstream utility in the ICI efficacy predictor of Ref. [11]. Our contributions are to (i) analyse mutational signatures in a real-world clinical population, including non-highly mutated samples, with a particular focus on SBS4 and SBS87 (ii) validate their clinical impact for use in an AI-powered ICI efficacy predictor, and (iii) estimate the reliability of the TSO500 panel, the Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) 468 gene panel, the FoundationOne CDx (F1 CDx) panel, and the pan cancer panel of Xu et al. [23].

2. Methods

2.1. Patients and tumor tissue

Two cohorts of patients with (predominantly advanced) NSCLC were selected who were treated with PD(L)1 inhibiting immunotherapy (or combination therapy). Pre-ICI treatment biopsies were taken and retrospectively analysed. In the Cuppens et al. cohort [22], 126 formalin-fixed paraffin-embedded (FFPE) tumour-only samples were analysed with TSO500. The Donker et al. discovery cohort, comprising 101 fresh frozen samples (derived from matched tumour-normal tissue), were analysed with whole genome sequencing (WGS) [11,24].

2.2. Upstream bioinformatics

TSO500 data from the cohort of Cuppens et al. [22] were analysed using the Illumina DRAGEN TruSight Oncology 500 Analysis Software version 2.1.0. Analysis was carried out on an Illumina DRAGEN server version 3. Variants eligible for tumour mutational burden (TMB) calculation (see the software manual for details [25]) were taken from the *_*TMB_Trace.tsv* where the values in the column *IncludedInTMBNumerator* were equal to one. Coverage size was extracted from the *.*tmb.metrics.csv* files and used to normalise estimates per megabase (Mb).

In silico analysis was carried out on the whole genome sequencing data of the Donker et al. cohort [11]. Mutations (in variant call format) were projected onto the *in silico* panel with *BEDtools* [26]. For comparison with the TSO500, the *TST500C_manifest.bed* file of the DRAGEN TruSight Oncology 500 Analysis Software was used. To study if our findings were specific to panel size or composition, we performed *in silico* analysis on other targeted panels. Specifically, the Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) 468 gene panel, the FoundationOne CDx (F1 CDx) panel, and the pan cancer panel (PANCAN) of Xu et al. [23]. To analyse these panels we used the BED files that were part of CaMutQC v0.5.0 [27]. Panel sizes were estimated by merging genomic regions using BEDOPS [28] and then summing the region sizes. The resulting panel sizes (TSO500: 1.27 Mb, MSK-IMPACT: 1.84 Mb, F1 CDx: 1.34 Mb, Xu et al. PANCAN: 4.0 Mb) were used to normalise *in silico* estimates per Mb.

2.3. Downstream analysis

Mutation spectrum, $\boldsymbol{X}$, computation and signature deconvolution [using non-negative matrix factorisation (NMF) [29]] was done as previously described [11]. Briefly, non-synonymous somatic singlets (i.e., single base substitutions) are partitioned by substitution (six in total) and flanking context (four-by-four nucleotides) and counted per patient (m in total), resulting in a mutation spectrum with $n = 96$ mutation channels. COSMIC v3.3 mutational signatures $\boldsymbol{H}$ (a p-by-n non-negative matrix of $p = 78$ signatures) were used to deconvolute the (n = 96) mutation channels into mutational signature attributions $\boldsymbol{W}$ (so that $\boldsymbol{X} \approx \boldsymbol{WH}$). After deconvolution, mutational signatures flagged as possible sequencing artifacts by COSMIC were not further considered (except for the reconstruction error, see below). We validated the previously reported durable benefit (DB, defined as progression-free survival $\geq$ six months) naïve Bayes classifier [11]. This machine learning classifier modelled features SBS4 and SBS87 as zero-inflated exponentials.

To measure the reconstruction errors introduced by the deconvolution process, we analysed deviations from the conservation of singlets. That is, for all samples i (rows), the sum across the mutation spectrum $\boldsymbol{X}$ channels j (columns) equals the attributed mutations $\boldsymbol{W}$ across all signatures k (columns), $\sum_{j=1}^{96} X_{ij} \approx \sum_{k=1}^{78} W_{ik}$. We define violations as deviations thereof $\epsilon_i = \frac{\sum_{j=1}^{96} X_{ij} - \sum_{k=1}^{78} W_{ik}}{\sum_{j=1}^{96} X_{ij}}$.

2.4. Statistical aspects

Predicted DB probabilities are dichotomised at 43.8 % (following Ref. [11]) to compute the accuracy. Average precision and area under the receiver operating characteristic curve (ROC AUC) are directly computed from the probabilities (i.e., not dichotomised). Correlations were analysed using Kendall-tau. Mutational spectra were compared using a multinomial test, where we pooled the counts across samples. We set the probability distribution of the null hypothesis to be the frequencies of the whole genome sequencing data and tested deviations thereof using the χ^2 distribution. To account for multiple testing, p-values were adjusted using the Benjamini-Hochberg (BH) method and denoted by q. All conducted statistical tests were two sided. Ninety-five per cent confidence intervals were obtained by 1000 bootstrap permutations unless indicated otherwise. Throughout the text, we denote estimate a and ninety-five per cent confidence intervals $[a-b, a+c]$ by a^{+c}_{-b}.

2.5. Theoretical aspects false negative analysis

For each panel of size l, we extrapolated the number of false negatives to a hypothetical panel of size $2l, 3l, 4l, ..$ but similar composition. To this end, patients with a similar signature profile were clustered using k-means, where the number of clusters ($k = 5$) was determined by eye using the elbow method. To achieve a panel size of nl, the dataset was n-fold resampled with replacement and mutations of n patients were aggregated while stratifying by cluster (so that the total number of patients remained the same before and after coarse graining). For all panels we extrapolated up to $10l$ except the larger pan cancer panel of Xu et al. [23], where we analysed up to and including $5l$, to cover a similar size in Mb.

To theoretically model the non-synonymous mutations, we used a uniform background probability that scales with size (similar to [30], although our mutation model, below, differs). Specifically, we estimated the probability to detect a mutation on a panel of size l Mb to be $p_{\text{detect}} = \frac{l}{47.9}$, assuming an exome size of 47.9 Mb [31]. We denote $n_{\text{NS}}^{(s)}$ as the number of non-synonymous mutations attributed to signature s (e.g., SBS4 or SBS87). The probability of a false negative (FN, no attributed mutations detected) for signature s is then $p\left(n_{\text{NS}}^{(s)}\right) = (1 - p_{\text{detect}})^{n_{\text{NS}}^{(s)}}$. Finally, we determined the false negative rate (FNR), $p\left(\text{FN}^{(s)}\right)$, in the Discovery cohort of Ref. [11] for each signature s. To this end, we estimated, $n_{\text{NS}}^{(s)}$ for each sample and then averaged

$$p\left(\text{FN}^{(s)}\right) = \left\langle p\left(n_{\text{NS}}^{(s)}\right) \right\rangle_{n_{NS}^{(s)}} \tag{1}$$

across all positive (i.e., $n_{\text{NS}}^{(s)} > 0$) samples using the empirical distribution [i.e., $p_{\text{data}}(n) = \frac{1}{m}\sum_{i=1}^{m}\delta(n, n^{(i)})$, with $\delta(x, y)$ the Kronecker delta function and *naïve* indicating positive samples in the Discovery set].

3. Result

3.1. TSO500-derived mutational signature attributions are poor predictors

In total, pre-ICI treatment samples from $M = 126$ lung cancer patients of Cuppens et al. [22] were re-analysed (Fig. 1 and Table 1). After deconvoluting the non-synonymous point mutations, we inferred on average, 0.65 Mb^{-1} substitutions attributed to SBS4 [interquartile range (IQR): 0.00 to 0.00 Mb^{-1}] and 0.75 Mb^{-1} to SBS87 (IQR: 0.00 to 1.26 Mb^{-1}) (Fig. S1, Supplementary Material). Reassuringly, FFPE damage-related signatures SBS1 and SBS30 [32] were mostly absent with an average attribution of 0.26 Mb^{-1} (median: 0.00; IQR: 0.00 to 0.00 Mb^{-1}) and 0.22 Mb^{-1} (median: 0.00; IQR: 0.00 to 0.00 Mb^{-1}), respectively.

Next, we set out to validate the classifier of Donker et al. [11] which uses signatures SBS4 and SBS87 to predict DB. Overall, the predictor performed poorly with an accuracy of $0.51^{+0.09}_{-0.09}$, average precision of $0.52^{+0.11}_{-0.11}$, and a ROC AUC of $0.50^{+0.10}_{-0.09}$. To show that this was not merely attributable to differences in feature distributions, we completely re-trained the model on the TSO500 data. Leave-one-out cross validation showed similar poor performance with an accuracy of $0.30^{+0.08}_{-0.07}$, average precision of $0.53^{+0.12}_{-0.10}$, and a ROC AUC of $0.44^{+0.11}_{-0.11}$. This lack of model fit hints at a problem with data quality.

3.2. Poor performance relates to panel

The absence of a signature is an important contributor to the DB prediction. Given the small panel size (1.27 Mb), we reasoned that the poor performance is attributable to "false negative" signature attributions. That is, the panel fails to detect the presence of a mutational signature in the sample. This prompted us to validate the mutational signatures using *in silico* analysis. To this end, the whole genome sequencing Discovery dataset from [11] was re-analysed (Table 1). Based on the number of non-synonymous signature attributions in Ref. [11] we theoretically estimated (based on panel size alone) a FNR of $0.23^{+0.07}_{-0.06}$ for SBS4 and $0.67^{+0.04}_{-0.05}$ for SBS87. The FNR of other mutational signatures were estimated to be at least as high as that of SBS4

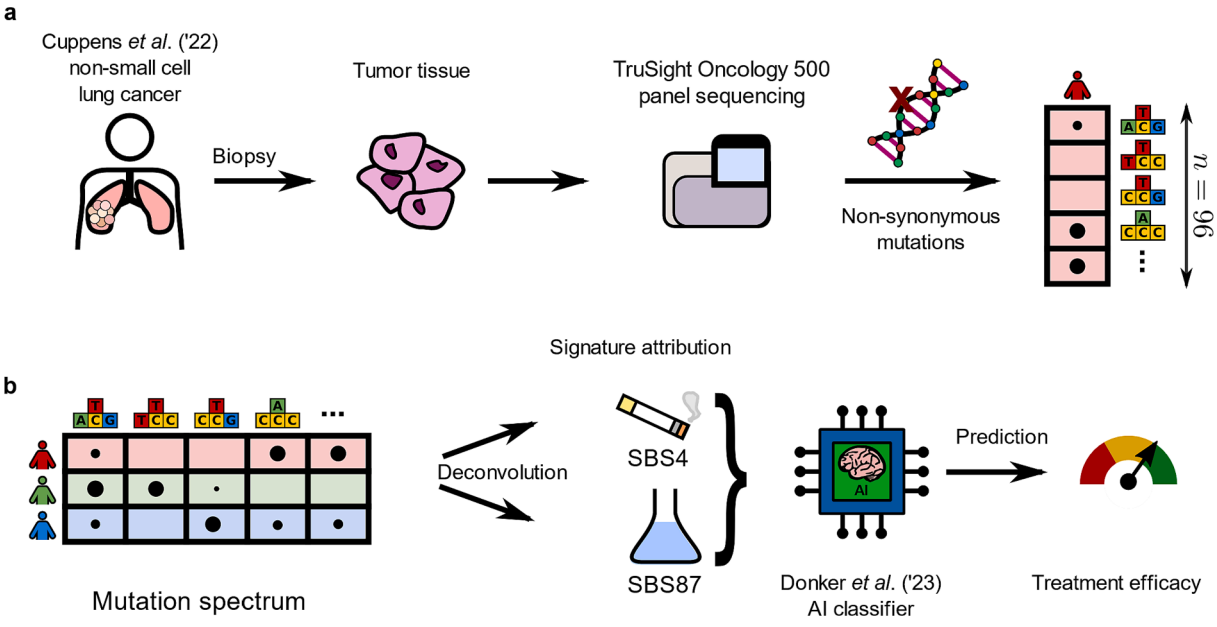


Fig. 1. An immune checkpoint inhibitor (ICI) efficacy classifier was tested on panel-based mutational signatures. **a**, Pre-treatment tumour tissue of a (previously published [22]) non-small cell lung cancer population were sequenced using the TruSight Oncology 500 assay. Per patient, non-synonymous singlets were aggregated by flanking context (indicated by pyramids) into a mutation spectrum. **b**, The panel-derived mutation spectra were deconvoluted into mutational signature attributions. Signature SBS4 and SBS87 attributions were used to predict treatment efficacy using a previously published artificial intelligence (AI) classifier [11].

Table 1
Patient demographics and outcome of immune checkpoint inhibition treated non-small cell lung cancer patients. Symbols and abbreviations: αCTLA-4, cytotoxic T lymphocyte associated antigen-4 inhibitor; αPD-1, programmed death-1 inhibitor; αPD-L1, programmed death-ligand 1 inhibitor; αTNFRSF7, tumour necrosis factor receptor superfamily type 7 inhibitor; αVEGF-A, vascular endothelial growth factor inhibitor; FFPE, formalin-fixed and paraffin embedded; TSO500, Illumina TruSight Oncology 500.

		Cuppens *et al.* ('22)	Donker *et al.* ('23)
M		126	101
Age, median (range)		66 (48–85)	64 (44–82)
Gender, *m* (%)	female	52 (41.3)	52 (51.5)
	male	74 (58.7)	49 (48.5)
Smoker, *m* (%)	unknown		101 (100.0)
	current or former	112 (88.9)	
	never	14 (11.1)	
Extend of disease, *m* (%)	stage IV	112 (88.9)	101 (100.0)
	stage II/III	14 (11.1)	
Prior therapy, *m* (%)	1 ≥ previous lines of therapy	81 (64.3)	83 (82.2)
	naïve	45 (35.7)	18 (17.8)
Treatment, *m* (%)	αCTLA-4 + αPD-1		1 (1.0)
	αCTLA-4 + αPD-1 with chemotherapy		1 (1.0)
	αPD-1	123 (97.6)	77 (76.2)
	αPD-1 with chemotherapy		13 (12.9)
	αPD-L1	3 (2.4)	3 (3.0)
	αPD-L1 with chemotherapy		2 (2.0)
	αPD-L1/cabozantinib		1 (1.0)
	αTNFRSF7 + αPD-1		1 (1.0)
	αVEGF-A + αPD-1		2 (2.0)
Durable benefit, *m* (%)	no	64 (50.8)	57 (56.4)
	yes	62 (49.2)	36 (35.6)
	unknown		8 (7.9)
Sequencing, *m* (%)	TSO500	126 (100.0)	
	whole genome		101 (100.0)
Tissue, *m* (%)	FFPE	126 (100.0)	
	fresh frozen		101 (100.0)

(Supplementary Table 1). Projecting the WGS non-synonymous mutations onto an *in silico* TSO500 panel revealed that the observed FNR were $0.67^{+0.10}_{-0.11}$ and $0.68^{+0.09}_{-0.11}$ for SBS4 and SBS87, respectively. That is, smoking signature SBS4 attributable singlets were highly underrepresented compared to what was theoretically expected based on panel size alone. These large FNRs were not specific to SBS4 and SBS87. Other mutational signatures also had a large FNR (Figs. S2 and S3, Supplementary Material).

Retraining the classifier on the *in silico* dataset showed similar performance degradation (compared to the original WGS mutations) as on the Cuppens et al. cohort with an accuracy of $0.59^{+0.10}_{-0.11}$, average precision of $0.30^{+0.12}_{-0.08}$, and a ROC AUC of $0.15^{+0.10}_{-0.09}$. (For reference, the accuracy, average precision, and ROC AUC for the original, non-projected WGS data, were $0.74^{+0.09}_{-0.09}$, $0.63^{+0.18}_{-0.16}$, and $0.74^{+0.10}_{-0.12}$, respectively [11]). These results provide further confirmation that the lack of predictive power is related to the size of the panel.

3.3. Small panels miss signatures

Next, we studied the FNR as a function of panel size and composition. Comparing the panel's mutation spectra $\boldsymbol{X}$ with WGS, indicated that the Xu et al. pan cancer panel differed [$p = 1.8{\cdot}10^{-12}$, $k = 95$ degrees of freedom χ^2-test] but other panels [TSO500, $p = 0.24$; F1 Cdx, $p = 0.40$; MSK-IMPACT , $p = 0.16$, $k = 95$, χ^2-test] had a similar allocation of mutations across the $n = 96$ channels. Signatures with > 10 positive cases (since we coarse grain over up to ten signature positive patients) and an average attribution of $> 2.9\%$ (i.e., above the expected reconstruction error, see Sec. 3.4, below) were selected for further analysis. This left a total of nine signatures (SBS1, 2, 4, 13, 15, 24, 35, and 39) (Fig. S4, Supplementary Material) and we highlight the results of SBS4 and SBS87 (Fig. 2). For SBS87 the FNR was high across all compositions and (extrapolated) panel sizes, with the decline in FNR saturating starting from ~10 Mb (Fig. 2a). The estimated FNRs were, within the 95 % confidence margin, largely consistent with theory. We do observe a slight underestimation of the theory (especially for the F1 CDx and pan cancer panel) for larger panel sizes. For SBS4, the FNR remained high across different compositions up to a panel size of ~10 Mb (Fig. 2b). Interestingly, these FNRs were much higher than predicted theoretically. Other mutational signatures showed qualitatively similar trends (Fig. S4, Supplementary Material), although the FNR remained high for DNA mismatch repair-associated signature SBS15 [33], cisplatin-associated signature SBS35 [34], and SBS39. This was attributed to specific singlets being missed on the panel, which then remained missing in the extrapolated FNRs. Indeed, on the non-extrapolated TSO500 *in silico* mutations, the addition of one GCG:CGC → GTG:CAC singlet for SBS15, CTT:GAA → CAT:GTA singlet for SBS35, and ACA:TGT → AGA:TCT singlet for SBS39 led to the largest reduction in FNR (out of all channels) for the respective signature. Incidentally, from the samples that were whole-genome wide positive for SBS15 ($m = 44$), SBS35 ($m = 51$), or SBS39 ($m = 71$), those that were negative on TSO500 were almost exclusively negative for respectively GCG:CGC → GTG:CAC (36/36, $p = 5.6{\cdot}10^{-9}$, Fisher exact test), CTT:GAA → CAT:GTA (41/41, $p = 1.1{\cdot}10^{-4}$, Fisher exact test) or ACA:TGT → AGA:TCT (60/62, $p = 0.076$, Fisher exact test).

Conversely, *in silico* TSO500 analysis indicated an average 1.8-fold enrichment (median: 1.7; IQR: 1.2–2.3) for G:C → T:A singlets (Fig. S5, Supplementary Material). That is, the panel was enriched for singlets that are linked to damage from smoke-related aromatic compounds [35]), but depleted for SBS4-attributed singlets. These findings suggest that the signature deconvolution process may itself have introduced spurious attributions.

3.4. Deconvolution of small sets of mutations lead to anomalies

Signature deconvolution can contribute to the discrepancy in two ways: (i) by introducing reconstruction errors or (ii) by signature misattributions (where mutations are attributed to the wrong signature). We therefore further validated the signatures and found that both factors played a role. On the TSO500 the average violation of the conservation of singlets was 30.2% (median: 24.9%; IQR: 16.2–36.1%). For reference, the average violation without a panel was only 2.90% (median: 2.07%; IQR: 0.98–3.92%), significantly less than the TSO500 panel ($p = 6.9{\cdot}10^{-27}$ independent t-test). Other panels had similar violations on average as TSO500: 39.9% (IQR: 22.8–51.6%), 31.4% (IQR: 18.5–41.5%), and 14.7% (IQR: 6.61–18.2%) for the F1 CDx, MSK-IMPACT, and PANCAN panels, respectively. We verified that these errors were not related to lack of convergence in the NMF algorithm, by decreasing the error tolerance and increasing the number of iterations. We then looked for misattributions. To this end, attributions were normalised (i.e.,$\widetilde{W}_{ij} = W_{ij}/|W_i|_1$, where $|\cdot|_1$ is the L_1 norm) and we compared the *in silico* projected signatures on the TSO500 with the signatures derived from all (non-synonymous) singlets (Fig. 3). On average, signature SBS4 had the largest difference (16.5%) followed by aflatoxin-associated signature [36] SBS24 (6.53%), SBS39 (3.85%), and SBS35 (1.95%). Incidentally, these were the signatures that all had higher FNRs than theoretically predicted (Fig. S4, Supplementary Material). Finally, we determined which signatures, with at least one attribution ($m = 57$), were misattributed. Signatures SBS3 ($q = 0.03$,

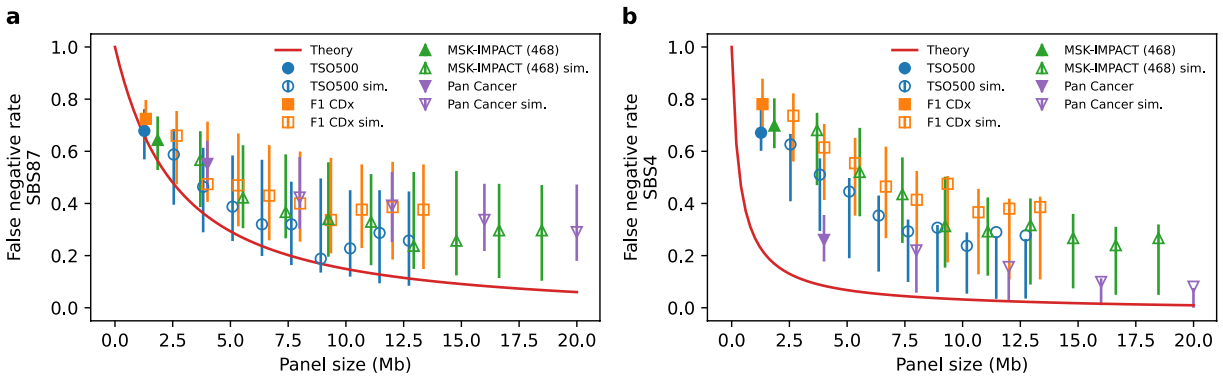


Fig. 2. Relation between (extrapolated) panel size, composition, and the probability that an assay fails to detect a single base substitution (SBS) signature. **a**, False negative rate (FNR, compared to whole genome sequencing) of signature SBS87. Filled markers are observed *in silico* FNRs (panels are indicated in the legend). Open markers (indicated by sim. In the legend) are FNRs extrapolated to larger (integral multiples) panel sizes by coarse graining over (resampled) patients with similar signature profile. The ab initio theory (solid line) corresponds to a substitution detection probability proportional to panel size. **b**, Like in **a**, but for signature for SBS4. Error bars indicate the 95 % confidence interval region obtained by bootstrapping 100 iterations. Abbreviations: F1 CDx, FoundationOne CDx; Mb, megabases; MSK-IMPACT, Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets; sim., simulated;TSO500, TruSight Oncology 500.

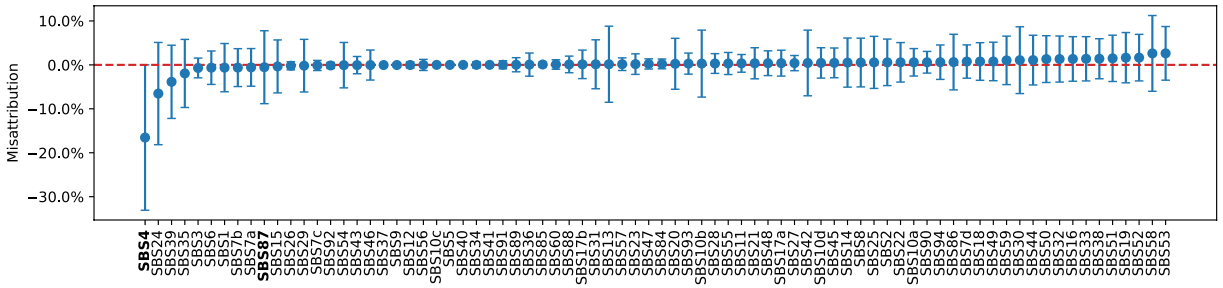


Fig. 3. *In silico* analysis of the TruSeq Oncology 500 panel shows that mutational signature SBS4, but not SBS87, is misattributed. Average (normalised) misattributions (and standard deviation) per mutational signature (ranked from low to high).

$p = 0.0023$), SBS4 ($q = 4.5{\cdot}10^{-15}, p = 7.9{\cdot}10^{-17}$), SBS19 ($q = 0.046$, $p = 0.0049$), SBS24 ($q = 4.5{\cdot}10^{-6}$,p $= 1.6{\cdot}10^{-7}$), SBS38 ($q = 0.03, p = 0.0026$), and SBS39 ($q = 1.9{\cdot}10^{-4}, p = 1.0{\cdot}10^{-5}$) were all found to be significantly misattributed [one sample BH corrected *t*-test]. SBS87 was not misattributed ($q = 0.69, p = 0.52$, one sample BH corrected *t*-test). Analysis of the SBS4 depletion linked a (single) signature to misattributions: the cisplatin-associated signature [34] SBS35, as it was consistently increased with decreasing SBS4 ($\tau = -0.25$, $q = 0.036$, $p = 7.0{\cdot}10^{-4}$, BH corrected Kendall tau correlation). No significant misattributions were observed for SBS87.

4. Discussion

Our (parameter free) *ab initio* theory relied on panel size l only and ignored composition. *In silico* analyses showed that for SBS87 this was justified. SBS4 attributions were, however, relatively depleted compared to our theoretical predictions. We ascribed this to reconstruction errors and misattributions in the deconvolution process of relatively small panels. SBS4's susceptibility to misattributions is not surprising in view of its relatively high entropy (3.8 nats compared to an average SBS signature entropy of 3.0 nats). Since SBS4 is most similar to SBS94 (cosine distance 0.15, or $0.089{\cdot}2\pi$ rad) and SBS45 (cosine distance 0.17, or $0.094{\cdot}2\pi$ rad) we expected misattributions primarily to these signatures. However, SBS4 misattributions were found to be linked to cisplatin-associated signature [34] SBS35 (cosine distance 0.32, or $0.13{\cdot}2\pi$ rad). We have no direct explanation for this observation other than that this signature also had a high probability of smoke-related [35] G:C → T:A substitutions.

Several limitations apply to our *in silico* analysis. First, we extrapolated existing panels to hypothetical larger-sized panels of fixed composition by coarse graining resampled patients. This was a convenient but unrealistic idealisation: changes in panel size can only be accompanied by changes in composition. The high extrapolated FNR for SBS15, SBS35, and SBS39, which were ascribed to the missing of specific singlets, are assumed to be related to this oversimplification. The FNR of other mutational signatures maybe affected by the same mechanism, and therefore overestimate the FNR, specifically for the larger extrapolated panel sizes (where FNR saturation as a function of panel size was observed). Second, the coarse graining approach tacitly assumes that a panel double the size (but a similar composition) detects on average twice as many mutations. Third, although we coarse grained over similar patients (by stratifying by signature cluster), the aggregation procedure may underestimate the variability of the false negative rate for a small set of $m = 101$ patients. So the"true" error bars in the FNR may be wider than those calculated. Fourth, we implicitly assumed that the mutations detected using whole genome sequencing of matched tumour-normal pairs with different sample handling and bioinformatics pipeline, faithfully represent mutations that would be detected using the tumour-only TSO500 assay. Reassuringly, analytical validation of TSO500 shows, at least for TMB, high correlation with whole exome or genome sequencing [16,17,37–39]. Whether this also applies more generally for the three-nucleotide singlets is, to the best of our knowledge, still an open problem. We can therefore not rule out that germline mutations from the of Cuppens et al. cohort [22] could affect the inferred mutational signatures.

Effects related to formalin fixation and paraffin embedding (e.g., C:G → T:A cytosine deaminations [32]) on the samples of Cuppens et al. [22] were not fully addressed with our *in silico* analysis (which were based on fresh-frozen material). Earlier work found that reliable panel-based variant calling may be achieved by imposing a variant allele

frequency limit of detection $\geq$ 5% together with molecular barcodes to suppress deamination artefacts [38]. Both are part of the TSO500 workflow. Moreover, the average attribution to FFPE-related signatures SBS1 and SBS30 [32] in the samples of Cuppens et al. was found to be low. Finally, there is a slight difference in population between those reported by Cuppens et al. [22] and Donker et al. [11]. In particular, the latter solely consists of metastatic lung cancer while the former also contained stage II or III lung cancer (11.1%). In addition, there were more treatment naïve patients in the cohort of Cuppens et al. Both factors could affect the (for Cuppens et al. unobserved whole genome) distribution of mutations across patients. The mutation distribution, in turn, affects the theoretically inferred FNR [cf. Eq. (1)]. However, with a FNR around 70% this is unlikely to change the conclusions regarding the utility of these panel-based mutational signature attributions, where larger sizes of mutational panels will be relevant as also indicated by the better performance of the pan-cancer panel.

We expect that our findings, that mutational signature attributions from panel sequencing have a large FNR, carry over to most other cancer types, since lung cancer is among the highest mutated cancers [40]. Possible exceptions are melanoma and microsatellite-unstable colorectal cancer—both are characterised by a high mutation load. While different mutational processes are operative in other types of cancer, they would presumably be affected by panel size in a similar way.

Finally, we (i) considered non-synonymous mutations only unlike [18–21] and (ii) different from [18–20] we also analysed samples without a high mutation count. Both factors would contribute to a reduction in the number of false negative signature attributions compared to our work. Heuristically speaking, these factors can be thought of as inducing a larger "effective panel size". Indeed, both our *in silico* analysis cohort (based on samples from [11]) and that of [21] were originally collected in Ref. [24]. As such, our results can be considered complementary to the work of [21].

In toto, our findings suggest that larger panels are needed to be reliable enough (to detect the presence of mutational signatures in a sample) for use in ICI treatment-efficacy prediction models. Compared to size, panel composition plays a minor role.

5. Conclusion

This paper studied if mutational signatures derived from targeted gene panel sequencing can be used to predict immune checkpoint inhibition efficacy in non-small cell lung cancer (NSCLC). Our results came out negative. Based on theoretical arguments, experimental data, and *in silico* simulations we showed that mutational signature attributions from panel sequencing are unreliable. Specifically, mutational signatures in general and SBS4 and SBS87 in specific are characterised by a large false negative rate (FNR around 70%). Not only are commercially available targeted gene panels too small to detect sufficient mutations. We also observed a secondary effect, where small ensembles of singlets lead to reconstruction errors and misattributions in the deconvolution process. We therefore suggest that, for downstream classification tasks in NSCLC, signature attributions should be derived from whole exome or genome instead of targeted panel sequencing.

CRediT authorship contribution statement

H.C. Donker: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **K. Cuppens:** Resources, Writing – review & editing. **G. Froyen:** Resources, Writing – review & editing. **H.J.M. Groen:** Writing – review & editing. **T.J.N. Hiltermann:** Writing – review & editing. **B. Maes:** Resources, Writing – review & editing. **E. Schuuring:** Writing – review & editing. **P.-J. Volders:** Resources, Writing – review & editing. **G.A. Luntera:** Writing – review & editing. **B. van Es:** Conceptualization, Validation, Formal analysis, Data curation, Writing – review & editing, Visualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: HCD: None to declare; KC: None to declare; GF: None to declare; HJMG: Consulting or Advisory Role: Novartis, Lilly. TJNH: Advisory/consultancy fees from AstraZeneca, Bristol-Myers-Squibb, Illumina, Merck Sharp Dohme, Roche, and research grants/funding from AstraZeneca, Hoffmann-La Roche. BM: None to declare; ES: Honoraria/speakers fee: Bio-Rad, Roche, Agena Bioscience, Illumina, Lilly; Consulting or Advisory Role: MSD/Merck, Astellas, Bayer, BMS, Agena Bioscience, Janssen Cilag (Johnson & Johnson), Novartis, Roche, AstraZeneca, Amgen, Lilly; Research Funding: Biocartis, Bio-Rad, Roche, Agena Bioscience, AstraZeneca, InVitae/Archer (all paid to UMCG); Travel, Accommodations, Expenses: Roche Molecular Diagnostics, Bio-Rad. PJV: None to declare; GAL: Research grant from Boehringer-Ingelheim. Owns shares in Genomics PLC.; BvE: None to declare.

Acknowledgements

This publication and the underlying research are partly facilitated by Hartwig Medical Foundation and the Center for Personalized Cancer Treatment (CPCT) which have generated, analysed and made available data for this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lungcan.2023.107286.

References

[1] U. Dafni, Z. Tsourti, K. Vervita, S. Peters, Immune checkpoint inhibitors, alone or in combination with chemotherapy, as first-line treatment for advanced non-small cell lung cancer. A systematic review and network meta-analysis, Lung Cancer 134 (2019) 127–140.
[2] L.M. Sholl, Biomarkers of response to checkpoint inhibitors beyond PD-L1 in lung cancer, Mod. Pathol. 35 (Suppl 1) (2022) 66–74.
[3] L.M. Sholl, F.R. Hirsch, D. Hwang, J. Botling, F. Lopez-Rios, L. Bubendorf, M. Mino-Kenudson, A.C. Roden, M.B. Beasley, A. Borczuk, E. Brambilla, G. Chen, T.-Y. Chou, J.-H. Chung, W.A. Cooper, S. Dacic, S. Lantuejoul, D. Jain, D. Lin, Y. Minami, A. Moreira, A.G. Nicholson, M. Noguchi, M. Papotti, G. Pelosi, C. Poleri, N. Rekhtman, M.-S. Tsao, E. Thunnissen, W. Travis, Y. Yatabe, A. Yoshida, J. B. Daigneault, A. Zehir, S. Peters, I.I. Wistuba, K.M. Kerr, J.W. Longshore, The promises and challenges of tumor mutation burden as an immunotherapy biomarker: A perspective from the International Association for the Study of Lung Cancer Pathology Committee, J. Thorac. Oncol. 15 (9) (2020) 1409–1424.
[4] B. Zhang, Y.i. Liu, S. Zhou, H. Jiang, K.e. Zhu, R. Wang, Predictive effect of PD-L1 expression for immune checkpoint inhibitor (PD-1/PD-L1 inhibitors) treatment for non-small cell lung cancer: A meta-analysis, Int. Immunopharmacol. 80 (2020) 106214.
[5] S.W. Brady, A.M. Gout, J.H. Zhang, Therapeutic and prognostic insights from the analysis of cancer mutational signatures, Trends Genet. 38 (2) (2022) 194–208.
[6] N.A. Rizvi, M.D. Hellmann, A. Snyder, P. Kvistborg, V. Makarov, J.J. Havel, W. Lee, J. Yuan, P. Wong, T.S. Ho, M.L. Miller, N. Rekhtman, A.L. Moreira, F. Ibrahim, C. Bruggeman, B. Gasmi, R. Zappasodi, Y. Maeda, C. Sander, E.B. Garon, T. Merghoub, J.D. Wolchok, T.N. Schumacher, T.A. Chan, Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer, Science 348 (6230) (2015) 124–128.
[7] D. Miao, C.A. Margolis, N.I. Vokes, D. Liu, A. Taylor-Weiner, S.M. Wankowicz, D. Adeegbe, D. Keliher, B. Schilling, A. Tracy, M. Manos, N.G. Chau, G.J. Hanna, P. Polak, S.J. Rodig, S. Signoretti, L.M. Sholl, J.A. Engelman, G. Getz, P.A. Jänne, R. I. Haddad, T.K. Choueiri, D.A. Barbie, R. Haq, M.M. Awad, D. Schadendorf, F. S. Hodi, J. Bellmunt, K.-K. Wong, P. Hammerman, E.M. Van Allen, Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors, Nat Genet 50 (9) (2018) 1271–1281.
[8] W. Chong, Z. Wang, L. Shang, S. Jia, J. Liu, Z. Fang, F. Du, H. Wu, Y. Liu, Y. Chen, H. Chen, Association of clock-like mutational signature with immune checkpoint inhibitor outcome in patients with melanoma and NSCLC, Mol. Ther.-Nucleic Acids 23 (2021) 89–100.
[9] V. Anagnostou, N. Niknafs, K. Marrone, D.C. Bruhm, J.R. White, J. Naidoo, K. Hummelink, K. Monkhorst, F. Lalezari, M. Lanis, S. Rosner, J.E. Reuss, K. N. Smith, V. Adleff, K. Rodgers, Z. Belcaid, L. Rhymee, B. Levy, J. Feliciano, C. L. Hann, D.S. Ettinger, C. Georgiades, F. Verde, P. Illei, Q.K. Li, A.S. Baras, E. Gabrielson, M.V. Brock, R. Karchin, D.M. Pardoll, S.B. Baylin, J.R. Brahmer, R. B. Scharpf, P.M. Forde, V.E. Velculescu, Multimodal genomic features predict

outcome of immune checkpoint blockade in non-small-cell lung cancer, Nature Cancer 1 (1) (2020) 99–111.
[10] S. Wang, M. Jia, Z. He, X.-S. Liu, APOBEC3B and APOBEC mutational signature as potential predictive markers for immunotherapy response in non-small cell lung cancer, Oncogene 37 (29) (2018) 3924–3936.
[11] H.C. Donker, B. van Es, M. Tamminga, G.A. Lunter, L.C.L.T. van Kempen, E. Schuuring, T.J.N. Hiltermann, H.J.M. Groen, Using genomic scars to select immunotherapy beneficiaries in advanced non-small cell lung cancer, Sci. Rep. 13 (1) (2023).
[12] A. Ravi, et al., Genomic and transcriptomic analysis of checkpoint blockade response in advanced non-small cell lung cancer, Nat. Genet. (2023) 807–819.
[13] L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, S.A.J.R. Aparicio, S. Behjati, A. V. Biankin, G.R. Bignell, N. Bolli, A. Borg, A.-L. Børresen-Dale, S. Boyault, B. Burkhardt, A.P. Butler, C. Caldas, H.R. Davies, C. Desmedt, R. Eils, J.E. Eyfjörd, J.A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jäger, D.T.W. Jones, D. Jones, S. Knappskog, M. Kool, S.R. Lakhani, C. López-Otín, S. Martin, N.C. Munshi, H. Nakamura, P.A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J.V. Pearson, X.S. Puente, K. Raine, M. Ramakrishna, A.L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P.N. Span, J.W. Teague, Y. Totoki, A.N.J. Tutt, R. Valdés-Mas, M. M. van Buuren, L. van `t Veer, A. Vincent-Salomon, N. Waddell, L.R. Yates, J. Zucman-Rossi, P. Andrew Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S.M. Pfister, P.J. Campbell, M. R. Stratton, Signatures of mutational processes in human cancer (vol 500, pg 415, 2013), Nature 500 (7463) (2013) 415–421.
[14] J.E. Kucab, X. Zou, S. Morganella, M. Joel, A.S. Nanda, E. Nagy, C. Gomez, A. Degasperi, R. Harris, S.P. Jackson, V.M. Arlt, D.H. Phillips, S. Nik-Zainal, A compendium of mutational signatures of environmental agents, Cell 177 (4) (2019) 821–836.e16.
[15] B. Li, S.W. Brady, X. Ma, S. Shen, Y. Zhang, Y. Li, K. Szlachta, L.i. Dong, Y.u. Liu, F. Yang, N. Wang, D.A. Flasch, M.A. Myers, H.L. Mulder, L. Ding, Y. Liu, L. Tian, K. Hagiwara, K.e. Xu, X. Zhou, E. Sioson, T. Wang, L. Yang, J. Zhao, H. Zhang, Y. Shao, H. Sun, L. Sun, J. Cai, H.-Y. Sun, T.-N. Lin, L. Du, H. Li, M. Rusch, M. N. Edmonson, J. Easton, X. Zhu, J. Zhang, C. Cheng, B.J. Raphael, J. Tang, J. R. Downing, L.B. Alexandrov, B.-B. Zhou, C.-H. Pui, J.J. Yang, J. Zhang, Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia, Blood 135 (1) (2020) 41–55.
[16] V. Pestinger, M. Smith, T. Sillo, J.M. Findlay, J.-F. Laes, G. Martin, G. Middleton, P. Taniere, A.D. Beggs, Use of an integrated pan-cancer oncology enrichment next-generation sequencing assay to measure tumour mutational burden and detect clinically actionable variants, Mol. Diagn. Ther. 24 (3) (2020) 339–349.
[17] C. Zhao, et al., TruSight Oncology 500: Enabling Comprehensive Genomic Profiling and Biomarker Reporting with Targeted Sequencing. bioRxiv, 2020: p. 2020.10.21.349100.
[18] V. Endris, I. Buchhalter, M. Allgäuer, E. Rempel, A. Lier, A.-L. Volckmar, M. Kirchner, M. Winterfeld, J. Leichsenring, O. Neumann, R. Penzel, W. Weichert, H. Glimm, S. Fröhling, H. Winter, F. Herth, M. Thomas, P. Schirmacher, J. Budczies, A. Stenzinger, Measurement of tumor mutational burden (TMB) in routine molecular diagnostics: in silico and real-life analysis of three larger gene panels, Int. J. Cancer (2019).
[19] L.I. Kroeze, R.M. de Voer, E.J. Kamping, D. von Rhein, E.A.M. Jansen, M.J. W. Hermsen, M.C.P. Barberis, J. Botling, E.M. Garrido-Martin, F. Haller, L. Lacroix, B. Maes, S. Merkelbach-Bruse, V. Pestinger, N. Pfarr, A. Stenzinger, M.M. van den Heuvel, K. Grünberg, M.J.L. Ligtenberg, Evaluation of a hybrid capture-based pan-cancer panel for analysis of treatment stratifying oncogenic aberrations and processes, J. Mol. Diagn. 22 (6) (2020) 757–769.
[20] G.R.M. van den Heuvel, L.I. Kroeze, M.J.L. Ligtenberg, K. Grünberg, E.A.M. Jansen, D. von Rhein, R.M. de Voer, M.M. van den Heuvel, Mutational signature analysis in non-small cell lung cancer patients with a high tumor mutational burden, Respir. Res. 22 (1) (2021).
[21] S.M. Ernst, J.M. Mankor, J. van Riet, J.H. von der Thüsen, H.J. Dubbink, J.G.J. V. Aerts, A.J. de Langen, E.F. Smit, A.-M. Dingemans, K. Monkhorst, Tobacco smoking-related mutational signatures in classifying smoking-associated and nonsmoking-associated NSCLC, J. Thorac. Oncol. 18 (4) (2023) 487–498.
[22] K. Cuppens, P. Baas, E. Geerdens, B. Cruys, G. Froyen, L. Decoster, M. Thomeer, B. Maes, HLA-I diversity and tumor mutational burden by comprehensive next-generation sequencing as predictive biomarkers for the treatment of non-small cell lung cancer with PD-(L)1 inhibitors, Lung Cancer 170 (2022) 1–10.
[23] Z.W. Xu, et al., Assessment of tumor mutation burden calculation from gene panel sequencing data, Oncotarget. Ther. 12 (2019) 3401–3409.
[24] P. Priestley, J. Baber, M.P. Lolkema, N. Steeghs, E. de Bruijn, C. Shale, K. Duyvesteyn, S. Haidari, A. van Hoeck, W. Onstenk, P. Roepman, M. Voda, H. J. Bloemendal, V.C.G. Tjan-Heijnen, C.M.L. van Herpen, M. Labots, P.O. Witteveen, E.F. Smit, S. Sleijfer, E.E. Voest, E. Cuppen, Pan-cancer whole-genome analyses of metastatic solid tumours, Nature 575 (7781) (2019) 210–216.
[25] Illumina *DRAGEN TruSight Oncology 500 Analysis Software v2.1 (Local), User Guide*. 2022.
[26] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6) (2010) 841-842.
[27] X. Wang, J. Ren, Q. Zhao, *Abstract 5004: Integrative quality control of cancer somatic mutations with CaMutQC.* Cancer Res., 82(12_Supplement) (2022) 5004-5004.
[28] S. Neph, et al., *BEDOPS: high-performance genomic feature operations.* Bioinformatics, 28(14) (2012) 1919-1920.
[29] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
[30] J. Budczies, M. Allgäuer, K. Litchfield, E. Rempel, P. Christopoulos, D. Kazdal, V. Endris, M. Thomas, S. Fröhling, S. Peters, C. Swanton, P. Schirmacher, A. Stenzinger, Optimizing panel-based tumor mutational burden (TMB) measurement, Ann. Oncol. 30 (9) (2019) 1496–1506.
[31] A.J. Coffey, F. Kokocinski, M.S. Calafato, C.E. Scott, P. Palta, E. Drury, C.J. Joyce, E.M. LeProust, J. Harrow, S. Hunt, A.-E. Lehesjoki, D.J. Turner, T.J. Hubbard, A. Palotie, The GENCODE exome: sequencing the complete human exome, Eur. J. Hum. Genet. 19 (7) (2011) 827–831.
[32] Q. Guo, E. Lakatos, I.A. Bakir, K. Curtius, T.A. Graham, V. Mustonen, The mutational signatures of formalin fixation on the human genome, Nat. Commun. 13 (1) (2022).
[33] B. Meier, N.V. Volkova, Y.e. Hong, P. Schofield, P.J. Campbell, M. Gerstung, A. Gartner, Mutational signatures of DNA mismatch repair deficiency in C. elegans and human cancers, Genome Res. 28 (5) (2018) 666–675.
[34] A. Boot, M.N. Huang, A.W.T. Ng, S.-C. Ho, J.Q. Lim, Y. Kawakami, K. Chayama, B. T. Teh, H. Nakagawa, S.G. Rozen, In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors, Genome Res. 28 (5) (2018) 654–665.
[35] S., Nik-Zainal, et al., The genome as a record of environmental exposure. Mutagenesis, 30(6) (2015) 763-770.
[36] S. Chawanthayatham, C.C. Valentine, B.I. Fedeles, E.J. Fox, L.A. Loeb, S.S. Levine, S.L. Slocum, G.N. Wogan, R.G. Croy, J.M. Essigmann, Mutational spectra of aflatoxin B 1 in vivo establish biomarkers of exposure for human hepatocellular carcinoma, Proc. Natl. Acad. Sci. U.S.A. 114 (15) (2017).
[37] B.o. Wei, J. Kang, M. Kibukawa, G. Arreaza, M. Maguire, L. Chen, P. Qiu, L. Lang, D. Aurora-Garg, R. Cristescu, D. Levitan, Evaluation of the TruSight Oncology 500 assay for routine clinical testing of tumor mutational burden and clinical utility for predicting response to pembrolizumab, J. Mol. Diagn. 24 (6) (2022) 600–608.
[38] A. Stenzinger, et al., Harmonization and standardization of panel-based tumor mutational burden measurement: Real-world results and recommendations of the quality in pathology study, J. Thorac. Oncol. 15 (7) (2020) 1177–1189.
[39] C. Heydt, J. Rehker, R. Pappesch, T. Buhl, M. Ball, U. Siebolts, A. Haak, P. Lohneis, R. Büttner, A.M. Hillmer, S. Merkelbach-Bruse, Analysis of tumor mutational burden: correlation of five large gene panels with whole exome sequencing, Sci. Rep. 10 (1) (2020).
[40] L.B. Alexandrov, et al., The repertoire of mutational signatures in human cancer, Nature 578 (7793) (2020) 94–101.