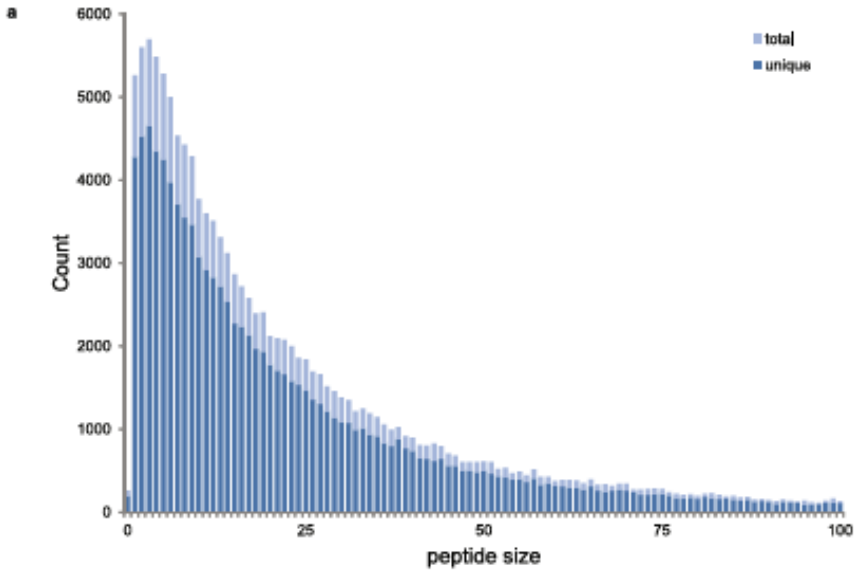
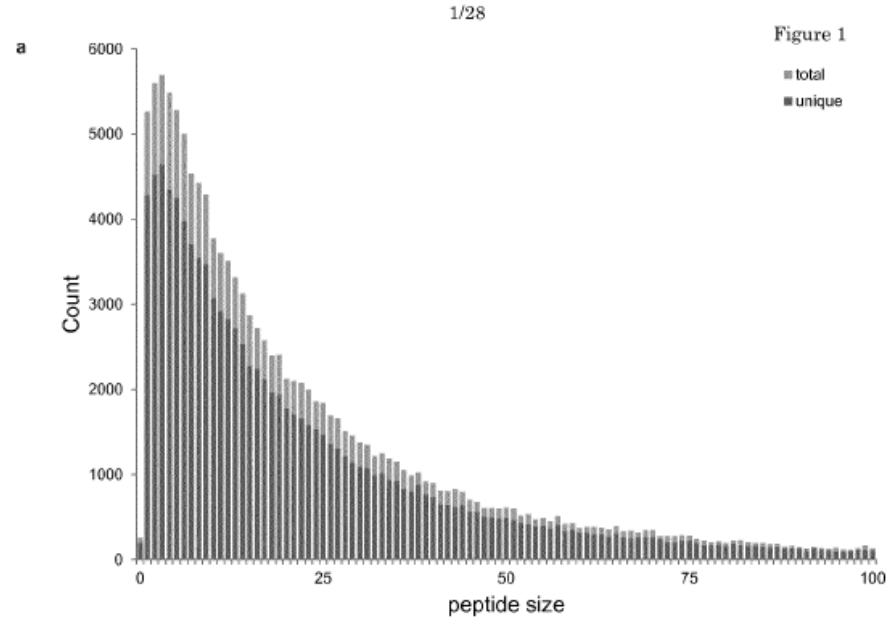
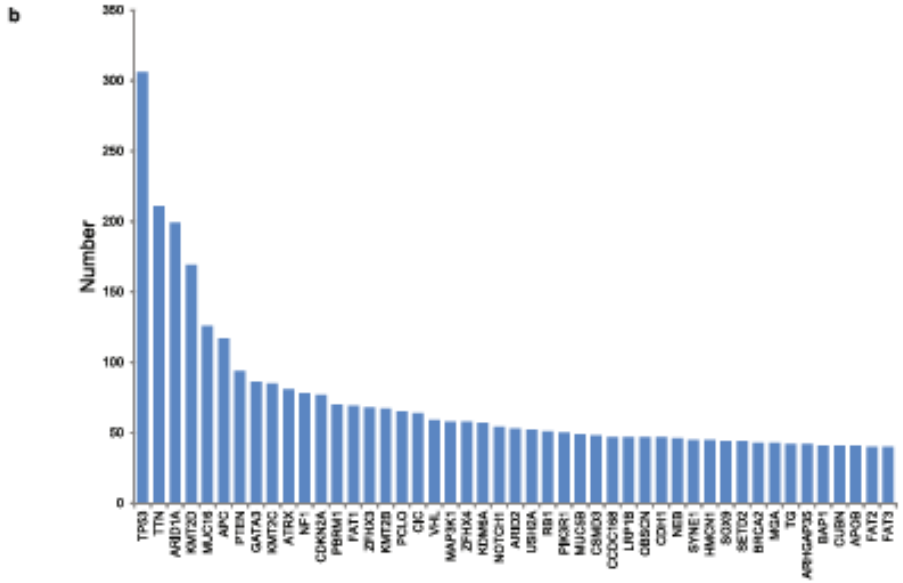
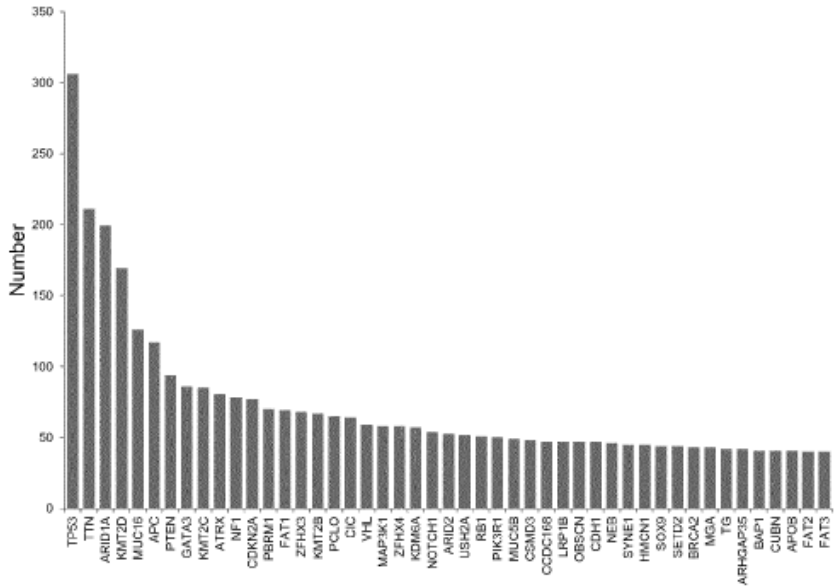
Bijlage bij publicatie A. Tsoutsanis, “Over patenten en integriteit. De affaire Plasterk in vier vragen”, *NJB* 2024, p. 1233-1235.

#	**Artikel Koster en Plasterk (ingediend 12.10.2018, goedgekeurd 14.02.2019 en gepubliceerd 29.04.2019)**	**Octrooiaanvraag met aanvraagnummer PCT/NL/2019/050496 resp. EP191676620 als ingediend op 05.04.2019 met uitsluitend naamsvermelding Plasterk als uitvinder. De tekst van deze aanvraag is grotendeels gelijk aan de later op 25.07.2019 ingediende octrooiaanvraag WO2020022903A1.**
1	Somatic mutations in cancer can result in neoantigens against which patients can be vaccinated. The quest for tumor specific neoantigens has yielded no targets that are common to all tumors, yet foreign to healthy cells. Single base pair substitutions (SNVs) at best can alter 1 amino acid which can result in a neoantigen; with the exception of rare site-specific oncogenic driver mutations (such as RAS) such mutations are private. Here, we describe a source of common neoantigens induced by frame shift	Somatic mutations in cancer can result in neoantigens against which patients can be vaccinated. Unfortunately, the quest for tumor specific neoantigens has yielded no targets that are common to all tumors, yet foreign to healthy cells. Single base pair substitutions (SNVs) at best can alter 1 amino acid which can result in a neoantigen. However, with the exception of rare site-specific oncogenic driver mutations (such as RAS or BRAF) such mutations are private and thus not generalizable.
2	mutations are private. Here, we describe a source of common neoantigens induced by frame shift mutations, based on analysis of 10,186 TCGA tumor samples. We find that these frame shift mutations can produce long neoantigens. These are completely new to the body, and indeed recent evidence suggests that frame shifts can be highly immunogenic. We report that many different frame shift	In the present disclosure we provide a source of common neoantigens induced by frame shift mutations, based on analysis of 10,186 TCGA tumor samples and 568 and 1918 tumor samples (see Priestley et al. 2019 at https://doi.org/10.1101/415133 and Razavi et al. 2018 Cancer Cell 34:427-438, respectively). We find that these frame shift mutations can produce long neoantigens. These neoantigens are typically new to the body, and can be highly immunogenic. The heterogeneity in the mutations that are found in tumors of
3	suggests that frame shifts can be highly immunogenic. We report that many different frame shift mutations converge to the same small set of 3' neo open reading frame peptides (NOPs), all encoded by the Neo-ORFeome. We find that a fixed set of only 1,244 neo-peptides in as much as 30% of all TCGA cancer patients. For some tumor classes this is higher; e.g. for colon and cervical cancer, peptides derived from only ten genes (saturated at 90 peptides) can be applied to 39% of all patients. 50% of all TCGA patients can be achieved at saturation (using all those peptides in the library found more than once). A pre-fabricated library of vaccines (peptide, RNA or DNA) based on this set can provide off the shelf, quality certified, 'personalized' vaccines within hours, saving months of vaccine preparation. This is crucial for critically ill cancer patients with short average survival expectancy after diagnosis.	treatments. In the present disclosure it was found that many of the possible different frame shift mutations in a gene converge to the same small set of 3' neo open reading frame peptides (neopeptides or NOPs). We find a fixed set of only 1,244 neopeptides in as much as 30% of all TCGA cancer patients. For some tumor classes this is higher; e.g. for colon and cervical cancer, peptides derived from only ten genes (saturated at 90 peptides) can be applied to 39% of all patients. 50% of all TCGA patients can be targeted at saturation (using all those peptides in the library found more than once). A pre-fabricated library of vaccines (peptide, RNA or DNA) based on this set can provide off the shelf, quality certified, 'personalized' vaccines within hours, saving months of vaccine preparation. This is important for critically ill cancer patients with short average survival expectancy after diagnosis.

Bijlage bij publicatie A. Tsoutsanis, “Over patenten en integriteit. De affaire Plasterk in vier vragen”, *NJB* 2024, p. 1233-1235.

	Artikel	Octrooiaanvraag
4	We have analyzed 10,186 cancer genomes from 33 tumor types of the TCGA (The Cancer Genome Atlas[22]) and focused on the 143,444 frame shift mutations represented in this cohort (see Table S1). Translation of these mutations after re-annotation to a RefSeq annotation, starting in the protein reading frame, can lead to 70,439 unique peptides that are 10 or more amino acids in length (a cut-off we have set at a size sufficient to shape a distinct epitope in the context of MHC (Fig. 1a). The list of genes most commonly represented in the cohort and containing such frame shift mutations is headed nearly exclusively by tumor driver genes, such as NF1, RB, BRCA2 (Fig. 1b and Table S2) whose whole or partial loss of function apparently contributes to tumorigenesis. Note that a priori frame shift mutations are expected to result in loss of gene function more than a random SNV, and more independent of the precise position. In conclusion, NOPs initiated from a frameshift mutation and of a significant size are prevalent in tumors, and are enriched in cancer driver genes. Alignment of the translated NOP products onto the protein sequence reveals that a wide array of different frame shift mutations translate in a common downstream stretch of neo open reading frame peptides ('NOPs'), as dictated by the −1 and +1 alternative reading frames. While we initially screened for NOPs of ten or more amino acids, their open reading frame in the out-of-frame genome often extends far beyond that search window. As a result we see (Fig. 2) that hundreds of different frame shift mutations all at different sites in the gene nevertheless converge on only a handful of NOPs. Similar patterns are found in other common driver genes (Supplementary Fig. S1).	We have analyzed 10,186 cancer genomes from 33 tumor types of the 40 TCGA (The Cancer Genome Atlas[22]) and focused on the 143,444 frame shift mutations represented in this cohort. Translation of these mutations after re-annotation to a RefSeq annotation, starting in the protein reading frame, can lead to 70,439 unique peptides that are 10 or more amino acids in length (a cut off we have set at a size sufficient to shape a distinct epitope in the context of MHC (figure 1a). The list of BRCA2 (figure 1b) whose whole or partial loss of function apparently contributes to tumorigenesis. Note that a priori frame shift mutations are expected to result in loss of gene function more than a random SNV, and more independent of the precise position. NOPs initiated from a frameshift mutation and of a significant size are prevalent in tumors, and are enriched in cancer driver genes. Alignment of the translated NOP products onto the protein sequence reveals that a wide array of different frame shift mutations translate in a common downstream stretch of neo open reading frame peptides ('NOPs'), as dictated by the -1 and +1 alternative reading frames. While we initially screened for NOPs of ten or more amino acids, their open reading frame in the out-of-frame genome often extends far beyond that search window. As a result we see (figure 2) that hundreds of different frame shift mutations all at different sites in the gene nevertheless converge on only a handful of NOPs. Similar patterns are found in other common driver genes (figure 5).
5		 Figure 1 

Bijlage bij publicatie A. Tsoutsanis, “Over patenten en integriteit. De affaire Plasterk in vier vragen”, *NJB* 2024, p. 1233-1235.

	Artikel	**Octrooiaanvraag**
6		
7	Figuren 2A, 2B, 2C, 2D, 2E en 2F zijn identiek	Figuren 2A, 2B, 2C, 2D, 2E en 2F zijn identiek

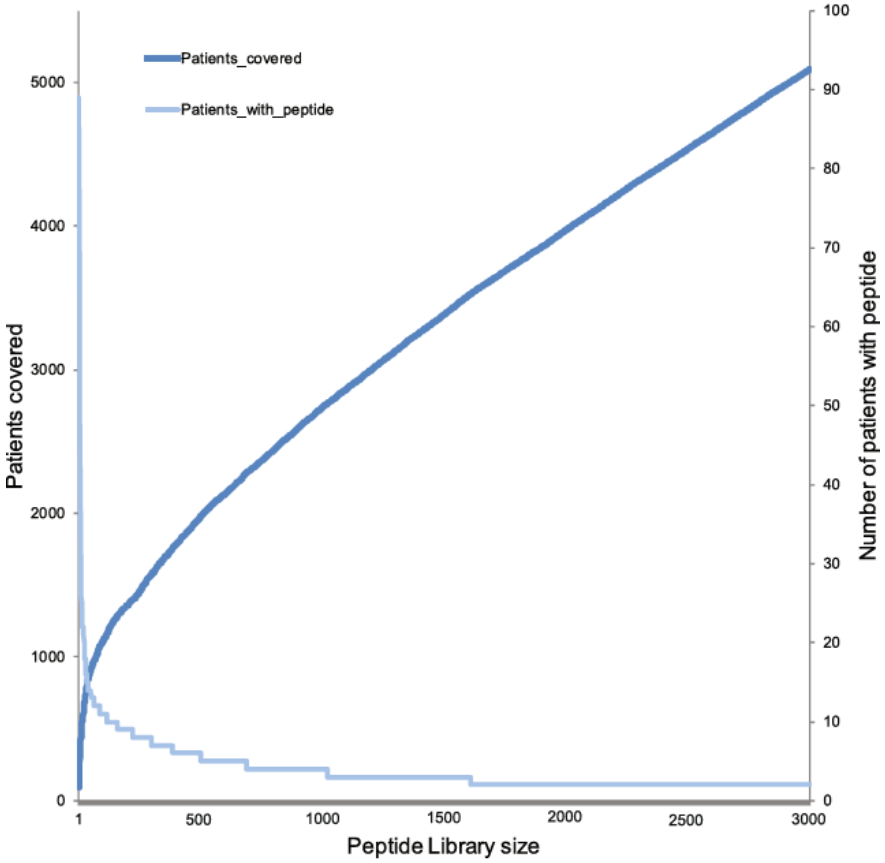
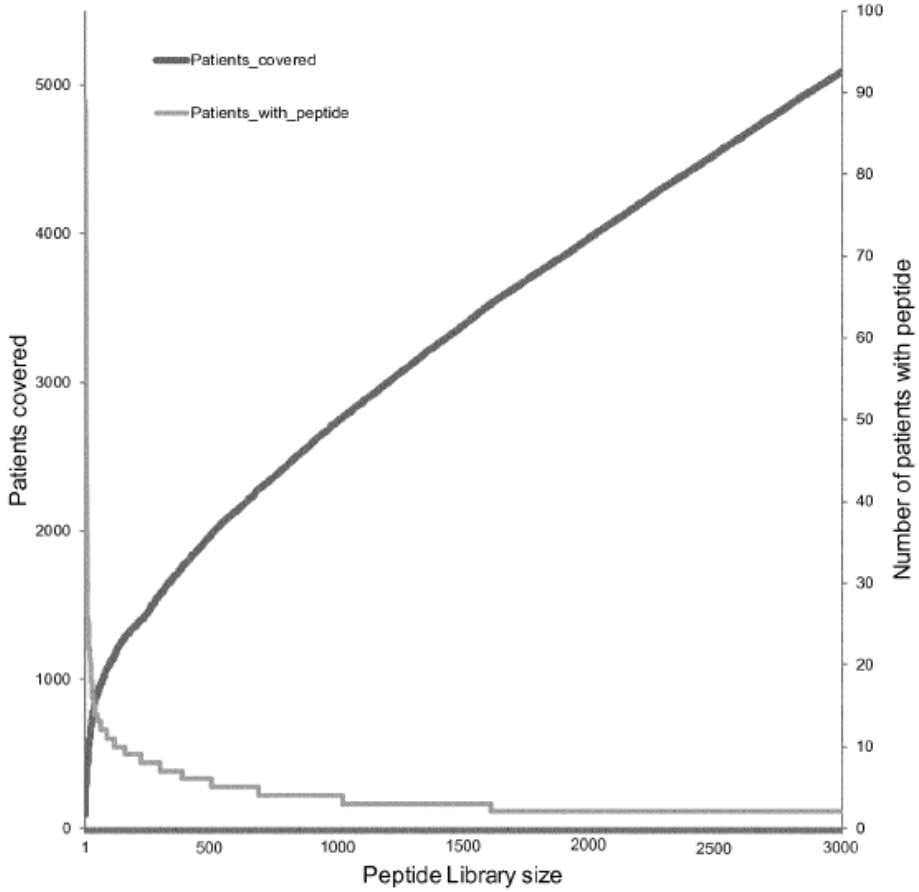
Bijlage bij publicatie A. Tsoutsanis, “Over patenten en integriteit. De affaire Plasterk in vier vragen”, *NJB* 2024, p. 1233-1235.

	Artikel	Octrooiaanvraag
8	Figure 2 illustrates that the precise location of a frame shift does not seem to matter much; the more or less straight slope of the series of mutations found in these 10,186 tumors indicates that it is not relevant for the biological effect (presumably reduction/loss of gene function) where the precise frame shift is, as long as translation stalls in the gene before the downstream remainder of the protein is expressed. As can also be seen in Fig. 2, all frame shift mutations alter the reading frame to one of the two alternative frames. Therefore, for potential immunogenicity the relevant information is the sequence of the alternative ORFs and more precisely, the encoded peptide sequence between 2 stop codons. We term these peptides ‘proto Neo Open Reading Frame peptides’ or pNOPs, and generated a full list of all thus defined out of frame protein encoding regions in the human genome, of 10 amino acids or longer. We refer to the total sum of all Neo-ORFs as the Neo-ORFeome. The Neo-ORFeome contains all the peptide potential that the human genome can generate after simple frame-shift induced mutations. The size of the Neo-ORFeome is 25.6 Mb. To investigate whether or not Nonsense Mediated Decay would wipe out frame shift mRNAs, we turned to a public repository containing read coverage for a large collection of cell lines (CCLE). We processed the data in a similar fashion as for the TCGA, identified the locations of frame shifts and subsequently found that, in line with the previous literature[23–25], at least a large proportion of expressed genes also contained the frame shift mutation within the expressed mRNAs (Supplementary Fig. S2). On the mRNA level, NOPs can be detected in RNAseq data. We next investigated how the number of patients relates to the number of NOPs. We sorted 10-mer peptides from NOPs by the number of new patients that contain the queried peptide. Assessed per tumor type, frame shift mutations in genes with very low to absent mRNA expression were removed to avoid overestimation. Of note NOP sequences are sometimes also encountered in the normal ORFeome, presumably as result of naturally occuring isoforms (e.g. Supplementary Fig. S3). Also these peptides were excluded. Taking into account the rules described above, and demanding that the addition of a peptide adds at least 1 new patient, we can create a library of possible ‘vaccines’ that is optimally geared towards covering the TCGA cohort, a cohort large enough that, also looking at the data presented here, it is representative of future patients (Table S3). Using this strategy 30% of all patients can be covered with a fixed collection of only 1,244 peptides of length 10 (Fig. 3). Since tumors will regularly have more than 1 frame shift mutation, one can use a ‘cocktail’ of different NOPs to optimally attack a tumor. Indeed, given a library of 1,244 peptides, 27% of the covered TCGA patients contain 2 or more ‘vaccine’ candidates. We ran the pNOPs through the NetMHC4.0 algorithm to predict	Figure 2 illustrates that the precise location of a frame shift does not seem to matter much; the more or less straight slope of the series of mutations found in these 10,186 tumors indicates that it is not relevant for the biological effect (presumably reduction/loss of gene function) where the precise frame shift is, as long as translation stalls in the gene before the downstream remainder of the protein is expressed. As can also be seen in figure 2, all frame shift mutations alter the reading frame to one of the two alternative frames. Therefore, for potential immunogenicity the relevant information is the sequence of the alternative ORFs and more precisely, the encoded peptide sequence between 2 stop codons. We term these peptides 'proto Neo Open Reading Frame peptides' or pNOPs, and generated a full list of all thus defined out of frame protein encoding regions in the human genome, of 10 amino acids or longer. We refer to the total sum of all Neo-ORFs as the Neo-ORFeome. The Neo-ORFeome contains all the peptide potential that the human genome can generate after simple frame-shift induced mutations. The size of the Neo-ORFeome is 46.6 Mb. To investigate whether or not Nonsense Mediated Decay would wipe out frame shift mRNAs, we turned to a public repository containing read coverage for a large collection of cell lines (CCLE). We processed the data in a similar fashion as for the TCGA, identified the locations of frame shifts and subsequently found that, in line with the previous literature[23-25], at least a large proportion of expressed genes also contained the frame shift mutation within the expressed mRNAs (figure 6). On the mRNA level, NOPs can be detected in RNAseq data. We next investigated how the number of patients relates to the number of NOPs. We sorted 10-mer peptides from NOPs by the number of new patients that contain the queried peptide. Assessed per tumor type, frame shift mutations in genes with very low to absent mRNA expression were removed to avoid overestimation. Of note NOP sequences are sometimes also encountered in the normal ORFeome, presumably as result of naturally occuring isoforms (e,g, figure 7). Also these peptides were excluded. We can create a library of possible ‘vaccines’ that is optimally geared towards covering the TCGA cohort, a cohort large enough that, also looking at the data presented here, it is representative of future patients (figure 10). Using this strategy 30% of all patients can be covered with a fixed collection of only 1,244 peptides of length 10 (figure 3). Since tumors will regularly have more than 1 frame shift mutation, one can use a ‘cocktail’ of different NOPs to optimally attack a tumor. Indeed, given a library of 1,244 peptides, 27% of the covered TCGA patients contain 2 or more ‘vaccine’ candidates.

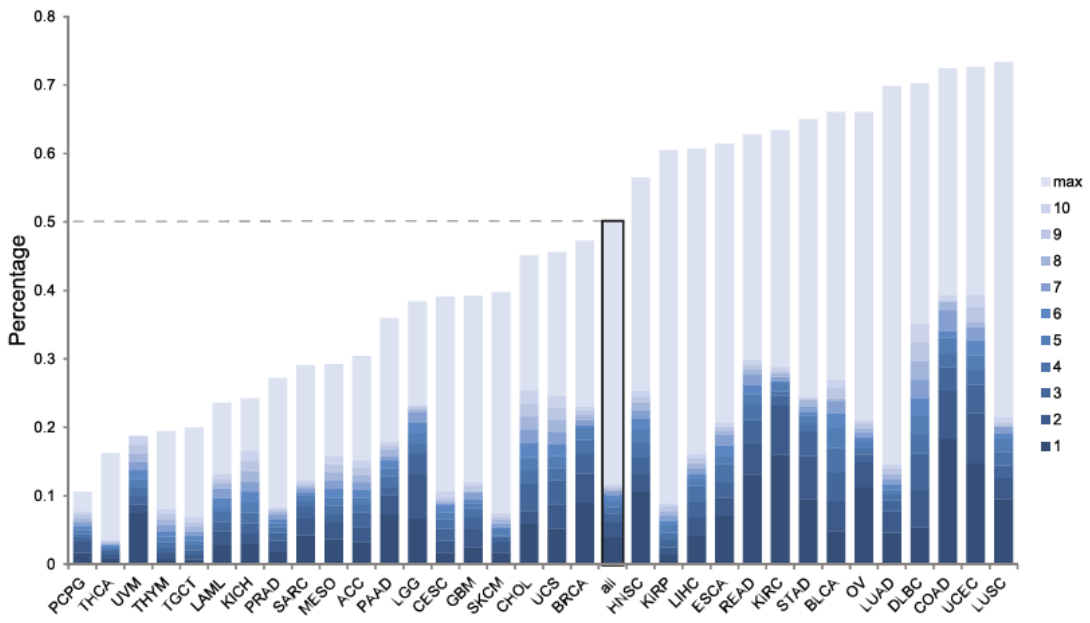
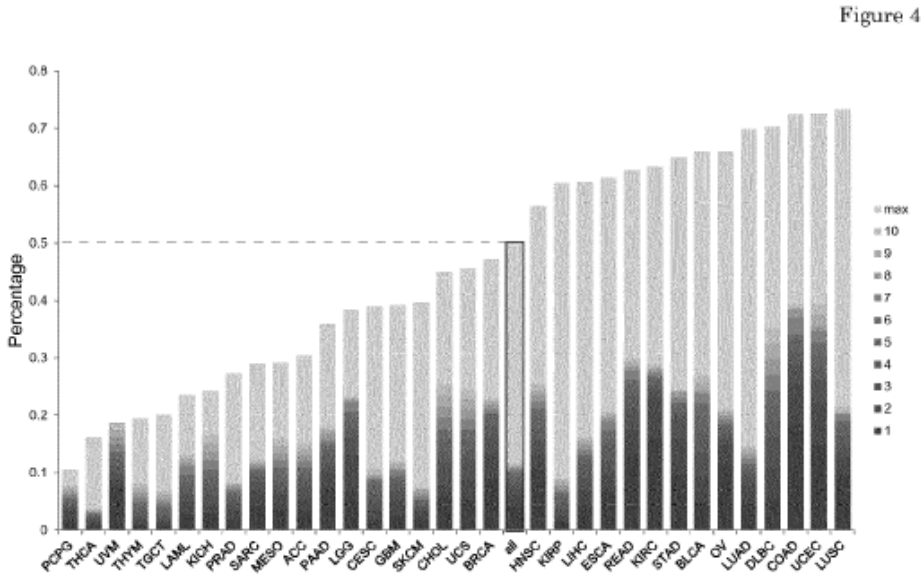
Bijlage bij publicatie A. Tsoutsanis, “Over patenten en integriteit. De affaire Plasterk in vier vragen”, *NJB* 2024, p. 1233-1235.

	Artikel	Octrooiaanvraag
9	(Fisher's Exact test $p < 2.2\ 10^{-16}$). In conclusion, using a limited pool with optimal patient inclusion of vaccines, a large proportion of patients is covered. Strikingly, using only 6 genes (TP53, ARID1A, KMT2D, GATA3, APC, PTEN), already 10% of the complete TCGA cohort is covered (Supplement Table S4). Separating this by the various tumor types, we find that for some cancers (like Pheochromocytoma and Paraganglioma (PCPG) or Thyroid carcinoma (THCA)) the hit rate is low, while for others up to 39% can be covered even with only 10 genes (Colon adenocarcinoma (COAD) using 60 peptides, Uterine Corpus Endometrial Carcinoma (UCEC) using 90 peptides), Fig. 4 and Table S4. At saturation (using all peptides encountered more than once) 50% of TCGA is covered and more than 70% can be achieved for specific cancer types (COAD, UCEC, Lung squamous cell carcinoma (LUSC) 72%, 73%, 73% respectively). As could be expected, these roughly follow the mutational load in the respective cancer types (Table S1). In addition some frame shifted genes are highly enriched in specific tumor types (e.g. VHL, GATA3. Supplementary Fig. S4). We conclude that at saturating peptide coverage, using only very limited set of genes, a large cohort of patients can be provided with off the shelf vaccines.	In conclusion, using a limited pool with optimal patient inclusion of vaccines, a large proportion of patients is covered. Strikingly, using only 6 genes (TP53, ARID1A, KMT2D, GATA3, APC, PTEN), already 10% of the complete TCGA cohort is covered. Separating this by the various tumor types, we find that for some cancers (like Pheochromocytoma and Paraganglioma (PCPG) or Thyroid carcinoma (THCA)) the hit rate is low, while for others up to 39% can be covered even with only 10 genes (Colon adenocarcinoma (COAD) using 60 peptides, Uterine Corpus Endometrial Carcinoma (UCEC) using 90 peptides), figure 4. At saturation (using all peptides encountered more than once) 50% of TCGA is covered and more than 70% can be achieved for specific cancer types (COAD, UCEC, Lung squamous cell carcinoma (LUSC) 72%, 73%, 73% respectively). As could be expected, these roughly follow the mutational load in the respective cancer types. In addition some frame shifted genes are highly enriched in specific tumor types (e.g. VHL, GATA3. figure 8). We conclude that at saturating peptide coverage, using only very limited set of genes, a large cohort of patients can be provided with off the shelf vaccines.
10	To validate the presence of NOPs, we used the targeted sequencing data on 10,129 patients from the MSK-IMPACT cohort[26]. For the 341–410 genes assessed in this cohort, we obtained strikingly similar results in terms of genes frequently affected by frame shifts and the NOPs that they create (Supplementary Fig. S5). Even within this limited set of genes, 86% of the library peptides (in genes targeted by MSK-IMPACT) were encountered in the patient set. Since some cancers, like glioblastoma or pancreatic cancer, show survival expectancies after diagnosis measured in months rather than years (e.g. see[27]), it is of crucial importance to move as much of the work load and time line to the moment before diagnosis. Since the time of whole exome sequencing after biopsy is currently technically days, and since the scan of a resulting sequence against a public database describing these NOPs takes seconds, and the shipment of a peptide of choice days, a vaccination can be done theoretically within days and practically within a few weeks after biopsy. This makes it attractive to generate a stored and quality controlled peptide vaccine library based on the data presented here, possibly with replicates stored on several locations in the world. The synthesis in advance will - by economics of scale - reduce costs, allow for proper regulatory oversight,	To validate the presence of NOPs, we used the targeted sequencing data on 10,129 patients from the MSK-IMPACT cohort 26. For the 341-410 genes assessed in this cohort, we obtained strikingly similar results in terms of genes frequently affected by frame shifts and the NOPs that they create (figure 9). Even within this limited set of genes, 86% of the library peptides (in genes targeted by MSK-IMPACT) were encountered in the patient set. Since some cancers, like glioblastoma or pancreatic cancer, show survival expectancies after diagnosis measured in months rather than years (e.g. see 27), it is of importance to move as much of the work load and time line to the moment before diagnosis. Since the time of whole exome sequencing after biopsy is currently technically days, and since the scan of a resulting sequence against a public database describing these NOPs takes seconds, and the shipment of a peptide of choice days, a vaccination can be done theoretically within days and practically within a few weeks after biopsy. This makes it attractive to generate a stored and quality controlled peptide vaccine library based on the data presented here, possibly with replicates stored on several locations in the world.
11	De onderzoeksmethoden beschreven op p. 5-8 in het artikel komen verregaand overeen met de methoden beschreven op p. 74-77 van de octrooiaanvraag	
12	Voetnoten en bronnen 1-36 beschreven op p. 7 in het artikel zijn identiek met de bronnen beschreven op p. 77-79 van de octrooiaanvraag.	

Bijlage bij publicatie A. Tsoutsanis, “Over patenten en integriteit. De affaire Plasterk in vier vragen”, *NJB* 2024, p. 1233-1235.

	Artikel	**Octrooiaanvraag**
13		Figure 3 

Bijlage bij publicatie A. Tsoutsanis, “Over patenten en integriteit. De affaire Plasterk in vier vragen”, *NJB* 2024, p. 1233-1235.

	Artikel	**Octrooiaanvraag**
14		

* * * * *