CeGaT GmbH | Paul-Ehrlich-Str. 23 | 72076 Tübingen | Germany

Dr. med. Richard Roe
CeGaT GmbH
Paul-Ehrlich-Straße 23
72076 Tübingen

Name	Doe, Jane (*MM.DD.YYYY)
Sex	Female
Patient-ID	12345
Sample receipt	02.01.2023 (EDTA blood) 02.01.2023 (Tumor FFPE)
Report date	13.01.2022
Report-ID	R154321

Neoepitope prediction and peptide selection – Doe, Jane (*MM.DD.YYYY)

Indication	**Adenocarcinoma of the pancreas (ID 10/2022)**
Material	**Tumor tissue: Sample of the liver** Sample collection 10/2022 DNA and RNA isolation from tumor in FFPE (FFPE ID 12345-22 A) after macrodissection with estimated tumor content of 60 % Diagnostically estimated tumor content 60% **Normal tissue: EDTA blood**
Order	**1.** Somatic molecular genetic analysis of a tumor tissue sample: Tumor exome analysis TUM02, evaluation of somatic variants of potential clinical relevance in 749 tumor-related genes (TUM01) **2. Tumor transcriptome analysis, neoepitope prediction and peptide selection**

Dear Dr. Roe,

Thank you for your request for molecular genetic analysis, neoepitope prediction and peptide selection.

RESULTS

24 peptides were selected based on neoepitope prediction results of your patient's tumor sequencing data. The selected peptides are predicted to activate not only cytotoxic T cells but also T helper cells. Therefore, in addition to short peptides (8-12 amino acids) potentially binding to HLA class I molecules also long peptides (~17 amino acids) potentially binding to HLA class II molecules were included.

Individual peptide selection based on neoepitope prediction:

No.	Peptide	Gene and Coding info	NAF DNA	NAF RNA	FPKM	HLA allele
1	RDFTKGMWYGS	TP53:NM_000546:c.180C>T:p.P90S	0.48	0.67	51.42	A*02:01
2	HSQGFGKTLTL	HEATR1:NM_018072.5:c.3725T>G:p.R1235T	0.31	0.35	32.78	C*07:02
3	TRNSYEIGIRR	KMT2C:NM_170606.3:c.4030dupG: p.I1344N*fs*13	0.26	0.37	25.42	C*07:02
4	EVESRYILYN	PTDSS2:NM_030783:c.874C>A:p.A292E	0.17	0.11	19.39	A*02:01
5	MRCQKNIHENSR	SPAG16:NM_024532:c.1200G>T:p.G400C	0.52	0.22	18.69	C*07:02
6	YNDIRGEVIH	XPC:NM_004628:c.1517T>C:p.R506Y	0.17	0.19	15.76	A*02:01
7	VWPKHERYMVLL	ZNF770:NM_014106:c.452A>G:p.I151L	0.13	0.25	11.95	A*04:01
8	DVTTQNGYWG	KMT2C:NM_170606.3:c.4953G>T:p.A1545T	0.29	0.26	10.20	B*07:02
9	KHRDKHFSH	TNIP1:NM_001252385.1:c.9461A>C:p.M452H	0.33	0.53	5.50	C*07:04
10	RSWRLKHPQ	DERL3: NM_198440.4:c.245G>C:p.V77Q	0.44	0.62	1.27	B*15:18
11	IMFSRPECHDVH	PRKD1:NM_002742:c.1601A>C:p.V534S	0.36	0.81	3.63	A*02:01, B*07:02, C*07:02
12	TWTIPVRWDECMLHKTP	KRAS:NM_004985:c.35G>A:p.G12D	0.27	0.43	132.39	HLA class II
13	FMKYTHNSNWLIQVRSG	TP53:NM_000546: c.180C>T:p.P90S	0.48	0.67	25.78	HLA class II
14	ETFKKRHFRQHISSING	SLIT2:NM_004787:c.300G>T:p.Y100T	0.45	0.69	21.68	HLA class II
15	WLITMHPMWDHWSCSRG	KMT2C:NM_170606.3:c.1859G>T:p.K543P	0.21	0.43	20.22	HLA class II
16	TWVYLLFDEVWIWQGSF	SPAG16:NM_024532:c.1200C>T:p.W400V	0.52	0.22	18.69	HLA class II
17	VWMKFDPIQDHIHTDDR	XPC:NM_004628:c.1517C>T:p.R506D	0.17	0.19	15.76	HLA class II
18	GNQKIHKGHWDHWNGWV	ZNF770:NM_014106:c.452G>A:p.L151W	0.13	0.25	11.95	HLA class II
19	SHFPIMNQNDEKYSGP	KMT2C:NM_170606.3:c.4953G>T:p.H1545S	0.29	0.26	10.20	HLA class II
20	RIGPIQYEPKHIFWEHR	TNIP1:NM_001252385.1:c.9461A>C:p.C452E	0.33	0.53	5.50	HLA class II
21	NGYKQNGSSGFPSFVMY	PPP2R5B:NM_006244.4:c.99A>C:p.N33K	0.12	0.25	5.43	HLA class II
22	KVKKMGMYTKEKTYGSH	NFE2L2:NM_020909.4:c.1119_1120delinsTT: p.S373M	0.56	0.48	2.83	HLA class II
23	KWNIERGHRHQPTHSGT	MKRN2:NM_014160.5:c.614A>G:p.E205G	0.10	0.09	3.76	HLA class II
24	TRLNIGNFLITPWPKSQ	ENPP2:NM_006209.5:c.1562G>C:p.W521S	0.21	0.11	2.16	HLA class II

NAF: Novel allele frequency, the frequency with which the mutated allele occurs in the sequencing (1 is 100%). The observed frequencies are influenced by the tumor content and do not correlate directly to the variant frequency in the tumor. **FPKM:** Fragments Per Kilobase Million, expression level metric.

Based on the DNA sequencing analysis of the blood sample the HLA genotype was determined to be:

HLA-A*02:01, HLA-A*04:01, HLA-B*07:02, HLA-B*15:18, HLA-C*07:02, HLA-C*07:04

Please do not hesitate to contact us if you have any questions.

Medical report written by: Dr. rer. nat. Forename Surname

Proofread by: Dr. rer. nat. Forename Surname, Dr. rer. nat. Forename Surname

With kind regards,

Dr. med. Dr. rer. nat. Saskia Biskup
Dr. med. Friedmar Kreuz, M.A.

Consultant for Human Genetics

ADDITIONAL INFORMATION

Methods

DNA and RNA isolation: The isolation of tumor and normal DNA as well as tumor RNA was performed at CeGaT GmbH. Macrodissection prior to isolation was performed, if necessary. The tumor material was assessed by a pathology specialist.

The pathological services (confirmation of the histological diagnosis and determination of the tumor content) were carried out on our behalf by a specialist in pathology. Pathology services are not within the scope of the ISO 15189 accreditation.

NGS-laboratory DNA: Protein-coding regions, as well as flanking intronic regions and additional disease-relevant non-coding regions, were enriched using in-solution hybridization technology, and were sequenced using the Illumina NovaSeq6000 system.

NGS-laboratory RNA: Library preparation was performed using the TruSeq Total RNA (RiboZero rRNA removal Kit) or- SMARTer Stranded Total RNA Library Kit and subsequently analyzed using high-throughput sequencing on the HiSeq/NovaSeq system (Illumina). Mapping of sequencing reads to the hg19 reference genome was performed with STAR (Version 2.5.2b). Gene expression analysis (counting of aligned reads per gene, calculation of normalized read counts and calculation of FPKM values) was done with DESeq2 (Love et al., 2014, PMID: 25516281) in R (R Core Team 2015)). FFPM (Fusion Fragments Per Million) is used as a normalized measure of the fusion-supporting rna-seq fragments in the dataset. It is calculated as the number of deduplicated reads that support a specific fusion divided by the number of deduplicated reads that could successfully be aligned to the reference genome.

Computational analysis DNA: Illumina bcl2fastq2 was used to demultiplex sequencing reads. Adapter removal was performed with Skewer. The trimmed reads were mapped to the human reference genome (hg19) using the Burrows Wheeler Aligner. Reads mapping to more than one location with identical mapping score were discarded. Read duplicates that likely result from PCR amplification were removed. The remaining high-quality sequences were used to determine sequence variants (single nucleotide changes and small insertions/deletions). The variants were annotated based on several internal as well as external databases.

Computational analysis RNA: Sequencing data was demultiplexed using bcl2fastq2. Adapter sequences were removed using Skewer and the resulting reads were mapped to the human reference genome hg19 using STAR aligner.

Genetic data evaluation DNA: Only variants (SNVs/small indels) with a novel allele frequency (NAF) of ≥ 5% in the tumor sample within the coding regions and their adjacent intronic regions (-/+ 8 base pairs) were evaluated. A list of all the variants with an allele frequency of 5% considered in the genetic data evaluation can be requested at any time. The clinical interpretation of variants is based on different external and internal databases and on information from scientific literature. The sensitivity of the test is dependent on the tumor content of the analyzed material, the sample quality, and the sequencing depth. In this case, 97.55% of the targeted regions were covered by a minimum of 70 high-quality sequencing reads per base. The diagnostic tumor content (expert estimate) was 60%. A theoretical sensitivity

of >99% can be obtained for variants with a NAF ≥30% when a coverage of 35 reads per base is achieved. Variants are named according to the HGVS recommendations without any information regarding the cis or trans configuration.

Genetic data evaluation RNA: The sensitivity of the test is dependent on the tumor content of the analyzed material, the sample quality, and the amount of transcripts sequenced.

Prediction of neoepitopes: Somatic single nucleotide variants and small InDels (in frame or frameshift insertions and/or deletions) only present in the tumor but absent in the normal tissue of the patient were identified by sequencing analysis. Typing of HLA class I is performed on sequencing results from the patient's normal tissue with an extension of OptiType (Szolek et al., 2014, PMID: 25143287). Identified somatic variants and in-phase germline variants are translated into peptide sequences and sent to the Center for Bioinformatics Tübingen, Dept. of Computer Science, University Tübingen for epitope prediction. MHC class I epitopes are predicted using SYFPEITHI, netMHC-4.0 and netMHCpan-3.0 (Rammensee et al., 1999, PMID: 10602881; Andreatta et al., 2016, PMID: 26515819; Nielsen et al., 2016, PMID:27029192). Peptides containing somatic variants that are classified as binder by at least one prediction method are further evaluated. The respective thresholds for classification as binder are defined as <500 nM for netMHC and netMHCpan as well as >50% of maximal score for SYFPEITHI. Peptides resembling a wildtype sequence in the human proteome (based on UniProtKB/Swiss-Prot, human, 9/7/14) are excluded.

Selection of peptides: Peptides derived from genes most probably not expressed in the patient's tumor entity were excluded. For this purpose, expression data for the respective variant were analyzed using RNA sequencing data of the tumor sample. Putative HLA class I epitopes with a high HLA class I binding prediction score derived from variants with high allele frequencies were selected. Peptides predicted to bind to different HLA class I molecules of the patient were prioritized. Peptides which are predicted to bind to several HLA types were further prioritized. Putative HLA class II epitopes with a length of +/-17 amino acids were designed to contain variants with high allele frequencies. Peptides spanning variants in possible tumor drivers were prioritized. Peptides with a high percentage of hydrophobic amino acids, peptides with a high probability for gelation or dimerization were excluded to avoid solubility problems in an aqueous solution and problems during synthesis. However, since it is impossible to reliably predict these characteristics, it cannot be guaranteed that all selected peptides can be synthesized and solubilized.
The bioinformatically identified somatic variants corresponding to all selected peptides were manually reviewed in the sequencing data and filtered for false positives.
A file containing all peptides predicted to bind to your patient's HLA class I molecules can be sent to you by E-mail upon request.

The procedure described above was developed and validated in-house (Laboratory developed test; LDT). A minimal tumor content of 20% was taken as a basis.

CeGaT GmbH | Paul-Ehrlich-Str. 23 | 72076 Tübingen | Germany
Phone: + 49 7071 565 44 55 | Fax: + 49 7071 565 44 56 | info@cegat.com | www.cegat.com
Court District Stuttgart - HRB 729958 | VAT No: DE265504070
Managing Directors: Dr. Dr. Saskia Biskup, Dr. Dirk Biskup, Dr. Detlef Schumann
Volksbank in der Region eG | IBAN: DE73 6039 1310 0543 4480 02 | SWIFT / BIC: GENODES1VBH