

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

Frau
Dr. med. Jane Doe
CeGaT GmbH
Paul-Ehrlich-Straße 23
72076 Tübingen

Name	Roe, Richard (*TT.MM.JJJJ)
Geschlecht	männlich
Patienten-ID	12345
Probeneingang	02.01.2023 (EDTA-Blut) 02.01.2023 (Tumor-FFPE)
Befunddatum	13.01.2023
Befund-ID	R154321

Neopitop-Vorhersage und Auswahl von Peptiden – Roe, Richard (*TT.MM.JJJJ)

Indikation	Pankreaskarzinom (ED 10/2022)
Material	Tumorgewebe: Leberstanzbiopsat Probenentnahme 10/2022 DNA- und RNA-Isolierung aus FFPE-Material (FFPE-ID 12345-22 A) nach Makrodissektion mit geschätztem Tumorgehalt von 60 % (HE Färbung) Diagnostisch bestimmter Tumorgehalt 60 % Normalgewebe: EDTA-Blut
Auftrag	1. Somatische molekulargenetische Analyse aus Tumorgewebe: Tumor-Exom-Analyse TUM02, Bewertung der therapierelevanten Varianten in 749 Panelgenen (TUM01) 2. Tumor Transkriptom-Analyse, Neopitop-Vorhersage und Auswahl von Peptiden

Sehr geehrte Frau Dr. Doe,

herzlichen Dank für die Zusendung der Proben Ihres Patienten zur molekulargenetischen Tumordiagnostik, Neopitop-Vorhersage und Peptidauswahl.

ERGEBNIS

Basierend auf den Neopitop-Vorhersage-Ergebnissen zum Tumor Ihres Patienten wurden 24 Tumorspezifische Peptide ausgewählt.

Die ausgewählten Peptide sollen sowohl zytotoxische T-Zellen als auch T-Helfer Zellen aktivieren. Daher wurden neben kurzen, HLA-Klasse-I-restringierten Peptiden (8-12 Aminosäuren), auch lange Peptide (~ 17 Aminosäuren), die möglicherweise an HLA-Klasse II-Moleküle binden, in die Auswahl mit eingeschlossen.

Auswahl Tumor-spezifischer Peptide nach Neoepitop-Vorhersage:

Nr.	Peptid	Gen, Transkript und Variante	NAF DNA	NAF RNA	FPKM	HLA-Allel
1	RDFTKGMWYGS	TP53:NM_000546:c.180C>T;p.P90S	0.48	0.67	51.42	A*02:01
2	HSQGFGLTL	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.I1344Nfs*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	VWPKHERYMLL	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	TWTIPVRWDECMHLKTP	KRAS:NM_004985:c.35G>A;p.G12D	0.27	0.43	132.39	HLA Klasse II
13	FMKYTHNSNWLIVRSG	TP53:NM_000546: c.180C>T;p.P90S	0.48	0.67	25.78	HLA Klasse II
14	ETFKKRHRFHSSING	SLIT2:NM_004787:c.300G>T;p.Y100T	0.45	0.69	21.68	HLA Klasse II
15	WLITMHPMWDHWSCSRG	KMT2C:NM_170606.3:c.1859G>T;p.K543P	0.21	0.43	20.22	HLA Klasse II
16	TWVYLLFDEVWIWQGSF	SPAG16:NM_024532:c.1200C>T;p.W400V	0.52	0.22	18.69	HLA Klasse II
17	VWMKFDPIQDHIHTDDR	XPC:NM_004628:c.1517C>T;p.R506D	0.17	0.19	15.76	HLA Klasse II
18	GNQKIHKGHWDHWNGWV	ZNF770:NM_014106:c.452G>A;p.L151W	0.13	0.25	11.95	HLA Klasse II
19	SHFPIMNQNDKYS GP	KMT2C:NM_170606.3:c.4953G>T;p.H1545S	0.29	0.26	10.20	HLA Klasse II
20	RIGPIQYEPKHIFWEHR	TNIP1:NM_001252385.1:c.9461A>C;p.C452E	0.33	0.53	5.50	HLA Klasse II
21	NGYKQNGSSGFPSFVMY	PPP2R5B:NM_006244.4:c.99A>C;p.N33K	0.12	0.25	5.43	HLA Klasse II
22	KVKKMGMYTKEKTYGSH	NFE2L2:NM_020909.4:c.1119_1120delinsTT:p.S373M	0.56	0.48	2.83	HLA Klasse II
23	KWNIERGHRHQPTHSGT	MKRN2:NM_014160.5:c.614A>G;p.E205G	0.10	0.09	3.76	HLA Klasse II
24	TRLNIGNFLITPWPKSQ	ENPP2:NM_006209.5:c.1562G>C;p.W521S	0.21	0.11	2.16	HLA Klasse II

NAF: Novel allele frequency, entspricht der Frequenz, mit der das mutierte Allel in der Sequenzierung detektierbar war (1 entspricht 100 %). Die beobachteten Frequenzen werden durch den Tumorgehalt und durch Kopienzahlveränderungen beeinflusst und entsprechen nicht direkt der Häufigkeit der Variante im Tumor. **FPKM:** Fragments Per Kilobase Million, Expressionswert.

Aus den Sequenzierdaten der Blutprobe wurde folgender HLA-Typ bestimmt:

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

Für Rückfragen stehen wir Ihnen gerne zur Verfügung.

Befund erstellt von: Dr. rer. nat. Vorname Nachname

Geprüft durch: Dr. rer. nat. Vorname Nachname, Dr. rer. nat. Vorname Nachname

Mit freundlichen Grüßen

Dr. med. Dr. rer. nat. Saskia Biskup

Dr. med. Friedmar Kreuz, M.A.

Fachärztin/Facharzt für Humangenetik

ERGÄNZENDE INFORMATIONEN

Methoden DNA- und RNA-Isolierung: Die Isolierung der Tumor- und Normal-DNA sowie der Tumor-RNA wurde von der CeGaT GmbH durchgeführt. Falls nötig wurde eine Makrodissektion durchgeführt. Die Begutachtung des Tumormaterials erfolgte durch einen Facharzt für Pathologie.

Die pathologischen Leistungen (Bestätigung der histologischen Diagnose, Bestimmung des Tumorgehaltes) erfolgten in unserem Auftrag durch einen Facharzt für Pathologie. Die Leistungen der Pathologie gehören nicht zum Akkreditierungsumfang der ISO 15189.

Sequenzierung DNA: Es wurde DNA aus Normalgewebe und Tumorgewebe sequenziert. Die kodierenden Bereiche sowie die angrenzenden Intronbereiche und weitere nicht kodierende, krankheitsrelevante Regionen wurden mittels in-solution-hybridization Technologie angereichert und anschließend mittels Hochdurchsatz-Sequenzierung auf dem Illumina NovaSeq6000 System analysiert.

Sequenzierung RNA: Die Library-Herstellung erfolgte mit dem TruSeq Total RNA (RiboZero rRNA removal Kit) oder SMARTer Stranded Total RNA Library Kit und wurde anschließend mittels Hochdurchsatz-Sequenzierung auf dem Illumina HiSeq/NovaSeq System analysiert. Mapping der Sequenzierreads zum Referenzgenom hg19 wurde mit STAR (Version 2.5.2b) durchgeführt. Die Genexpressionsanalyse (Zahl der alignierten Reads pro Gen, Berechnung der normalisierten Read-Counts und der FPKM-Werte) wurde mit DESeq2 (Love et al., 2014, PMID: 25516281) in R (R Core Team 2015) durchgeführt. FFPM (Fusion Fragments Per Million) ist ein normalisiertes Maß für die Reads im Datensatz, die die Fusion unterstützen. Der Wert wird berechnet als die Anzahl der Reads, die die Fusion zeigen (nach Duplikatentfernung) geteilt durch die Anzahl der Reads (nach Duplikatentfernung), die auf das Referenzgenom gemappt werden konnten.

Bioinformatik DNA: Die Sequenzierdaten wurden mit Illumina bcl2fastq2 aufbereitet. Mitsequenzierte Adaptersequenzen wurden mit Skewer entfernt und die so erhaltenen Sequenzen durch den Burrows Wheeler Aligner gegen das humane Referenzgenom (hg19) aligniert. Sequenzen, die nicht eindeutig einer genomischen Position zugeordnet werden konnten, wurden entfernt, ebenso Sequenzduplikate, die wahrscheinlich auf die Amplifikation zurückzuführen sind. Anhand der verbleibenden Sequenzen hoher Qualität wurden Sequenzvarianten (Einzelnukleotidaustausche und kurze Insertionen/Deletionen) bestimmt. Diese wurden mit verschiedenen internen und externen Datenbanken annotiert.

Bioinformatik RNA: Die Sequenzierdaten wurden mit Illumina bcl2fastq2 aufbereitet. Mitsequenzierte Adaptersequenzen wurden mit Skewer entfernt und die so erhaltenen Sequenzen durch den STAR aligner gegen das humane Referenzgenom (hg19) aligniert.

Genetische Datenauswertung DNA: Bewertet werden alle somatischen Varianten (SNVs/Small Indels) innerhalb der kodierenden Regionen, sowie in deren flankierenden intronischen Regionen (-/+ 8 Basenpaare) mit einer Novel Allelfrequenz (NAF) von > 5 % in der Tumorprobe. Eine klinische Interpretation erfolgt anhand unterschiedlicher externer und interner Datenbanken sowie einer Literaturrecherche. Jederzeit kann die Liste aller Varianten angefordert werden. Die Sensitivität des Tests ist abhängig vom Tumorgehalt des Untersuchungsmaterials, der

Probenqualität sowie der Sequenziertiefe. Im vorliegenden Fall wurde mittels Hochdurchsatz-Sequenzierung eine Sequenziertiefe von min. 70X für 97,55% der kodierenden Bereiche erreicht. Der diagnostische Tumorgehalt (Expertenschätzung) der Probe beträgt 60 %. Bei einer Sequenziertiefe von 35 Reads pro Base wird eine theoretische Sensitivität von > 99 % für die Detektion von Varianten mit einer NAF \geq 30 % erreicht. Die Nomenklatur gefundener Varianten erfolgt nach den Richtlinien der HGVS.

Genetische Datenauswertung RNA: Die Sensitivität des Tests ist abhängig vom Tumorgehalt des Untersuchungsmaterials, der Probenqualität sowie der Datenmenge an sequenzierten Transkripten.

Vorhersage von Neoepitopen: Mittels Exom-Analyse der Tumor- und Normalprobe wurden somatische missense Varianten (single nucleotide variants) sowie kleine InDels (inframe und frameshift) identifiziert. Die HLA-Typisierung erfolgte anhand der Sequenzier-Daten des Normalgewebes mittels einer Weiterentwicklung der Software OptiType (Szolek et al., 2014, PMID: 25143287). Die identifizierten somatischen Varianten sowie Keimbahnvarianten innerhalb desselben Exons auf demselben Transkript (in phase) wurden in Peptidsequenzen übersetzt und zur Epitopvorhersage an das Zentrum für Quantitative Biologie (Universität Tübingen) gesendet. Die Vorhersage für HLA-Klasse-I-Epitope erfolgte mit Hilfe von SYFPEITHI, netMHC-4.0 und netMHCpan-3.0 (Rammensee et al., 1999, PMID: 10602881; Andreatta et al., 2016, PMID: 26515819; Nielsen et al., 2016, PMID: 27029192). Für Peptide, die eine somatische Variante enthalten und von mindestens einem der Vorhersageprogramme als bindend eingestuft wurden, wurden weiter für die Auswahl herangezogen. Die Kriterien, als Binder eingestuft zu werden sind eine ermittelte Bindeaffinität von <500 nM für netMHC und netMHCpan, sowie ein Bindungs-Score von über 50% des möglichen Maximal-Scores für SYFPEITHI. Ausgeschlossen wurden Peptide mit somatischen Varianten, die zu 100% einer natürlich vorkommenden Wildtypsequenz im menschlichen Proteom entsprechen (basierend auf UniProtKB/Swiss-Prot, human, 9/7/14).

PeptidAuswahl: Ausgeschlossen wurden Peptide von Varianten, die höchstwahrscheinlich im Tumorgewebe des Patienten nicht exprimiert werden. Hierfür wurden die generierten Daten der RNA-Sequenzierung herangezogen. Ausgewählt wurden mögliche HLA-Klasse-I-Epitope mit hohem vorhergesagten Bindungs-Score, hoher Allelfrequenz und hohen Expressionswerten. Wenn möglich, wurden vorhergesagte Neoepitope für alle HLA-Klasse-I-Moleküle des Patienten ausgewählt. Des Weiteren wurden Peptide bevorzugt, die möglicherweise an mehrere HLA-Klasse-I-Moleküle des Patienten binden. Zusätzlich wurden mögliche HLA-Klasse-II-Epitope mit einer Länge von etwa 17 Aminosäuren errechnet, die somatische single nucleotide missense Varianten, oder kleine Insertionen oder Deletionen beinhalten. Daraus ausgewählt wurden mögliche HLA-Klasse-II-Epitope mit hoher Allelfrequenz und hohen Expressionswerten. Peptide von Varianten in Genen, die in der Tumorbioogie eine Rolle spielen (tumor driver) wurden priorisiert. Peptide mit einem hohen Anteil hydrophober Aminosäuren, sowie Peptide mit einer hohen Wahrscheinlichkeit zur Gelbildung oder Dimerisierung wurden ausgeschlossen, um Löslichkeitsprobleme in wässriger Lösung und Probleme während der Synthese zu vermeiden. Leider sind diese Parameter nicht mit Sicherheit vorhersehbar, weswegen weder für spätere Löslichkeit noch für erfolgreiche Synthese garantiert werden kann. Die bioinformatisch identifizierten Varianten, die unserer PeptidAuswahl zu Grunde liegen, wurden in den Sequenzierdaten nochmals manuell überprüft.

Auf Wunsch können wir Ihnen eine Datei mit allen Peptiden, für die eine Bindung an die HLA-Klasse-I-Moleküle Ihres Patienten vorhergesagt wurde, per E-Mail zuschicken.

Bei dem oben beschriebenen Verfahren handelt es sich um einen inhouse entwickelten und validierten Test (Laboratory developed test; LDT). Dabei wurde ein Mindesttumorgehalt von 20 % zugrunde gelegt.