Study-20-611: NeoE Identification and Prioritization for Patient 0411

Version 1.5

Study #	Study-20-611
Study ID	NeoE Identification and Prioritization for Patient 0411
Responsible Scientist	Sam Pan
Date	14th January, 2021
Study Objective	To conduct Neoepitope (NeoE) Identification and Prioritization for Patient PACT482C based on WES of the patient's PBMC, and WES & RNA-Seq of the patient's tumor biopsy according to PD-002_004, which is governed by RTR-007_003. To endorse patient tumor-specific neoepitope (neoE)-HLA complexes and to deliver oligo sequences to Protein Science for reagent production. To profile the tumor microenvironment (TME).
Initial Study Reviewed By	Data Missing / 14th January, 2021
Final Study Reviewed By	Data Missing / 14th January, 2021
Readouts	• Summary of patient-specific somatic mutations identified • HLA typing • Mutation truncality and cellular prevalence category (CPC) of somatic nonsynonymous mutations • NeoE-annotations • NeoE exactMatch summary • Oligos for neoE-HLA production
For Information Only (FIO)	• TIL analysis for TCR in tumor microenvironment (TME) • Microsatellite Stability (MSS) analysis • Tumor biomarker expression by RNASeq

1.0 MATERIAL AND METHODS

1.1 Patient Sample Information

Table 1. Patient sample information.

Description	Information
Cancer type at initial diagnosis (Icon/Rave)	colorectal adenocarcinoma
TGCA acronym for primary cancer type (manifest)	COAD+READ
Biopsy tissue tumor type (primary or metastatic - HGX Inventory Tracker)	METASTATIC
Tumor subtype (i.e. tissue of primary cancer - HGX Inventory Tracker)	COLON
Sample Origin(i.e. tissue the biopsy was collected from - HGX Inventory Tracker)	LIVER
Biopsy tissue localization(HGX Result Tracker)	Liver
Patient ID number	0411
PACT ID for patient	PACT482C
PACT Specimen ID for tumor biopsy	PACT482C_T_PP001611
PACT Specimen ID for PBMC	PACT482C_N_PP001607_0801

Patient sample information is presented in Table 1. For this Melanoma patient, PBMC and a biopsy from a Premalignant tumor legs were sequenced.

1.2 Methods

Patient samples were registered, then NeoE Identification and Prioritization was performed according to the Personalized Product Development Protocol-PD-002 Version 004 (PD-002004), which is governed by the Research Technical Report-RTR-007 Version 003 (RTR-007003). See **Table 2** for details related to the performed experiment.

Table 2. Experiment Benchling registration.

Date	Experiment ID	Experiment Title	Benchling Link	Executed by (initials)
28th November, 2020	EXP20002238	NeoE identification and prioritization for patient 0411 (PACT482C_T_PP001611, 1st sample)	https://pact.benchling.com/pact/f/lib_Fidzq3TM-registry/bfi_Q0SGYxJs-study-20-611/edit	DB(Benchling registration) SP (Pipeline execution)

1.3 Software

Software programs used during execution of PD-002_004 and links to source code are listed in Appendix B.

1.4 Data

1.4.1 Input

Location(s) of raw sequence data, generated by Personalis:

-
-
-

1.4.2 Output

Locations of data generated by performing the NeoE Identification and Prioritization experiment.

Data location on the cloud

-

Data location on premise (support data is required for downstream processes; supplemental and additional data are for information only):

- Support data: P:\PACT-0101\0411_PACT482C\2_BINF\4_Required_Data
- Supplemental data: P:\PACT-0101\0411_PACT482C\2_BINF\5_Supplemental_Data\
- Additional data: P:\ PACT-0101\0411_PACT482C \ \2_BINF\6_Additional_Data\PACT482C_T_PP001611_12860106F

2.0 RESULTS

2.1 Overview

NeoE Identification and Prioritization for Patient 0411 (PACT482C) was performed according to PD-002_004. All standard results are presented in this results section and, if performed, the resutls of any additional analyses are presented in the Amendments section.

2.2 Mutation Identification

The first step in PD-002_004 is to detect somatic mutations present in the tumor WES data. Mutation locations were annotated and classified using Oncotator. **Figure 1** shows the variant counts by category. A list of possible classifications is here. (Note that Funcotator (linked) and Oncotator (used) classify variants into the same categories). The absence of a category in **Figure 1** means that no variants from that category were detected.

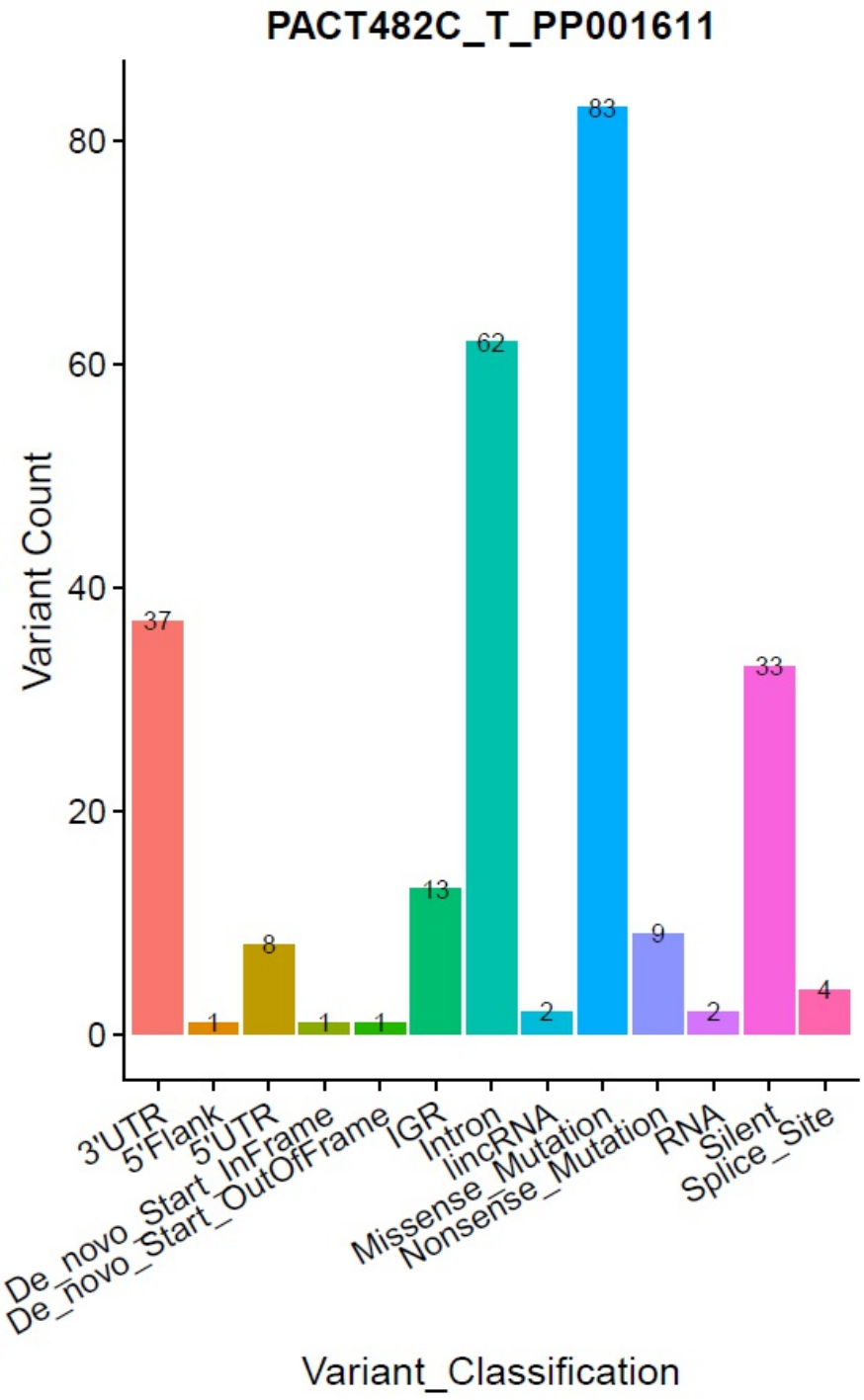

Figure 1. Variants identified from PACT482C within different variant classifications (source data: here). Mutations were classified using Oncotator. Variant count is the number of variants detected within the indicated categories. IGR: intergenic region. Missense_Mutation: somatic nonsynonymous coding mutation.

Within this tumor biopsy, **83** missense mutations (somatic nonsynonymous coding mutations; NSM) were found (Figure 1 & Figure 4). Coding region tumor mutation burden (TMB) is calculated by dividing the number of missense mutations detected by the exome sequencing genomic footprint for the coding regions, which was about 35 Mb. Therefore, the coding region TMB was estimated to be **2. 4** mutations/Mb.

2. 3 Fusion Detection

The algorithm developed to detect fusion proteins in tumor samples is based on detection of an RNA sequene that is a list of known fusion sequences, which were compiled from TCGA and Cosmic databases. Many known circular RNA sequences, which are not predicted to produce protein, were removed from the list during development. However, the list is likely to contain sequences that are not predicted to produce proteins.

No fusion proteins were predicted after analysis of this patient's tumor biopsy RNA-Seq data. This is tumor fusion Detected Comment

2. 4 Tumor Content Analysis

Tumor content (TC; cellularity in Figure 2) was determined by a pathologist from Histogenex (HGX) and by NGS (PD-002_004). TC is the percent of tumor cells in the specified area of a biopsy. Histologists at HGX estimated that the TC of the tumor area was 70% (Row 9, Table 3). The Sequenza package was used to estimate the TC of the macro-dissected *NGS-sequenced* tissue area from a tumor biopsy slide. Sequenza TC estimation is based on WES from both tumor biopsy and matched normal PBMC samples. For this tumor biopsy sample, NGS-estimated TC was *38%* (Row 17 of *Table 3*, also shown in *Figure 2*).

Table 3: Tumor biopsy information. Tumor content estimation by HGX pathology and NGS.

Row #	Description	Information
1	HGX Identifier	12860106F
2	Study ID	PACT-0101
3	Sample ID	PP001611
4	Patient ID	0411
5	Tissue localization	Liver
6	Biopsy Type	Small biopsy (<5mm)
7	Estimated number of cancer cells	1000-10000
8	Primary Tumor	N
9	TC by HGX (%)	70
10	Necrosis in tumor area (%)	1
11	Infiltrate in tumor area (%)	1
12	Stroma TILs (%)	1
13	Overall Tumor Content	24.5
14	Tissue size (mm2)	14.9
15	Tumor Size (mm2)	4.0
16	Tumor Size/ Tissue size	0.27
17	TC by NGS (%)	38

TC=Tumor Content, NGS= Next Generation Sequencing, NOE=Not Otherwise Evaluable

2. 5 Mutation Truncality Analysis

Mutation truncality analysis was performed according to the PACT-developed Minimal-subclone method This step of PD-002_004 assigns each mutation to a cellular prevalence category (CPC). In Figure 2, the cellular prevalence plot shows somatic small nucleotide variants (SNV) (second row of short vertical sticks) and, by the color of their median cellular prevalence dots, the categories they were assigned. The SNV of this patient were assigned to the following categories: CPC 1-truncal (blue), CPC 1-hiPrevalence (cyan) and CPC 2 (gray) (data).

This patient had 1 public mutation(s) present in a gene in the Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) database (Figure 2 and Table 4).

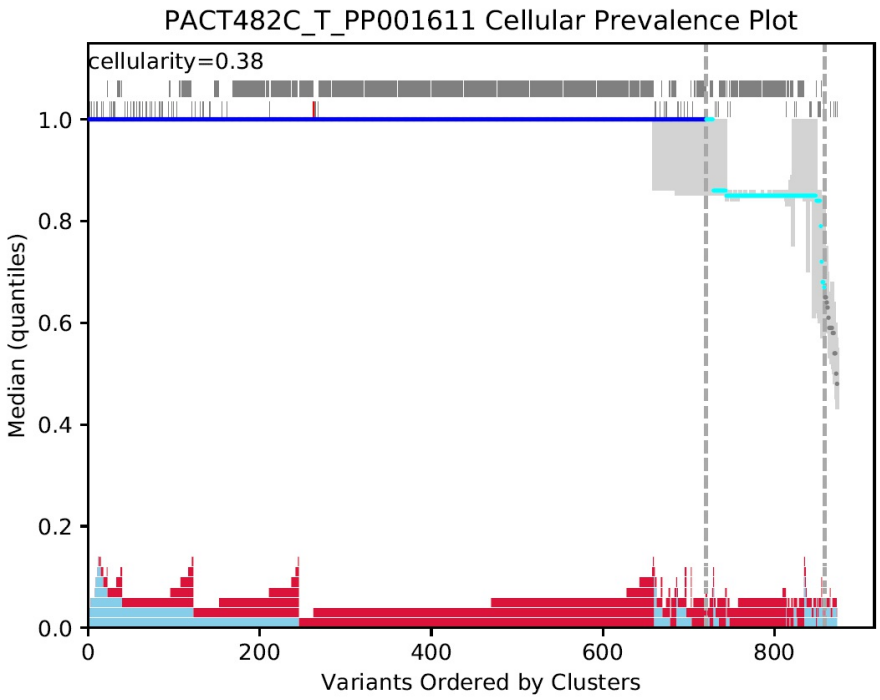

Figure 2. Cellular prevalence plot of results from Minimal-subclone analysis. *Quantiles of cellular prevalence are represented by vertical solid grey lines on Y axis. Median cellular prevalence of CPC I-truncal is marked by blue dots. Median cellular prevalence of CPC I-high prevalence is marked by cyan dots. Median cellular prevalence of CPC II is marked by grey dots. (Note the dots often visually merge into horizontal lines.) Vertical sticks near the top of plot are variants (SNP are the first row, SNV are the second row). Red vertical sticks indicate COSMIC CGC mutations. Stack blocks near the bottom of the plot reflect the genotype results from associated Sequenza solution. Red stacks are the number of A alleles, light blue stacks are the number of B alleles. Vertical dotted grey lines separate clusters. Clusters are ordered by their median CP value and variants within a cluster are ordered by their estimated CP value on the X axis. SNP: small nucleotide polymorphisms. Cellularity: tumor content.*

Table 4: Patient mutations present in the COSMIC CGC database.

Id	hugoSymbol	varClassification	medianCp	q3Cp	q1Cp	CPC	xAxis
chr17:7577094:G>A	TP53	Missense_Mutation	1.00	1.00	1.00	1-truncal	263

*medianCp: median cellular prevalence. q3Cp: third quantile of cellular prevalence. q1Cp: first quantile of cellular prevalence. CPC: cellular prevalence category (refer to RTR-007_003). xAxis: X axis position in the corresponding cellular prevalence graph (**Figure 2**). If "None" is present, it indicates no COSMIC CGC mutations were detected.*

2. 6 HLA Typing and coverage by PACT 66-HLA panel

HLA types were estimated according to PD-002_004. The protocol first uses OptiType to determine HLA types from DNA WES Normal, DNA WES Tumor, and RNA-Seq Tumor data. If the HLA types by OptiType were concordant then the concordant alleles are listed below. If they were discordant, then OptiType was run on the WES Normal sample and those results are presented below.

- HLA alleles by OptiType: C*06:02, C*07:02, B*07:02, B*57:01, A*02:01, A*01:01

The HLA types detected by OptiType which exist in the PACT-66 panel were used for neoE-HLA binding affinity prediction in the next step.
0 of the HLA alleles determined by ATHLATES were homozygous, so this patient has 6 unique HLA alleles and were covered by the PACT 66-HLA panel. They are:

- HLA alleles covered: C*06:02, C*07:02, B*07:02, B*57:01, A*02:01, A*01:01

2. 7 Detection of RNA expression

RNA sequencing identified **32 (38. 55421686746988%)** expressed NSM within the tumor (>=1 mRNA absolute read count) (**Figure 4**). Zero RNA reads were detected for **51** of the NSM

2. 8 NeoE-HLA Binding Affinity Prediction and Selection

NeoE-HLA elution (EL) and binding affinity (BA) ranks (EL Rank and BA Rank) for the **6** unique covered HLA alleles and **32** expressed NSM were predicted using netMHCpan4.1. Both EL Rank and BA Rank were The flowchart in **Figure 4** summarizes the results of each step in the neoE identification and prioritization process when EL BA Rank was used.
genetated for each neoE-HLA by netMHCpan4.1. An algorhiym was applied to te netMHCpan4.1 results to combine the EL Rank and BA Rank into an EL BA Rank based on whichever rank (EL or BA) was highest. NeoE-HLA complexes were selected by the Average-Out method according to their netMHCpan4.1 EL BA ank. After Average-Out selection, **0** of the 32 expressed NSM were covered by 352 neoE-HLA candidates and the selected candidates included **0** unique tumor-specific neoEs. Predicted binding affinities of the neoE-HLA candidates are shown in **Figure 3**.

To verify that the neoE candidate sequences did not exist in the normal human proteome an "exactMatch" test was performed (NeoE column, **Table 5**). Meanwhile, the wild-type peptides of neoE candidates were confirmed as exact matches to sequences in the normal human proteome (Wild Type epitopes column, Table 5). The exactMatch test was passed by **0** of the **0** selected unique neoEs.

The flowchart in **Figure 4** summarizes the results of each step in the neoE identification and prioritization process when EL BA Rank was used.

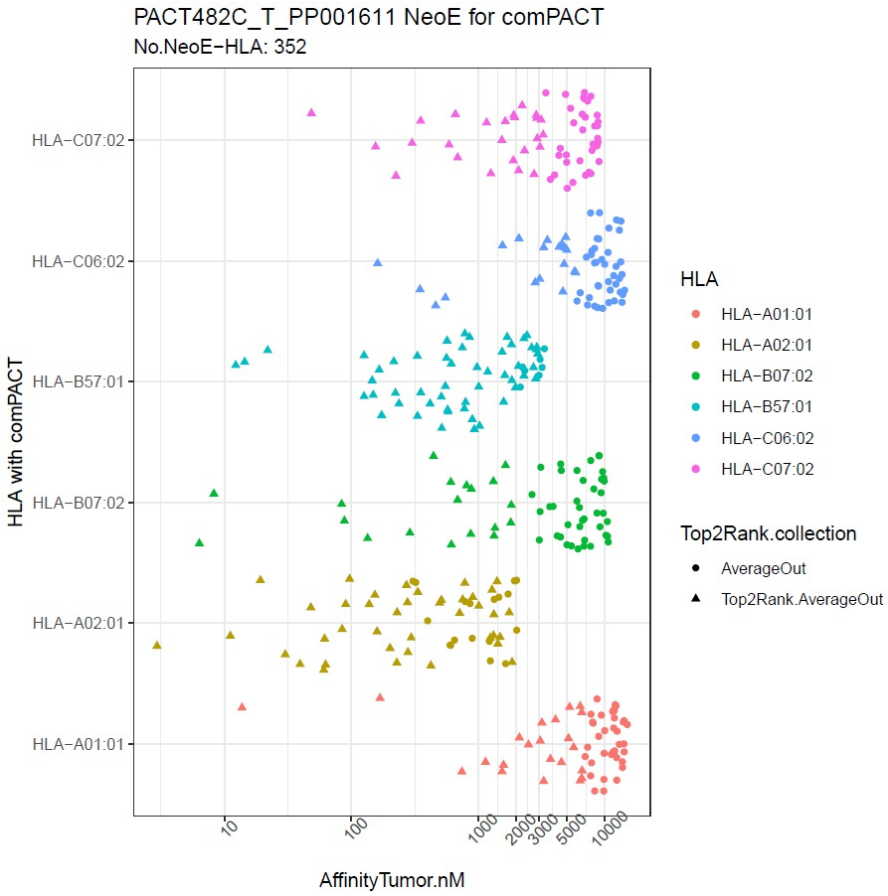

Figure 3. The 352 neoE-HLA candidates selected by Average-Out for the patient HLA alleles covered by the PACT 66-HLA panel. The x-axis is the predicted binding affinities (Kd) for each neoE-HLA(lower kd => stronger binding). The y-axis groups comPACTs by the patient's HLA alleles covered by the PACT 66-HLA panel. The size of the circle correlates with the absolute number of RNA reads detected.

Table 5: NeoE exactMatch.

Patient ID	NeoE (peptideTumor) exactMatch test (Verification)	Wild Type epitopes (peptideNormal) exactMatch test (Verification)
0411	0 NeoE without exact match	0 corresponding wild-type peptides with exact match (positive control)

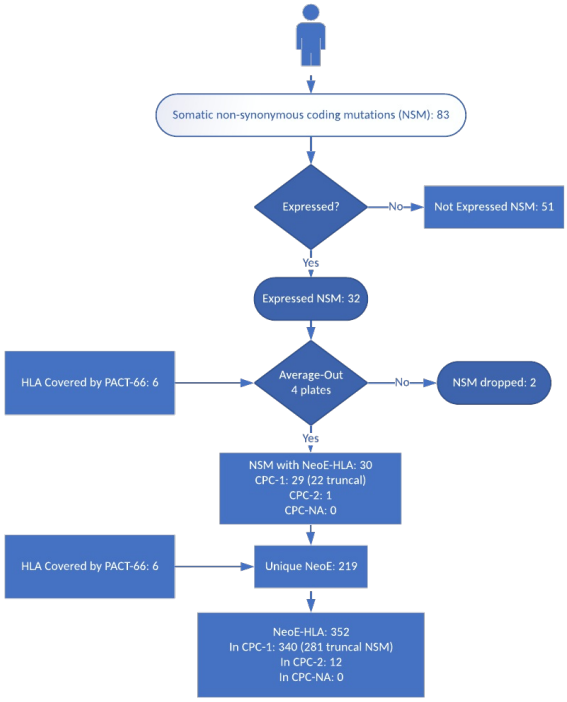

Figure 4. Results from somatic non-synonymous coding mutations (NSM) to neoE-HLA using the Minimal-subclone CPC and Average-Out methods to prioritize and select neoE-HLA complexes recommended for production. The flowchart shows the number of NSM, HLA alleles, and neoE-HLA candidates at each stage of the process with the Average-Out method. NA: not available.

NeoE-HLA candidate selection using EL BA Rank resulted in each of the **6** unique covered HLA alleles presenting **58-59** NeoEs (see "NeoE per HLA" row in **Table 6.**) Furthermore, each HLA alleles presented neoEs from **15-20** of the **30** NSM with oligos (see "NSM per HLA" row in **Table 6.**). The median number of neoE per HLA was **1. 00** per NSM(range 0-17). This patient had missense mutations in **8** COSMIC CGC genes (bolded rows in Table 6). Some neoE candidates were selected for presentation by multiple HLA alleles. Candidate neoE-HLA pairs and corresponding neoE sequences are here.

Table 6: NeoE-HLA per NSM per HLA and total with absolute number of RNA reads and CPC for each NSM.

Gene Protein Change	HLA-A01:01	HLA-A02:01	HLA-B07:02	HLA-B57:01	HLA-C06:02	HLA-C07:02	neoE-HLA per NSM	RNA Reads	CPC
ACE:p.V535G	4	5	2	6	10	10	37	44	1-truncal
ALPPL2:p.V419A	0	2	2	0	3	3	10	2	1-truncal
ANKRD52:p.R118W	0	0	4	2	0	0	6	9	1-truncal
ATM:p.E586Q	**0**	**2**	**0**	**0**	**1**	**2**	**5**	**18**	**1-truncal**
ATM:p.K2811Q	**2**	**3**	**0**	**3**	**2**	**2**	**12**	**37**	**2**
CISD1:p.R73C	0	0	0	1	0	0	1	6	1-truncal
CRAMP1L:p.A334D	1	0	4	0	2	2	9	9	1-truncal
FAM172A:p.L8F	6	15	2	17	4	3	47	6	1-truncal
FLCN:p.R168H	**3**	**3**	**2**	**9**	**12**	**12**	**41**	**15**	**1-truncal**
HSP90B1:p.V528L	4	1	0	0	0	1	6	274	1-truncal
ITGBL1:p.D281Y	6	0	0	1	6	5	18	4	1-truncal
KMT2C:p.E531V	**1**	**3**	**0**	**0**	**2**	**2**	**8**	**58**	**1-truncal**
MAP3K1:p.R248Q	**0**	**0**	**3**	**0**	**0**	**0**	**3**	**2**	**1-truncal**
MAT2A:p.R264C	1	2	0	0	0	0	3	71	1-truncal
MCM4:p.Q822H	0	0	3	4	1	1	9	17	1-truncal
MED23:p.R765T	0	0	0	1	0	0	1	10	1-hiPreval
POLQ:p.K109N	5	3	0	0	0	0	8	4	1-hiPreval
POT1:p.K427E	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**4**	**1-hiPreval**
RASSF6:p.T365R	0	1	1	0	1	1	4	1	1-truncal
SATB2:p.P468L	3	3	8	6	3	1	24	103	1-truncal
TCF25:p.E374K	1	0	3	0	0	0	4	50	1-truncal
TGFBI:p.M502V	3	4	11	0	3	4	25	120	1-truncal
TP53:p.R282W	**1**	**0**	**1**	**2**	**0**	**0**	**4**	**220**	**1-truncal**
TRRAP:p.L694I	**4**	**8**	**1**	**2**	**2**	**3**	**20**	**114**	**1-hiPreval**
TTC7B:p.A437S	1	0	6	3	1	2	13	3	1-hiPreval
USP12:p.K96R	0	0	0	1	1	0	2	10	1-truncal

UVSSA:p.Q172E	1	0	0	0	1	1	3	32	1-truncal
WDTC1:p.E357Q	2	2	6	0	2	2	14	27	1-hiPreval
XRN1:p.P540Q	9	0	0	1	1	1	12	22	1-truncal
ZNF3:p.Q100P	0	0	0	0	1	1	2	16	1-hiPreval
NSM per HLA	19	16	16	15	20	20	30		
NeoE per HLA	58	58	59	59	59	59	352		

NSM = somatic nonsynonymous mutations (Gene.Protein_Change). RNA reads = absolute number of RNA reads that contained the mutation. CPC = cellular prevalence category. The red/blue heatmap shows the highest (red) and lowest (blue) number of neoE per NSM per HLA. Baby blue rows and column are totals. Gray-filled NSM (1st column) were also detected in other samples from this patient and non-filled were only detected in this sample. Bold = COSMIC CGC genes.

2.9 Generation of DNA Oligo Sequences to make NeoE-HLA Complexes
The sequences of oligos for 352 neoE-HLA candidates, selected using PD-002_004 methods, were generated. The oligo file for ordering is here.

3.0 RECOMMENDATION

The number of expressed NSM (32) in this sample was above the interim cutoff of 10. The tumor content by NGS (38.00%) of this sample was above the interim cutoff of 30%. Interim cutoffs are stated in the Acceptance Criteria of PD-002_004 (Section 8).
This is Low Expressed Nsm Comment

Low Tc By Ngs Pct Comment

This is test recommendation

4.0 APPENDIX A: ANALYSES FOR INFORMATION ONLY

Appendix A presents further analyses of the patient's data, which are for information only.

4.1 Microsatellite Instability (MSI) Analysis:
MSI analysis was conducted with the MANTIS program as described in PD-002_004. The analysis compared the presence and absence of mutations in microsattelite regions in the WES data and determined their relative frequency in tumor compared to normal specimens.
The MANTIS results indicated that the biopsy was microsatellite **stable** (the result file is here).

4.2 Tumor Infiltrated Lymphocyte (TIL) Analysis

TCR analysis was conducted based on RNA-Seq data with the MiXCR program as described in PD-002_004. Due to limited sequencing coverage in the TCR regions, the purpose of this analysis is to provide qualitative, but not quantitative, assessment of T cells in the tumor microenvironment. In addition, it is important to note that the TCRs detected in the patient's RNA-Seq data may not be from tumor-specific T cells. These T cells could be immune cells in normal blood circulation rather than being true TILs.

In this tumor biopsy, **164** and **132** CDR3 from TCRα and TCRβ, respectively, were detected. For comparison, the averages of **15** TCRα and **23** TCRβ CDR3 detected in the biopsies of individuals within a TCGA cohort of **COLON** patients were graphed adjacent to the patient's total numbers of each receptor in Figure 5. TCGA cohort data is from Li et al. (Nat Genet. 2016 Jul; 48(7): 725—732.) and summary tables are in slides here. Note that TCGA biopsies are from the primary cancer site and this patient's biopsy sample was also from a premalignant legs tumor . Of the total, **121** and **103** unique CD3 from TCRα and TCRβ were detected, respectively. MiXCR groups identical sequences into clonotypes, since the presence of multiple identical reads suggests that multiple T cell clones are present in the TIL. In the TIL of this patient, multiple identical reads of **31 (26% of unique)** TRA and **24 (23% of unique)** TRB CDR3 sequences were detected (Figure 6). The TIL TCR data is here.

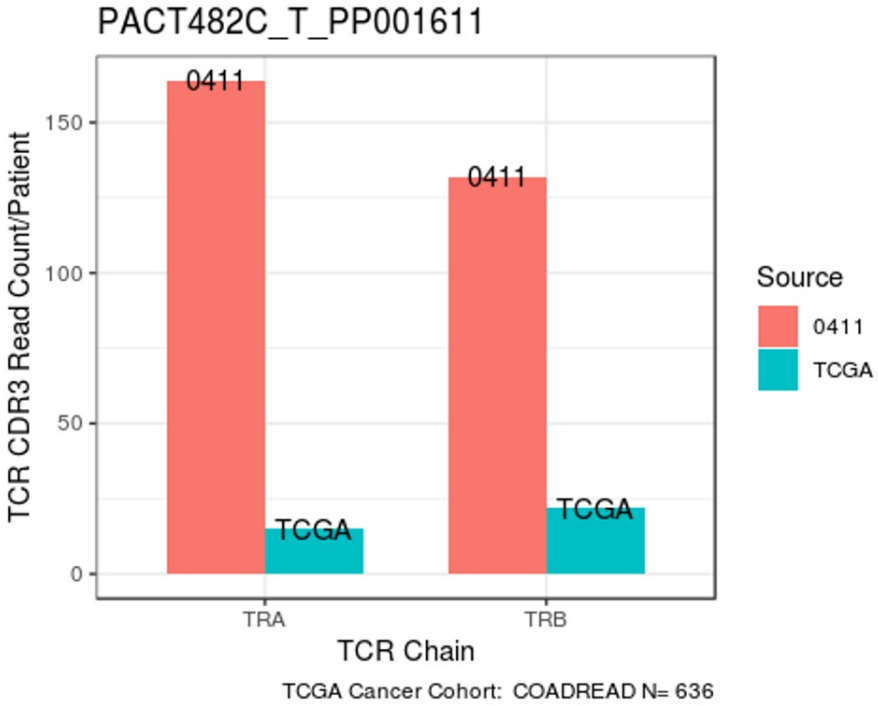

Figure 5. Comparison of patient versus TCGA cohort total CDR3 read counts of TCR alpha and beta detected in TIL. *The sources of numbers of TRA and TRB are either the number present in the patient's TIL (0405) or the average number in a TCGA cohort with the same type of cancer as the patient's original diagnosis.*

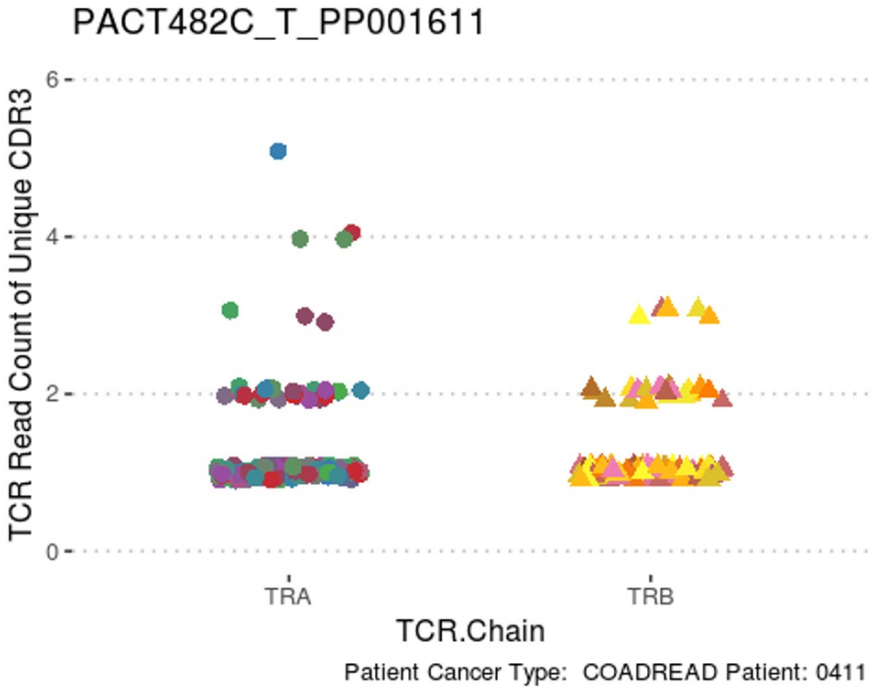

Figure 6. RNA read count of unique CDR3 in the TIL. *Each dot represents a unique CDR3 sequence. Along the x-axis, dots are clustered by TCR chain (TRA or TRB). The y-axis shows the number of identical CDR3 RNA reads detected for each sequence. The dots are colored by V-gene in the TCR chains.*

4. 3 Tumor Microenvironment (TME) Analysis

RNA-seq data from this patient was analyzed to provide a snapshot of the TME. Gene expression levels were normalized by transcript length, library size, and transcript per million (TPM) transcripts. Normalization by length allows comparison between transcripts within a sample, normalization by library size allows comparison between multiple samples, and TPM normalization converts the values to a human-digestible number.

On average, there are approximately 360,000 transcripts/cell in a mammalian cell (see reference). The rule of thumb is, that if a gene in a bulk

sequencing sample has less than 1 TPM, then that indicates there are less than 1 transcripts/cell and it is generally considered not expressed. If TPM is between 1 and 2, the transcript is low or has close to no expression. However, it is important to remember that cells in tumor biopsies are not evenly distributed. Therefore, TPM values are more qualitative than quantitative. In this report, genes with TPM values lower than 2 are considered to have low expression and are colored green in **Table 7**.

Table 7 holds a list of common immunity-related genes and their TPM values. RNA expression of "HLA-A, HLA-B, and HLA-C genes were detected" .

Table 7. Biomarker expression based on tumor RNA-seq. (Data source: here)

Gene_Name	Alias	Category	TPM
B2M	B2M	HLA-I Complex	305.85
CD4	CD4	TCR co-R.	148.73
CD8A	CD8	CD8 T cells	4.67
FOXP3	FOXP3	TReg master regulator	3.77
HLA-A	**HLA_A**	**HLA-I Complex**	**756.08**
HLA-B	**HLA_B**	**HLA-I Complex**	**1056.73**
HLA-C	**HLA_C**	**HLA-I Complex**	**886.04**
HLA-DRA	HLA-DR	Ag presentation	880.08
IDO1	IDO	Tryptophan metabolism	3.42
IDO2	INDOL1	Tryptophan metabolism	1.61
TAP1	TAP1	HLA-I Antigen Presentation	40.25
TAP2	TAP2 start-32781544	HLA-I Antigen Presentation	0.86
TAP2	TAP2 start-32789610	HLA-I Antigen Presentation	39.24
ACTB	ActinB	housekeeping	3838.16
GAPDH	GAPDH	housekeeping	118.66
CALR	CALR	HLA-I Antigen Presentation	1172.08
CD81	CD81	Tetraspanin	1137.62
PDIA3	ERp57	HLA-I Antigen Presentation	505.07
POMP	HSPC014	Immunoproteasome	283.62
CD44	CD44		247.87
FAS	CD95		232.44
ANPEP	CD13	Metalloprotease	227.83
CD68	CD68	Macrophage	196.37
PTPRC	CD45RO		168.44
TPP2	TPP2	Peptide trimming in ER	149.62
ITGAX	CD11c	DC cells	144.50
TFRC	CD71	Transferrin R.	137.08
MSR1	CD204	Class A S.R.	120.53
NRP1	CD304	VEGF co-R.	119.19
TAPBP	Tapasin	HLA-I Antigen Presentation	106.54
ERAP1	ARTS-1	Peptide trimming in ER	104.77
PSME1	PA28alpha	Immunoproteasome	94.28
ARG1	ARG1	Arginase 1	93.35
ALCAM	CD166	Glycoprotein	91.95
PSMB8	beta5i	Immunoproteasome	90.80
IL7R	IL7		85.33
TMEM173	STING	STING	84.51
IFNGR1	CD119	IFNg R.	83.25
PSMB9	beta1i	Immunoproteasome	79.03
BCL6	BCL-6		77.23
ITGAM	CD11b	Complement R. 3	73.88
FCGR3A	CD16	Low affinity FCGR3a	73.44
CD82	CD82	Glycoprotein	70.79
CD163	CD163	High affinity S.R.	65.57
SLC2A1	GLUT1		63.47
SLC1A5	ASCT2		58.48
FCGR2A	CD32	Low affinity FCGR2a	55.03
CD14	CD14	LPS co-R.	45.74
ICAM1	CD54	ICAM-1	44.87
THOP1	THOP1	Peptide cleavage in ER	42.05
CD36	CD36	Class B S.R.	37.96

ERAP2	LRAP	Peptide trimming in ER	37.31
MRC1	CD206	Mannose receptor	34.92
PSME2	PA28beta	Immunoproteasome	33.25
PSMB10	beta2i	Immunoproteasome	31.99
NT5E	CD73	Adenosine pathway	30.26
CD40	CD40	Costimulatory R.	26.47
CD3E	CD3	TCR co-R.	23.36
TLR4	CD284	Toll-Like Receptor	14.70
SIGLEC1	CD169	Sialoadhesin	12.69
IL3RA	CD123	IL-3 R.	11.53
IL2RB	CD122		11.27
CD27	CD27		10.45
CD86	CD86	Costimulatory ligand	9.38
PRDM1	Blimp-1		8.82
SLAMF7	SLAMF7	Costimulatory ligand	8.10
IL2RA	CD25	IL-2 R.	8.00
CXCR4	CXCR4	Chemokine R. 4	7.06
HAVCR2	Tim-3		6.56
CD80	CD80	Costimulatory ligand	6.47
FCGR1A	CD64	High affinity FCGR1a	5.21
TNFRSF4	OX40	Costimulatory R.	5.10
CD38	CD38	cADP ribose hydrolase	5.03
IL15	IL15		4.98
SELL	CD62L		4.97
CD28	CD28		4.94
KLRG1	KLRG-1		4.43
TNFRSF18	GITR	Costimulatory R.	4.41
C5AR1	CD88	Complement R. 5	4.30
TBX21	Tbet		3.67
TNFRSF9	4-1BB	Costimulatory R.	3.61
PDCD1LG2	PD-L2	Coinhibitory ligand	3.42
CD7	CD7	NK cells	3.41
GZMB	GrzB		3.15
LAG3	Lag3		2.90
MKI67	Ki-67	Proliferation marker	2.67
ADORA2A	A2aR	Adenosine pathway	2.52
EOMES	Eomes		2.44
SPN	CD43		1.81
PDCD1	PD-1		1.71
PRF1	Preforin		1.70
ICOS	ICOS	Costimulatory R.	1.58
IL12A	IL12		1.57
CTLA4	CTLA-4	Coinhibitory R.	1.26
MS4A1	CD20	B cells	1.06
TIGIT	VSIG9	T cell response	0.96
CD274	PD-L1	Coinhibitory ligand	0.89
CCR7	CCR7		0.71
FUT4	CD15	Granulocyte	0.59
IFNG	IFNG	Interferon Type II	0.30
IL2	IL2		0.29
IL21	IL21		0.00

TPM=transcript per million. Genes of interest (those involved in antigen presentation, CD4, CD8, FOXP3, IDO1, IDO2, and housekeeping genes) are colored with a heatmap and at the top of the table. Other genes are organized by TPM value. Genes with TPM lower than 2 are considered low or not expressed and those outside of the heatmap are highlighted in light green.

5.0 APPENDIX B: SOFTWARE, REFERENCES, & HLA PANNEL

5.1 References to PACT Internal Documentation

Document	Version	Title	Location

Study-20-288	001	Application of Minimal-subclone estimation method to determine cellular prevalence category for neoepitope targets	Q drive
Study-20-234	001	Cellular prevalence module release test	Q drive
PD-002	004	NeoE Prediction and Identification	Q drive
RTR-007	003	NeoE Selection and Prioritization	Q drive
Study-20-171	001	NeoE-HLA candidate selection with Average-out method	Sharepoint
Study-20-172	001	Testing of implementation of Average-out method for neoE-HLA selection	Benchling

5. 2 Software packages used while following PD-002_004

Program Name	Version	Category
ATHLATES	v1.0	HLA calling from WES
bam-readcount	latest	Utility Tool
BWA	0.7.9	WES
FastQC	latest	Sequence QC
GATK	3.6	WES
HISAT2	2.0.4	RNASeq
MANTIS	V1.0.3	Microsatellite instability (MSI) status
muTect/muTect2	1.1.7	WES variant calling
netMHC	3.4	Neoantigen prediction
netMHCpan	4	Neoantigen prediction
Oncotator	1.9.2	WES mutation annotation
OptiType	1.3.4	HLA calling from WES and RNASeq
picard-tools	1.115	WES
PMHC-I immunogenicity predictor	1	PMHC Class I immunogenicity prediction
Pyclone	0.13.1	Truncality
python	2.7/3.5	Utility Tool
R	3.4.4	Utility Tool
Razers	3.5.3	HLA calling from WES (OptiType)
RStudio		Utility Tool
samtools	1.3.1	Utility Tool
Sequenza	3.0.0	Genotyping/Cellularity
STAR	2.5.3	RNASeq
StringTie	1.2.2	RNASeq
VarDict	1.4.7	WES
VarScan2	2.3.9	WES variant calling

5. 3Reference databases used while following PD-002_004

Database Name	Version	Location
dbSNP	v146	ftp://ftp.broadinstitute.org/bundle
Ensembl	release 75	http://ensembl.org/
ExAc	3.1	http://exac.broadinstitute.org
GATK Resource Bundle	hg19/Grch37	ftp://ftp.broadinstitute.org/bundle
GTEx	V4	https://wwwgtexportal.org/
Human Proteome	Homo_sapiens.GRCh37.75.pep.all.fa	http://ensembl.org/
IMGT(TCR/HLA)	3.1.17	http://wwwimgt.org/
Oncotator Datasource	Dec112014 release	ftp://ftp.broadinstitute.org/bundle
RefSeq	1052019	ftp://hgdownload.cse.ucsc.edu/goldenPath
TCGA	Version 1.0	https://portal.gdc.cancergov/
UCSC	hg19	ftp://ftp.broadinstitute.org/bundle

5. 4 PACT 66-HLA panel

HLA-A	HLA-B	HLA-C
HLA-A*25:01	HLA-B*15:07	HLA-C*03:03
HLA-A*26:01	HLA-B*27:05	HLA-C*07:04
HLA-A*29:02	HLA-B*35:03	HLA-C*08:01
HLA-A*68:02	HLA-B*37:01	HLA-C*08:02
HLA-A*11:01	HLA-B*38:01	HLA-C*12:02
HLA-A*23:01	HLA-B*41:02	HLA-C*12:03
HLA-A*30:01	HLA-B*44:05	HLA-C*14:02
HLA-A*33:03	HLA-B*49:01	HLA-C*15:02
HLA-A*11:01	HLA-B*52:01	HLA-C*17:01
HLA-A*23:01	HLA-B*55:01	HLA-C*03:03

HLA-A*30:01	HLA-B*13:02	HLA-C*07:04
HLA-A*01:01	HLA-B*15:07	HLA-C*08:01
HLA-A*33:03	HLA-B*27:05	HLA-C*08:02
HLA-A*02:01	HLA-B*35:03	HLA-C*12:02
HLA-A*03:01	HLA-B*37:01	HLA-C*12:03
HLA-A*24:02	HLA-B*38:01	HLA-C*14:02
HLA-A*30:02	HLA-B*41:02	HLA-C*01:02
HLA-A*31:01	HLA-B*44:05	HLA-C*15:02
HLA-A*32:01	HLA-B*49:01	HLA-C*04:01
HLA-A*33:01	HLA-B*08:01	HLA-C*17:01
HLA-A*68:01	HLA-B*52:01	HLA-C*06:02
HLA-A*01:01	HLA-B*15:01	HLA-C*07:02
HLA-A*02:01	HLA-B*55:01	HLA-C*16:01
HLA-A*03:01	HLA-B*15:03	HLA-C*01:02
HLA-A*24:02	HLA-B*35:01	HLA-C*04:01
HLA-A*30:02	HLA-B*40:02	HLA-C*06:02
HLA-A*31:01	HLA-B*42:01	HLA-C*07:02
HLA-A*32:01	HLA-B*44:03	HLA-C*16:01
HLA-A*33:01	HLA-B*51:01	HLA-C*02:02
HLA-A*68:01	HLA-B*53:01	HLA-C*03:04
HLA-A*25:01	HLA-B*08:01	HLA-C*05:01
HLA-A*26:01	HLA-B*15:01	HLA-C*07:01
HLA-A*29:02	HLA-B*15:03	HLA-C*02:02
HLA-A*68:02	HLA-B*35:01	HLA-C*03:04
HLA-A*25:01	HLA-B*40:02	HLA-C*05:01
HLA-A*26:01	HLA-B*42:01	HLA-C*07:01
HLA-A*29:02	HLA-B*44:03	HLA-C*03:03
HLA-A*68:02	HLA-B*51:01	HLA-C*07:04
HLA-A*11:01	HLA-B*07:02	HLA-C*08:01
HLA-A*23:01	HLA-B*53:01	HLA-C*08:02
HLA-A*30:01	HLA-B*14:02	HLA-C*12:02
HLA-A*33:03	HLA-B*18:01	HLA-C*12:03
HLA-A*11:01	HLA-B*27:02	HLA-C*14:02
HLA-A*23:01	HLA-B*39:01	HLA-C*15:02
HLA-A*30:01	HLA-B*40:01	HLA-C*17:01
HLA-A*33:03	HLA-B*44:02	HLA-C*03:03
HLA-A*01:01	HLA-B*46:01	HLA-C*07:04
HLA-A*02:01	HLA-B*50:01	HLA-C*08:01
HLA-A*03:01	HLA-B*57:01	HLA-C*08:02
HLA-A*24:02	HLA-B*58:01	HLA-C*12:02
HLA-A*30:02	HLA-B*07:02	HLA-C*12:03
HLA-A*25:01	HLA-B*14:02	HLA-C*14:02
HLA-A*31:01	HLA-B*18:01	HLA-C*15:02
HLA-A*26:01	HLA-B*27:02	HLA-C*17:01
HLA-A*32:01	HLA-B*39:01	HLA-C*01:02
HLA-A*29:02	HLA-B*40:01	HLA-C*04:01
HLA-A*33:01	HLA-B*44:02	HLA-C*06:02
HLA-A*68:02	HLA-B*46:01	HLA-C*07:02
HLA-A*68:01	HLA-B*50:01	HLA-C*16:01
HLA-A*25:01	HLA-B*57:01	HLA-C*01:02
HLA-A*25:01	HLA-B*58:01	HLA-C*04:01
HLA-A*26:01	HLA-B*13:02	HLA-C*06:02
HLA-A*26:01	HLA-B*15:07	HLA-C*07:02
HLA-A*01:01	HLA-B*27:05	HLA-C*16:01
HLA-A*29:02	HLA-B*35:03	HLA-C*03:03
HLA-A*29:02	HLA-B*37:01	HLA-C*02:02
HLA-A*02:01	HLA-B*38:01	HLA-C*07:04
HLA-A*68:02	HLA-B*41:02	HLA-C*03:04
HLA-A*68:02	HLA-B*44:05	HLA-C*08:01

HLA-A*03:01	HLA-B*49:01	HLA-C*05:01
HLA-A*25:01	HLA-B*52:01	HLA-C*08:02
HLA-A*24:02	HLA-B*55:01	HLA-C*07:01
HLA-A*26:01	HLA-B*13:02	HLA-C*03:03
HLA-A*30:02	HLA-B*15:07	HLA-C*12:02
HLA-A*29:02	HLA-B*27:05	HLA-C*03:03
HLA-A*31:01	HLA-B*35:03	HLA-C*07:04
HLA-A*68:02	HLA-B*37:01	HLA-C*12:03
HLA-A*32:01	HLA-B*38:01	HLA-C*07:04
HLA-A*33:01	HLA-B*41:02	HLA-C*08:01
HLA-A*68:01	HLA-B*44:05	HLA-C*14:02
HLA-A*25:01	HLA-B*49:01	HLA-C*08:01
HLA-A*26:01	HLA-B*52:01	HLA-C*08:02
HLA-A*29:02	HLA-B*55:01	HLA-C*15:02
HLA-A*68:02	HLA-B*08:01	HLA-C*08:02
HLA-A*11:01	HLA-B*15:01	HLA-C*12:02
HLA-	HLA-	HLA-

6.0 AMENDMENTS

This is test amendments

Supporting Documents:

- 51930_dummy.pdf