Report - Sample **CESC1**

NeoDisc_v1.7.0 - run Paper

July 6, 2024

1 Sample summary

Main information about the sample **CESC1**.

1.1 Tumor and Germline information

Project(s): **AgDisc_NeoTIL**
WES Sequencing facility: **Microsynth**
RNASeq Sequencing facility: **Health 2030 Genome Center-Geneva**
The sample ID is: **CESC1**
The tumor ID is: **CESC1-A**
The germline ID is: **CESC1-PBMC**
The RNASeq ID is: **CESC1-A**
Diagnosis: **Epidermoid cervical carcinoma**
TCGA Cancer type: **Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma** (Matched TCGA Cancer type)
Tumor purity estimated by Sequenza: **31.0%** (reliable for samples with more than 30% of tumor content)
Mutational load estimation: **25.497 SM/Mb** (based on high-confidence somatic mutations identified)

2 Aim of the analysis

NeoDisc Paper analysis

2.1 Bait used for library preparation

The bait used for the library preparation was: **SureSelect_Exome_V7**

3 Sequencing data Quality Controls

Analysis of the quality of the sequencing data, from FastQC. See sequencing.qcfail.com for more information.

3.1 Overview

Brief overview of the FastQC report (full germline and tumor reports available in the zip file, in html format). Only the most important parameters are shown. The full FastQC report contain additional data and should be examinated if any of the reported parameters failed.
✓ : **Pass**, ◇ : **Warning**, × : **Fail**

3.1.1 Germline

- ✓ **Basic Statistics**
- ✓ **Per base sequence quality**
- ✓ **Per tile sequence quality**
- ✓ **Per sequence quality scores**
- ✓ **Overrepresented sequences**
- ✓ **Adapter Content**

3.1.2 Tumor

✓ **Basic Statistics** ✓ **Per tile sequence quality** ✓ **Overrepresented sequences**

✓ **Per base sequence quality** ✓ **Per sequence quality scores** ✓ **Adapter Content**

4 HLA Typing

Summary of HLA typing data. Molecular and NGS (DNA and/or RNA) derived HLA typing on germline and tumor samples are reported in the table below. In column **Selected**, HLA alleles available for predictions are highlighted in **bold** and alleles not available are highlighted in *italic*. Alleles determined from molecular and NGS data in germline/tumor sample(s) are shown in column(s) **Molecular DNA** and/or **RNA**, where color indicates if allele is identical or different from selected HLA alleles. LOH means Loss of heterozygosity, i.e. the allele is lost.

Class	Selected Typing	Molecular Typing	DNA		RNA
			CESC1-PBMC	CESC1-A	CESC1-A
A	**A*24:02**	A*24:02	A*24:02	A*24:02 [2]	A*24:02 [0.58]
A	**A*31:01**	A*31:01	A*31:01	A*31:01 [1]	A*31:01 [0.42]
B	**B*35:43**	B*35:43	B*35:43	B*35:43 [2]	B*35:43 [0.49]
B	**B*51:01**	B*51:01	B*51:12	B*51:01 [1]	B*51:01 [0.51]
C	**C*01:02**	C*01:02	C*01:02	C*01:02 [1]	C*01:02 [0.42]
C	**C*15:02**	C*15:02	C*15:02	C*15:02 [2]	C*15:02 [0.58]
DPA1	**DPA1*01:03**	DPA1*01:03	DPA1*01:03	DPA1*01:03 [2]	DPA1*01:03 [1.00]
DPA1	-	-	-	-	-
DPB1	**DPB1*04:01**	DPB1*04:01	DPB1*04:01	DPB1*04:01 [1]	DPB1*04:01 [0.41]
DPB1	**DPB1*04:02**	DPB1*04:02	DPB1*04:02	DPB1*04:02 [1]	DPB1*04:02 [0.59]
DQA1	**DQA1*03:01**	DQA1*03:01	DQA1*03:01	DQA1*03:01 [2]	DQA1*03:01 [0.52]
DQA1	-	DQA1*03:03	DQA1*03:03	DQA1*03:03 [0] (LOH)	DQA1*03:03 [0.48]
DQB1	**DQB1*03:01**	DQB1*03:01	DQB1*03:01	DQB1*03:01 [1]	DQB1*03:01 [0.55]
DQB1	**DQB1*03:02**	DQB1*03:02	DQB1*03:02	DQB1*03:02 [2]	DQB1*03:02 [0.45]
DRB1	**DRB1*04:04**	DRB1*04:04	DRB1*04:04	DRB1*04:04 [3]	DRB1*04:04 [0.47]
DRB1	**DRB1*04:08**	DRB1*04:08	DRB1*04:08	DRB1*04:08 [1]	DRB1*04:08 [0.53]
DRB3	-	-	-	-	-
DRB3	-	-	-	-	-
DRB4	**DRB4*01:03**	DRB4*01:03	DRB4*01:03	DRB4*01:03 [2]	DRB4*01:03 [1.00]
DRB4	-	-	-	-	-
DRB5	-	-	-	-	-
DRB5	-	-	-	-	-

5 Antigens presentation machinery

Analysis of presentation-machinery associated pathways. Bulk-RNASEQ derived expression and mutations in each gene associated to each pathway is summarized in the table below. Genes expression from bulk RNASEQ is compared to GTEx (Cervix - Ectocervix) and TCGA (Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma) expression and is reported in the associated figure and table. In the table, GTEx and TCGA 25-75 percentiles TPM expression values are reported

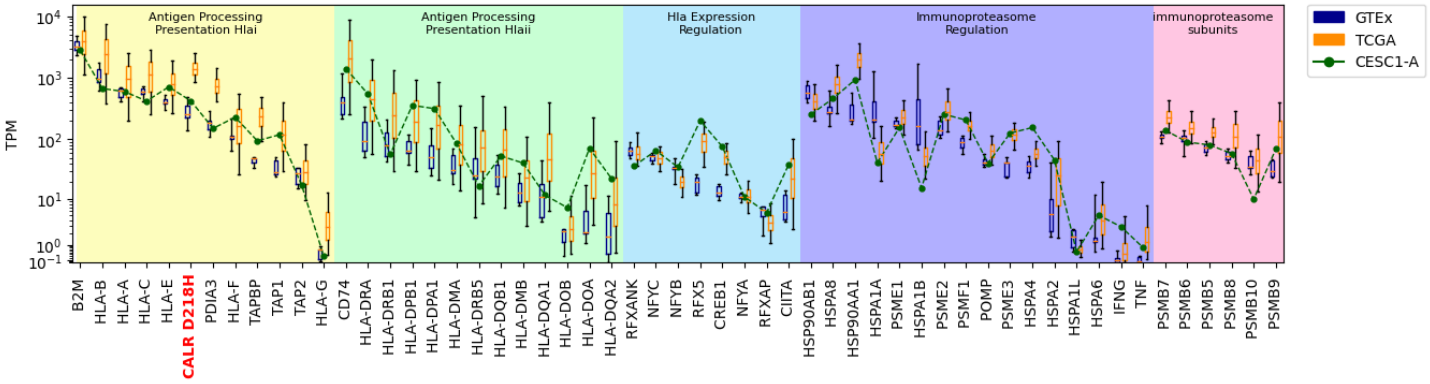


Figure 1: Genes expression comparison Separated per pathway. GTEx and TCGA gene expression is represented as boxplots. Sample-specific gene expression is determined from bulk RNAseq and is shown as connected dots.

Pathway	Gene	GTEx TPM (5-95 pc)	TCGA TPM (5-95 pc)	Sample(s) TPM (min-max)	Expression Regulation	CopyNumber	Mutation(s)
antigen processing presentation hlaI	B2M	2315.37-4847.04	1143.63-9781.03	2797.4428	Normal	1,1	-
	HLA-B	630.83-1733.21	376.44-7365.52	671.022	Normal	1,1	-
	HLA-A	409.71-713.71	204.53-2516.29	598.2836	Normal	1,1	-
	HLA-C	423.71-731.58	256.54-2873.67	420.9264	DOWN	1,1	-
	HLA-E	298.77-532.33	270.58-1921.47	713.2668	UP	1,1	-
	CALR	139.21-491.11	870.93-2527.28	422.0277	Normal	1,1	D218H
	PDIA3	112.25-284.62	409.28-1463.51	150.9959	Normal	1,1	-
	HLA-F	66.50-216.31	27.00-543.06	229.5595	UP	1,1	-
	TAPBP	34.23-50.92	95.72-479.00	94.6267	UP	1,1	-
	TAP1	23.87-50.15	29.79-392.57	119.3348	UP	1,1	-
	TAP2	15.03-34.60	9.48-83.18	17.5557	Normal	1,1	-
	HLA-G	0.10-0.90	0.37-12.76	0.2935	Normal	1,1	-
antigen processing presentation hlaII	CD74	215.40-1191.38	259.26-8881.38	1409.7586	UP	2,0	-
	HLA-DRA	50.08-346.06	58.26-1944.48	543.6026	UP	1,1	-
	HLA-DRB1	44.22-211.72	29.74-1327.32	56.6893	Normal	1,1	-
	HLA-DPB1	38.14-120.98	29.62-923.33	358.2771	UP	1,1	-
	HLA-DPA1	25.49-153.39	21.31-847.17	316.7573	UP	1,1	-
	HLA-DMA	18.09-70.84	13.87-356.99	85.0783	UP	1,1	-
	HLA-DRB5	4.65-161.16	8.35-513.99	16.7571	Normal	1,1	-
	HLA-DQB1	12.29-54.65	6.87-337.93	53.3596	Normal	1,1	-
	HLA-DMB	7.56-28.13	3.03-111.07	40.8756	UP	1,1	-
	HLA-DQA1	3.67-44.50	6.17-405.81	11.6069	Normal	1,1	-
	HLA-DOB	0.31-2.56	0.42-11.36	6.9341	UP	1,1	-
	HLA-DOA	1.10-17.67	3.23-229.32	71.4541	UP	1,1	-
	HLA-DQA2	0.03-11.14	0.49-95.15	22.3287	UP	1,1	-
hla expression regulation	RFXANK	49.38-90.16	37.14-128.19	36.2357	DOWN	1,1	-
	NFYC	39.73-59.19	29.38-77.22	65.3967	UP	2,1	-
	NFYB	17.80-49.01	10.78-32.54	34.8471	Normal	1,1	-
	RFX5	11.60-25.94	35.02-192.15	201.3611	UP	4,1	-
	CREB1	9.42-18.31	26.74-86.61	76.0672	UP	1,1	-
	NFYA	8.54-13.06	5.44-20.79	10.897	Normal	1,1	-
	RFXAP	1.83-7.41	1.14-8.42	5.5746	Normal	1,1	-
	CIITA	3.76-14.09	2.53-102.19	38.834	UP	1,1	-
immunoproteasome regulation	HSP90AB1	400.23-881.60	199.89-797.52	255.4216	DOWN	1,1	-
	HSPA8	163.67-610.86	264.02-1590.91	460.5813	Normal	1,1	-
	HSP90AA1	176.80-574.30	955.57-3683.52	935.0248	UP	1,1	-
	HSPA1A	104.20-1277.20	20.52-166.06	41.237	DOWN	1,1	-
	PSME1	153.45-231.83	121.33-428.93	155.4548	Normal	2,1	-
	HSPA1B	67.35-1697.40	22.40-135.39	15.5745	DOWN	1,1	-
	PSME2	104.99-236.69	134.88-684.18	257.1103	UP	2,1	-
	PSMF1	58.48-114.86	100.26-274.19	206.6131	UP	3,1	-
	POMP	34.64-56.97	37.73-114.11	40.6486	Normal	1,1	-
	PSME3	22.91-51.20	67.74-189.41	126.5609	UP	1,1	-
	HSPA4	23.41-53.78	37.45-94.31	155.9186	UP	2,0	-
	HSPA2	1.65-41.92	1.61-95.54	45.9859	UP	1,1	-
	HSPA1L	0.56-2.62	0.28-1.39	0.5239	DOWN	1,1	-
	HSPA6	0.57-11.99	0.75-19.63	4.9723	Normal	4,1	-
	IFNG	0.00-0.63	0.02-4.70	2.9106	UP	1,1	-
	TNF	0.00-0.33	0.17-7.68	0.8001	UP	1,1	-
immunoproteasome subunits	PSMB7	86.23-132.81	131.78-435.92	141.3196	UP	2,1	-
	PSMB6	52.45-139.22	85.64-284.83	90.3871	Normal	1,1	-
	PSMB5	55.11-92.51	84.60-221.83	81.0965	Normal	2,1	-
	PSMB8	42.10-70.19	33.48-285.26	58.7714	Normal	1,1	-
	PSMB10	26.53-65.96	13.30-119.86	9.8851	DOWN	1,1	-
	PSMB9	23.10-46.24	19.44-393.20	69.6451	UP	1,1	-

Table 1: Gene expression and mutations in presentation-machinery associated pathways. Genes highlithted in Red have potential defect in gene expression, mutations or CNA. Expression regulation is marked as **DOWN**, **UP** or **Normal** if sample-specific gene expression is lower than GTEx 5th percentile, higher than GTEx 95th percentile or in the GTEx 5th-95th percentile expression range, respectively.

6 Copy Number analysis

Copy number analysis of the sample **CESC1**. Genome-wide tumor allele frequency and tumor/germline sequencing depth ratio are shown in (Figure 1). Genome-wide estimated allele-specific copy number is shown in (Figure 2), annotated with key oncogenes amplifications, tumor supressor genes defects and APM defects. HLA allele specific alleles copy numbers, with their relative expression, as determined from RNAseq, are shown in (Figure 3).

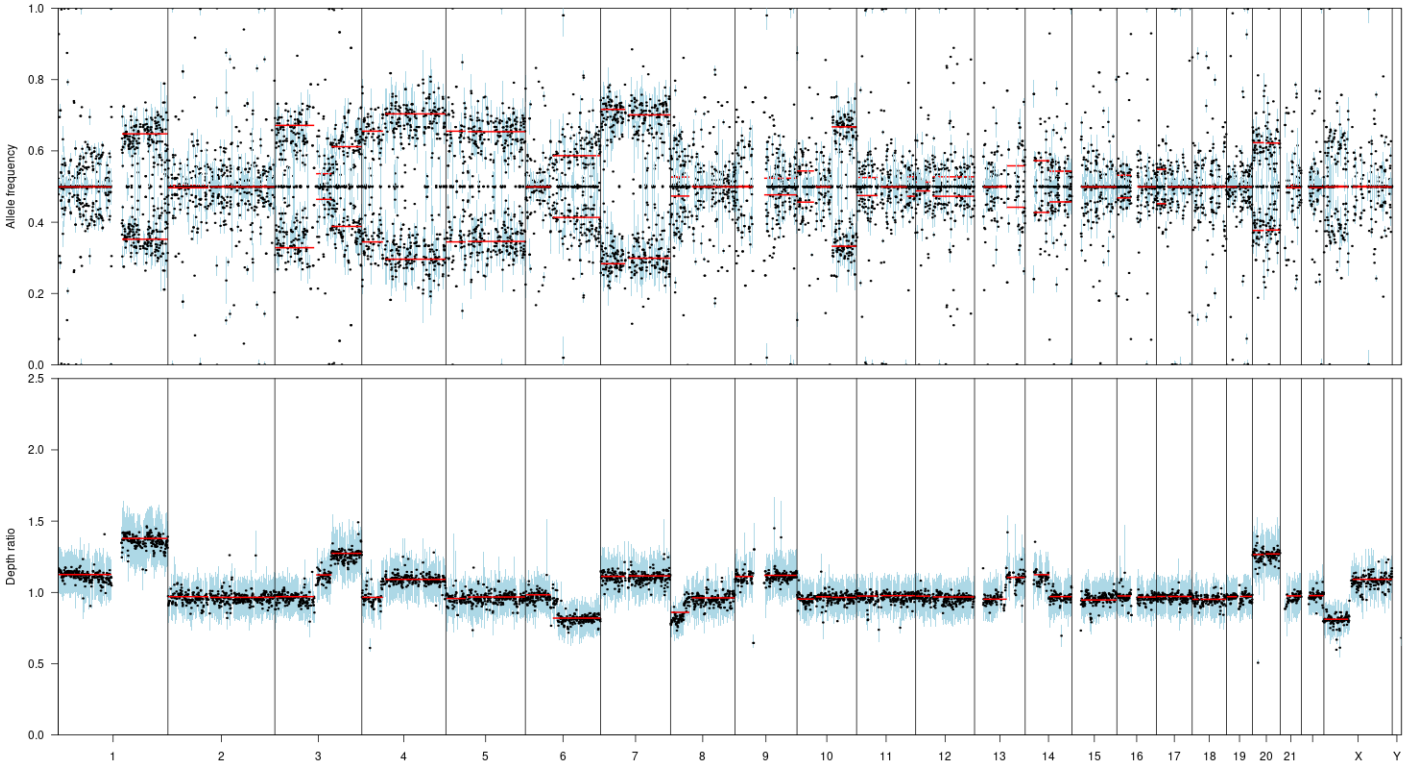


Figure 2: variant allele frequency and sequencing depth ratio

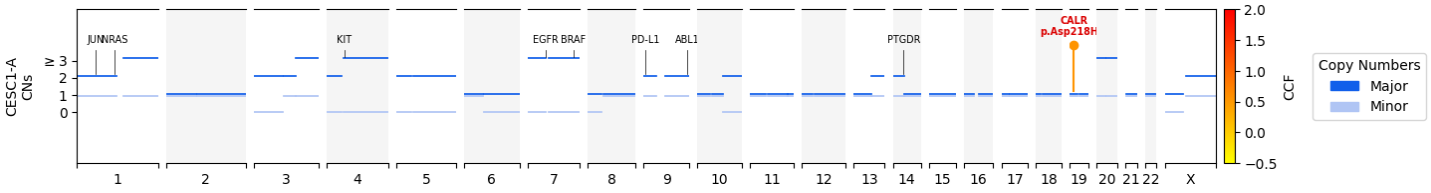


Figure 3: Tumor allele-specific copy number

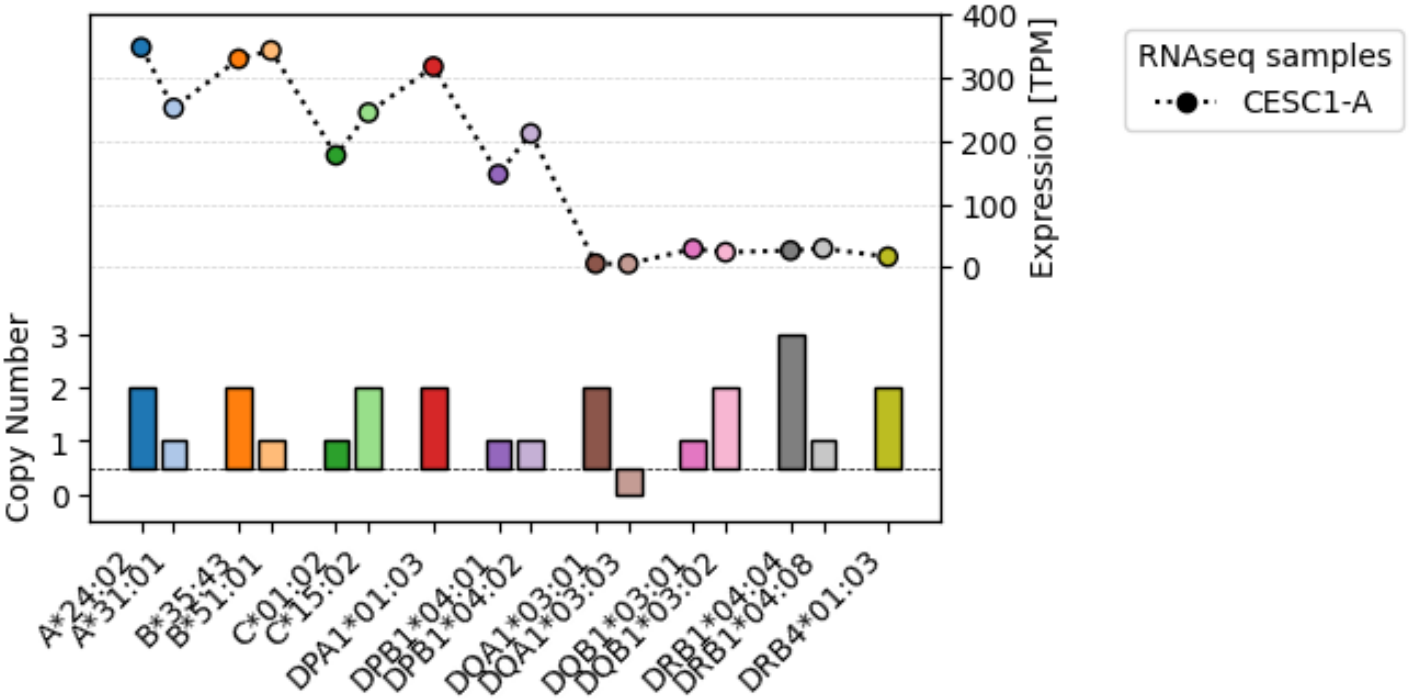


Figure 4: Tumor HLA allele-specific copy numbers

7 Variant data

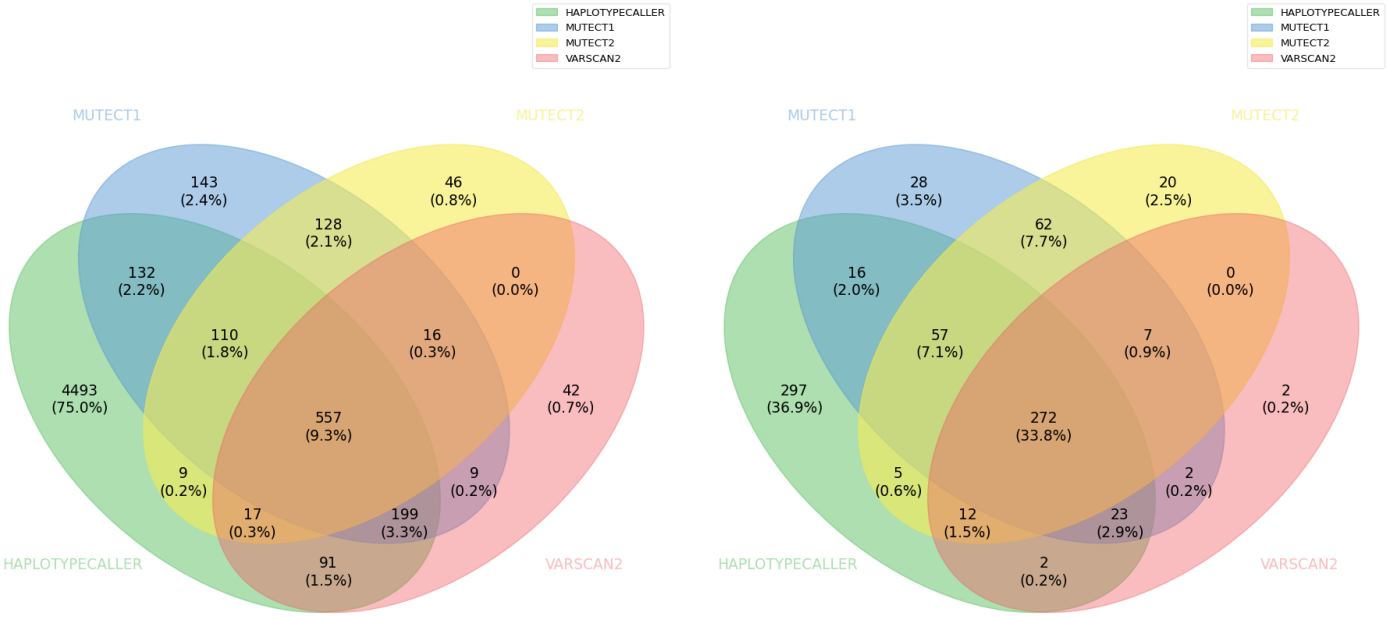
Variant calling information about the sample **CESC1**.

7.1 Variant caller overlap

Summary of the variant calls by the different callers and their overlap.

7.1.1 Somatic mutations (SMs)

Number of mutations affecting the protein sequence: **804** in total (union of variant callers mutations) and **458** passing neoDisc filters.
Venn diagrams of somatic mutations detected by all variant calling algorithms. Coding and non-coding refers to all somatic mutations called without regard to their effect at the protein level. Coding-only SMs are all SMs inducing an amino-acid change in a protein (Figure 3).



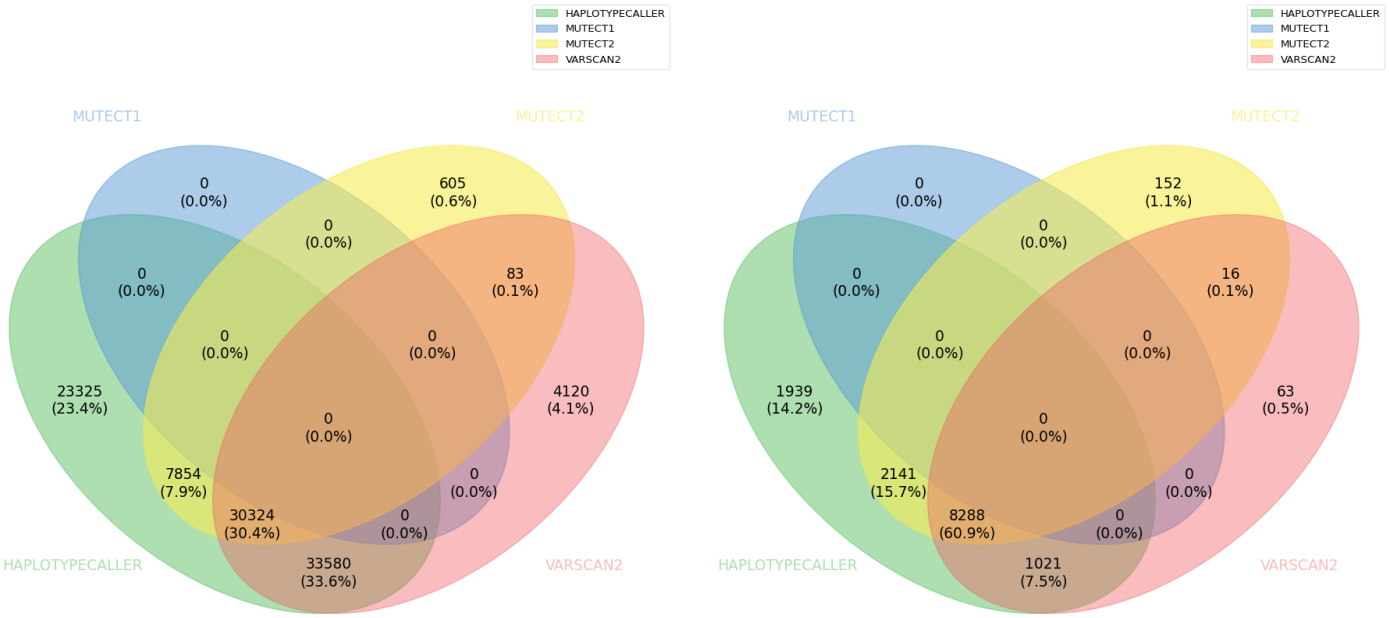
(a) Coding and non-coding SMs

(b) Coding-only SMs

Figure 5: Venn diagrams of all (a) and final (b) SMs

7.1.2 Single Nucleotide Polymorphisms (SNPs)

Venn diagrams of Single Nucleotide Polymorphisms detected by all variant calling algorithms. Coding and non-coding refers to all SNPs called without regard to their effect at the protein level. Coding-only SNPs are all SNPs linked to an SM and which induce an amino-acid change in a protein (Figure 4).



(a) Coding and non-coding SNPs

(b) Coding-only SNPs

Figure 6: Venn diagrams of all (a) and final (b) SNPs

7.2 Variant type distribution

Variant types are calculated as the fraction of the specific variant type over the total number of variants called. All the variants are taken into account. (Table 3)

Group	Nb of variants	Percentage
GLvar	4440	4.225
SM	5607	5.335
SNP	95051	90.440

Table 2: Variant types repartition table

7.3 Variant read statistics

The table shows the percentage of reads supporting the alternative variant for each somatic mutation in the tumor sample. Note that each variant is in a single category (e.g. a variant called by Mutect and Varscan doesn't count in the Varscan-only category). (Table 4)

Caller(s)	Nb of SMs	Mean	Median
HC—M1—M2—VS	557	17.2	15.6
HC—M1—M2	110	14.7	12.9
HC—M1—VS	199	20.9	18.8
HC—M2—VS	17	18.9	15.6
M1—M2—VS	16	11.9	11.3
HC—M1	132	20.5	17.7
HC—M2	9	19.0	17.5
HC—VS	79	31.7	30.8
M1—M2	128	8.5	8.5
M1—VS	9	14.4	13.3
HC	4116	49.1	42.9
M1	143	15.1	12.1
M2	46	9.5	7.4
VS	42	25.9	18.1

Table 3: Variant reads support statistics table

7.4 Mutational load

Mutational load was calculated as the total number of variants called and the size of the genome covered by the library preparation kit. See LB Alexandrov et al. Nature (2013) for more information. (Table 1). Driver mutations and mutations in driver genes were annotated from Intogen database (Table 2). Variants Cancer Cell Fraction (CCF) was calculated from the VAF, tumor content and chromosomes copy numbers. CCF represents the proportion of cancer cells that carry a variant, and is used to classify clonal and subclonal somatic mutations (clonal SM having a CCF + standard-deviation ≥ 0.95) (Figure 5).

Subset	SM / Mb
Union of all Callers	113.821
Intersection 2 or more Callers	25.497
Intersection All Callers	11.307

Table 4: Mutational load Table

gene	mutation	gene_driver_Intogen	mutation_driver_statement_Intogen
TJP1	p.His299Asp	TUMOR DRIVER	NA
ABCB1	p.Asp642Asn	OTHER TUMOR DRIVER	NA
BCLAF1	p.Ser718Arg	OTHER TUMOR DRIVER	NA
CALR	p.Asp218His	OTHER TUMOR DRIVER	NA
DDX3X	p.Arg376Cys	OTHER TUMOR DRIVER	NA
ERBB3	p.Gly419Ala	OTHER TUMOR DRIVER	NA
FAT1	p.His1342Tyr	OTHER TUMOR DRIVER	NA
FBXW7	p.Glu287Lys	OTHER TUMOR DRIVER	NA
FH	p.Ile262Leu	OTHER TUMOR DRIVER	NA
FRG1	p.Asn153Asp	OTHER TUMOR DRIVER	NA
FRG1	p.Asn25Asp	OTHER TUMOR DRIVER	NA
FXR1	p.Arg386Lys	OTHER TUMOR DRIVER	NA
FXR1	p.Arg415Lys	OTHER TUMOR DRIVER	NA
HERC2	p.Glu515Lys	OTHER TUMOR DRIVER	NA
KANSL1	p.Glu379Gln	OTHER TUMOR DRIVER	NA
KLF4	p.Arg299Trp	OTHER TUMOR DRIVER	NA
MAP3K1	p.Ser704Leu	OTHER TUMOR DRIVER	NA
MMP2	p.Glu525Gln	OTHER TUMOR DRIVER	NA
MUC4	p.His2573Asp	OTHER TUMOR DRIVER	NA
MUC4	p.Leu2977Ile	OTHER TUMOR DRIVER	NA
MUC4	p.Leu578Phe	OTHER TUMOR DRIVER	NA
MUC4	p.Ser1032Leu	OTHER TUMOR DRIVER	NA
MUC4	p.Ser1560Leu	OTHER TUMOR DRIVER	NA
MUC4	p.Ser941Leu	OTHER TUMOR DRIVER	NA
RAD50	p.Gly1199Val	OTHER TUMOR DRIVER	NA
STAG2	p.Glu444Lys	OTHER TUMOR DRIVER	NA
SVEP1	p.Asp476His	OTHER TUMOR DRIVER	NA
TNPO2	p.Met306Ile	OTHER TUMOR DRIVER	PREDICTED DRIVER
ZNF292	p.Ser1586Phe	OTHER TUMOR DRIVER	NA

Table 5: Mutations in Driver genes

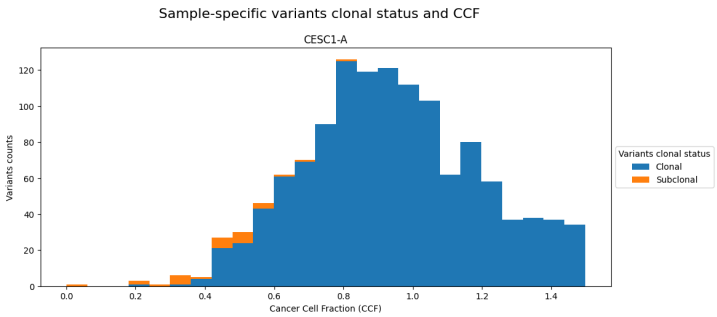


Figure 7: SM corrected VAF and clonal status CESC1-A

7.5 Expression of mutated genes based on GTEx and TCGA

A heatmap of the expression of genes mutated in the sample **CESC1** is shown in Figure 7. Gene expression data come from the GTEx database (Healthy RNAseq datasets). If available, genes are sorted by their expression from RNASEQ data, if not, by there expression in TCGA **Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma** cohort, in decreasing order. Only mutations passing our filters are used (mutations detected by a minimum of 2 variant callers).

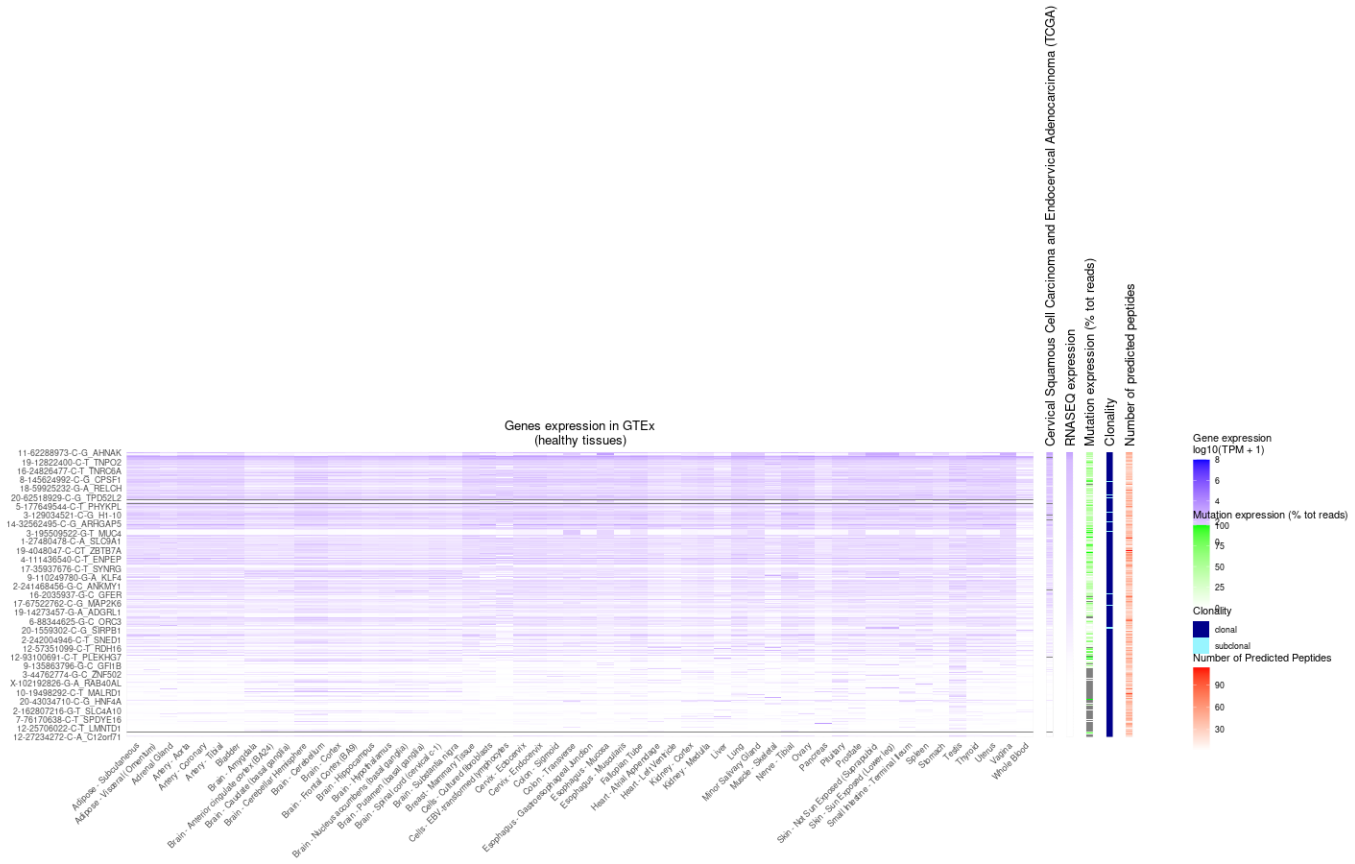


Figure 8: Heatmap - Gene expression in TPM

8 Inflammation Status

Inflammation status was calculated from RNASEQ data (if available), on each RNAseq sample. Genes were scored as described by Gajewski et al. across all samples in TCGA **Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma** cohort.

- **CESC1-A : T-cell-inflamed**

A heatmap of the inflammation score of all genes included in the Gajewski genes list, across all samples in TCGA **Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma** cohort is shown below. Patient sample(s) are shown at the top of the heatmap, above the inflammation status bar. (Figure 8)

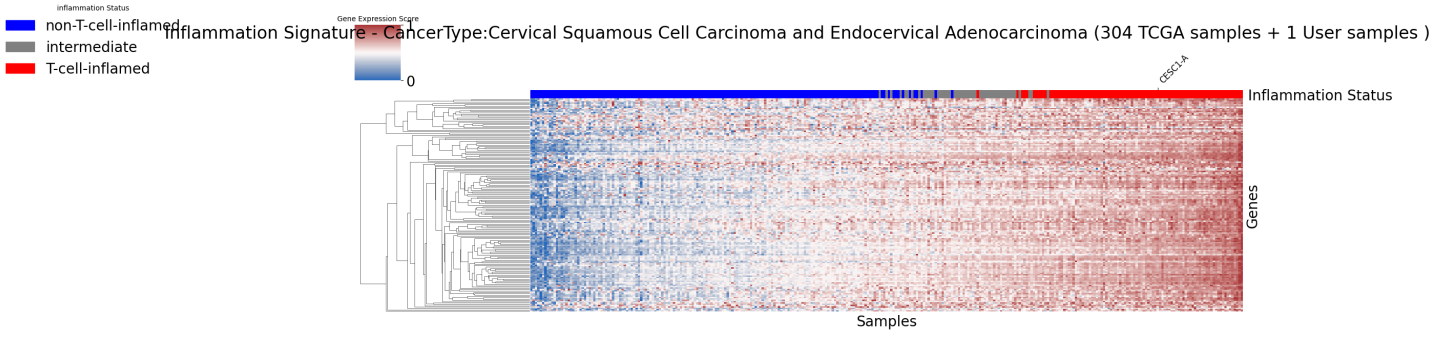


Figure 9: Heatmap - Inflamation status, relative to TCGA

9 Expressed TAAs

TAA genes expressed at RNASEQ level are reported in the table below. Only high quality TAAs are taken into account (no/very low expression of the gene in healthy tissues (GTEx) except in testis). For each identified TAA source gene, known (from litterature) immunogenic TAAs were compared to patient HLA alleles. The total number of TAAs tested immunogenic on sample's HLA alleles

is reported under textbfPeptides - Tested and TAAs predicted (were tested on alleles that the patient does not necessarly carry) to bind sample's HLA alleles are reported in textbfPeptides - Predicted.

Type	Gene	Expression (TPM)	Peptides - Litterature	Peptides - Predicted
HC-TSA	MLANA	1.0954	1	22
HC-TSA	MAGEA10	1.4704	0	131
HC-TSA	MAGEA11	3.4654	0	115
TAA	SLCO6A1	1.4648	0	183
TAA	POU4F1	1.0749	0	154
TAA	IQCM	1.1951	0	110
TAA	IL22RA2	1.6488	0	84
TAA	VCX	2.7823	0	28
TAA	GAGE2A	12.3771	0	7
TAA	GAGE1	3.8365	0	7
TAA	GAGE12J	2.4142	0	4
TAA	GNGT1	4.7597	0	1

Table 6: Expressed TAAs table

10 Virus infection

Viruses identified from RNAseq data are reported here. Evidence of virus infection can be: High confidence (HC) / Low confidence (LC). The virus score is a value calculated as: Sum of all virus CPMs, Normalized by the number of transcript per virus. (CPMs: counts per transcript, normalized by the median transcript length and multiplied by 1'000'000). The human endogenous retrovirus K113 is not reported here.
Viruses associated with cancer types are listed in the table below:

Cancer Type	Virus	Frequency
Cervical cancer	HPV	100%
Vulvar cancer	HPV	Frequent in youngers
Anal squamous neoplasia	HPV	Frequent in youngers
Head and Neck cancer	HPV	20%
Stomach cancer	EBV	20%
Nasopharyngeal carcinoma	EBV	Unknown
Merkel cell carcinoma	MCV	50%

Table 7: Virus - cancer types association table

Detected Viruses are reported in the table below, if any.

Virus	Score	Confidence
ref—NC_001357.1—Human_papillomavirus-18—gi—9626069	1.66e+02	HC
ref—NC_003663.2—Cowpox_virus—gi—30844336	1.08e-03	LC
ref—NC_045512.2—Wuhan_seafood_market_pneumonia_v...	1.01e-03	LC
ref—NC_012783.2—Cercopithecine_herpesvirus_5_str...	7.48e-04	LC
ref—NC_001731.1—Molluscum_contagiosum_virus_subt...	6.70e-04	LC
ref—NC_014649.1—Acanthamoeba_polyphaga_mimivirus...	2.93e-05	LC

Table 8: Virus identification table

11 Mass Spectrometry

Mass Spectrometry run date(s): **20210723**
HLA-I DDA raw files:

1) 20210723_CTE-BIO-19519-HLAIp_R01.raw
2) 20210723_CTE-BIO-19519-HLAIp_R02.raw

HLA-II DDA raw files:

1) 20210723_CTE-BIO-19519-HLAIIp_R01.raw
2) 20210723_CTE-BIO-19519-HLAIIp_R02.raw

HLA-I DIA raw files:

1) 20210723_DIA_CTE-BIO-19519-HLAIp_R03.raw
2) 20210723_DIA_CTE-BIO-19519-HLAIp_R04.raw

HLA-II DIA raw files:

1) 20210723_DIA_CTE-BIO-19519-HLAIIp_R03.raw
2) 20210723_DIA_CTE-BIO-19519-HLAIIp_R04.raw

11.1 MS statistics and Quality control data

Summary of peptidomics analysis. Comet was used for peptide search and NewAnce was used for the calculation of group-specific FDR. If available, FragPipe analysis was used for identification and mergeed with NewAnce results. Thresholds of 3% and 1% FDR were used for peptide selection with NewAnce and Fragpipe, respectively. The following table summarizes the results of the MS analysis.

All peptides (passing the FDR threshold and non-decoy) were taken into account for the Identifications, Lenghts, XCorr, TAAs, SNPs and SMs lines. The percentage of binders was calculated from the 9mers only for HLA-I peptides and from 15mers only for HLA-II peptides.
The histograms below show the distribution of peptides length for HLA-I and/or HLA-II identified peptides.
The pie charts summarize the distribution of HLA alleles binding the identified peptides. MixMHCpred and/or MixMHC2pred were used to assign to each peptide the best patient-specific HLA allele. Only high-confidence (rank <= 2) binders were selected for the plots. Venn diagrams show the overlap in peptide identifications.

	HLA-I				HLA-II			
	PC	**NC**	**NH**	**NEO**	**PC**	**NC**	**NH**	**NEO**
Identifications (tot)	23955	8	1	3	5952	0	0	1
Identifications (DDA and DIA)	18573	1	0	3	3918	0	0	0
Identifications (DDA Only)	2347	5	1	0	1781	0	0	1
Identifications (DIA Only)	3035	2	0	0	253	0	0	0
Lenghts (mean)	9.24	9.12	10	8.67	15.5	NA	NA	12
XCorr (mean)	2.11	2.32	1.78	1.85	2.84	NA	NA	1.25
PP Prob (mean)	0.98	0.99	NA	1.0	0.97	NA	NA	NA
Percentage binders	91.94	62.5	0.0	100.0	81.35	NA	NA	0.0
SNPs (tot)	125				145			
SMs (tot)	3				1			
TAAs (tot)	4				0			
Median gene expression	42.83				84.24			
Median peptide quantity	1.13E+06				3.98E+05			

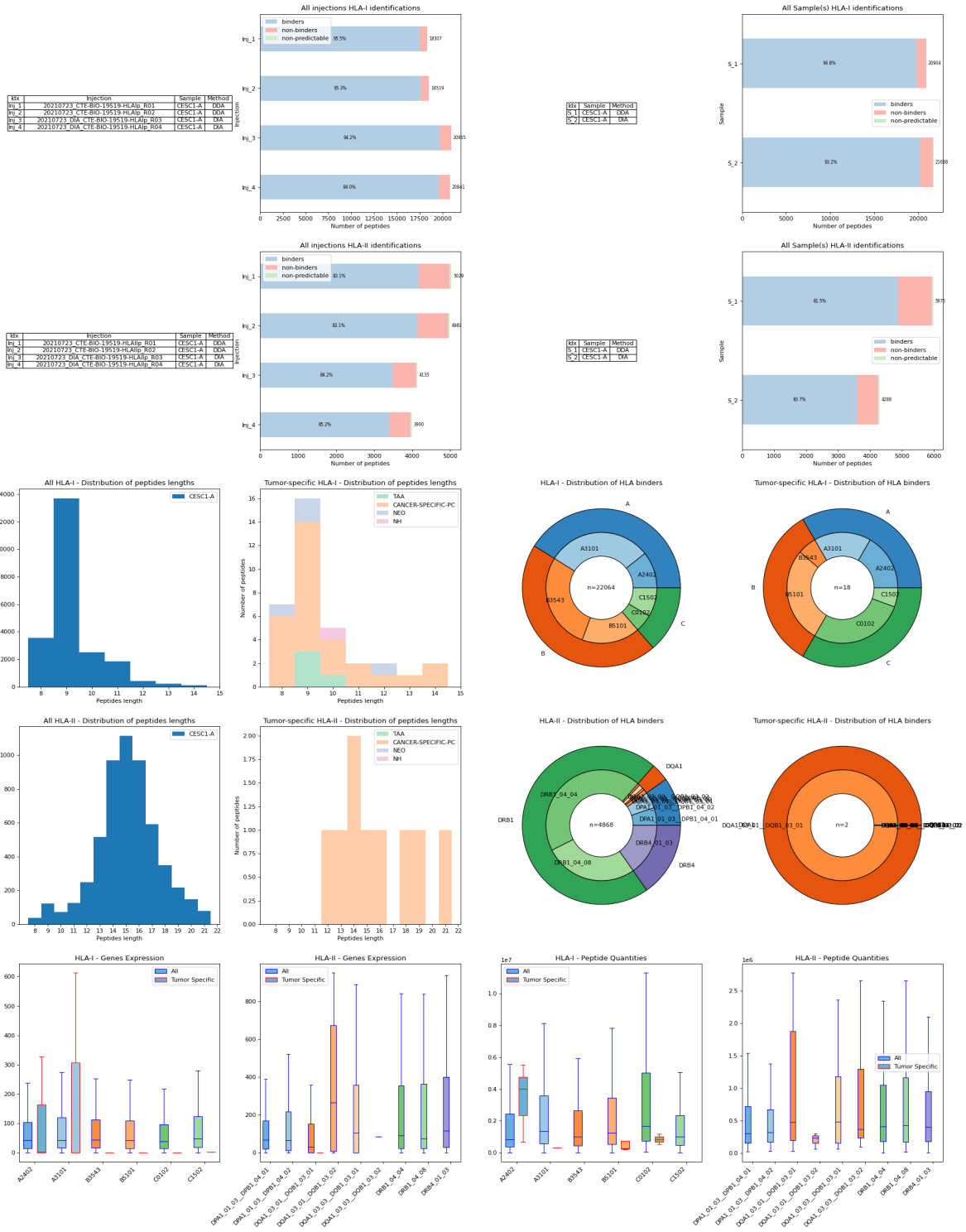
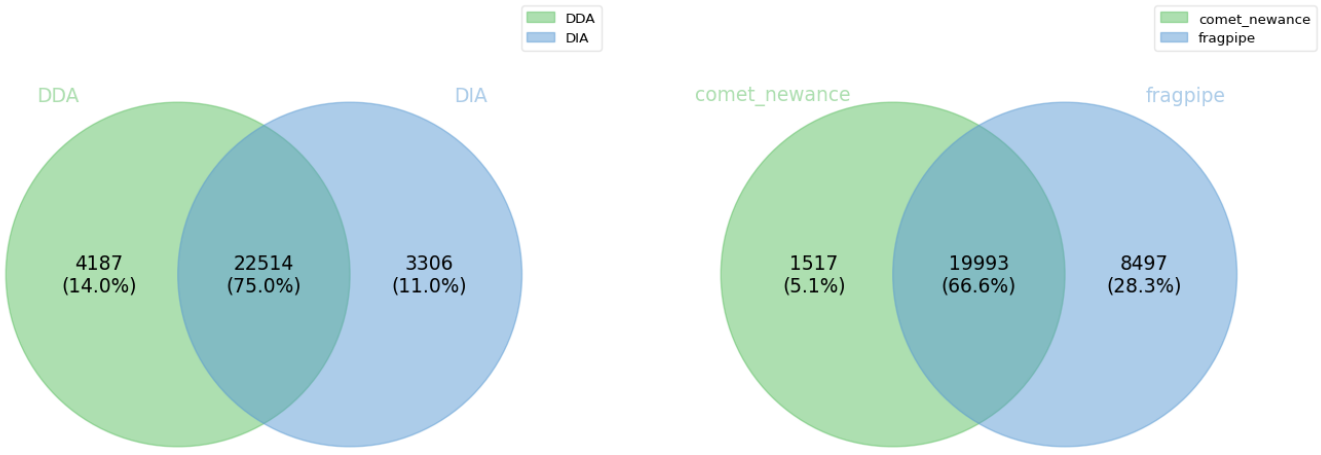

All injections HLA-I identifications
binders
non-binders
non-predictable
Number of peptides
All Sample(s) HLA-I identifications
Sample
All injections HLA-II identifications
All Sample(s) HLA-II identifications
All HLA-I - Distribution of peptides lengths
Tumor-specific HLA-I - Distribution of peptides lengths
HLA-I - Distribution of HLA binders
Tumor-specific HLA-I - Distribution of HLA binders
All HLA-II - Distribution of peptides lengths
Tumor-specific HLA-II - Distribution of peptides lengths
HLA-II - Distribution of HLA binders
Tumor-specific HLA-II - Distribution of HLA binders
HLA-I - Genes Expression
HLA-II - Genes Expression
HLA-I - Peptide Quantities
HLA-II - Peptide Quantities

Figure 10: MS



(a) Measurment Methods

(b) MS pipelines

Figure 11: Venn diagrams of measurment methods (a) and MS pipelines (b) peptide identifications overlap

All peptides (passing the FDR threshold and non-decoy) were taken into account for motif deconvolution. Peptides were separated between class-I and class-II and between canonical, non-canonical and viral for motif deconvolution, using MoDec. Upon deconvolution, each motif was assigned to a patient HLA (number of stars indicates confidence in motif attribution). The number of motifs was determined based on MoDec AIC value.

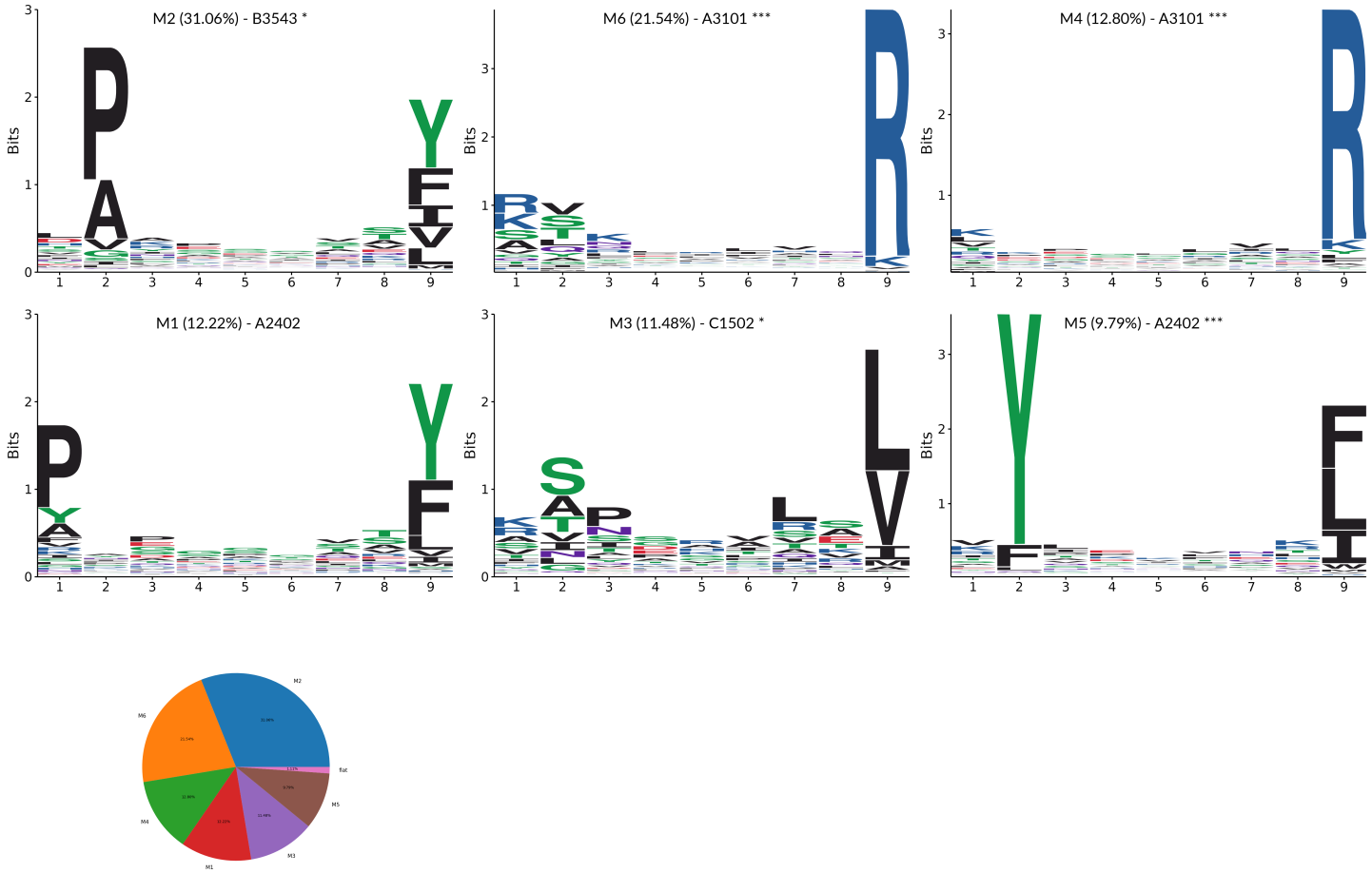


Figure 12: Canonical Class-I

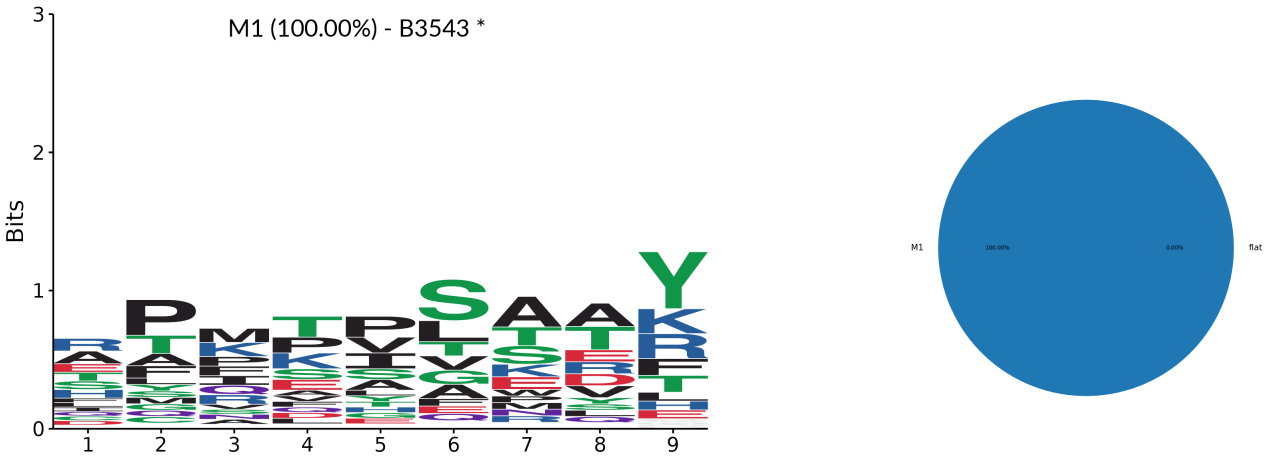


Figure 13: Non-Canonical Class-I

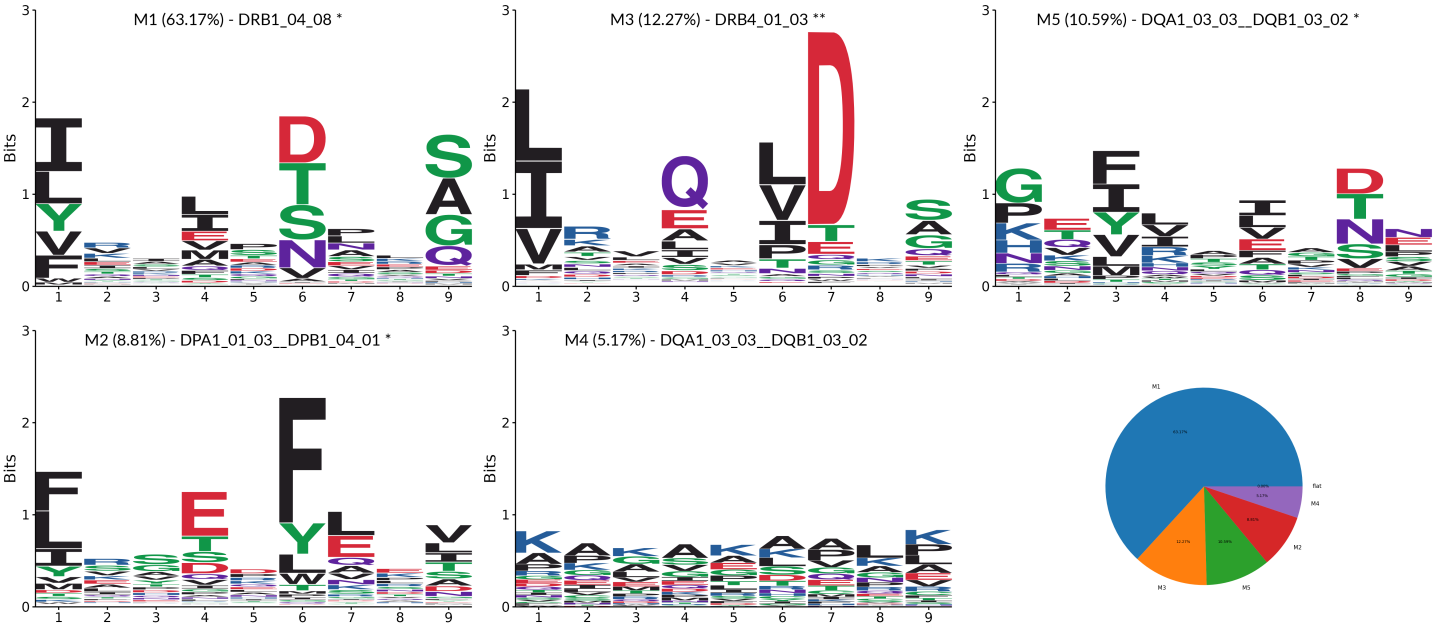


Figure 14: Canonical Class-II

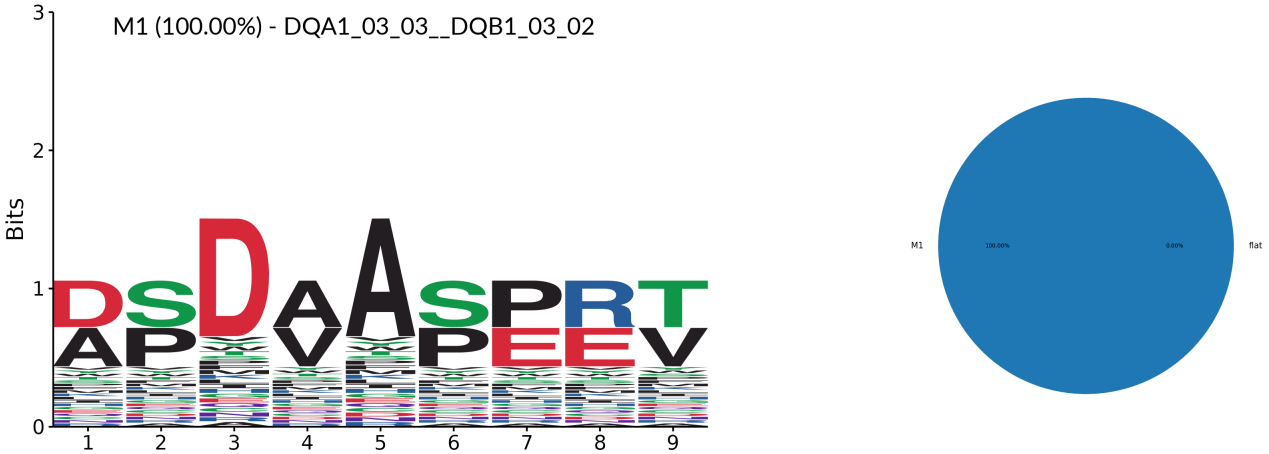


Figure 15: Non-Canonical Class-II

12 Files

The following files are attached to this report:

1. NeoDisc output tables
2. Long peptides plot