# TNBC report

Authors: Chantisa Keeratipusana and Harald Grove

Date: 2019/02/21

## Materials:

Six FFPE tumor samples were sequenced using whole exome sequencing (WES). Library preparation was done with Agilent SureSelect V5+UTR. There were no matched normal samples available. The six samples were selected based on the recommendation from Aj. Bhoom from a set of 61. The results from the remaining samples will be made available when the analysis is finished.

## Methods:

We followed Broadinstitute’s GATK pipeline for detection of somatic variants, using version GRCh38 of the human reference genome as the mapping reference. The best practices guideline was followed, except that Mutect2 was run in tumor only mode, using a panel-of-normals (PoN) and a list of known allele frequencies provided by Broadinstitute to estimate the probability of each variant being a germline variant. It was indicated that the PoN should ideally be from samples sequenced using the same protocol as the target, but that a generic PoN would still be better than nothing.

Each variant was annotated using VEP with version 95 of the Ensembl annotation database.

HLA alleles was estimated for each sample using the Kourami software and used as input for predicting the epitopes using the Pvactools software.

## Results:

The HLA-alleles of each sample is listed in table 1. For input into Pvactools, we only used the two first groups of numbers for each allele. We also noticed that one sample, BTN-075, was missing one Class II locus and was homozygous for all the remaining loci.

**Table 1: HLA typing results of each sample using the Kourami software.**

|  |  |  |
| --- | --- | --- |
| **Sample ID** | **HLA typing result** | |
| **MHC Class I** | **MHC Class II** |
| **BTN-075** | HLA-A\*24:02:01G,  HLA-B\*46:01:01G,  HLA-C\*01:02:01G | DQB1\*03:03:02G,  DRB1\*09:01:02G |
| **BTN-077** | HLA-A\*11:50Q,  HLA-A\*33:03:01G,  HLA-B\*18:01:01G,  HLA-B\*18:58,  HLA-C\*07:01:01G,  HLA-C\*07:04:01G | DQA1\*06:01:01G,  DQA1\*01:03:01G,  DQB1\*03:01:01G,  DQB1\*06:01:01G,  DRB1\*12:02:01,  DRB1\*15:02:01G |
| **BTN-080** | HLA-A\*11:121,  HLA-A\*02:07:01G,  HLA-B\*46:01:01G,  HLA-C\*01:02:01G,  HLA-C\*01:04 | DQA1\*01:01:01G,  DQA1\*03:01:01G,  DQB1\*05:02:01G,  DQB1\*03:02:01G,  DRB1\*04:05:01,  DRB1\*04:05:01 |
| **BTN-089** | HLA-A\*11:01:01G,  HLA-A\*11:50Q,  HLA-B\*40:01:01G,  HLA-C\*04:01:01G,  HLA-C\*07:02:01G | DQA1\*05:01:01G,  DQA1\*03:01:01G,  DQB1\*03:02:01G,  DQB1\*03:01:01G,  DRB1\*04:03:01,  DRB1\*11:01:01G |
| **BTN-091** | HLA-A\*02:03:01G,  HLA-A\*02:07:01G,  HLA-B\*46:01:01G,  HLA-B\*40:02:01G,  HLA-C\*01:02:01G,  HLA-C\*15:02:01G | DQA1\*01:01:01G,  DQA1\*03:01:01G,  DQB1\*05:03:01G,  DQB1\*03:03:02G,  DRB1\*09:01:02G,  DRB1\*15:02:01G |
| **BTN-092** | HLA-A\*11:50Q,  HLA-A\*31:01:02G,  HLA-B\*15:01:01G,  HLA-B\*55:02:01G,  HLA-C\*01:02:01G,  HLA-C\*03:03:01G | DQA1\*01:02:01G,  DQA1\*06:01:01G,  DQB1\*03:01:01G,  DQB1\*06:02:01G,  DRB1\*15:01:01G,  DRB1\*12:42 |

The Pvactools provides output for each combination of detected variant, gene transcript and assigned HLA-allele. This results file is then filtered down to the most likely epitopes, using the ic50 scores before reducing the number of output columns down the most relevant (see attached Excel file). The number of entries at each stage can be found in table 2.

**Table 2: Number of epitopes predicted by Pvactools for each sample.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample ID** | **Number of epitopes** | | |
| **All epitopes** | **Unique variants from total epitopes** | **Filtered epitopes** |
| BTN-075 | 257,716 | 11,807 | 327 |
| BTN-077 | 698,844 | 13,620 | 604 |
| BTN-080 | 622,885 | 14,168 | 601 |
| BTN-089 | 630,509 | 14,032 | 637 |
| BTN-091 | 760,596 | 14,622 | 650 |
| BTN-092 | 671,784 | 13,274 | 580 |

The condensed, filtered results from Pvactools were compared between samples, numbers of shared epitopes between each pair of samples are presented in table 3. The number of samples having the same predicted epitope are shown in table 4 for the most prevalent epitopes.

The same analysis was also performed using only the name of the affected gene, not including the position of the mutation (see attached Excel file).

**Table 3: Number of epitopes shared between each pair of samples.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | BTN-075 | BTN-077 | BTN-080 | BTN-089 | BTN-091 | BTN-092 |
| BTN-075 | 327 | 17 | 12 | 19 | 9 | 13 |
| BTN-077 | 17 | 603 | 18 | 26 | 13 | 24 |
| BTN-080 | 12 | 18 | 600 | 31 | 22 | 24 |
| BTN-089 | 19 | 26 | 31 | 634 | 27 | 32 |
| BTN-091 | 9 | 13 | 22 | 27 | 649 | 32 |
| BTN-092 | 13 | 24 | 24 | 32 | 32 | 579 |

**Table 4: The top 8 most commonly shared epitopes among the six samples.**

|  |  |  |
| --- | --- | --- |
| **Gene name** | **Mutation position (aa)** | **Number of samples** |
| HOXA4 | 140 | 6 |
| IER5 | 202 | 5 |
| IGFBP3 | 32 | 5 |
| BEST3 | 383 | 4 |
| EOMES | 120 | 4 |
| IRF2BP2 | 78 | 4 |
| MDGA1 | 947-948 | 4 |
| MMP17 | 20 | 4 |

## Supplementary file:

**pvactools\_condensed\_filtered.xlsx**

This file contains data gathered from the condensed and filtered output files from the six samples. Pvactools has used various filtering criteria, primarily predicted ic50 score, to determine the most likely antigens. The number of columns are also reduced, seeking to retain only the most important information.

Sheets:

“gene-pos\_count”: The final results from each sample presented as “gene-name\_position” (1st column) and the number of times this epitope was found among all samples (2nd column).

“gene-pos\_intersection”: Matrix showing the number of shared epitopes between each pair of samples.

“gene\_count” and “gene\_intersection”: As above, but only using the gene name and not the position of the mutation in the gene.

“BTN-0XX”: The final report from Pvactools of predicted epitopes for each sample.