

# Rat DCIS study

Anne Trinh

2021-02-18



# Contents

<b>Preface</b>	<b>7</b>
0.1 Packages and Software . . . . .	7
0.2 External software . . . . .	7
0.3 Annotations . . . . .	7
<b>1 Cohort summary</b>	<b>9</b>
1.1 Size information . . . . .	9
1.2 Calculating growth rates . . . . .	9
1.3 FACS data (DN/CD45/EpCAM) . . . . .	9
1.4 FACS data . . . . .	9
1.5 Summary Table . . . . .	9
<b>2 Whole-slide imaging</b>	<b>11</b>
2.1 Associate the frequencies with other data types . . . . .	11
2.2 Cellular composition . . . . .	11
2.3 Associate composition with other covariates . . . . .	11
2.4 Estimate tumor size . . . . .	11
2.5 Correlations between different subpopulations . . . . .	11
2.6 Associations between CD8 counts with other clinical variables . .	11
<b>3 Spatial statistics</b>	<b>13</b>
3.1 knn-Distances: . . . . .	13
3.2 The interacting fraction . . . . .	13
3.3 M-H distances . . . . .	13
3.4 Comparison between metrics . . . . .	13
3.5 Other cell types . . . . .	13
<b>4 Expression data</b>	<b>15</b>
4.1 Running alignment . . . . .	15
4.2 RNA Initial QC . . . . .	15
4.3 Normalisation . . . . .	15
4.4 Processing files for external software . . . . .	15
<b>5 RNA data: preliminary plots</b>	<b>17</b>

5.1	PCA plots . . . . .	17
5.2	Expression patterns by cell type . . . . .	17
<b>6</b>	<b>DESeq analysis: Progression/Immunotherapy cohort</b>	<b>19</b>
6.1	DN vs Ep . . . . .	19
6.2	No. samples in comparisons . . . . .	19
6.3	Set-up cell-type specific the comparisons . . . . .	19
6.4	PCA plots . . . . .	19
6.5	Pearson correlation plots of samples . . . . .	19
<b>7</b>	<b>Collating results and running GSEA</b>	<b>21</b>
<b>8</b>	<b>ER/Pgr Subtyping</b>	<b>23</b>
8.1	Gene List . . . . .	23
8.2	Progression cohort . . . . .	23
8.3	Characterisation cohort . . . . .	23
8.4	DN samples . . . . .	23
8.5	Comparison with staining . . . . .	23
8.6	Summary of expression markers for each subtype/cell fraction . .	23
<b>9</b>	<b>DESeq analysis: Immunotherapy/Growth comparisons</b>	<b>25</b>
9.1	Summary of comparisons . . . . .	25
9.2	DESeq: any treatment vs control . . . . .	25
9.3	Growing vs stable emphasis . . . . .	25
<b>10</b>	<b>Summary of GSEA runs</b>	<b>27</b>
10.1	Stable vs growing: all samples . . . . .	27
10.2	Comparisons based on treatment . . . . .	27
10.3	Pathways of Interest 2 . . . . .	27
<b>11</b>	<b>Epcam+ Inflammatory vs non-inflammatory samples</b>	<b>29</b>
11.1	Comparison of enriched pathways across samples . . . . .	29
11.2	DeSeq comparison . . . . .	29
11.3	Analyse the non-inflammatory samples . . . . .	29
11.4	Same analysis with only luminal samples . . . . .	29
<b>12</b>	<b>Associate Epcam+ inflammatory with survival</b>	<b>31</b>
12.1	Associating CD74 with phenotype and outcome . . . . .	31
12.2	Signature: Lum cases non-inflammatory: growing vs stable . . .	31
<b>13</b>	<b>DESeq analysis: Characterisation cohort (big vs small)</b>	<b>33</b>
13.1	CD45 samples . . . . .	33
13.2	Epithelial samples . . . . .	33
13.3	Check expression of checkpoint proteins . . . . .	33
<b>14</b>	<b>Signature analysis</b>	<b>35</b>
14.1	MHC signature analysis . . . . .	35

<b>15 Immune estimation</b>	<b>37</b>
15.1 Overview of the cell types . . . . .	37
15.2 characterisation cohort: assoc with size . . . . .	37
15.3 Comparison with FACS data . . . . .	37
15.4 Progression cohort . . . . .	37
15.5 Clinical associations . . . . .	37
15.6 Summary of the outcome . . . . .	37
<b>16 BCR clonotype analysis</b>	<b>39</b>
16.1 Processing summary . . . . .	39
16.2 Summary Stats . . . . .	39
16.3 Diversity metrics . . . . .	39
16.4 Compare the characterisation cohort . . . . .	39
16.5 Associate with clinicopathological data (progression) . . . . .	39
16.6 Associate with signature scores . . . . .	39
<b>17 Whole Genome Sequencing Mutations</b>	<b>41</b>
17.1 Data . . . . .	41
17.2 Plots . . . . .	41
17.3 Coding variants . . . . .	41
17.4 Overview of the mutations . . . . .	41
17.5 Metacore analysis of commonly mutated pathways . . . . .	41
<b>18 Mutations in RNA</b>	<b>43</b>
18.1 Haplotype caller . . . . .	43
18.2 Load files . . . . .	43
18.3 Identifying polymorphisms: compare frequencies in CD45 and Ep data . . . . .	43
18.4 Find coding mutations which are specific to ep cells . . . . .	43
18.5 Filtering WGS data based on CD45 data . . . . .	43
<b>19 Mutations in progression cohort</b>	<b>45</b>
19.1 Summary of common mutations: . . . . .	45
19.2 Look at the common mutations (cosmic ones) . . . . .	45
19.3 Also look at the top RNA-mutations with are Ep/DN specific . . . . .	45
19.4 Mutation load? . . . . .	45
<b>20 Trichrome staining</b>	<b>47</b>
20.1 Associations with cellular fraction (wsi) . . . . .	47
20.2 Associations with growth and spatial patterns . . . . .	47
<b>A Main Figures</b>	<b>49</b>
Figure 1 . . . . .	49
Figure 2: Immune system of NMU-rat . . . . .	49
Figure 3: Immunotherapy applied to NMU rats . . . . .	49
<b>B Extended Figures</b>	<b>51</b>



# Preface

Placeholder

## 0.1 Packages and Software

## 0.2 External software

## 0.3 Annotations

### 0.3.1 Genomic properties

### 0.3.2 Gene name homologs between organisms

### 0.3.3 Gene signatures and data-bases

#### 0.3.3.1 Human gene homologs

#### 0.3.3.2 GSEA compendiums





# Chapter 1

## Cohort summary

Placeholder

### 1.1 Size information

### 1.2 Calculating growth rates

### 1.3 FACS data (DN/CD45/EpCAM)

### 1.4 FACS data

### 1.5 Summary Table

#### 1.5.1 Compare the characterisation vs progression cohort

#### 1.5.2 Summary of the RNA data



## Chapter 2

# Whole-slide imaging

Placeholder

- 2.1 Associate the frequencies with other data types
- 2.2 Cellular composition
- 2.3 Associate composition with other covariates
- 2.4 Estimate tumor size
- 2.5 Correlations between different subpopulations
- 2.6 Associations between CD8 counts with other clinical variables



## Chapter 3

# Spatial statistics

Placeholder

### 3.1 knn-Distances:

3.1.1 Comparison to manual classification

3.1.2 Associations with outcome to treatment

3.1.3 Growth

### 3.2 The interacting fraction

3.2.1 Comparison to manual & select optimal r

3.2.2 Growth

3.2.3 Treatment

### 3.3 M-H distances

3.3.1 Comparison to Manual Scoring

3.3.2 Growth

3.3.3 Treatment

### 3.4 Comparison between metrics

### 3.5 Other cell types



## Chapter 4

# Expression data

Placeholder

### 4.1 Running alignment

### 4.2 RNA Initial QC

### 4.3 Normalisation

#### 4.3.1 preliminary visualisation (to remove outliers)

### 4.4 Processing files for external software





## Chapter 5

# RNA data: preliminary plots

Placeholder

### 5.1 PCA plots

### 5.2 Expression patterns by cell type



## Chapter 6

# DESeq analysis: Progression/Immunotherapy cohort

Placeholder

### 6.1 DN vs Ep

### 6.2 No. samples in comparisons

### 6.3 Set-up cell-type specific the comparisons

### 6.4 PCA plots

#### 6.4.1 EpCAM

#### 6.4.2 CD45

#### 6.4.3 DN

### 6.5 Pearson correlation plots of samples



## Chapter 7

# Collating results and running GSEA

Placeholder



## Chapter 8

# ER/Pgr Subtyping

Placeholder

### 8.1 Gene List

### 8.2 Progression cohort

### 8.3 Characterisation cohort

### 8.4 DN samples

### 8.5 Comparison with staining

### 8.6 Summary of expression markers for each subtype/cell fraction

#### 8.6.0.1 DN samples:

#### 8.6.0.2 CD45 samples:

#### 8.6.0.3 Epithelial samples:





## Chapter 9

# DESeq analysis: Immunotherapy/Growth comparisons

Placeholder

### 9.1 Summary of comparisons

### 9.2 DESeq: any treatment vs control

#### 9.2.1 EPCAM samples

### 9.3 Growing vs stable emphasis

#### 9.3.1 DN fraction

#### 9.3.2 CD fraction

#### 9.3.3 Ep Fraction



## Chapter 10

# Summary of GSEA runs

Placeholder

### 10.1 Stable vs growing: all samples

#### 10.1.1 barplots of enriched pathways

### 10.2 Comparisons based on treatment

### 10.3 Pathways of Interest 2



## Chapter 11

# Epcam+ Inflammatory vs non-inflammatory samples

Placeholder

### 11.1 Comparison of enriched pathways across samples

### 11.2 DeSeq comparison

#### 11.2.1 Compare inflammatory vs non-inflammatory

#### 11.2.2 Finding a signature for each branch

### 11.3 Analyse the non-inflammatory samples

#### 11.3.1 Non-inflammatory branch: Ep only samples

#### 11.3.2 GSEA

### 11.4 Same analysis with only luminal samples



## Chapter 12

# Associate Epcam+ inflammatory with survival

Placeholder

- 12.1 Associating CD74 with phenotype and outcome
- 12.2 Signature: Lum cases non-inflammatory:  
growing vs stable





## Chapter 13

# DESeq analysis: Characterisation cohort (big vs small)

Placeholder

### 13.1 CD45 samples

#### 13.1.1 PCA plot

#### 13.1.2 Differential Gene Expression

#### 13.1.3 GSEA

### 13.2 Epithelial samples

#### 13.2.1 PCA plot

#### 13.2.2 GSEA

### 13.3 Check expression of checkpoint proteins



## Chapter 14

# Signature analysis

Placeholder

### 14.1 MHC signature analysis

#### 14.1.1 MHC-I

#### 14.1.2 MHC-II

#### 14.1.3 MHC presentation proteins

#### 14.1.4 inflammation related genes: IL6-JAK-STAT and TNF/NFKB



## Chapter 15

# Immune estimation

Placeholder

**15.1 Overview of the cell types**

**15.2 characterisation cohort: assoc with size**

**15.3 Comparison with FACS data**

**15.4 Progression cohort**

**15.5 Clinical associations**

**15.5.1 Associate with Treatment**

**15.5.2 Association with Growth**

**15.6 Summary of the outcome**



## Chapter 16

# BCR clonotype analysis

Placeholder

**16.1 Processing summary**

**16.2 Summary Stats**

**16.3 Diversity metrics**

**16.4 Compare the characterisation cohort**

**16.5 Associate with clinicopathological data  
(progression)**

**16.6 Associate with signature scores**





## Chapter 17

# Whole Genome Sequencing Mutations

Placeholder

### 17.1 Data

17.1.1 Extract mutation signatures

17.1.2 Annotate the data with human common variants

### 17.2 Plots

17.2.1 Quick overview

17.2.2 Coding variants & TMB

### 17.3 Coding variants

17.3.1 Sites which are commonly mutated?

### 17.4 Overview of the mutations

17.4.1 All mutations

17.5 Metacore analysis of commonly mutated  
pathways



## Chapter 18

# Mutations in RNA

Placeholder

18.1 Haplotype caller

18.2 Load files

18.3 Identifying polymorphisms: compare frequencies in CD45 and Ep data

18.4 Find coding mutations which are specific to ep cells

18.5 Filtering WGS data based on CD45 data



## Chapter 19

# Mutations in progression cohort

Placeholder

- 19.1 Summary of common mutations:
- 19.2 Look at the common mutations (cosmic ones)
- 19.3 Also look at the top RNA-mutations with are Ep/DN specific
- 19.4 Mutation load?



## Chapter 20

# Trichrome staining

Placeholder

**20.1 Associations with cellular fraction (wsi)**

**20.2 Associations with growth and spatial patterns**

Here we have a links to the list of the figures in this study





## Appendix A

# Main Figures

### Figure 1

### Figure 2: Immune system of NMU-rat

Fig?? : GSEA of

### Figure 3: Immunotherapy applied to NMU rats

This figure concerns analyses run in @ref(#gPreface)

Fig?? : Growth rates of tumors Fig?? : Growth rates of tumors w.r.t. treatment

Fig?? : Histogram of growth rates of tumors Fig?? : Contingency table growth rate and treatment



## Appendix B

# Extended Figures

Ext Fig??