Sorbonne University

Master 2 BMC - Systems Immunology

Characterization of neutrophil populations associated with osteosarcoma and association with the tumor's inflammatory status

By

HUYNH Minh-Anh

September 2022

Table of Contents

1 Introduction **1**

2 Objective and experimental principle **4**

3 Methods **5**

3.1 Data collection 5

3.2 Normalization 5

3.3 Construction of gene signatures 5

3.4 MDS Clustering 6

3.5 Determination of cell population abundance 6

3.5.1 MCP Counter 6

3.5.2 Cybersort 6

3.6 Differential Expression Gene analysis 6

3.6.1 Differentially Expressed Genes (DEG) 6

3.6.2 Gene enrichment 6

3.7 Statistical Testing 6

4 Results **7**

4.1 Determination of inflammatory groups 7

4.1.1 MDS visualization and k-means clustering 7

4.2 Characterization of osteosarcomas associated to inflammatory status 8

4.3 Characterization of intra-tumor inflammation associated to inflammatory status . . . 10

5 Discussion **11**

6 Bibliography **12**

7 Appendix: code

List of Abbreviations

- DEG : Differentially Expressed Genes
- MDS : MultiDimensional Scaling
- scRNA-seq : single-cell RNA sequencing
- TANs: Tumor-associated Neutrophils
- TME: Tumor MicroEnvironment
- TPM : Transcripts Per Million
- UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction

Acknowledgements

Thank you for following this tutorial!

I hope you'll find it useful to write a very professional dissertation.

1 Introduction

- introduce the reader to the subject area and clarify the knowledge gap that the dissertation research will fill.
- set the context for the dissertation by reviewing the relevant literature.
- include relevant references to general (theoretical papers and reviews) and specific (specific to the particular question addressed) literature, to justify the research that has been undertaken and define the questions being addressed.
- state the primary research questions and hypotheses in the final paragraph.
- follow an 'inverted triangle' format, progressing from general scientific ideas and why they matter to the specific research questions addressed in the dissertation project.

The introduction should not be just a 'Literature Review.'

My 4.5 month research internship was carried out under the supervision of Dr. Dominique MODROWSKI at INSERM U1132 "BIOSCAR" directed by Dr. Martine COHEN-SOLAL, at the Lariboisière Hospital, a research unit that has been dedicated to the pathophysiology of bone and cartilage diseases. My work has focused on elucidating the mechanisms related to osteosarcoma and the inflammatory state, with particular emphasis on neutrophils.

Recently, immunotherapy has been particularly successful in treating previously difficult and lethal cancers, and was promoted as primary care therapy in many of them. However osteosarcomas have remained resilient to immunotherapy, instead relying on classical chemotherapy which is mildly effective and has plethora of side effects. In order to find a cure for difficult to treat cancers, better understanding of the tumor microenvironment (TME) could potentially lead to effective and novel targeted TME therapies.

Current progress in RNA sequencing in bulk RNA-Seq and single-cell RNA sequencing (scRNA-seq) has shown its potential in exploring the tumor microenvironment (TME), in order to explore intra-tumor heterogeneity and cellular dialogue between tumor cells and inflammatory cells.

Neutrophils are the first innate immune cells recruited during an inflammation and multiple papers have already elucidated that tumor-associated neutrophils (TANs) or circulating neutrophils are associated with worse patient survival therapy and chemoresistance (Faget et al., 2021; Long et al., 2021). Neutrophils support cancer development through 3 pathways : they are able to promote cancer initiation, assistance of metastasis and increase of tumor growth (Faget et al., 2021).

Neutrophils are first recruited to the TME, through CXCR2 ligands chemokines (CXCL1, CXCL2, CXCL5). Pro-tumor neutrophils support cancer cells initiation cells by releasing reactive oxygen species (ROS), RNS and proteases, and enabling a variety of action, such as angiogenesis, protecting tumor cells from antimicrobial factors via NETs, preventing CD8+ T cell activity through iNOS, arginase 1 secretion (Long et al., 2021), and facilitation of metastasis through NK cell suppression and escorting tumor cells.

Epidemiology Extraskeletal osteosarcoma accounts for < 1% of all soft tissue sarcomas and approximately 4% of all osteosarcomas . It typically arises during midlife and late adulthood, with most patients being in the fifth to seventh decades of life at diagnosis; occurrence in children is uncommon. Males may be affected more frequently than females (M:F ratio: 0.8-19:1) .

Conventional osteosarcomas (COSs) can arise in any bone, but the vast majority originate in the long bones of the extremities, most commonly in the distal femur (30%), followed by the proximal tibia (15%) and the proximal humerus (15%), i.e. sites of the most proliferative growth plates. In long bones, the tumour is usually metaphyseal (90%) and only infrequently develops in the diaphysis (9%) or rarely in the epiphysis. The jaws are the fourth most common site of origin . Involvement of the small bones of the extremities and multifocal osteosarcoma, either synchronous or metachronous, are rare, the latter representing metastatic spread rather than multiple independent primary tumours . Telangiectatic osteosarcomas (TAEOSs) also frequently develop around the knee (-60%) and in the proximal humerus (-20%) (116). They occur in the metaphysis, commonly with direct extension into the adjacent epiphysis and diaphysis. Small cell osteosarcoma (SCOS) has a similar disMbution but more commonly develops in the diaphysis of long bones (10-15%) .

syndrome, who have an increased incidenee of osteosarcoma. Patients with hereditary retinoblastoma also have a high risk of developing osteosarcoma , in particular after receiving ionizing radiation therapy. The genes causing these syndromes are also the most commonly mutated genes in sporadic osteosarcoma JP53 in > 90% and RB1 in as many as 56% of cases) . Germline mutations in various RECQ helicases underlie another group of rare syndromes associated with COS, including Bloom syndrome (BLM [FIECQL3]), Werner syndrome (WRN), and Rothmund—Thomson syndrome (RECQL4) . Acquiring chromosomal instability is also the hallmark of sporadic COS and probably the most crucial step for initiating and driving tumour development. Syndrome-related COSs have been recognized for a long time, but the increasing use of DNA sequencing for genotyping neoplasms and also the germline of individuals has identified pathogenic germline mutations in as many as 17.9% of COSs in larger

studies , a figure that is likely to increase in sequencing studies to come.

There is general consensus about recurring amplifications for some regions, such as gains of chromosome arms 6p (40- 50% of cases, harbouring RUNX2, VEGFA, E2F3, and CDC5L [CDC5]), 8q (45-55% of cases, harbouring MYC), and 17p, which have been detected by classic karyotyping and conventional comparative genomic hybridization, as well as deep sequencing. However, studies are difficult to compare because the definition of a recurrent alteration varies . The TP53 antagonist MDM2 is amplified in about 10% of cases, suggesting a pre-existing central low-grade osteosarcoma that underwent dedifferentiation in at least a subset of cases . FGFR1 amplifications have been demonstrated in 18.5% of cases; alterations in the IGF1R signalling pathway were observed 14% of COSs . Homozygous loss of CDKN2A occurs in 10% of COSs, is associated with an adverse outcome, and has been implicated

in osteosarcoma development from a mesenchymal progenitor . RB1 is deleted in about 50% of osteosarcomas . Other recurrently deleted genes include LSAMP, DLG2, and WWOX . Distinct patterns of large-scale transitions and loss of heterozygosity reminiscent of that seen in BRCA1I BFlCA2-deV have been identified in COS, suggesting a deficiency in homologous recombination repair (socalled BRCAness) . These findings indicate a potential sensitivity to poly (ADP-ribose) polymerase (PARP) inhibitors

Gene expression profiling demonstrated an association between macrophage expression profiles and lack of metastases, suggesting a beneficial effect of macrophage infiltration. These findings have also been confirmed by immunohistochemical analysis . A number of genes have been demonstrated to be hypermethylated in COS, affecting transcriptional activity HIC1, WIF1, PHLDA2 (TSSC3), RASSF1 (RASSF1A), GADD45. and RUNX2 {1571}. Methylation of ER (ESR1) seems particularly interesting, because this steroid receptor is involved in osteoblastic differentiation. Relieving ESR1 hypermethylation by DNA methyltransferases resulted in growth inhibition of tumour cells

COS has a broad immunoprofile that lacks diagnostic specificity. Commonly expressed antigens include SATB2, osteocalcin (BGLAP), osteonectin (SPARC), osteoprotegerin (TNFRSF11B), RUNX2, S100, actins, and CD99 . Importantly (because it is a diagnostic pitfall), osteosarcomas may also express keratin and EMA . Tumour cells are generally negative for CD31, CD45, and FOS, with FOS representing a relatively recent surrogate marker for the FOSgene rearrangements typically observed in osteoid osteoma and osteoblastoma (1012,91A,1739A). FOS immunohistochemistry might also be helpful in d ifferentiating osteoblastoma and osteoblastomalike osteosarcoma. TAEOS and

SCOS have an immunophenotype similar to that of COS. SATB2 is regarded as a very sensitive marker for osteoblastic differentiation but lacks specificity

2 Objective and experimental principle

Previously, the lab has demonstrated that tumor stem cells have specific properties inside the tumor, characterized by the calpaïne 6 biomarker. This cell is capable of coordinating invasion of distant tissues and confers to the whole tumor a specific phenotype.

The study aims to analyze public dataset and to make a custom analysis using NicheNet in order to understand the cellular crosstalk and interactions in the TME, between neutrophils and tumor cells, more specifically, calpaïne 6 tumor stem cells.

Collaborating with Dr Jean-Marc Schwartz's team in Manchester and a M2 in systems biology, bioinformatics, collecting data from available open-source studies, we will perform a customized downstream analysis on GSE87686 and TargetOS, PEMRBOSARC and a scRNAseq dataset in order to identify the relationships between neutrophils and osteosarcoma through the use of genetic signatures available from MSigDB and customized gene lists.

NicheNet to identify ligand activity and identify receptors, and predict their targeted consequences.

3 Methods

3.1 Data collection

Analyses were performed on two osteosarcoma bulk RNA-seq datasets, TARGET-OS Osteosarcoma and GSE87686. Data was collected from TARGET-OS using GDC Data Transfer Tool UI (v1.0.0), returning 19493 protein coding genes and 88 samples for TARGET-OS, containing both raw data and TPM data. GSE87686 data was obtained through the lab's previously pre-processed kallisto files, downloaded through SRA Run Selector to obtain SRA run files. Data was then imported from kallisto files via *tximport (v1.22.0)* R package.

Genes were converted to ENST and ENSG and finally HUGO gene symbols through *biomaRt (v2.50.3)*.

3.2 Normalization

Raw data was converted to TPM as it is the best performing normalization method than FPKM or RPKM, based on its perservation of biological signal as compared to other methods (Abrams et al., 2019). Calculation was performed using counts and lengths for each gene, returned from the tximport to kallisto process.

Z-score was used to normalize the data for each gene, in order to be able to visualize the data in corresponding heatmaps.

Z-score of the mean of tpm data was used to construct grouped heatmaps for a given gene or gene signature.

Inflammatory groups characterizing the intensity of inflammatory status in tumors were created by choosing the lowest and highest mean of Z-score of the Hallmark Inflammatory Response signature from MSigDB, containing 200 genes. ICAM4 was notably not present in the dataset in TARGET-OS cohort. The groups were cut off evenly using the *ntile* function in *dplyr (v1.0.8)* R package.

3.3 Construction of gene signatures

Gene signatures relevant to the research topic were obtained from MSigDB via *msigdbr* (v7.5.1)
HAY_BONE_MARROW_NEUTROPHIL.v7.5.1

Canonical markers for osteosarcoma markers were adapted from Zhou et al. (2020) 's analysisand

their canonical markers generated from the literature in the Supplementary Table 2.

3.4 MDS Clustering

3.5 Determination of cell population abundance

3.5.1 MCP Counter

Bulk RNA-seq cell populations can be estimated using Microenvironment Cell Populations-counter (MCP-counter) (Becht et al., 2016). Non-metric Kruskal-Wallis testing was used to determine significant differences between cell populations between Low, Medium and High groups with $P < 0.05$ were considered as statistically significant.

3.5.2 Cybersort

3.6 Differential Expression Gene analysis

3.6.1 Differentially Expressed Genes (DEG)

Using *DESeq2 (v1.34.0)* (Love et al., 2014), standard DEG pipeline using raw data was performed between inflammatory groups (Low, Medium, High). DEGs were identified with *adjusted p.value* < *0.05* and *log2FoldChange* > *1*.

3.6.2 Gene enrichment

enrichR v(3.0) was used (Xie et al., 2021; **R-enrichR?**) for functional gene enrichment/pathway analysis. Following database were queried : *"GO_Molecular_Function_2021," "Human_Gene_Atlas," "BioPlanet_2019," "GO_Biological_Process_2021," "GO_Cellular_Component_2021"*

3.7 Statistical Testing

Statistical testing was performed using *R software 4.1.2 (2021-11-01)*. Kruskal-Wallis testing was performed for non-metric comparative analysis between groups. Post-hoc analysis was performed using Dunn's test as opposed to Wilcoxon due to the test taking into account Kruskal-Wallis's rank. P < 0.05 and P.adj < 0.05 was considered statistically significant.

4 Results

Some more guidelines from the School of Geosciences.

This section should summarize the findings of the research referring to all figures, tables and statistical results (some of which may be placed in appendices).

- include the primary results, ordered logically - it is often useful to follow the same order as presented in the methods.
- alternatively, you may find that ordering the results from the most important to the least important works better for your project.
- data should only be presented in the main text once, either in tables or figures; if presented in figures, data can be tabulated in appendices and referred to at the appropriate point in the main text.

Often, it is recommended that you write the results section first, so that you can write the methods that are appropriate to describe the results presented. Then you can write the discussion next, then the introduction which includes the relevant literature for the scientific story that you are telling and finally the conclusions and abstract – this approach is called writing backwards.

4.1 Determination of inflammatory groups

4.1.1 MDS visualization and k-means clustering

Functional clusters of osteosarcoma samples were created using k-means clustering of an MDS visualization of the Hallmark Inflammatory Response signature.

```
## Warning in as_grob.default(plot): Cannot convert object of class
## knit_image_pathsknit_asis into a grob.
```

```
## Warning in as_grob.default(plot): Cannot convert object of class
## knit_image_pathsknit_asis into a grob.
```

Inflammatory groups characterizing the intensity of inflammatory status in tumors were created by choosing the lowest and highest mean of Z-score of the hallmark inflammatory response signature

Figure 1: MDS of Hallmark Inflammatory Signature

from MSigDB, containing 200 genes. ICAM4 was notably not present in the dataset in TARGET-OS cohort. The groups were cut off evenly using the *ntile* function in *dplyr* R package.

Those manually defined groups are relevant as they correspond fairly well to functional groups, defined by k-means clustering based on MDS visualization **(Figure 2)**. Each sample is thus attributed to its inflammatory status and this group will be subsequently used for the following results.

4.2 Characterization of osteosarcomas associated to inflammatory status

Comparison of the mean of *Hp Osteosarcoma* gene signature and gene relating to types of osteosarcomas associated to inflammatory groups were performed, represented through a heatmap. Despite high heterogeneity between samples and inflammatory status, the Z-score of the mean of genes in Low versus High group is statistically significatively different.

However *Hp Osteosarcoma* in GSEA is not significatively differrent ($p = 0.163$).

Comparison of the expression of specific osteosarcoma markers relating to osteoblastic, chondroblastic, fibroblastic markers through a heatmap representation. Hierarchical clustering of the samples does not appear to be associated with corresponding inflammatory status. However it does reveal that there are groups of osteoblastic, chondroblastic and fibroblastic osteosarcomas which is expected.

The mean of markers of proliferation (MKI67, PCNA, TOP2A) associated to osteosarcomas have been compared to inflammatory status, along with the mean of the three markers. Kruskal-Wallis testing is significant ($p = 0.00968$) and post-hoc Dunn analysis reveals that the mean of the proliferation markers between low and high group is significatively different ($p = 0.016$). The data suggests that proliferation is hindered when inflammatory status is high in the osteosarcoma samples.

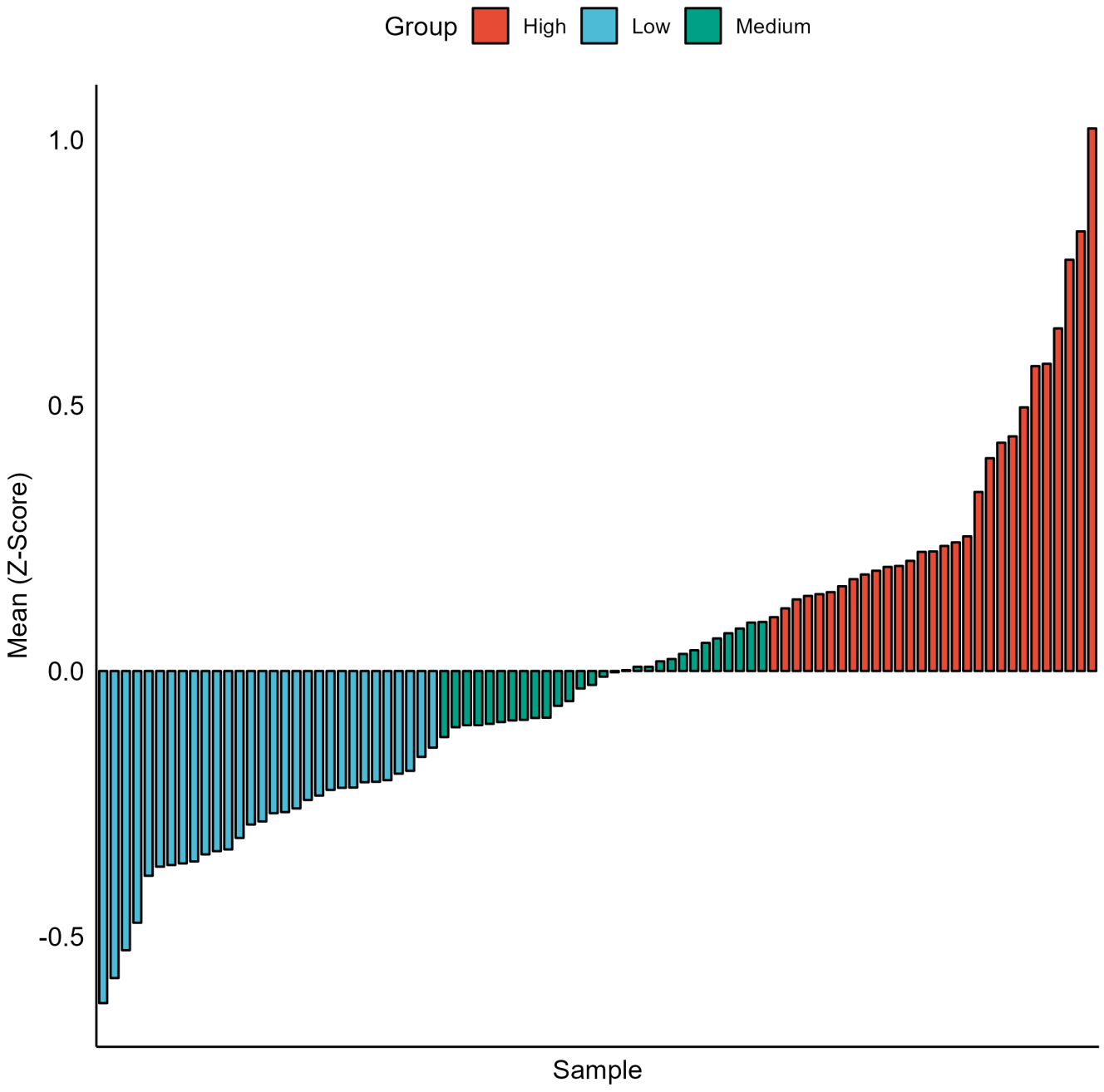

Figure 2: Histogram

4.3 Characterization of intra-tumor inflammation associated to inflammatory status

5 Discussion

the purpose of the discussion is to summarise your major findings and place them in the context of the current state of knowledge in the literature. When you discuss your own work and that of others, back up your statements with evidence and citations.

- The first part of the discussion should contain a summary of your major findings (usually 2 – 4 points) and a brief summary of the implications of your findings. Ideally, it should make reference to whether you found support for your hypotheses or answered your questions that were placed at the end of the introduction.
- The following paragraphs will then usually describe each of these findings in greater detail, making reference to previous studies.
- Often the discussion will include one or a few paragraphs describing the limitations of your study and the potential for future research.
- Subheadings within the discussion can be useful for orienting the reader to the major themes that are addressed.

6 Bibliography

Abrams, Z.B., Johnson, T.S., Huang, K., Payne, P.R.O., and Coombes, K. (2019). A protocol to evaluate RNA sequencing normalization methods. BMC Bioinformatics *20*, 679.

Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biology *17*, 218.

Faget, J., Peters, S., Quantin, X., Meylan, E., and Bonnefoy, N. (2021). Neutrophils in the era of immune checkpoint blockade. Journal for Immunotherapy of Cancer *9*, e002242.

Long, W., Chen, J., Gao, C., Lin, Z., Xie, X., and Dai, H. (2021). Brief review on the roles of neutrophils in cancer development. Journal of Leukocyte Biology *109*, 407–413.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550.

Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Current Protocols *1*, e90.

Zhou, Y., Yang, D., Yang, Q., Lv, X., Huang, W., Zhou, Z., Wang, Y., Zhang, Z., Yuan, T., Ding, X., et al. (2020). Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. Nature Communications *11*, 6322.

7 Appendix: code

Analyses were conducted using the R Statistical language (version 4.1.2; R Core Team, 2021) on Windows 10 x64 (build 22000). Code is available at _https://github.com/Minh-AnhHuynh/Osteosarcoma-Project_.

Code philosophy follows “The tidyverse style guide” written by Hadley Wickham. using *styler (v1.7.0)* to restyle code, and tidyverse environment as much as possible, substituting base R functions to their *dplyr* and *purrr* equivalent in order to ensure consistent code readability, usage and naming.

```
## - Session info ---------------------------------------------------------------
##  setting  value
##  version  R version 4.1.2 (2021-11-01)
##  os       Windows 10 x64 (build 22000)
##  system   x86_64, mingw32
##  ui       RTerm
##  language (EN)
##  collate  French_France.1252
##  ctype    French_France.1252
##  tz       Europe/Paris
##  date     2022-06-07
##  pandoc   2.14.0.3 @ C:/Program Files/RStudio/bin/pandoc/ (via rmarkdown)
##
## - Packages -------------------------------------------------------------------
##  package   * version date (UTC) lib source
##  knitr     * 1.39    2022-04-26 [1] CRAN (R 4.1.3)
##  librarian * 1.8.1   2021-07-12 [1] CRAN (R 4.1.3)
##
##  [1] C:/Users/Minh-Anh/Documents/R/win-library/4.1
##  [2] C:/Program Files/R/R-4.1.2/library
##
## ------------------------------------------------------------------------------
```

package	version	source
knitr	1.39	CRAN (R 4.1.3)
librarian	1.8.1	CRAN (R 4.1.3)

```
## Analyses were conducted using the R Statistical language (version 4.1.2; R Core Team,
##
## References
## ----------
##   - Desi Quintans (2021). librarian: Install, Update, Load Packages from CRAN, 'GitHu
##   - R Core Team (2021). R: A language and environment for statistical computing. R Fo
##   - Yihui Xie (2022). knitr: A General-Purpose Package for Dynamic Report Generation
```

Summary

English Summary