

Contents

Summary	ii
Introduction	1
Human Mesenchymal Stem/Stromal Cells	2
Multiple Myeloma	3
Myeloma-hMSC Interactions	3
Myeloma Bone Disease	3
Dissemination of Myeloma Cells	4
Multidimensional Data in Biomedical Research	5
Nontransparencies in Biomedical Data Analyses	6
Semi-Big Data: Big Enough to Cause Problems	7
The Shortcomings of Common Biomedical Analysis Tools	8
Modern Standards of Software Development	10
What makes Python an “Easy” Programming Language?	12
The Potential of Python Data Science Packages for Biomedicine	16
Aims	20
Chapter 1: Modelling Myeloma Dissemination <i>in vitro</i>	21
Abstract	21
Introduction	22
Materials and Methods	24
Results	28
Discussion	43
Chapter 2: Semi-Automating Data Analysis with <code>plotastic</code>	47
Abstract	47
Introduction	48
Statement of Need	50
Example	51
Overview	53
Discussion	55
Summarising Discussion	58
Time Lapse	58
Myeloma	58
Semi-Automated Analysis Improves Agility During Establishing new <i>in vitro</i> Methods	58
Plotastic and Jupyter Notebooks as a Standard for Research Documentation	59
How Biosciences can Benefit from Standards of Software Development	59
References	60
Appendices	73
A Supplementary Data & Methods	73
A.1 Figures	73
A.2 Tables	90
A.3 Materials & Methods	98
B Documentation of <code>plotastic</code>	115

B.1	Class Diagram	116
B.2	Readme	118
B.3	Example Analysis “qpcr”	131
C	Submission Forms & Documents	138
C.1	Author Contributions	138
C.2	Affidavit	145
C.3	Curriculum Vitae	147

Introduction

To provide a comprehensive background for the following chapters that focus on the interaction of human mesenchymal stromal cells (hMSCs) with multiple myeloma (MM) cells, this

Aims

This project defines these aims:

- Characterize the interaction between myeloma cells and mesenchymal stromal cells
- Develop methods
- Face the challenge of time-dependent cell adhesion through
- Provide tools to analyze multidimensional cell adhesion data

Chapter 2: Semi-Automating Data Analysis with `plotastic`

Abstract

`plotastic` addresses the challenges of transitioning from exploratory data analysis to hypothesis testing in Python's data science ecosystem. Bridging the gap between `seaborn` and `pingouin`, this library offers a unified environment for plotting and statistical analysis. It simplifies the workflow with user-friendly syntax and seamless integration with familiar `seaborn` parameters (`y`, `x`, `hue`, `row`, `col`). Inspired by `seaborn`'s consistency, `plotastic` utilizes a `DataAnalysis` object to intelligently pass parameters to `pingouin` statistical functions. Hence, statistics and plotting are performed on the same set of parameters, so that the strength of `seaborn` in visualizing multidimensional data is extended onto statistical analysis. In essence, `plotastic` translates `seaborn` parameters into statistical terms, configures statistical protocols based on intuitive plotting syntax and returns a `matplotlib` figure with known customization options and more. This approach streamlines data analysis, allowing researchers to focus on correct statistical testing and less about specific syntax and implementations.

Introduction

The reproducibility crisis in research highlights a significant challenge in contemporary biosciences, where a substantial portion of studies faces reproducibility issues (Baker, 2016; Begley & Ioannidis, 2015; Gosselin, 2021). One critical yet often overlooked aspect contributing to this crisis is data management. The literature most often refers to *big data* as the main challenge (Gomez-Cabrero et al., 2014). However, these challenges are also present in smaller datasets, which the author refers to as *semi-big data*. This term describes datasets that – while not extensive enough to necessitate advanced computational tools typically reserved for *big data* – are sufficiently large to render manual analysis very time-intensive. Semi-big data is often generated by methods like automated microscopy or multiplex qPCR, which produce volumes of data that are manageable on a surface level, but pose substantial barriers for in-depth, manual reproducibility (Bustin, 2014; Incerti et al., 2019). This is further complicated by the complexity inherent in multidimensional datasets (Krzywinski & Savig, 2013): Modern biosciences describe processes (e.g. cell-adhesion) that are highly dependent on multiple experimental parameters (factors), like ‘time’ or ‘kinds of treatments’ (Rebl et al., 2010; McKay et al., 1997). Manually grouping the data by multiple factors (facetting) is challenging and error-prone, especially when the data is not structured in a way that is immediately compatible with statistical tests. Without a clearly documented data analysis protocol and standardized data formats, analysis of multidimensional data becomes nontransparent and too overwhelming for reproduction (Bustin, 2014).

The evolving standards in data analysis advocate for the standardization of analytical pipelines, rationalization of sample sizes, and enhanced infrastructure for data storage, addressing some of these challenges (Goodman et al., 2016; Wilkinson et al., 2016). However, these advancements can place undue pressure on researchers, particularly those with limited training in statistics, underscoring the need for intuitive, user-friendly analytical tools (Federer et al., 2016; Lakhlifi et al., 2023; Armstrong, 2014; Gómez-López et al., 2019; Leek & Peng, 2015).

In this context, `plotastic` emerges as a tool designed to democratize access to sophisticated statistical analysis, offering a user-centric interface that caters to researchers across varying levels of statistical proficiency. `plotastic` simplifies inferential statistics based on the idea that statistical analyses are often performed based on how the data is visualized. This principle is not only intuitive but also statistically sound, because the parameters that structure the figure (e.g. facetting) are often times re-used for statistical testing (e.g. independent variables or factors). By integrating robust statistical methodologies within an accessible framework, `plotastic` could contribute to enhancing the reproducibility of research in the biosciences (Gomez-Cabrero et al., 2014).

The user-centric approach of plotastic distinguishes itself from the fully automated pipelines used for big data, which are designed to handle extensive computational tasks. Instead, plotastic focuses on ease-of-use and structures its commands to enable an interactive review of intermediate outputs, a concept the author refers to as *semi-automation* (Tab. 1).

Table 1: Key Principles of Semi-Automation and their Implementation in Plotastic

No.	Principle	Implementation in plotastic
1	Standardized input: The data to-be-analyzed follows a strict standardized format. The user should be able to convert their data into that format.	Long-format pandas DataFrames are used as input
2	Automation over flexibility: If there is an obvious way to do things, automate it and minimize user input. User options should be added with good reason. Avoid situations where the user is asked to pass the same parameter twice. This reduces the risk of human error, confusion and time spent on configuration.	E.g. passing the parameter “subject” once makes the rest of the pipeline switch automatically to the paired versions of statistical tests.
3	Out of the box functionality: The software’s default configuration should provide acceptable (but potentially sub-optimal) results. Beginners should be invited to experiment without the need to learn custom configurations. Options are still available to allow feature-rich adaptations according to the needs of both data and user.	Default tests are standard unpaired t-tests and ANOVA
4	Focus on intermediate outputs: The user composes the analysis pipeline using smaller commands that are each designed to provide human-readable output of an intermediate result. Each step is a stage to control quality, allowing quick error detection and troubleshooting.	Processing steps are separated into main steps: assumption tests, factor analysis, post-hoc analysis and plotting
5	Highly useful error messages: Never leave the user hanging. Tell him what went wrong <i>and</i> what the software was expecting.	E.g.: <code>ValueError: User passed 'subect' as subject, please choose one of ['subject', 'event', 'region']</code>

The need for plotastic in this specific project arose from two main challenges (for further details, see summarizing discussion). The first is the author’s need for a tool that could handle the complex, multidimensional data generated by e.g. qPCR experiments. These experiments involved the analysis of multiple outcomes across multiple genes, timepoints, method variations, cell-types, biological replicates, technical replicates etc., resulting in datasets that are challenging to analyse manually. Such complexity was necessary, since establishing new methods required extensive controls and creative variation of the experimental setup. Data analysis had to be automated somehow, since the lab-work itself was already time-intensive. The second challenge was to accept the potential of plot-configured statistical analyses. The author believes that the way data is visualized is often the way it should be analyzed. This vision is not limited to biomedical application, but a general principle that could benefit the scientific community overall. Making plotastic a generalized tool was a conscious decision to maximize its adoption rate and ensure its long-term relevance and quality, of which biomedical research will also benefit.

Statement of Need

Python’s data science ecosystem provides powerful tools for both visualization and statistical testing. However, the transition from exploratory data analysis to hypothesis testing can be cumbersome, requiring users to switch between libraries and adapt to different syntaxes. `seaborn` has become a popular choice for plotting in Python, offering an intuitive interface. Its statistical functionality focuses on descriptive plots and bootstrapped confidence intervals (Waskom, 2021). The library `pingouin` offers an extensive set of statistical tests, but it lacks integration with common plotting capabilities (Vallat, 2018). `statannotations` integrates statistical testing with plot annotations, but uses a complex interface and is limited to pairwise comparisons (Charlier et al., 2022).

`plotastic` addresses this gap by offering a unified environment for plotting and statistical analysis. With an emphasis on user-friendly syntax and integration of familiar `seaborn` parameters, it simplifies the process for users already comfortable with `seaborn`. The library ensures a smooth workflow, from data import to hypothesis testing and visualization.

Example

The following code demonstrates how plotastic analyzes the example dataset “fmri”, similar to Waskom (2021) (Fig. 1).

```
1 ### IMPORT PLOTASTIC
2 import plotastic as plst
3
4 # IMPORT EXAMPLE DATA
5 DF, _dims = plst.load_dataset("fmri", verbose = False)
6
7 # EXPLICITLY DEFINE DIMENSIONS TO FACET BY
8 dims = dict(
9     y = "signal",      # y-axis, dependent variable
10    x = "timepoint",    # x-axis, independent variable (within-subject factor)
11    hue = "event",      # color, independent variable (within-subject factor)
12    col = "region"      # axes, grouping variable
13 )
14 # INITIALIZE DATAANALYSIS OBJECT
15 DA = plst.DataAnalysis(
16     data=DF,           # Dataframe, long format
17     dims=dims,         # Dictionary with y, x, hue, col, row
18     subject="subject", # Datapoints are paired by subject (optional)
19     verbose=False,     # Print out info about the Data (optional)
20 )
21 # STATISTICAL TESTS
22 DA.check_normality()  # Check Normality
23 DA.check_sphericity() # Check Sphericity
24 DA.omnibus_rm_anova() # Perform RM-ANOVA
25 DA.test_pairwise()   # Perform Posthoc Analysis
26
27 # PLOTTING
28 (DA
29 .plot_box_strip()     # Pre-built plotting function initializes plot
30 .annotate_pairwise(    # Annotate results from DA.test_pairwise()
31     include="__HUE"    # Use only significant pairs across each hue
32 )
33 )
```

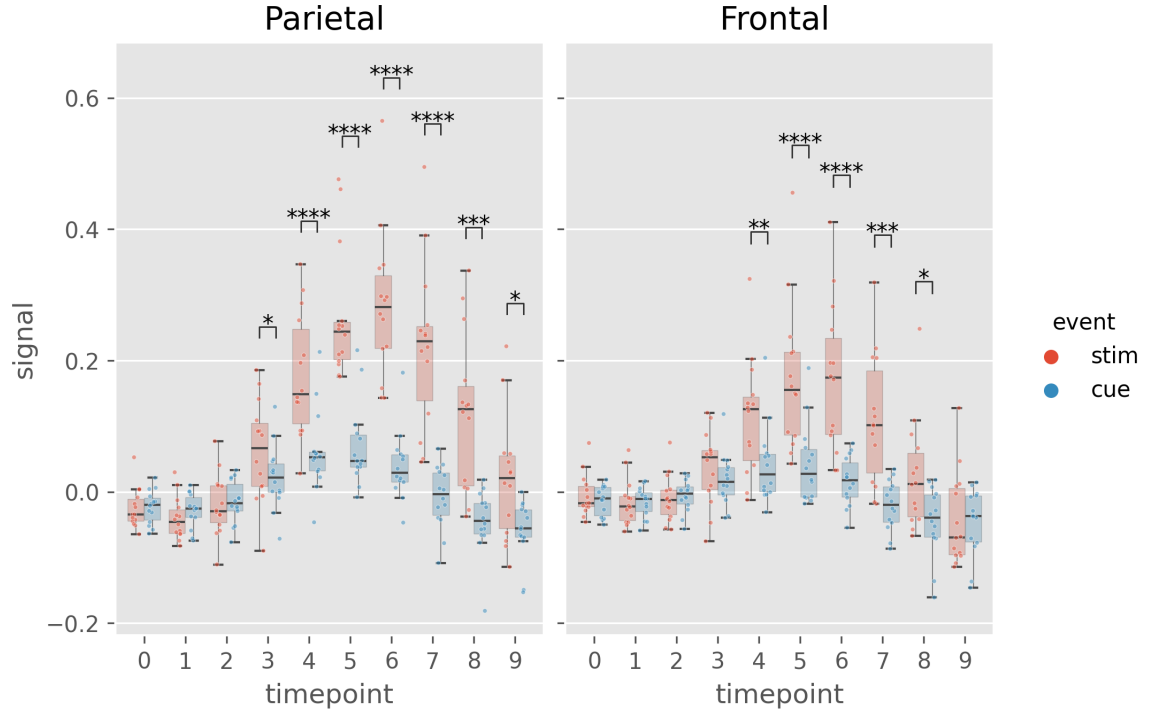


Figure 1: Example figure of plotastic (version 0.1). Image style was set by `plt.style.use('ggplot')`

Table 2: Results from `DA.check_sphericity()`. plotastic assesses sphericity after grouping the data by all grouping dimensions (hue, row, col). For example, `DA.check_sphericity()` grouped the 'fmri' dataset by "region" (col) and "event" (hue), performing four subsequent sphericity tests for four datasets.

'region', 'event'	spher	W	chi2	dof	pval	group count	n per group
'frontal', 'cue'	True	3.26e+20	-462.7	44	1	10	[14]
'frontal', 'stim'	True	2.45e+17	-392.2	44	1	10	[14]
'parietal', 'cue'	True	1.20e+20	-452.9	44	1	10	[14]
'parietal', 'stim'	True	2.44e+13	-301.9	44	1	10	[14]

Table 3: Results of `DA.omnibus_rm_anova()`. plotastic performs one two-factor RM-ANOVA per axis (grouping the data by row and col dimensions) using x and hue as the within-factors. For this example, `DA.omnibus_rm_anova()` grouped the 'fmri' dataset by "region" (col), performing two subsequent two-factor RM-ANOVAs. Within-factors are "timepoint" (x) and "event" (hue). For conciseness, GG-Correction and effect sizes are not shown.

'region'	Source	SS	ddof1	ddof2	MS	F	p-unc	stars
'parietal'	timepoint	1.583	9	117	0.175	26.20	3.40e-24	****
'parietal'	event	0.770	1	13	0.770	85.31	4.48e-07	****
'parietal'	timepoint * event	0.623	9	117	0.069	29.54	3.26e-26	****
'frontal'	timepoint	0.686	9	117	0.076	15.98	8.28e-17	****
'frontal'	event	0.240	1	13	0.240	23.44	3.21e-4	***
'frontal'	timepoint * event	0.242	9	117	0.026	13.031	3.23e-14	****

Overview

The functionality of `plotastic` revolves around a seamless integration of statistical analysis and plotting, leveraging the capabilities of `pingouin`, `seaborn`, `matplotlib` and `statannotations` (Vallat, 2018; Waskom, 2021; Hunter, 2007; Charlier et al., 2022). It utilizes long-format pandas DataFrames as its primary input, aligning with the conventions of `seaborn` and ensuring compatibility with existing data structures (Wickham, 2014; Team, 2020; McKinney, 2010).

`plotastic` was inspired by `seaborn` using the same set of intuitive and consistent parameters (`y`, `x`, `hue`, `row`, `col`) found in each of its plotting functions (Waskom, 2021). These parameters intuitively delineate the data dimensions plotted, yielding ‘facetted’ subplots, each presenting `y` against `x`. This allows for rapid and insightful exploration of multidimensional relationships. `plotastic` extends this principle to statistical analysis by storing these `seaborn` parameters (referred to as dimensions) in a `DataAnalysis` object and intelligently passing them to statistical functions of the `pingouin` library. This approach is based on the impression that most decisions during statistical analysis can be derived from how the user decides to arrange the data in a plot. This approach also prevents code repetition and streamlines statistical analysis. For example, the subject keyword is specified only once during `DataAnalysis` initialisation, and `plotastic` selects the appropriate paired or unpaired version of the test. Using `pingouin` alone requires the user to manually pick the correct test and to repeatedly specify the subject keyword in each testing function.

In essence, `plotastic` translates plotting parameters into their statistical counterparts. This translation minimizes user input and also ensures a coherent and logical connection between plotting and statistical analysis. The goal is to allow the user to focus on choosing the correct statistical test (e.g. parametric vs. non-parametric) and worry less about specific implementations.

At its core, `plotastic` employs iterators to systematically group data based on various dimensions, aligning the analysis with the distinct requirements of tests and plots. Normality testing is performed on each individual sample, which is achieved by splitting the data by all grouping dimensions and also the x-axis (`hue`, `row`, `col`, `x`). Sphericity and homoscedasticity testing is performed on a complete sampleset listed on the x-axis, which is achieved by splitting the data by all grouping dimensions (`hue`, `row`, `col`) (Tab. 2). For omnibus and posthoc analyses, data is grouped by the `row` and `col` dimensions in parallel to the `matplotlib` axes, before performing one two-factor analysis per axis using `x` and `hue` as the within/between-factors. (Tab. 3).

`DataAnalysis` visualizes data through predefined plotting functions designed for drawing multi-layered plots. A notable emphasis within `plotastic` is placed on showcasing individual

datapoints alongside aggregated means or medians. In detail, each plotting function initializes a `matplotlib` figure and axes using `plt.subplots()` while returning a `DataAnalysis` object for method chaining. Axes are populated by `seaborn` plotting functions (e.g., `sns.boxplot()`), leveraging automated aggregation and error bar displays. Keyword arguments are passed to these `seaborn` functions, ensuring the same degree of customization. Users can further customize plots by chaining `DataAnalysis` methods or by applying common `matplotlib` code to override `plotastic` settings. Figures are exported using `plt.savefig()`.

`plotastic` also focuses on annotating statistical information within plots, seamlessly incorporating p-values from pairwise comparisons using `statannotations` (Charlier et al., 2022). This integration simplifies the interface and enables options for pair selection in multidimensional plots, enhancing both user experience and interpretability.

For statistics, `plotastic` integrates with the `pingouin` library to support classical assumption and hypothesis testing, covering parametric/non-parametric and paired/non-paired variants. Assumptions such as normality, homoscedasticity, and sphericity are tested. Omnibus tests include two-factor RM-ANOVA, ANOVA, Friedman, and Kruskal-Wallis. Posthoc tests are implemented through `pingouin.pairwise_tests()`, offering (paired) t-tests, Wilcoxon, and Mann-Whitney-U.

To sum up, `plotastic` stands as a unified and user-friendly solution catering to the needs of researchers and data scientists, seamlessly integrating statistical analysis with the power of plotting in Python. It streamlines the workflow, translates `seaborn` parameters into statistical terms, and supports extensive customization options for both analysis and visualization.

Discussion

As awareness of the complexities associated with multidimensional data in biomedical research increases, there is a growing demand for tools that not only simplify analysis but also enhance its intuitiveness and effectiveness (Dunn et al., 2017). `plotastic` is designed to meet this demand by seamlessly integrating data visualization with inferential statistics, making sophisticated statistical methods accessible to researchers of all expertise levels. This integration could be pivotal as it allows the visualization of data —how it is grouped and presented— to directly guide the statistical analysis, reducing the need for in-depth statistical knowledge and ensuring that the analyses are intuitively aligned with the visual aspects of the data. This approach could not only simplify the analytical process but also enhance the transparency and reproducibility of research findings.

Statistical Features: A detailed list of implemented and planned features is provided on the GitHub page of the project (Kuric, 2024). `plotastic` is comprehensive in its current scope, incorporating a robust suite of statistical tests that cater to a wide range of research needs. It includes assumption tests for normality, homoscedasticity, and sphericity, alongside classical statistical tests such as ANOVA and t-tests, available in both parametric and non-parametric forms, as well as paired and unpaired variants. However, its reliance on the `pingouin` library means that `plotastic` is subject to the same limitations as `pingouin` itself. For instance, it does not yet support survival analysis tools like log-rank tests and Kaplan-Meier plots, which are critical for certain biomedical applications. While there are external packages that offer these capabilities, integrating them into `plotastic` could significantly expand its utility and provide a more unified user experience (Davidson-Pilon, 2019).

One known issue in `plotastic` is its handling of multiple testing corrections. Currently, `plotastic` might not correctly apply these corrections when the data is split across different facets with their own y-axes (facetted by `row` and `col` keywords), which can lead to potentially incorrect statistical inferences. This is a fixable issue, and plans are in place to address it in upcoming versions to ensure that corrections for multiple testing are appropriately applied across the complete dataset. Additionally, bivariate analysis tools like correlation and regression are not yet implemented, since `plotastic` focused on data with a categorical x-axis, which is more common in biomedical research.

Plotting Features: The plotting capabilities of `plotastic` employ all of `seaborn`'s non-facetgrid plotting functions (e.g. `sns.boxplot()`), which include a wide range of plot types but may not cover all possible visualizations (Waskom, 2021). Future versions could expand the range of specialized plots, for example QQ-plots. `plotastic` focuses on offering both high- and low-level plotting configuration: `Multiplots` automate overlaying multiple plot types, which

is extremely useful for displaying raw data points alongside aggregated statistics (barplots, boxplots, etc.), a feature that can be cumbersome to implement manually. Low-level plotting configuration is supported just like in `seaborn`, since `plotastic` uses `matplotlib` as its backend. This level of flexibility is unique to `plotastic`, serving both beginners and advanced users.

Annotating statistical results into plots (e.g. `***` above barplots) is a key requirement in modern biomedical journals and could be key feature why researchers choose proprietary software like *GraphPad Prism* over other solutions. `plotastic` automates this process as well, making it a strong competitor to other statistical software.

Software Testing: The development of `plotastic` adheres to modern software engineering principles to ensure reliability and maintainability. The project utilizes continuous integration practices, which means that with every change to the codebase, a comprehensive test suite is automatically run to identify potential bugs and ensure that new contributions do not disrupt existing functionalities. This test suite covers approximately 79 % of the testable lines of code, a statistic tracked automatically by an independent service called `codecov`, highlighting a strong commitment to software quality (*Codecov*, 2024).

Documentation: Documentation serves as a critical resource for enhancing user experience and adoption, especially for software aimed at users with varying levels of expertise. Currently, `plotastic`'s documentation is focused on basic functionalities. These include detailed installation instructions, example analyses using five test datasets from `seaborn` that are commonly used in teaching statistics, guidelines on dimension switching with commands like `DataAnalysis.switch()`, and tutorials on constructing and configuring plots, annotating statistical data, and utilizing multiplot capabilities.

However, the documentation of `plotastic` could be significantly enhanced. Currently, it lacks a dedicated website, relying instead on GitHub-hosted Jupyter notebooks. While useful, these notebooks are not the most user-friendly or maintainable format for documentation as they can be challenging to navigate and don't update synchronously with software changes. A more robust approach would involve leveraging services like `Read the Docs` or `Sphinx` to generate and host documentation directly from the codebase (*Read the Docs*, 2024; *Sphinx*, 2024). This would not only ensure that the documentation remains up-to-date with the latest software developments but also provide a more accessible and navigable user experience, meeting the expectations of users who prefer a dedicated website for software documentation.

Usability for Non-Statisticians: `plotastic` aims to make statistical analysis more accessible to researchers without extensive statistical training by intuitively mapping plotting concepts to statistical operations. To the author's knowledge, this approach is unique to `plotastic` and has great potential to make statistics easier. Still, the software requires responsible and self-critical usage, as emphasized by the thorough disclaimer on its GitHub page regarding the

software’s statistical robustness, (Kuric, 2024). The disclaimer highlights that while `plotastic` can facilitate gaining practical experience with statistics and provide a preliminary analysis, it is not a substitute for professional statistical consultation. It is designed to aid users in generating publication-grade figures and performing statistical tests, provided they have a basic understanding of the procedures involved or have their results verified by a statistician. To enhance usability for non-statisticians, `plotastic` could incorporate a system to suggest appropriate statistical tests based on data characteristics, like parametric tests for normally distributed data. This feature would guide users in selecting the correct tests, thereby augmenting the tool’s functionality and broadening its appeal. Additionally, the GitHub page provides critical guidelines for responsible statistical practice, urging users to document their work in detail, understand the limitations of the tests applied, and consult professionals to validate their findings, ensuring that `plotastic` supports but does not replace thorough statistical analysis (Sandve et al., 2013; Kuric, 2024).

Usability for Non-Programmers: Most biomedical researchers prefer software with a graphical user interface (GUI) and are not confident in using a command-line interface (CLI), suggesting a low adoption rate for `plotastic` among non-bioinformaticians. However, the dawn of Artificial intelligence provides a powerful argument for non-programmers to give CLIs a chance: GUIs are not compatible with ChatGPT, whereas code is. In fact, both ChatGPT-3.5 and -4 show impressive performance in python (Arefin et al., 2023). This is a game changer, since researchers can now use similar tools as programmers and are limited by their methodological expertise to formulate a correct prompt (Qureshi et al., 2023)⁴. Concretely, when a software is not working as intended, users of a GUI are likely to be stuck without help or further research. Users of a CLI however, can use ChatGPT to ask for code-corrections, implementations of new features (e.g. editing a plot), explanations of the code line-by-line, but also for advice on how to proceed with a statistical analysis with their specific data and needs in mind. Still, ChatGPT requires responsible use, as ChatGPT-4, is not sufficient as a standalone tool for statistical analysis (Ordak, 2023). It is important to note that `plotastic` is not yet known to ChatGPT, but could be included in future versions, depending on the popularity of `plotastic`.

Together, the transition to a new data analysis software, especially one that incorporates coding, presents a learning curve. However, the advantages of `plotastic` in terms of analytical clarity, speed, and depth are anticipated to outweigh these initial challenges.

Adoption and Open-Source Contributions: The adoption rate of `plotastic` is a critical factor for its sustainability, particularly in the open-source environment where community contributions can significantly support the author in improving and maintaining the software.

⁴Kelleher (2024): “You can now recognize and learn the language of almost anything with structure, and you can translate it to anything with structure — so text-protein, protein-text. [...] Everybody is a programmer, and the programming language of the future is called ‘human.’”

Since its publication in the Journal of Open Source Software on March 9, 2024, `plotastic` has garnered attention with 41 visits and 8 *stars* (similar to a ‘like’ on social media platforms) on its GitHub page. This level of engagement, while modest, shows initial interest and potential for growth. Active involvement from the community is essential for ongoing improvements; hence, efforts are being made to enhance the software’s documentation and structure to attract more contributors: `plotastic`’s GitHub page shows a detailed outline of the software’s architecture as a class diagram in unified modeling language (UML) format, helping potential contributors orient themselves within `plotastic`’s several modules and classes (shown in Appendix B.1). But further efforts are required, e.g. only few functions are documented with docstrings, which help understanding the purpose and usage of each function. Still, `plotastic` is a general-purpose data analysis software designed not only for biologists but for a broad range of scientific disciplines, making it a versatile tool with promising potential for wider adoption.

Contributions to Methodological Transparency and Biomedicine: `plotastic` contributes significantly to methodological transparency in biomedical research by ensuring that statistical analyses are performed alongside visual representations. This integrated approach not only enhances the interpretation of results but also facilitates the replication of analyses, promoting a higher degree of transparency. Although it streamlines the data analysis process, it is not a complete solution to the reproducibility crisis in scientific research. Researchers must still possess a basic understanding of data analysis principles and be cautious about their reliance on scripting solutions like Python, which might be less familiar to some biomedical researchers.

Statistical literacy and lack of training is a well-documented challenge among clinicians and biomedical researchers, that could decrease the confidence in presenting their analyses in detail (Lakhlifi et al., 2023; Federer et al., 2016). Since `plotastic` alleviates some need for statistical knowledge by automating the configuration of statistical tests, the room for error is reduced, and the user can lay off some responsibility to the software, gaining confidence in presenting their analysis transparently.

Furthermore, `plotastic`’s compatibility with the Jupyter ecosystem leverages “simple, intuitive ways to both capture and embed computational work directly into our papers,” as advocated by Mesirov (2010). This integration makes `plotastic` not just a tool for analysis but also a means of enhancing the accessibility and reproducibility of scientific work. As Peng (2011) suggests, the exploration of data and analysis code may often be sufficient to verify the quality of scientific claims. This seems plausible, given that statistical tests themselves pose rigorous requirements on the data, and the results are often not interpretable without the context data. Combining `plotastic` with Jupyter Notebooks provides a compelling solution to transparently integrate and document both intermediate results and analytical processes, thus

furthering scientific rigor and replicability.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2016, March). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems* (No. arXiv:1603.04467). arXiv. Retrieved 2024-03-07, from <http://arxiv.org/abs/1603.04467> doi: 10.48550/arXiv.1603.04467
- Aggarwal, R., Ghobrial, I. M., & Roodman, G. D. (2006, October). Chemokines in multiple myeloma. *Experimental hematology*, 34(10), 1289–1295. Retrieved 2023-04-02, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3134145/> doi: 10.1016/j.exphem.2006.06.017
- Akhmetzyanova, I., McCarron, M. J., Parekh, S., Chesi, M., Bergsagel, P. L., & Fooksman, D. R. (2020). Dynamic CD138 surface expression regulates switch between myeloma growth and dissemination. *Leukemia*, 34(1), 245–256. Retrieved 2023-04-04, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6923614/> doi: 10.1038/s41375-019-0519-4
- Alcorta-Sevillano, N., Macías, I., Infante, A., & Rodríguez, C. I. (2020, December). Deciphering the Relevance of Bone ECM Signaling. *Cells*, 9(12), 2630. Retrieved 2023-12-20, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7762413/> doi: 10.3390/cells9122630
- Alsayed, Y., Ngo, H., Runnels, J., Leleu, X., Singha, U. K., Pitsillides, C. M., ... Ghobrial, I. M. (2007, April). Mechanisms of regulation of CXCR4/SDF-1 (CXCL12)-dependent migration and homing in multiple myeloma. *Blood*, 109(7), 2708–2717. doi: 10.1182/blood-2006-07-035857
- Anders, S., Pyl, P. T., & Huber, W. (2015, January). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 31(2), 166–169. doi: 10.1093/bioinformatics/btu638
- Andrews, S. (2010). *FASTQC. A quality control tool for high throughput sequence data*.
- Arefin, S. E., Heya, T. A., Al-Qudah, H., Ineza, Y., & Serwadda, A. (2023, July). *Unmasking the giant: A comprehensive evaluation of ChatGPT's proficiency in coding algorithms and data structures* (No. arXiv:2307.05360). arXiv. Retrieved 2024-05-03, from <http://arxiv.org/abs/2307.05360> doi: 10.48550/arXiv.2307.05360
- Armstrong, R. A. (2014, September). When to use the Bonferroni correction. *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 34(5), 502–508. doi: 10.1111/opo.12131
- Baker, M. (2016, May). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454. Retrieved 2024-04-22, from <https://www.nature.com/articles/533452a> doi: 10.1038/533452a
- Bao, L., Lai, Y., Liu, Y., Qin, Y., Zhao, X., Lu, X., ... Huang, X. (2013, September). CXCR4 is a good survival prognostic indicator in multiple myeloma patients. *Leukemia Research*, 37(9), 1083–1088. doi: 10.1016/j.leukres.2013.06.002
- Barzilay, R., Ben-Zur, T., Bulvik, S., Melamed, E., & Offen, D. (2009, May). Lentiviral delivery of LMX1a enhances dopaminergic phenotype in differentiated human bone marrow mesenchymal stem cells. *Stem cells and development*, 18(4), 591–601. doi: 10.1089/scd.2008.0138
- Begley, C. G., & Ioannidis, J. P. A. (2015, January). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, 116(1), 116–126. doi: 10.1161/CIRCRESAHA.114.303819
- Bianco, P. (2014). "Mesenchymal" stem cells. *Annual review of cell and developmental biology*, 30, 677–704. doi: 10.1146/annurev-cellbio-100913-013132
- Bladé, J., Beksac, M., Caers, J., Jurczyszyn, A., von Lilienfeld-Toal, M., Moreau, P., ... Richardson, P. (2022, March). Extramedullary disease in multiple myeloma: A systematic literature review. *Blood Cancer Journal*, 12(3), 1–10. Retrieved 2023-03-24, from <https://www.nature.com/articles/s41408-022-00643-3> doi: 10.1038/s41408-022-00643-3

- Blonska, M., Zhu, Y., Chuang, H. H., You, M. J., Kunkalla, K., Vega, F., & Lin, X. (2015, February). Jun-regulated genes promote interaction of diffuse large B-cell lymphoma with the microenvironment. *Blood*, 125(6), 981–991. Retrieved 2023-03-01, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4319238/> doi: 10.1182/blood-2014-04-568188
- Bolado-Carrancio, A., Rukhlenko, O. S., Nikonova, E., Tsyganov, M. A., Wheeler, A., Garcia-Munoz, A., ... Kholodenko, B. N. (2020, July). Periodic propagating waves coordinate RhoGTPase network dynamics at the leading and trailing edges during cell migration. *eLife*, 9, e58165. Retrieved 2024-04-25, from <https://elifesciences.org/articles/58165> doi: 10.7554/eLife.58165
- Bondi, A. B. (2000, September). Characteristics of scalability and their impact on performance. In *Proceedings of the 2nd international workshop on Software and performance* (pp. 195–203). New York, NY, USA: Association for Computing Machinery. Retrieved 2024-03-07, from <https://dl.acm.org/doi/10.1145/350391.350432> doi: 10.1145/350391.350432
- Bosch-Queralt, M., Tiwari, V., Damkou, A., Vaculčíaková, L., Alexopoulos, I., & Simons, M. (2022, March). A fluorescence microscopy-based protocol for volumetric measurement of lysolecithin lesion-associated de- and re-myelination in mouse brain. *STAR protocols*, 3(1), 101141. doi: 10.1016/j.xpro.2022.101141
- Boswell, D., & Foucher, T. (2011). *The Art of Readable Code: Simple and Practical Techniques for Writing Better Code*. "O'Reilly Media, Inc."
- Bou Zerdan, M., Nasr, L., Kassab, J., Saba, L., Ghossein, M., Yaghi, M., ... Chaulagain, C. P. (n.d.). Adhesion molecules in multiple myeloma oncogenesis and targeted therapy. *International Journal of Hematologic Oncology*, 11(2), IJH39. Retrieved 2023-02-01, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9136637/> doi: 10.2217/ijh-2021-0017
- Brandl, A., Solimando, A. G., Mokhtari, Z., Tabares, P., Medler, J., Manz, H., ... Beilhack, A. (2022, March). Junctional adhesion molecule C expression specifies a CD138low/neg multiple myeloma cell population in mice and humans. *Blood Advances*, 6(7), 2195–2206. Retrieved 2023-04-04, from <https://doi.org/10.1182/bloodadvances.2021004354> doi: 10.1182/bloodadvances.2021004354
- Brankatschk, R., Bodenhausen, N., Zeyer, J., & Bürgmann, H. (2012, June). Simple Absolute Quantification Method Correcting for Quantitative PCR Efficiency Variations for Microbial Community Samples. *Applied and Environmental Microbiology*, 78(12), 4481–4489. Retrieved 2023-05-27, from <https://journals.asm.org/doi/10.1128/AEM.07878-11> doi: 10.1128/AEM.07878-11
- Brooke, J. (1996, January). SUS – a quick and dirty usability scale. In (pp. 189–194).
- Bubendorf, L. (2001, August). High-throughput microarray technologies: From genomics to clinics. *European Urology*, 40(2), 231–238. doi: 10.1159/000049777
- Burger, R., Guenther, A., Bakker, F., Schmalzing, M., Bernand, S., Baum, W., ... Gramatzki, M. (2001). Gp130 and ras mediated signaling in human plasma cell line INA-6: A cytokine-regulated tumor model for plasmacytoma. *The Hematology Journal: The Official Journal of the European Haematology Association*, 2(1), 42–53. doi: 10.1038/sj.thj.6200075
- Burger, R., Günther, A., Bakker, F., Schmalzing, M., Bernand, S., Baum, W., ... Gramatzki, M. (2001, January). Gp130 and ras mediated signaling in human plasma cell line INA6: A cytokine-regulated tumor model for plasmacytoma. *Hematology Journal - HEMATOL J*, 2, 42–53. doi: 10.1038/sj.thj.6200075
- Bustin, S. A. (2014, December). The reproducibility of biomedical research: Sleepers awake! *Biomolecular Detection and Quantification*, 2, 35–42. Retrieved 2024-03-18, from <https://www.sciencedirect.com/science/article/pii/S2214753515000030> doi: 10.1016/j.bdq.2015.01.002
- Caplan, A. (1991). Mesenchymal stem cells. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, 9(5), 641–50. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1870029> doi: 10.1002/jor.1100090504

- Caplan, A. I. (1994, July). The mesengenic process. *Clinics in plastic surgery*, 21(3), 429–435.
- Carlson, M. (2016). Org.Hs.eg.db. *Bioconductor*. Retrieved 2023-06-09, from <http://bioconductor.org/packages/org.Hs.eg.db/> doi: 10.18129/B9.bioc.org.Hs.eg.db
- Chacon, S., & Straub, B. (2024, March). *Git - Book*. Retrieved 2024-03-07, from <https://git-scm.com/book/de/v2>
- Charlier, F., Weber, M., Izak, D., Harkin, E., Magnus, M., Lalli, J., ... Repplinger, S. (2022, October). *Trevismd/statannotatons: V0.5*. Zenodo. Retrieved 2023-11-16, from <https://zenodo.org/record/7213391> doi: 10.5281/ZENODO.7213391
- Chatterjee, M., Hönemann, D., Lentzsch, S., Bommert, K., Sers, C., Herrmann, P., ... Bargou, R. C. (2002, November). In the presence of bone marrow stromal cells human multiple myeloma cells become independent of the IL-6/gp130/STAT3 pathway. *Blood*, 100(9), 3311–3318. doi: 10.1182/blood-2002-01-0102
- Codecov. (2024). Retrieved 2024-05-02, from <https://github.com/codecov>
- Committee on Strategies for Responsible Sharing of Clinical Trial Data, Board on Health Sciences Policy, & Institute of Medicine. (2015). *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. Washington (DC): National Academies Press (US). Retrieved 2024-04-23, from <http://www.ncbi.nlm.nih.gov/books/NBK269030/>
- Cooper, G. M. (2000). The Cell: A Molecular Approach. 2nd Edition. *Sinauer Associates*, Proliferation in Development and Differentiation. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK9906/>
- da Silva Meirelles, L., Chagastelles, P. C., & Nardi, N. B. (2006, June). Mesenchymal stem cells reside in virtually all post-natal organs and tissues. *Journal of cell science*, 119(Pt 11), 2204–2213. doi: 10.1242/jcs.02932
- Davidson-Pilon, C. (2019, August). Lifelines: Survival analysis in Python. *Journal of Open Source Software*, 4(40), 1317. Retrieved 2024-05-02, from <https://joss.theoj.org/papers/10.21105/joss.01317> doi: 10.21105/joss.01317
- Ding, W., Goldberg, D., & Zhou, W. (2023, August). PyComplexHeatmap: A Python package to visualize multimodal genomics data. *iMeta*, 2(3), e115. doi: 10.1002/imt2.115
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013, January). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. Retrieved 2023-05-27, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530905/> doi: 10.1093/bioinformatics/bts635
- Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F., Krause, D., ... Horwitz, E. (2006). Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy*, 8(4), 315–317. doi: 10.1080/14653240600855905
- Dotterweich, J., Schlegelmilch, K., Keller, A., Geyer, B., Schneider, D., Zeck, S., ... Schütze, N. (2016, December). Contact of myeloma cells induces a characteristic transcriptome signature in skeletal precursor cells -Implications for myeloma bone disease. *Bone*, 93, 155–166. doi: 10.1016/j.bone.2016.08.006
- Dunn, W., Burgun, A., Krebs, M.-O., & Rance, B. (2017, November). Exploring and visualizing multidimensional data in translational research platforms. *Briefings in Bioinformatics*, 18(6), 1044–1056. Retrieved 2024-04-23, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5862238/> doi: 10.1093/bib/bbw080
- Duvall, P., Matyas, S., & Glover, A. (2007). *Continuous integration: Improving software quality and reducing risk*. Pearson Education. Retrieved from <https://books.google.de/books?id=PV9qfEdv9L0C>
- Dziadowicz, S. A., Wang, L., Akhter, H., Aesoph, D., Sharma, T., Adjero, D. A., ... Hu, G. (2022, January). Bone Marrow Stroma-Induced Transcriptome and Regulome Signatures of Multiple Myeloma. *Cancers*, 14(4), 927. Retrieved 2022-10-25, from <https://www.mdpi.com/2072-6694/14/4/927> doi: 10.3390/cancers14040927
- Ekmekci, B., McAnany, C. E., & Mura, C. (2016, July). An Introduction to Programming for Bioscientists: A Python-Based Primer. *PLOS Computational Biology*, 12(6), e1004867. Retrieved 2024-03-10, from <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004867> doi: 10.1371/journal.pcbi

- .1004867
- Evers, M., Schreder, M., Stühmer, T., Jundt, F., Ebert, R., Hartmann, T. N., ... Leich, E. (2023, March). Prognostic value of extracellular matrix gene mutations and expression in multiple myeloma. *Blood Cancer Journal*, 13(1), 43. doi: 10.1038/s41408-023-00817-7
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016, October). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. Retrieved 2023-06-09, from <https://doi.org/10.1093/bioinformatics/btw354> doi: 10.1093/bioinformatics/btw354
- Excel, M. (2023, August). *Announcing Python in Excel: Combining the power of Python and the flexibility of Excel*. Retrieved 2024-03-11, from <https://techcommunity.microsoft.com/t5/excel-blog/announcing-python-in-excel-combining-the-power-of-python-and-the/ba-p/3893439>
- Fazeli, P. K., Horowitz, M. C., MacDougald, O. A., Scheller, E. L., Rodeheffer, M. S., Rosen, C. J., & Klibanski, A. (2013, March). Marrow Fat and Bone—New Perspectives. *The Journal of Clinical Endocrinology and Metabolism*, 98(3), 935–945. Retrieved 2023-12-20, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3590487/> doi: 10.1210/jc.2012-3634
- Federer, L. M., Lu, Y.-L., & Joubert, D. J. (2016, January). Data literacy training needs of biomedical researchers. *Journal of the Medical Library Association : JMLA*, 104(1), 52–57. Retrieved 2024-04-24, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4722643/> doi: 10.3163/1536-5050.104.1.008
- Fernandez-Rebollo, E., Mentrup, B., Ebert, R., Franzen, J., Abagnale, G., Sieben, T., ... Wagner, W. (2017, July). Human Platelet Lysate versus Fetal Calf Serum: These Supplements Do Not Select for Different Mesenchymal Stromal Cells. *Scientific Reports*, 7, 5132. Retrieved 2023-05-02, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5506010/> doi: 10.1038/s41598-017-05207-1
- Flier, J. S. (2022). The Problem of Irreproducible Bioscience Research. *Perspectives in Biology and Medicine*, 65(3), 373–395. doi: 10.1353/pbm.2022.0032
- Frassanito, M. A., Cusmai, A., Iodice, G., & Dammacco, F. (2001, January). Autocrine interleukin-6 production and highly malignant multiple myeloma: Relation with resistance to drug-induced apoptosis. *Blood*, 97(2), 483–489. doi: 10.1182/blood.v97.2.483
- Friedenstein, A., & Kuralesova, A. I. (1971, August). Osteogenic precursor cells of bone marrow in radiation chimeras. *Transplantation*, 12(2), 99–108.
- Friedenstein, A. J., Piatetzky-Shapiro, I. I., & Petrakova, K. V. (1966, December). Osteogenesis in transplants of bone marrow cells. *Journal of embryology and experimental morphology*, 16(3), 381–390.
- Gabr, M. M., Zakaria, M. M., Refaie, A. F., Ismail, A. M., Abou-El-Mahasen, M. A., Ashamallah, S. A., ... Ghoneim, M. A. (2013). Insulin-producing cells from adult human bone marrow mesenchymal stem cells control streptozotocin-induced diabetes in nude mice. *Cell transplantation*, 22(1), 133–145. doi: 10.3727/096368912X647162
- Garcés, J.-J., Simicek, M., Vicari, M., Brozova, L., Burgos, L., Bezdekova, R., ... Paiva, B. (2020, February). Transcriptional profiling of circulating tumor cells in multiple myeloma: A new model to understand disease dissemination. *Leukemia*, 34(2), 589–603. doi: 10.1038/s41375-019-0588-4
- García-Ortiz, A., Rodríguez-García, Y., Encinas, J., Maroto-Martín, E., Castellano, E., Teixidó, J., & Martínez-López, J. (2021, January). The Role of Tumor Microenvironment in Multiple Myeloma Development and Progression. *Cancers*, 13(2). Retrieved 2021-02-02, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7827690/> doi: 10.3390/cancers13020217
- Gentleman. (n.d.). *Bioconductor - BiocViews*. Retrieved 2023-06-09, from <https://bioconductor.org/packages/3.17/BiocViews.html>
- Ghobrial, I. M. (2012, July). Myeloma as a model for the process of metastasis: Implications for therapy. *Blood*, 120(1), 20–30. Retrieved 2022-10-15, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3390959/> doi:

- 10.1182/blood-2012-01-379024
- Giorgi, F. M., Ceraolo, C., & Mercatelli, D. (2022, April). The R Language: An Engine for Bioinformatics and Data Science. *Life*, 12(5), 648. Retrieved 2024-04-21, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9148156/> doi: 10.3390/life12050648
- Glavey, S. V., Naba, A., Manier, S., Clauser, K., Tahri, S., Park, J., ... Ghobrial, I. M. (2017, November). Proteomic characterization of human multiple myeloma bone marrow extracellular matrix. *Leukemia*, 31(11), 2426–2434. Retrieved 2023-09-05, from <https://www.nature.com/articles/leu2017102> doi: 10.1038/leu.2017.102
- Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merckenschlager, M., Gisel, A., ... Tegnér, J. (2014, March). Data integration in the era of omics: Current and future challenges. *BMC Systems Biology*, 8(2), I1. Retrieved 2024-03-18, from <https://doi.org/10.1186/1752-0509-8-S2-I1> doi: 10.1186/1752-0509-8-S2-I1
- Gómez-López, G., Dopazo, J., Cigudosa, J. C., Valencia, A., & Al-Shahrour, F. (2019, May). Precision medicine needs pioneering clinical bioinformaticians. *Briefings in Bioinformatics*, 20(3), 752–766. doi: 10.1093/bib/bbx144
- Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016, June). What does research reproducibility mean? *Science Translational Medicine*, 8(341), 341ps12–341ps12. Retrieved 2024-03-18, from <https://www.science.org/doi/10.1126/scitranslmed.aaf5027> doi: 10.1126/scitranslmed.aaf5027
- Gorelick, M., & Ozsvald, I. (2020). *High Performance Python: Practical Performant Programming for Humans*. "O'Reilly Media, Inc."
- Gosselin, R.-D. (2021, February). Insufficient transparency of statistical reporting in preclinical research: A scoping review. *Scientific Reports*, 11(1), 3335. Retrieved 2024-03-11, from <https://www.nature.com/articles/s41598-021-83006-5> doi: 10.1038/s41598-021-83006-5
- Gramatzki, M., Burger, R., Trautman, U., Marschalek, R., Lorenz, H., Hansen-Hagge, T., ... Kalden, J. (1994). Two new interleukin-6 dependent plasma cell lines carrying a chromosomal abnormality involving the IL-6 gene locus. , *84 Suppl. 1*, 173a–173a. Retrieved 2023-03-24, from <https://www.cellosaurus.org/cellopub/CLPUB00060>
- Greenstein, S., Krett, N. L., Kurosawa, Y., Ma, C., Chauhan, D., Hideshima, T., ... Rosen, S. T. (2003, April). Characterization of the MM.1 human multiple myeloma (MM) cell lines: A model system to elucidate the characteristics, behavior, and signaling of steroid-sensitive and -resistant MM cells. *Experimental Hematology*, 31(4), 271–282. doi: 10.1016/s0301-472x(03)00023-7
- Gronthos, S., Graves, S. E., Ohta, S., & Simmons, P. J. (1994, December). The STRO-1+ fraction of adult human bone marrow contains the osteogenic precursors. *Blood*, 84(12), 4164–4173.
- Hannun, A., Digani, J., Katharopoulos, A., & Collobert, R. (2023). *MLX: Efficient and flexible machine learning on Apple silicon*. Retrieved from <https://github.com/ml-explore>
- Harrington, D. P., & Fleming, T. R. (1982). A Class of Rank Test Procedures for Censored Survival Data. *Biometrika*, 69(3), 553–566. Retrieved 2023-08-07, from <https://www.jstor.org/stable/2335991> doi: 10.2307/2335991
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020, September). Array programming with NumPy. *Nature*, 585(7825), 357–362. Retrieved 2023-08-09, from <https://www.nature.com/articles/s41586-020-2649-2> doi: 10.1038/s41586-020-2649-2
- Hideshima, T., Mitsiades, C., Tonon, G., Richardson, P. G., & Anderson, K. C. (2007, August). Understanding multiple myeloma pathogenesis in the bone marrow to identify new therapeutic targets. *Nature Reviews Cancer*, 7(8), 585–598. Retrieved 2023-02-07, from <https://www.nature.com/articles/nrc2189> doi: 10.1038/nrc2189

- Hose, D., Rème, T., Hielscher, T., Moreaux, J., Messner, T., Seckinger, A., ... Goldschmidt, H. (2011, January). Proliferation is a central independent prognostic factor and target for personalized and risk-adapted treatment in multiple myeloma. *Haematologica*, 96(1), 87–95. doi: 10.3324/haematol.2010.030296
- Hothorn, T., & Lausen, B. (n.d.). *Maximally Selected Rank Statistics in R*. Retrieved from <http://cran.r-project.org/web/packages/maxstat/index.html>.
- Howe, A., & Chain, P. S. G. (2015). Challenges and opportunities in understanding microbial communities with metagenome assembly (accompanied by IPython Notebook tutorial). *Frontiers in Microbiology*, 6, 678. doi: 10.3389/fmicb.2015.00678
- Hu, X., Villodre, E. S., Larson, R., Rahal, O. M., Wang, X., Gong, Y., ... Debeb, B. G. (2021, January). Decorin-mediated suppression of tumorigenesis, invasion, and metastasis in inflammatory breast cancer. *Communications Biology*, 4(1), 72. doi: 10.1038/s42003-020-01590-0
- Huang, S.-Y., Lin, H.-H., Yao, M., Tang, J.-L., Wu, S.-J., Hou, H.-A., ... Tien, H.-F. (2015). Higher Decorin Levels in Bone Marrow Plasma Are Associated with Superior Treatment Response to Novel Agent-Based Induction in Patients with Newly Diagnosed Myeloma - A Retrospective Study. *PloS One*, 10(9), e0137552. doi: 10.1371/journal.pone.0137552
- Hunter, J. D. (2007, May). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95. Retrieved 2023-11-15, from <https://ieeexplore.ieee.org/document/4160265> doi: 10.1109/MCSE.2007.55
- Incerti, D., Thom, H., Baio, G., & Jansen, J. P. (2019, May). R You Still Using Excel? The Advantages of Modern Software Tools for Health Technology Assessment. *Value in Health*, 22(5), 575–579. Retrieved 2024-03-11, from <https://www.sciencedirect.com/science/article/pii/S1098301519300506> doi: 10.1016/j.jval.2019.01.003
- Ioannidis, J. P. A. (2005, August). Why Most Published Research Findings Are False. *PLOS Medicine*, 2(8), e124. Retrieved 2024-04-22, from <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124> doi: 10.1371/journal.pmed.0020124
- Jansen, B. J. H., Gilissen, C., Roelofs, H., Schaap-Oziemlak, A., Veltman, J. A., Raymakers, R. A. P., ... Adema, G. J. (2010, April). Functional differences between mesenchymal stem cell populations are reflected by their transcriptome. *Stem cells and development*, 19(4), 481–490. doi: 10.1089/scd.2009.0288
- Kaplan, E. L., & Meier, P. (1958, June). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282), 457–481. Retrieved 2023-08-07, from <http://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501452> doi: 10.1080/01621459.1958.10501452
- Katz, B.-Z. (2010, June). Adhesion molecules—The lifelines of multiple myeloma cells. *Seminars in Cancer Biology*, 20(3), 186–195. Retrieved 2021-07-04, from <https://linkinghub.elsevier.com/retrieve/pii/S1044579X10000246> doi: 10.1016/j.semcancer.2010.04.003
- Kawano, M. M., Huang, N., Tanaka, H., Ishikawa, H., Sakai, A., Tanabe, O., ... Kuramoto, A. (1991, December). Homotypic cell aggregations of human myeloma cells with ICAM-1 and LFA-1 molecules. *British Journal of Haematology*, 79(4), 583–588. doi: 10.1111/j.1365-2141.1991.tb08085.x
- Kazman, R., Bianco, P., Ivers, J., & Klein, J. (2020, December). *Maintainability* (Report). Carnegie Mellon University. Retrieved 2024-03-07, from <https://kilthub.cmu.edu/articles/report/Maintainability/12954908/1> doi: 10.1184/R1/12954908.v1
- Kelleher, R. (2024, January). *NVIDIA CEO: ‘This Year, Every Industry Will Become a Technology Industry’*. Retrieved 2024-05-03, from <https://blogs.nvidia.com/blog/nvidia-ceo-ai-drug-discovery-jp-morgan-healthcare-2024/>
- Kelly, B. S., Kirwan, A., Quinn, M. S., Kelly, A. M., Mathur, P., Lawlor, A., & Killeen, R. P. (2023, May). The ethical matrix as a method for involving people living with disease and the wider public (PPI) in near-

- term artificial intelligence research. *Radiography (London, England: 1995)*, 29 Suppl 1, S103-S111. doi: 10.1016/j.radi.2023.03.009
- Kibler, C., Schermutzki, F., Waller, H. D., Timpl, R., Müller, C. A., & Klein, G. (1998, June). Adhesive interactions of human multiple myeloma cell lines with different extracellular matrix molecules. *Cell Adhesion and Communication*, 5(4), 307–323. doi: 10.3109/15419069809040300
- Kim, D., Langmead, B., & Salzberg, S. L. (2015, April). HISAT: A fast spliced aligner with low memory requirements. *Nature methods*, 12(4), 357–360. Retrieved 2024-04-26, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4655817/> doi: 10.1038/nmeth.3317
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., ... Jupyter Development Team (2016). *Jupyter Notebooks—a publishing format for reproducible computational workflows*. Retrieved 2024-04-20, from <https://ui.adsabs.harvard.edu/abs/2016ppap.book...87K> doi: 10.3233/978-1-61499-649-1-87
- Krekel, H., Oliveira, B., Pfannschmidt, R., Bruynooghe, F., Laughner, B., & Bruhin, F. (2004). *Pytest*. Retrieved from <https://github.com/pytest-dev/pytest>
- Krzywinski, M., & Savig, E. (2013, July). Multidimensional data. *Nature methods*, 10(7), 595. Retrieved 2024-04-22, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6092027/>
- Kuric, M. (2024, April). *Markur4/plotastic*. Retrieved 2024-05-02, from <https://github.com/markur4/plotastic>
- Kuric, M., & Ebert, R. (2024, March). Plotastic: Bridging Plotting and Statistics in Python. *Journal of Open Source Software*, 9(95), 6304. Retrieved 2024-03-11, from <https://joss.theoj.org/papers/10.21105/joss.06304> doi: 10.21105/joss.06304
- Lai, T.-Y., Cao, J., Ou-Yang, P., Tsai, C.-Y., Lin, C.-W., Chen, C.-C., ... Lee, C.-Y. (2022, April). Different methods of detaching adherent cells and their effects on the cell surface expression of Fas receptor and Fas ligand. *Scientific Reports*, 12(1), 5713. Retrieved 2023-06-01, from <https://www.nature.com/articles/s41598-022-09605-y> doi: 10.1038/s41598-022-09605-y
- Lakhlifi, C., Lejeune, F.-X., Rouault, M., Khamassi, M., & Rohaut, B. (2023, April). Illusion of knowledge in statistics among clinicians: Evaluating the alignment between objective accuracy and subjective confidence, an online survey. *Cognitive Research: Principles and Implications*, 8, 23. Retrieved 2024-04-24, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10118231/> doi: 10.1186/s41235-023-00474-1
- Leek, J. T., & Peng, R. D. (2015, April). Statistics: P values are just the tip of the iceberg. *Nature*, 520(7549), 612–612. Retrieved 2024-04-22, from <https://www.nature.com/articles/520612a> doi: 10.1038/520612a
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome Project Data Processing Subgroup (2009, August). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. Retrieved 2023-06-09, from <https://doi.org/10.1093/bioinformatics/btp352> doi: 10.1093/bioinformatics/btp352
- Localio, A. R., Goodman, S. N., Meibohm, A., Cornell, J. E., Stack, C. B., Ross, E. A., & Mulrow, C. D. (2018, June). Statistical Code to Support the Scientific Story. *Annals of Internal Medicine*, 168(11), 828–829. Retrieved 2024-04-23, from <https://www.acpjournals.org/doi/10.7326/M17-3431> doi: 10.7326/M17-3431
- Love, M. I., Huber, W., & Anders, S. (2014, December). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. Retrieved 2024-04-26, from <https://doi.org/10.1186/s13059-014-0550-8> doi: 10.1186/s13059-014-0550-8
- Maichl, D. S., Kirner, J. A., Beck, S., Cheng, W.-H., Krug, M., Kuric, M., ... Jundt, F. (2023, September). Identification of NOTCH-driven matrisome-associated genes as prognostic indicators of multiple myeloma patient survival. *Blood Cancer Journal*, 13(1), 1–6. Retrieved 2023-09-05, from <https://www.nature.com/articles/s41408-023-00907-6> doi: 10.1038/s41408-023-00907-6
- McKay, B. S., Irving, P. E., Skumatz, C. M., & Burke, J. M. (1997, November). Cell-cell adhesion molecules and

- the development of an epithelial phenotype in cultured human retinal pigment epithelial cells. *Experimental Eye Research*, 65(5), 661–671. doi: 10.1006/exer.1997.0374
- McKinney, W. (2010, January). Data Structures for Statistical Computing in Python. In (pp. 56–61). doi: 10.25080/Majora-92bf1922-00a
- McKinney, W. (2011, January). Pandas: A Foundational Python Library for Data Analysis and Statistics. *Python High Performance Science Computer*.
- Mesirov, J. P. (2010, January). Accessible Reproducible Research. *Science*, 327(5964), 415–416. Retrieved 2024-04-22, from <https://www.science.org/doi/10.1126/science.1179653> doi: 10.1126/science.1179653
- Moran, M. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, 100(2), 403–405. Retrieved 2024-04-24, from <https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0706.2003.12010.x> doi: 10.1034/j.1600-0706.2003.12010.x
- Muruganandan, S., Roman, A. A., & Sinal, C. J. (2009, January). Adipocyte differentiation of bone marrow-derived mesenchymal stem cells: Cross talk with the osteoblastogenic program. *Cellular and molecular life sciences : CMLS*, 66(2), 236–253. doi: 10.1007/s00018-008-8429-z
- Myers, G. J., Sandler, C., & Badgett, T. (2011). *The art of software testing* (3rd ed.). Wiley Publishing. Retrieved from <https://malenezi.github.io/malenezi/SE401/Books/114-the-art-of-software-testing-3-edition.pdf>
- Narzt, W., Pichler, J., Pirklbauer, K., & Zwintz, M. (1998, January). A Reusability Concept for Process Automation Software..
- Newville, M., Stensitzki, T., Allen, D. B., & Ingargiola, A. (2014, September). *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. Zenodo. Retrieved 2023-05-30, from <https://zenodo.org/record/11813> doi: 10.5281/zenodo.11813
- Nilsson, K., Bennich, H., Johansson, S. G., & Pontén, J. (1970, October). Established immunoglobulin producing myeloma (IgE) and lymphoblastoid (IgG) cell lines from an IgE myeloma patient. *Clinical and Experimental Immunology*, 7(4), 477–489.
- Nowotschin, S., & Hadjantonakis, A.-K. (2010, August). Cellular dynamics in the early mouse embryo: From axis formation to gastrulation. *Current opinion in genetics & development*, 20(4), 420–427. doi: 10.1016/j.gde.2010.05.008
- Okuno, Y., Takahashi, T., Suzuki, A., Ichiba, S., Nakamura, K., Okada, T., ... Imura, H. (1991, February). In vitro growth pattern of myeloma cells in liquid suspension or semi-solid culture containing interleukin-6. *International Journal of Hematology*, 54(1), 41–47.
- Ordak, M. (2023, September). ChatGPT’s Skills in Statistical Analysis Using the Example of Allergology: Do We Have Reason for Concern? *Healthcare (Basel, Switzerland)*, 11(18), 2554. doi: 10.3390/healthcare11182554
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019, December). *PyTorch: An Imperative Style, High-Performance Deep Learning Library* (No. arXiv:1912.01703). arXiv. Retrieved 2024-03-07, from <http://arxiv.org/abs/1912.01703> doi: 10.48550/arXiv.1912.01703
- Peng, R. D. (2011, December). Reproducible Research in Computational Science. *Science*, 334(6060), 1226–1227. Retrieved 2024-03-18, from <https://www.science.org/doi/10.1126/science.1213847> doi: 10.1126/science.1213847
- Perez, F., & Granger, B. E. (2007, May). IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, 9(3), 21–29. Retrieved 2024-04-20, from <https://ieeexplore.ieee.org/document/4160251> doi: 10.1109/MCSE.2007.53
- Perneger, T. V. (1998, April). What’s wrong with Bonferroni adjustments. *BMJ : British Medical Journal*, 316(7139), 1236–1238. Retrieved 2021-11-24, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1112991/>
-

- Pittenger, M. F., Mackay, A. M., Beck, S. C., Jaiswal, R. K., Douglas, R., Mosca, J. D., ... Marshak, D. R. (1999). Multilineage Potential of Adult Human Mesenchymal Stem Cells. , 284(April), 143–148. doi: 10.1126/science.284.5411.143
- Polager, S., & Ginsberg, D. (2009, October). P53 and E2f: Partners in life and death. *Nature Reviews Cancer*, 9(10), 738–748. Retrieved 2023-02-14, from <https://www.nature.com/articles/nrc2718> doi: 10.1038/nrc2718
- PyMOL. (2024). Retrieved 2024-04-30, from <https://pymol.org/>
- The Python Language Reference. (2024). Retrieved 2024-03-07, from <https://docs.python.org/3/reference/index.html>
- Qureshi, R., Shaughnessy, D., Gill, K. A. R., Robinson, K. A., Li, T., & Agai, E. (2023, April). Are ChatGPT and large language models “the answer” to bringing us closer to systematic review automation? *Systematic Reviews*, 12, 72. Retrieved 2024-05-03, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10148473/> doi: 10.1186/s13643-023-02243-z
- R Core Team. (2018). *R: A language and environment for statistical computing* [Manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.. Retrieved 2024-03-07, from <https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe>
- Rajkumar, S. V., Dimopoulos, M. A., Palumbo, A., Blade, J., Merlini, G., Mateos, M.-V., ... Miguel, J. F. S. (2014, November). International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet. Oncology*, 15(12), e538-548. doi: 10.1016/S1470-2045(14)70442-5
- Rajkumar, S. V., & Kumar, S. (2020, September). Multiple myeloma current treatment algorithms. *Blood Cancer Journal*, 10(9), 94. Retrieved 2023-07-03, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7523011/> doi: 10.1038/s41408-020-00359-2
- Ramakers, C., Ruijter, J. M., Deprez, R. H., & Moorman, A. F. (2003, March). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters*, 339(1), 62–66. Retrieved 2022-11-27, from <https://linkinghub.elsevier.com/retrieve/pii/S0304394002014234> doi: 10.1016/S0304-3940(02)01423-4
- Rayhan, A., & Gross, D. (2023). *The Rise of Python: A Survey of Recent Research*. doi: 10.13140/RG.2.2.27388.92809
- Read the Docs. (2024). Retrieved 2024-05-03, from <https://docs.readthedocs.io/en/stable/index.html>
- Rebl, H., Finke, B., Schroeder, K., & Nebe, J. B. (2010, October). Time-dependent metabolic activity and adhesion of human osteoblast-like cells on sensor chips with a plasma polymer nanolayer. *The International Journal of Artificial Organs*, 33(10), 738–748.
- Rigsby, R. E., & Parker, A. B. (2016, September). Using the PyMOL application to reinforce visual understanding of protein structure. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology*, 44(5), 433–437. doi: 10.1002/bmb.20966
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010, January). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139–140. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/19910308> doi: 10.1093/bioinformatics/btp616
- Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W. (2017, November). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18(1), 529. Retrieved 2024-04-25, from <https://doi.org/10.1186/s12859-017-1934-z> doi: 10.1186/s12859-017-1934-z
- Sacchetti, B., Funari, A., Remoli, C., Giannicola, G., Kogler, G., Liedtke, S., ... Bianco, P. (2016). No identical

- "mesenchymal stem cells" at different times and sites: Human committed progenitors of distinct origin and differentiation potential are incorporated as adventitial cells in microvessels. *Stem Cell Reports*, 6(6), 897–913. Retrieved from <http://dx.doi.org/10.1016/j.stemcr.2016.05.011> doi: 10.1016/j.stemcr.2016.05.011
- Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013, October). Ten Simple Rules for Reproducible Computational Research. *PLoS Computational Biology*, 9(10), e1003285. Retrieved 2024-03-07, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3812051/> doi: 10.1371/journal.pcbi.1003285
- Santos, B. S., Silva, I., Ribeiro-Dantas, M. d. C., Alves, G., Endo, P. T., & Lima, L. (2020, October). COVID-19: A scholarly production dataset report for research analysis. *Data in Brief*, 32, 106178. doi: 10.1016/j.dib.2020.106178
- Sanz-Rodríguez, F., Ruiz-Velasco, N., Pascual-Salcedo, D., & Teixidó, J. (1999, December). Characterization of VLA-4-dependent myeloma cell adhesion to fibronectin and VCAM-1: VLA-4-dependent Myeloma Cell Adhesion. *British Journal of Haematology*, 107(4), 825–834. Retrieved 2023-04-02, from <http://doi.wiley.com/10.1046/j.1365-2141.1999.01762.x> doi: 10.1046/j.1365-2141.1999.01762.x
- Sarin, V., Yu, K., Ferguson, I. D., Gugliemini, O., Nix, M. A., Hann, B., ... Wiita, A. P. (2020, October). Evaluating the efficacy of multiple myeloma cell lines as models for patient tumors via transcriptomic correlation analysis. *Leukemia*, 34(10), 2754–2765. doi: 10.1038/s41375-020-0785-1
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. In *Python in Science Conference* (pp. 92–96). Austin, Texas. Retrieved 2023-05-29, from <https://conference.scipy.org/proceedings/scipy2010/seabold.html> doi: 10.25080/Majors-92bf1922-011
- Seckinger, A., Delgado, J. A., Moser, S., Moreno, L., Neuber, B., Grab, A., ... Vu, M. D. (2017, March). Target Expression, Generation, Preclinical Activity, and Pharmacokinetics of the BCMA-T Cell Bispecific Antibody EM801 for Multiple Myeloma Treatment. *Cancer Cell*, 31(3), 396–410. Retrieved 2023-07-21, from [https://www.cell.com/cancer-cell/abstract/S1535-6108\(17\)30016-8](https://www.cell.com/cancer-cell/abstract/S1535-6108(17)30016-8) doi: 10.1016/j.ccell.2017.02.002
- Seckinger, A., Hillengass, J., Emde, M., Beck, S., Kimmich, C., Dittrich, T., ... Hose, D. (2018). CD38 as Immunotherapeutic Target in Light Chain Amyloidosis and Multiple Myeloma-Association With Molecular Entities, Risk, Survival, and Mechanisms of Upfront Resistance. *Frontiers in Immunology*, 9, 1676. doi: 10.3389/fimmu.2018.01676
- Shenghui, H., Nakada, D., & Morrison, S. J. (2009). Mechanisms of Stem Cell Self-Renewal. *Annual Review of Cell and Developmental Biology*, 25(1), 377–406. Retrieved from <https://doi.org/10.1146/annurev.cellbio.042308.113248> doi: 10.1146/annurev.cellbio.042308.113248
- Smith, A. M., Niemeyer, K. E., Katz, D. S., Barba, L. A., Githinji, G., Gymrek, M., ... Vanderplas, J. T. (2018). Journal of Open Source Software (JOSS): Design and first-year review. *PeerJ Preprints*, 4, e147. doi: 10.7717/peerj-cs.147
- Solimando, A. G., Malerba, E., Leone, P., Prete, M., Terragna, C., Cavo, M., & Racanelli, V. (2022, September). Drug resistance in multiple myeloma: Soldiers and weapons in the bone marrow niche. *Frontiers in Oncology*, 12, 973836. Retrieved 2022-10-23, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9533079/> doi: 10.3389/fonc.2022.973836
- Sphinx*. (2024). Retrieved 2024-05-03, from <https://docs.readthedocs.io/en/stable/intro/getting-started-with-sphinx.html>
- Sprynski, A. C., Hose, D., Caillot, L., Rème, T., Shaughnessy, J. D., Barlogie, B., ... Klein, B. (2009, May). The role of IGF-1 as a major growth factor for myeloma cell lines and the prognostic relevance of the expression of its receptor. *Blood*, 113(19), 4614–4626. Retrieved 2023-06-29, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2691749/> doi: 10.1182/blood-2008-07-170464
- Standal, T., Seidel, C., Plesner, T., Sanderson, R., Waage, A., Børset, M., & Sundan, A. (2002, November). Osteoprotegerin is bound, internalized, and degraded by multiple myeloma cells. *Blood*, 100, 3002–7. doi:

- 10.1182/blood-2002-04-1190
- Stock, P., Bruckner, S., Winkler, S., Dollinger, M. M., & Christ, B. (2014, April). Human bone marrow mesenchymal stem cell-derived hepatocytes improve the mouse liver after acute acetaminophen intoxication by preventing progress of injury. *International journal of molecular sciences*, 15(4), 7004–7028. doi: 10.3390/ijms15047004
- Sullivan, G. M., & Feinn, R. S. (2021, August). Facts and Fictions About Handling Multiple Comparisons. *Journal of Graduate Medical Education*, 13(4), 457–460. Retrieved 2024-03-10, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8370375/> doi: 10.4300/JGME-D-21-00599.1
- Tabolacci, C., De Martino, A., Mischiati, C., Feriotto, G., & Beninati, S. (2019, January). The Role of Tissue Transglutaminase in Cancer Cell Initiation, Survival and Progression. *Medical Sciences*, 7(2), 19. Retrieved 2023-03-17, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6409630/> doi: 10.3390/medsci7020019
- Tai, Y.-T., Li, X.-F., Breitkreutz, I., Song, W., Neri, P., Catley, L., ... Anderson, K. C. (2006, July). Role of B-cell-activating factor in adhesion and growth of human multiple myeloma cells in the bone marrow microenvironment. *Cancer Research*, 66(13), 6675–6682. doi: 10.1158/0008-5472.CAN-06-0190
- Tam, P. P., & Beddington, R. S. (1987, January). The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. *Development (Cambridge, England)*, 99(1), 109–126.
- Taskiran, I. I., Spanier, K. I., Dickmanken, H., Kempynck, N., Pančíková, A., Ekşi, E. C., ... Aerts, S. (2024, February). Cell-type-directed design of synthetic enhancers. *Nature*, 626(7997), 212–220. Retrieved 2024-04-21, from <https://www.nature.com/articles/s41586-023-06936-2> doi: 10.1038/s41586-023-06936-2
- Team, T. P. D. (2020, February). *Pandas-dev/pandas: Pandas*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.3509134> doi: 10.5281/zenodo.3509134
- Terpos, E., Ntanasis-Stathopoulos, I., Gavriatopoulou, M., & Dimopoulos, M. A. (2018, January). Pathogenesis of bone disease in multiple myeloma: From bench to bedside. *Blood Cancer Journal*, 8(1), 7. doi: 10.1038/s41408-017-0037-4
- Thompson, S., Dowrick, T., Ahmad, M., Xiao, G., Koo, B., Bonmati, E., ... Clarkson, M. J. (2020, July). SciKit-Surgery: Compact libraries for surgical navigation. *International Journal of Computer Assisted Radiology and Surgery*, 15(7), 1075–1084. doi: 10.1007/s11548-020-02180-5
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... Pachter, L. (2012, March). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3), 562–578. doi: 10.1038/nprot.2012.016
- Ullah, I., Subbarao, R. B., & Rho, G. J. (2015). Human mesenchymal stem cells - current trends and future prospective Bioscience Reports. doi: 10.1042/BSR20150025
- Urashima, M., Chauhan, D., Uchiyama, H., Freeman, G., & Anderson, K. (1995, April). CD40 ligand triggered interleukin-6 secretion in multiple myeloma. *Blood*, 85(7), 1903–1912. Retrieved 2021-02-01, from <https://ashpublications.org/blood/article/85/7/1903/123565/CD40-ligand-triggered-interleukin6-secretion-in> doi: 10.1182/blood.V85.7.1903.bloodjournal8571903
- Vallat, R. (2018, November). Pingouin: Statistics in Python. *Journal of Open Source Software*, 3(31), 1026. Retrieved 2023-05-29, from <https://joss.theoj.org/papers/10.21105/joss.01026> doi: 10.21105/joss.01026
- van Rossum, G., Lehtosalo, J., & Langa, L. (2014). *PEP 484 – Type Hints | peps.python.org*. Retrieved 2024-03-08, from <https://peps.python.org/pep-0484/>
- Van Valckenborgh, E., Croucher, P. I., De Raeve, H., Carron, C., De Leenheer, E., Blacher, S., ... Vanderkerken, K. (2004, September). Multifunctional role of matrix metalloproteinases in multiple myeloma: A study in the 5T2MM mouse model. *The American Journal of Pathology*, 165(3), 869–878. doi: 10.1016/S0002-9440(10)63349-4

- Viguet-Carrin, S., Garnero, P., & Delmas, P. D. (2006, March). The role of collagen in bone strength. *Osteoporosis International*, 17(3), 319–336. Retrieved 2023-12-20, from <https://doi.org/10.1007/s00198-005-2035-9> doi: 10.1007/s00198-005-2035-9
- Wadgaonkar, R., Phelps, K. M., Haque, Z., Williams, A. J., Silverman, E. S., & Collins, T. (1999, January). CREB-binding protein is a nuclear integrator of nuclear factor-kappaB and p53 signaling. *The Journal of Biological Chemistry*, 274(4), 1879–1882. doi: 10.1074/jbc.274.4.1879
- Waskom, M. L. (2021, April). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. Retrieved 2023-03-26, from <https://joss.theoj.org/papers/10.21105/joss.03021> doi: 10.21105/joss.03021
- Webster, G. A., & Perkins, N. D. (1999, May). Transcriptional Cross Talk between NF- κ B and p53. *Molecular and Cellular Biology*, 19(5), 3485–3495. Retrieved 2023-07-04, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC84141/>
- Weetall, M., Hugo, R., Maida, S., West, S., Wattanasin, S., Bouhel, R., ... Friedman, C. (2001, June). A Homogeneous Fluorometric Assay for Measuring Cell Adhesion to Immobilized Ligand Using V-Well Microtiter Plates. *Analytical Biochemistry*, 293(2), 277–287. Retrieved 2022-09-25, from <https://linkinghub.elsevier.com/retrieve/pii/S0003269701951401> doi: 10.1006/abio.2001.5140
- Weiss, C. J. (2022, September). Visualizing protein big data using Python and Jupyter notebooks. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology*, 50(5), 431–436. doi: 10.1002/bmb.21621
- Wickham, H. (2014, September). Tidy Data. *Journal of Statistical Software*, 59, 1–23. Retrieved 2023-11-15, from <https://doi.org/10.18637/jss.v059.i10> doi: 10.18637/jss.v059.i10
- Wilkins, A., Kemp, K., Ginty, M., Hares, K., Mallam, E., & Scolding, N. (2009, July). Human bone marrow-derived mesenchymal stem cells secrete brain-derived neurotrophic factor which promotes neuronal survival in vitro. *Stem cell research*, 3(1), 63–70. doi: 10.1016/j.scr.2009.02.006
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016, March). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. Retrieved 2024-03-18, from <https://www.nature.com/articles/sdata201618> doi: 10.1038/sdata.2016.18
- Witwer, K. W. (2013, February). Data submission and quality in microarray-based microRNA profiling. *Clinical chemistry*, 59(2), 392–400. Retrieved 2024-04-22, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4037921/> doi: 10.1373/clinchem.2012.193813
- Wong, A. D., & Searson, P. C. (2017, November). Mitosis-mediated intravasation in a tissue-engineered tumor-microvessel platform. *Cancer research*, 77(22), 6453–6461. Retrieved 2023-07-14, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5690825/> doi: 10.1158/0008-5472.CAN-16-3279
- Xu, W., Zhang, X., Qian, H., Zhu, W., Sun, X., Hu, J., ... Chen, Y. (2004, July). Mesenchymal stem cells from adult human bone marrow differentiate into a cardiomyocyte phenotype in vitro. *Experimental biology and medicine (Maywood, N.J.)*, 229(7), 623–631.
- Yang, A., Troup, M., & Ho, J. W. (2017, July). Scalability and Validation of Big Data Bioinformatics Software. *Computational and Structural Biotechnology Journal*, 15, 379–386. Retrieved 2024-03-07, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5537105/> doi: 10.1016/j.csbj.2017.07.002
- Zeissig, M. N., Zannettino, A. C. W., & Vandyke, K. (2020, December). Tumour Dissemination in Multiple Myeloma Disease Progression and Relapse: A Potential Therapeutic Target in High-Risk Myeloma. *Cancers*, 12(12). Retrieved 2021-02-03, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7761917/> doi: 10.3390/cancers12123643
- Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., ... Flicek, P. (2018, January).

- Ensembl 2018. *Nucleic Acids Research*, 46(D1), D754-D761. Retrieved 2023-05-27, from <https://doi.org/10.1093/nar/gkx1098> doi: 10.1093/nar/gkx1098
- Zhou, F., Meng, S., Song, H., & Claret, F. X. (2013, November). Dickkopf-1 is a key regulator of myeloma bone disease: Opportunities and challenges for therapeutic intervention. *Blood reviews*, 27(6), 261–267. Retrieved 2023-02-18, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4133945/> doi: 10.1016/j.blre.2013.08.002
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., ... Chanda, S. K. (2019, April). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, 10(1), 1523. Retrieved 2023-02-09, from <https://www.nature.com/articles/s41467-019-09234-6> doi: 10.1038/s41467-019-09234-6
- Ziemann, M., Eren, Y., & El-Osta, A. (2016, August). Gene name errors are widespread in the scientific literature. *Genome Biology*, 17(1), 177. Retrieved 2024-04-30, from <https://doi.org/10.1186/s13059-016-1044-7> doi: 10.1186/s13059-016-1044-7

Appendices

A Supplementary Data & Methods

A.1 Figures

A.2 Tables

A.3 Materials & Methods

B Documentation of `plotastic`