VRIJE UNIVERSITEIT AMSTERDAM

UNIVERSITEIT VAN AMSTERDAM

Master Thesis

Benchmarking transcript splicing variants tools and detection of novel drivers in sarcomas subtypes

by

Karlijn A. Elias

(VUNETID: kes540, Student number: 2706872.
Confidential)

LUMC

First supervisor:	Dr. MD Karoly Szuhai (K. Szuhai)
Daily supervisor:	Dr. Leon Mei (H. Mei)
Examinator:	Dr. K. Anton Feenstra (K.A. Feenstra)
Course ID:	XM_0103
Number of credit points:	60 EC
Research project's institute:	Leiden University Medical Center
Research project's institute's department:	Cell and Chemical Biology, section molecular cytogenetica

September 26, 2023

Submitted in partial fulfillment of the requirements for the joint UvA-VU degree of Master of Science in Bioinformatics and Systems Biology

Benchmarking transcript splicing variants tools and detection of novel drivers in sarcomas subtypes

Karlijn A. Elias
Vrije Universiteit Amsterdam and Universiteit van Amsterdam
Amsterdam, The Netherlands
k.a.elias@student.vu.nl

ABSTRACT

Important:
"Thesis style and length
Your thesis should be written in the format of an academic article. The main text (excluding title page and references) should be no longer than 3 pages (double-column) per month (6 EC), or 2500 words per month with 1-2 figures per month (6 EC); note that it may be shorter." → 30 pages (6 EC × 10 = 60 EC) or 2500 words × 10 with 10-20 figures/tables.

- This is a summary of the work
- Maximum of 300 word
- You should describe
  - The background of the research
  - Obtained results
  - The potential impact of the research on society and/or other research

1 INTRODUCTION

- Including a motivation: why is your work important?
- Start broad and narrow it down towards the research question
- State the research question explicitly
- Both the biological research question, as prior methodology that is similar should be introduced
  - You need give an overview and cite relevant earlier work with respect to both of these

Mesenchymal tumors are a collection of more than 160 different types and within them including benign and malignant forms (WHO Classification of Tumours Editorial Board, 2020). Sarcomas are malignant bone and soft tissue tumors (WHO Classification of Tumours Editorial Board, 2020) comprising more than 60 malignancies belong of bone and soft tissue tumors (Nacev et al., 2020; Carmagnani Pestana et al., 2019). They differ in biological and clinical features and are very rare (WHO Classification of Tumours Editorial Board, 2020; Nacev et al., 2020). There are about 1% of all adult human malignancies, and 20% of all childhood malignant cancers (WHO Classification of Tumours Editorial Board, 2020). One of these malignant sarcoma subtypes is Ewing sarcomas of tumors which is part of/includes small blue round cell tumors (SBRCTs) (Mendpara et al., 2023).

SBRCTs are undifferentiated mesenchymal tumors with uniform round nuclei and very little cytoplasm making them difficult to distinguish histologically (Lin et al., 2022; Antonescu, 2013). SBRCTs include Ewing sarcoma, rhabdomyosarcoma, mesenchymal chondrosarcoma and several rarer and emerging entities with distinctive molecular drivers, prognoses and responses to therapy (Mendpara et al., 2023; Lin et al., 2022; Rekhi et al., 2019). Ewing sarcomas shows gene fusions of mostly *EWSR1* with an *ETS* family of transcription factors (WHO Classification of Tumours Editorial Board, 2020). Another entity in SBRCTs is the round cell sarcoma with *EWSR1*-non-*ETS* fusions, also fusion with *FUS* occurs in this entity. Furthermore, *CIC*-rearranged sarcoma is sarcoma with *CIC*-related gene fusions. The gene fusion most occurring is *CIC-DUX4*. In SBRCTs, sarcoma with *BCOR* genetic alterations such as gene fusions occur (WHO Classification of Tumours Editorial Board, 2020), There is also a group of unclassified

small round cell sarcoma with no clear differentiation (Renzi et al., 2019). *EWSR1*-non-*ETS* sarcomas, *CIC*-rearranged sarcomas, *BCOR* genetic alterations sarcomas, and unclassified round cell sarcomas are called Ewing-like sarcoma; those entities are poorly understood but contain 11% of the cases of round cell sarcoma (Renzi et al., 2019). Figure 1 illustrates the gene fusions known so far in Ewing and Ewing-like sarcoma.

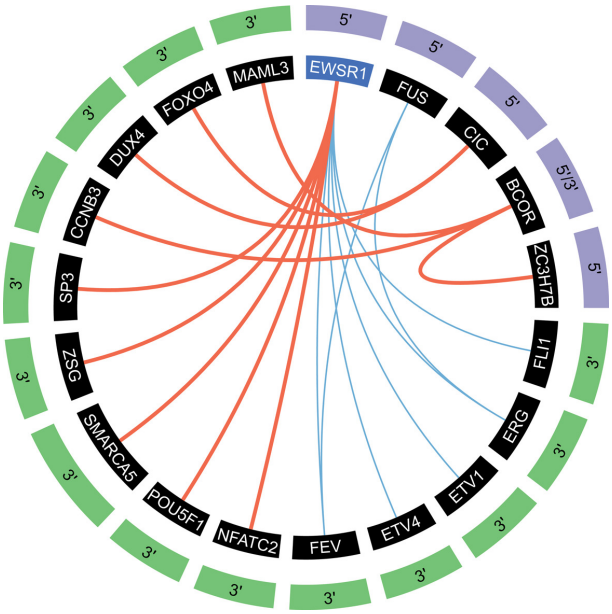

Figure 1: Gene fusions in Ewing and Ewing-like sarcoma. *Blue lines are the gene fusions occurring in Ewing sarcoma, and the red thicker lines are the Ewing-like sarcoma fusions. Green rectangles and purple are the 3' and 5' partners, respectively. Source: Renzi et al., 2019.*

The standard treatment for sarcomas on localized places The standard treatment for sarcomas on localized places is surgery in combination with pre-operative or post-operative chemotherapy and or radiation (Damerell et al., 2021). Survival chance and complete recovery after this treatment is high among patients with localized sarcomas. However, 10 to 20% and up to 50% of the patients have recurrence and metastasis development respectively. When metastasis happens, chemotherapy is the standard method as treatment, but the survival rate overall is low. This indicates that existing therapies are not always effective and require new therapies to treat sarcomas (Damerell et al., 2021). Therefore, targeted therapies are required. There are already existing these types of therapies (Damerell et al., 2021), but since sarcomas are very rare (WHO Classification of Tumours Editorial Board, 2020), it is important to distinguish molecular features.

An important molecular feature to distinguish these different sarcoma subtypes and to understand their biological difference is differential gene and transcript level expression. However, the currently used standard gene and transcript database still lacks rare and specific transcript annotations that might be crucial in sarcoma related tissues. An important molecular feature to distinguish these subtypes and to understand their biological difference is differential gene and transcript level expression. That creates the focus on the main research question to detect novel transcript splicing variants that can be potential key marker transcripts and further understand the underlying molecular mechanisms behind some of these sarcoma subtypes. First, bioinformatics tools and methods detecting novel splicing transcript variants will be benchmarked for known gene regions using Illumina short reads to identify novel transcript variants, thereby improving transcript annotation. Using this improved transcript annotation, differential transcript analysis methods will be performed and compared with literature to detect potential key marker transcripts that can be a driver molecular events or key passenger events. This research can be extended to find novel transcripts located in non-annotated gene regions.

2 MATERIAL AND METHODS

- 'The first aim of the methods should be that another researcher (think MSc student) can reproduce you work
- Be very precise and specific in the methods
- Nethertheless, try to explain why you make certain choices in the methodology

3 RESULTS

- Here you should show the evidence answering your research question.
- The results need to be present by showing the data using figures and tables.
- The results also need to be stated in the main text.
- The results are typically structured around strategically chosen figures and tables that represent your data and outcomes.
- You can briefly reflect if the results follow expectations, speculations should go into the discussion

4 DISCUSSION

- This should be centred around comparing your results to other research
- You need to cite other work, and compare this to your results.
- In addition you can make recommendations for future research in the same direction, and highlight any unresolved problems or uncertainties in your data
- No new results can be presented.

5 CONCLUSION

- No new facts!
- Can in some cases be combined with the discussion. In this case, the conclusions (answering your research question) should be the first one or two paragraphs of the discussion.

6 ACKNOWLEDGEMENTS

- Scientific work is never performed in isolation. This means you must acknowledge when and where you have been helped by others, which other work you have been building on, etc

7 AVAILABILITY

- Explain where the data, scripts and protocols are stored.
- If possible, provide a link to the repositories, otherwise provide a contact person

References:

- Citations in the text as (Author, Year)
- All items in the reference list must be cited in the main text

REFERENCES

Antonescu, C. (2013). Round cell sarcomas beyond e wing: Emerging entities. *Histopathology*, *64*(1), 26–37.

Carmagnani Pestana, R., Groisberg, R., Roszik, J., & Subbiah, V. (2019). Precision oncology in sarcomas: Divide and conquer. *JCO Precision Oncology*, *3*, 1–16.

Damerell, V., Pepper, M. S., & Prince, S. (2021). Molecular mechanisms underpinning sarcomas and implications for current and future therapy. *Signal transduction and targeted therapy*, *6*(1), 246.

Lin, Y.-W., Hsu, Y.-H., & Lee, M.-Y. (2022). Adult patient with neuroblastoma presenting as acute leukemia. *Cureus*, *14*(8).

Mendpara, V., Ali, A. S. M., Tango, T., Bhadana, R., Kanisetti, V., Tiwari, U., Sahu, S., & Pancholi, M. (2023). A case of extraskeletal ewing sarcoma treated with wide local excision with latissimus dorsi flap and systemic therapy. *Cureus*, *15*(3).

Nacev, B. A., Jones, K. B., Intlekofer, A. M., Yu, J. S., Allis, C. D., Tap, W. D., Ladanyi, M., & Nielsen, T. O. (2020). The epigenomics of sarcoma. *Nature Reviews Cancer*, *20*(10), 608–623.

Rekhi, B., Mridha, A., Kattoor, J., et al. (2019). Small round cell lesions of the bone: Diagnostic approach, differential diagnoses and impact on treatment. *Indian Journal of Pathology and Microbiology*, *62*(2), 199.

Renzi, S., Anderson, N. D., Light, N., & Gupta, A. (2019). Ewing-like sarcoma: An emerging family of round cell sarcomas. *Journal of cellular physiology*, *234*(6), 7999–8007.

WHO Classification of Tumours Editorial Board. (2020). *Soft tissue and bone tumours* (5th, Vol. Volume 3). World Health Organization.

SUPPLEMENTAL INFORMATION