A Review on Predictive Modeling for Adrenocortical Carcinoma

Daniel-Dragos Braghis
School of Electronics and Computer Science
University of Southampton
UoS ID: 32161204

1. Introduction

Adrenocortical cancer, also known as adrenal cortex carcinoma or adrenal cortical carcinoma (ACC), is a rare and aggressive malignancy that arises from the adrenal cortex, the outer layer of the adrenal glands. These glands are situated on top of each kidney and play a crucial role in producing hormones essential for regulating various bodily functions, including metabolism, blood pressure, and stress response.

Adrenocortical cancer accounts for only a small fraction of all cancer cases, with an estimated incidence of around 200 cases per year registered in the US [1]. Despite its rarity, adrenocortical cancer presents significant challenges in diagnosis, treatment, and prognosis due to its aggressive nature.

The clinical presentation of adrenocortical cancer can vary widely, with symptoms often related to hormone overproduction or effects on adjacent organs. Common symptoms include hormone excess (usually Hypercotisolism), abdominal pain with associated muscle weakness and weight loss, and hypertension, leading to symptoms such as hirsutism, virilization, or Cushing's syndrome [2].

This review aims to validate previous observations of Adrenocortical cancer in the context of the GDC TCGA Adrenocortical Cancer dataset provided by the University of California Santa Cruz through the Xena API and explore potential predictive modelling for this rare type of cancer.

2. Background

2.1. Phenotype

ACC exhibits notable gender biases in its phenotypic presentation across different age groups. In childhood, ACC typically manifests in the first 5 years with a higher incidence in females, showing a female-to-male ratio of 1.7:1 for patients younger than 3 years [2], [3]. This gender disparity becomes even more pronounced during adolescence. For patients over 13 years, females are significantly more affected than males with a ratio of 6.2:1, potentially indicating a hormonal or genetic predisposition that heightens that risk for females [3]. Even if less stark, this bias continues into adulthood with the ratio ranging from 1.5:1 to 2.5:1, consistently showing higher incidence in females [2].

This persistent gender difference across all age groups suggests an underlying biological mechanism in female development that facilitates ACC, highlighting the importance of considering gender as a significant factor in the clinical management and study of ACC.

Overall, ACC observes a 5-year survival rate of 38% [4]. Despite the generally unfavourable prognosis, individual variation in disease progression and recurrence is notable. Patients with stage 4 disease can survive anywhere from a few months to several years with some exceptionally long-term survival reported [2]. Prognostic factors for ACC have not been definitely established. While age at diagnosis is correlated with decreased overall survival, it's unclear if this applies to tumour-free survival [2]. Despite preoperative imaging, about 25% of stage 3 ACC cases are misclassified as stage 2, delaying appropriate treatment and reducing overall survivability [4].

2.2. Copy Number Variation

It is generally observed that copy number variations (CNVs) are closely related to the development and progression of various cancer types through alterations of gene expression levels [5]. Results (n=9159) have demonstrated a close correlation between CNV and differential gene expression with a positive linear influence on the majority of genes with clusters in disease-related pathways [5].

Genomic aberrations in 25 ACC clinical samples were analyzed, revealing commonly shared amplifications and deletions across the genome [6]. Amplifications were frequently observed in chromosomes 5, 7, 12, 16q, and 20, while deletions occurred in portions of chromosomes 1, 3p, 10q, 11, 14q, 15q, 17, and 22q. Additionally, genomic aberrations associated with differential survival were identified, including amplifications of 6q, 7q, 12q, and 19p, and deletions in specific regions of 3, 8, 10p, 16q, 17q, and 19q [6]. Findings suggest potential therapeutic options for ACC successfully clustering genomic aberrations into distinct groups associated with various survival outcomes [6].

Delving deeper, current research [2], [6] brings attention to several genes closely related to ACC: Beckwith-Wiedemann syndrome (BWS) involving alterations at the 11p15 locus affecting IGF2 - a key regulator of growth, CDKN1C also found on the 11p15 locus contributes to

tumour development in BWS, MEN1 at locus 11q13 which is usually associated with adrenal adenomas and less commonly ACC, Li Fraumeni Syndrome (LFS) associated with TP53 mutation highly prevalent in childhood ACC (50%-80%), rarely Neurofibromatosis type 1 (NF1 gene), PRKAR1A referenced in some rare case studies of Carney complex, and finally EPCAM mutations as part of Lynch syndrome involving mismatch repair genes (MSH2, MSH6, MLH1, PMS2, MSH6) that are commonly associated with a broad spectrum of cancers. Recent studies [7] have also confirmed recurrent homozygous deletions of ZNRF3, KREMEN1 and TERT amplification through whole-exome sequencing. These findings invite further research into targeted therapies tailored to the specific genetic profiles of ACC tumours, ultimately aiming to improve patient outcomes in this challenging cancer type.

2.3. Survival Prediction Models

Predictive models are essential for improving the management and outcomes of ACC patients. Several recent studies have contributed valuable insights into developing these predictive models.

In a single-centre analysis(n=67) using Kaplan-Meier methods and the Cox regression model it was found that the ENSAT stage is significantly associated with OS (HR 3.60; p=0.018) regardless of patients having stage 2 or 3 cancers [8]. Mitotane use improved clinical outcomes in stage 4 patients (HR 0.27; p=0.007) and venous invasion being associated with poor OS (HR 5.19; p=0.003) [8].

Another analysis of the American College of Surgeons National Cancer Database (NCDB)(n=3480) with SAS software and multivariate analysis on clinical and demographic factors has shown promise [9] with another study on the SEER Database(n=583) reinforcing that phenotype features showed good correlation between predicted and actual survival outcomes [11].

A multi-gene signature model identified 93 survival-associated genes and constructed a Cox proportional hazard regression model incorporating 14 genes, validated using three independent datasets. The resulting model showed high accuracy for predicting OS compared to other models and was independent of traditional clinical features like tumour stage [10].

3. Experimental Findings

3.1. Phenotype

We performed a phenotypic analysis in conjunction with survival statistics(n=95) to identify potential correlations that would help predict survival outcomes. First, we studied the age at diagnosis and gender distribution (Figure 1,2). Our dataset follows the general bias toward female ACC occurrence with a 2:1 ratio [2], [3], but the dataset does not provide much information on childhood ACC incidence (Figure 1) with only 7 samples attributed to the below 24-year-old ACC cases and sticks to the general statistic [4] with most of the sample representation concentrated at the older population.

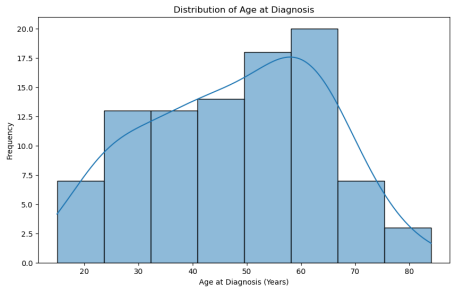


Figure 1. Gene copy number values and gene expression correlation table

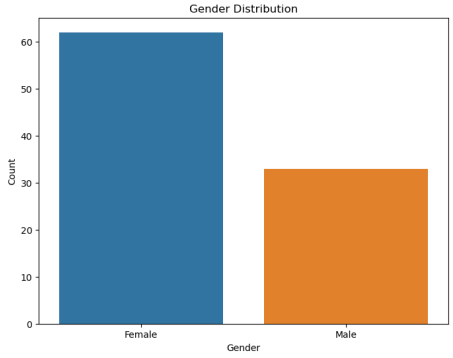


Figure 2. Gene copy number values and gene expression correlation table

Next, we plotted the distribution of overall survival time (Figure 3) and survival outcome by gender (Figure 4) to better understand any biases present and the general fatality rate of ACC. We note that, indeed, the majority of diagnosed patients have a below 5-year life expectancy after diagnosis [4] with approximately 30% mortality rate in our dataset. Interestingly, there doesn't seem to be any gender bias in survival outcomes with both genders registering the same survival rate thus excluding gender-specific disease progression in conjunction with an already established bias towards female occurences.

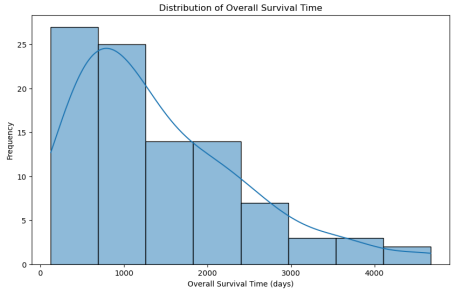


Figure 3. Gene copy number values and gene expression correlation table

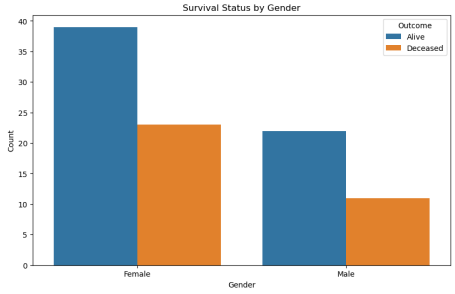


Figure 4. Gene copy number values and gene expression correlation table

As evidenced in background literature, stage 2 tumours are rare in ACC, frequently misclassified and later attributed to stage 3 [4]. This is true in our dataset too, with half the occurrences being classified as stage 1, a quarter as stage 0, and the rest as stage 3 or 4 with only a few cases receiving stage 2 classification (Figure 5). The dangers of ACC are apparent with mortality at 30% even with 75% detection in the early stages. This is exacerbated by occurrences in younger populations that do not frequently go through medical testing, leaving the diseases undetected until progression to the 4th stage, as depicted in Figure 6 which has a lower median age when diagnosed as stage 4 than the other stages. These statistics are strong evidence for survival outcome correlation with tumour stage at diagnosis, as usually expected in aggressive cancers like ACC.

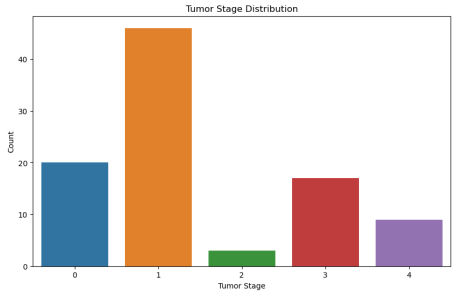


Figure 5. Gene copy number values and gene expression correlation table

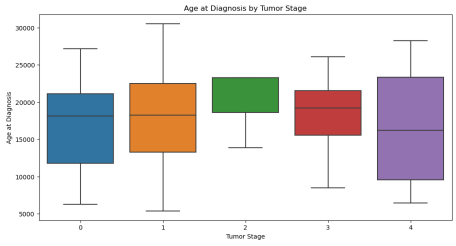


Figure 6. Gene copy number values and gene expression correlation table

This is further evidenced by the Kaplan-Meier curve for our dataset (Figure 7). There are 2 distinct areas in our plot: The first 2 years after diagnosis when the survival rate stays above 75% and the next 5 years when survivability drops to 50% before plateauing thereafter with confidence intervals very large past 7 years, indicating that most patients die or recover in this time frame with chances of survival probably dropping significantly.

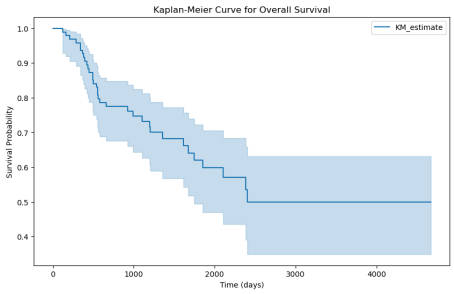


Figure 7. Gene copy number values and gene expression correlation table

Considering earlier observations, we selected a subset of phenotypic features that showed a relatively high correlation with survival time and outcome in our dataset (Appendix A). It is important to note that the tumour stage in our dataset(n=90) at discovery has little effect on survival time and outcome with even the Weiss score showing a weak correlation. Our findings show Mitotane Levels post-surgery as a better indication of the patient's survival chances which indicates that surgical intervention is relatively effective at curing ACC compared to Primary Therapy which mostly increases survival time. For a predictive model, we will still keep the features with less correlation, as they might have more predictive power in conjunction with other features.

3.2. Copy Number Variation

To experimentally validate previous study findings [5], [6], the Masked Copy Number Segment data (n=90) was binned and displayed on a heat map to inform further exploration into copy number alterations (Appendix B). Key observations from studying the heat map reveal:

- Chromosome 5: There is a prominent region of amplification indicated by dark red, suggesting a significant gain in copy number in this region across many samples. This might point to oncogenes or regions critical to ACC pathogenesis.
- Chromosome 22: A recurrent deletion is observed (blue region), which aligns with the findings of deletions at 22q12.1 [6], including the ZNRF3 locus.
- Chromosome 9: Shows a mix of deletions and amplifications, suggesting complex genomic alterations in this region.
- Chromosome 17: Displays a pattern that might correspond to deletions, possibly involving the TP53 locus, which is known to be deleted or mutated in ACC [2].
- Chromosome 12: Exhibits significant changes, potentially involving regions like 12q13-15, which have been implicated in other cancers.

The consistent amplification of regions like 5p15.33, including TERT, and recurrent deletions in regions like

22q12.1 highlight the loss of tumour suppressors such as ZNRF3 and KREMEN1 showing potential as therapeutic targets but a more detailed analysis was needed for specific gene regions. To achieve this, we plotted the start and end points of copied gene strands with their copy number value while highlighting clusters that showed aberrations in copy values through the majority of samples and labelling gene segments of particular interest from previous studies [2], [6], [7] (Appendix C).

As expected from background literature [6], we observe amplifications on chromosomes 5, 7, 12, partially 16 and 20 with some notable areas also on chromosomes 3 and 20. Deletions occurred, as predicted, on chromosomes 1 and 10 (with various aberrations), 11, and 17 but also on chromosomes 3 and 4. However, we haven't noted any particular clusters on chromosomes 14 and 15. We have also highlighted with yellow bars the segments that have been linked to cancers, confirming with strong evidence the commonality of aberrant copy number values in TP73, IGF2, CDKN1C, PRKARIA, MLH1, PMS2, TERT, KREMEN1 and ZNRF3 (last 3 evident in particular). MLH1, TP53, EPCAM, MSH6, NF1 and MEN1 also go through smaller clusters of copy variations supporting their connection to the development of ACC, but not as pronounced.

It was also hinted that there is a correlation between CNV and differential gene expression [5] which we tested by querying data(n=79) and calculating correlation for the previously identified genes associated with ACC with the following results:

	Gene	Probe ID	Correlation
0	TP73	ENSG00000078900.13	-0.013629
1	IGF2	ENSG00000167244.16	-0.055487
2	CDKN1C	ENSG00000129757.11	0.139110
3	MEN1	ENSG00000133895.13	0.511327
4	TP53	ENSG00000141510.14	0.321150
5	NF1	ENSG00000196712.15	0.173659
6	PRKAR1A	ENSG00000108946.13	0.221991
7	EPCAM	ENSG00000119888.9	-0.116947
8	MSH6	ENSG00000116062.13	0.197541
9	ZNRF3	ENSG00000183579.14	0.296148
10	KREMEN1	ENSG00000183762.11	NaN
11	MLH1	ENSG00000076242.13	-0.024675
12	TERT	ENSG00000164362.17	0.167550
13	PMS2	ENSG00000122512.13	-0.000458

Figure 8. Gene copy number values and gene expression correlation table

While TP53, MEN1 and ZNRF3 registered reasonable correlation, and CDKN1C, NF1, PRKAR1A, MSH6, and TERT had weak correlation, the rest of the genes showed no correlation whatsoever (or data was not available). These findings show that while there is a correlation between CNV and gene expression this is not always true and additional information is needed for this data to be used for predicting clinical outcomes.

Our findings corroborate existing literature on the genomic alterations involved in ACC and we additionally observe strong evidence of copy value clusters on chromosomes 3 and 4. We also confirmed a moderate correlation between CNV and gene expression for TP53, MEN1 and ZNRF3, even if not very strong, in our dataset. This might indicate the presence of yet-to-be-identified oncogenes or tumour suppressor genes that play a role in the development or progression of ACC. It is not excluded that these regions reflect the general instability of cancerous DNA which characterizes ACC and co-occurs with known alterations such as TP53, IGF3 and CDKN1C.

3.3. Survival Prediction Model

To explore predictive models based on our earlier findings we combined the phenotype data with CNV and gene expression datasets for training. The resulting dataset was pre-processed to train OS predictions - a classification problem thus any rows with NaN values for OS were removed leaving 94 samples for training and validation. We chose a 70/30 training/validation split because of the small amount of data available to make assessment of performance closer to reality with so few samples available. We chose several models to evaluate performance on our dataset based on previous research [9] [10]: Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and Support Vector Machines. We also evaluated them against a random guesser for sanity checking of our models. Figure 9 shows that Gradient Boosting achieved the best accuracy for our dataset with 0.8 accuracy. Decision Tree and Random Forest are slightly behind with 0.78 accuracy. SVM and Logistic Regression had the worst performance well below 0.7 accuracy. All models tested performed notably better than a random guesser and matched or exceeded (except Logistic Regression) the statistical performance for our 30% fatality rate dataset with a model that would just guess one survival outcome.

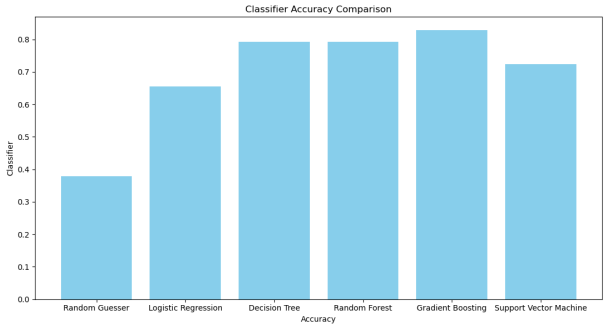


Figure 9. Classifier Accuracy Chart

While results are satisfactory, unfortunately, we can not conclude decisively that our trained models will perform equally well in the real world due to biases in our very small dataset like very few young patient samples.

However, we were interested in highlighting the features and biomarkers that improved the predictive power of our models the most. Found in Appendix D, are the feature weights at the end of the training cycle for the Decision Tree, Random Forest and Gradient Boosting. It can be observed that Post Surgical Procedure Carcinoma Status was the most

decisive feature in all three models. This is not surprising as surgical removal of tumours has long been the most effective method of curing cancer and the success of this operation plays a pivotal role in the success of the overall therapy. It is also important to note, that CNV and gene expression helped all three models even with low overall weighting, proving their predictive ability in ACC survival prognosis.

Following our experimental results, we identified the following areas for future research and improvements:

- A larger, more diverse dataset would help in training more robust models and potentially uncover new predictive features.
- Implementing advanced imputation techniques, such as multiple imputation by chained equations (MICE) or K-nearest neighbors (KNN) imputation, to handle missing data more effectively and reduce potential biases.
- Exploring hybrid models that integrate both statistical and machine learning approaches to leverage the strengths of each method.
- Testing the trained models on external datasets to assess their generalizability and ensure that they perform well across different populations and settings.
- Using techniques like LIME (Local Interpretable Model-agnostic Explanations) to interpret model predictions and provide actionable insights to clinicians.

4. Conclusion

In this study, we analyzed phenotypic data alongside survival statistics to predict overall survival (OS) outcomes in adrenocortical carcinoma (ACC) patients. Consistent with existing literature, our dataset expressed bias towards female patients and older age groups. Only a small fraction of the data represented young patients under 24 years old. Survival analysis indicated no gender bias in OS outcome with a 30% mortality rate predominantly within the first 7 years post-diagnosis. The tumour stage at diagnosis was a significant predictor of survival, emphasising the importance of early detection.

We explored various predictive models using a combined dataset of phenotype, copy number and gene expression data. Gradient Boosting emerged as the most accurate model with an 80% accuracy rate. While our models performed better than random guessing, the small dataset size limited the robustness of our conclusions.

Feature analysis indicated that post-surgical carcinoma status was an important predictor across multiple models. CNV and gene expression data, while less impactful, also contributed to predictive accuracy. Future research should focus on a larger sample size and validation with external datasets.

References

[1] American Cancer Society. Cancer Facts and Figures 2023. Atlanta: American Cancer Society, Inc. 2023.

[2] T. Else et al., "Adrenocortical carcinoma," (in eng), Endocr Rev, vol. 35, no. 2, pp. 282-326, Apr 2014, doi: 10.1210/er.2013-1029.

[3] A. Levasseur, T. Dumontet, and A. Martinez, ""Sexual dimorphism in adrenal gland development and tumorigenesis"," Current Opinion in Endocrine and Metabolic Research, vol. 8, 08/01 2019, doi: 10.1016/j.coemr.2019.07.008.

[4] M. Ayala-Ramirez et al., "Adrenocortical carcinoma: clinical outcomes and prognosis of 330 patients at a tertiary care center," (in eng), Eur J Endocrinol, vol. 169, no. 6, pp. 891-899, Dec 2013, doi: 10.1530/eje-13-0519.

[5] X. Shao et al., "Copy number variation is highly correlated with differential gene expression: a pan-cancer study," (in eng), BMC Med Genet, vol. 20, no. 1, p. 175, Nov 9 2019, doi: 10.1186/s12881-019-0909-5.

[6] E. A. Stephan et al., "Adrenocortical carcinoma survival rates correlated to genomic copy number variants," (in eng), Mol Cancer Ther, vol. 7, no. 2, pp. 425-31, Feb 2008, doi: 10.1158/1535-7163.Mct-07-0267.

[7] C. C. Juhlin et al., "Whole-Exome Sequencing Characterizes the Landscape of Somatic Mutations and Copy Number Alterations in Adrenocortical Carcinoma," The Journal of Clinical Endocrinology & Metabolism, vol. 100, no. 3, pp. E493-E502, 2015, doi: 10.1210/jc.2014-3282.

[8] L. H. Pan, C. C. Yen, C. J. Huang, X. N. Ng, and L. Y. Lin, "Prognostic predictors of adrenocortical carcinoma: A single-center thirty-year experience," (in eng), Front Endocrinol (Lausanne), vol. 14, p. 1134643, 2023, doi: 10.3389/fendo.2023.1134643.

[9] S. M. Zuber et al., "Validated predictive model for treatment and prognosis of adrenocortical carcinoma," (in eng), Surgery, vol. 175, no. 3, pp. 743-751, Mar 2024, doi: 10.1016/j.surg.2023.08.047.

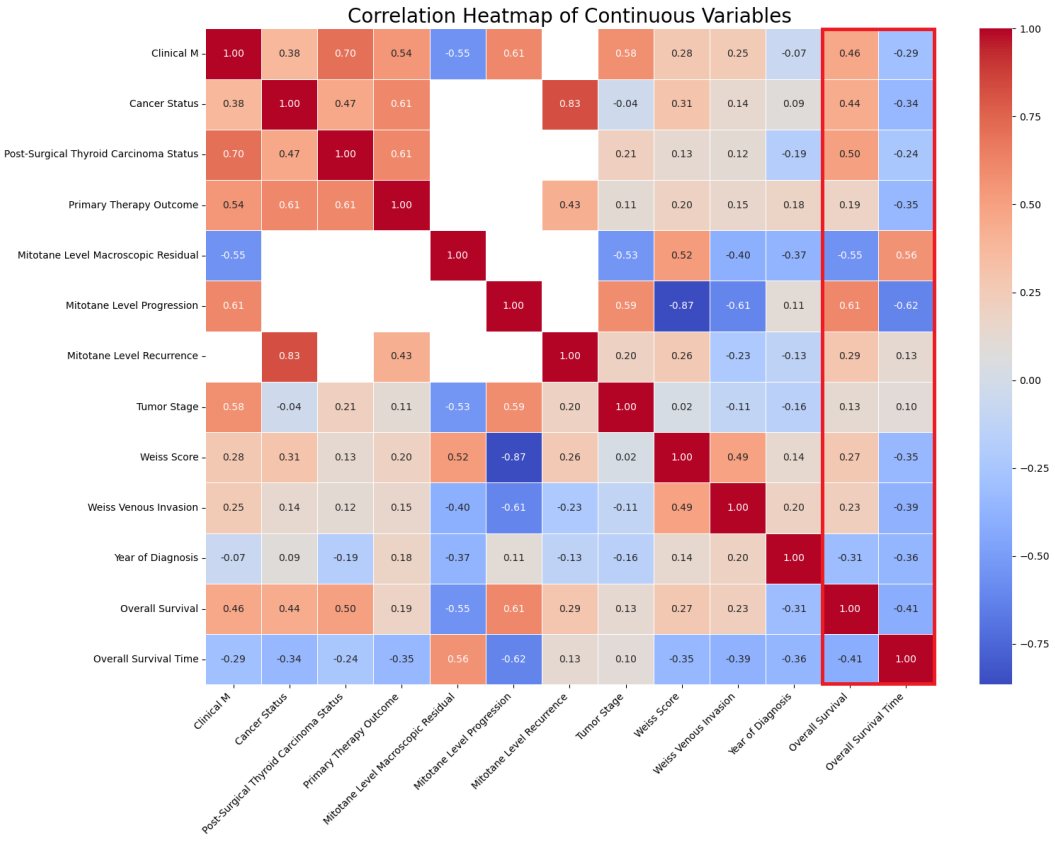
[10] X. Yan et al., "Identification and Validation of a Novel Prognosis Prediction Model in Adrenocortical Carcinoma by Integrative Bioinformatics Analysis, Statistics, and Machine Learning," (in eng), Front Cell Dev Biol, vol. 9, p. 671359, 2021, doi: 10.3389/fcell.2021.671359.

[11] S. Shebrain, "Prediction of Survival in Adrenocortical Carcinoma," Journal of Investigative Surgery, vol. 35, no. 5, pp. 1161-1162, 2022/05/04 2022, doi: 10.1080/08941939.2021.1991057.

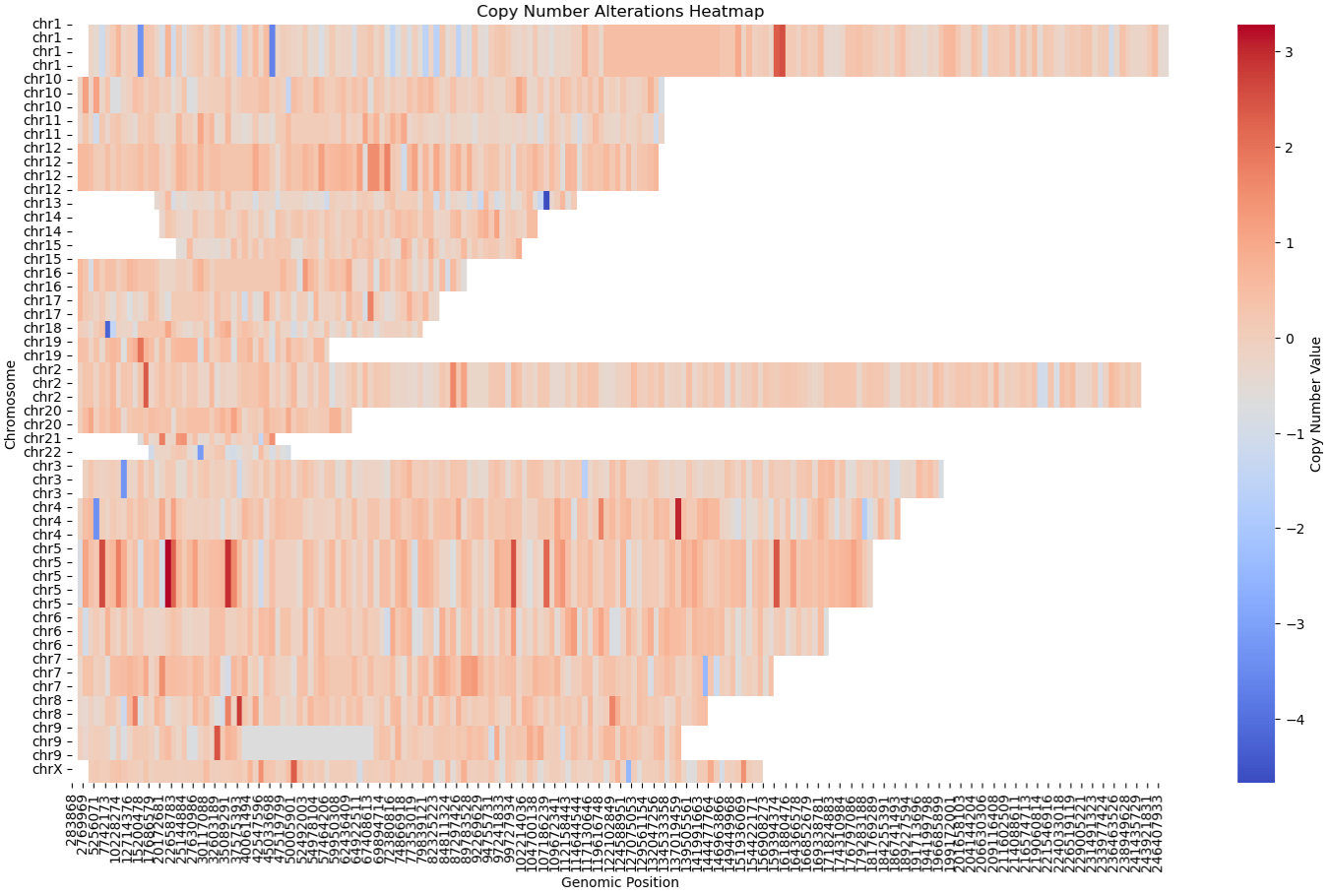
Appendix A.
Correlation heat map

Definitions:

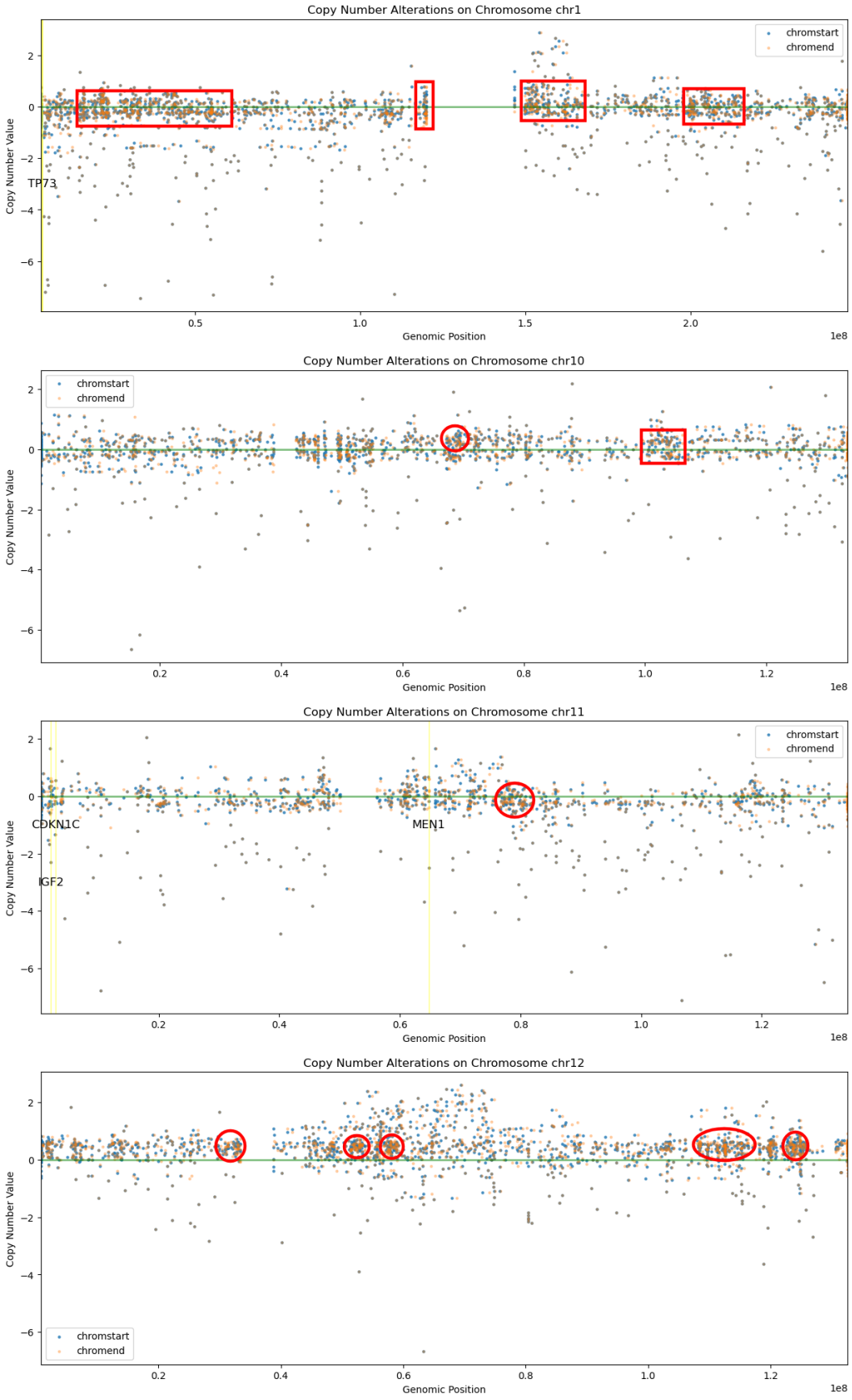
- Clinical M: Metastasis component of the TNM (Tumor, Node, Metastasis) cancer staging system, indicates whether the cancer has spread to other parts of the body.
- Cancer Status: eg. ”No evidence of disease”, ”Active disease”, ”Remission”.
- Post-Surgical Thyroid Carcinoma Status: Indicates whether there is residual disease left after surgical intervention.
- Mitotane Level Macroscopic Residual: Indicates the therapeutic level of mitotane (a drug used to treat adrenocortical carcinoma) related to visible tumours remaining after surgery.
- Mitotane Level Progression: Indicates the therapeutic level of mitotane in relation to disease progression.
- Mitotane Level Recurrence: Indicates the therapeutic level of mitotane in relation to disease recurrence.
- Tumour Stage: Value between 0 and 4.
- Weiss Score: This is a scoring system used to distinguish between benign and malignant adrenocortical tumours.
- Weiss Score Venous Invasion: This feature indicates whether venous invasion is present according to the Weiss criteria - one of the factors considered in the Weiss score.

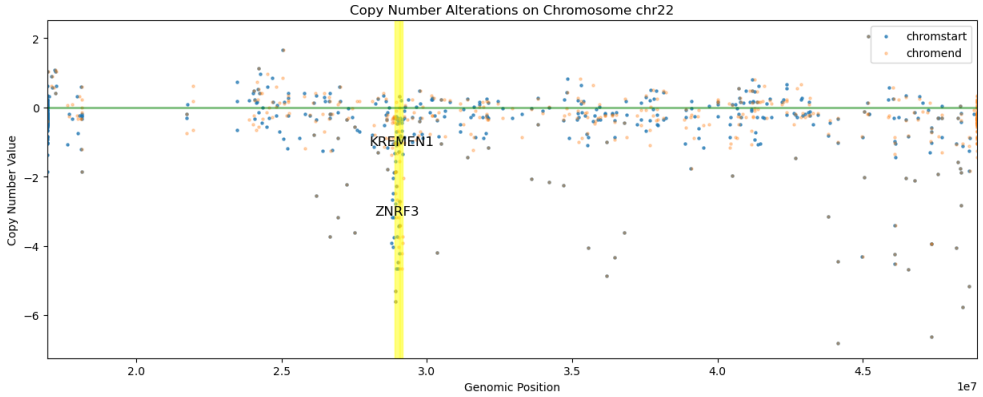


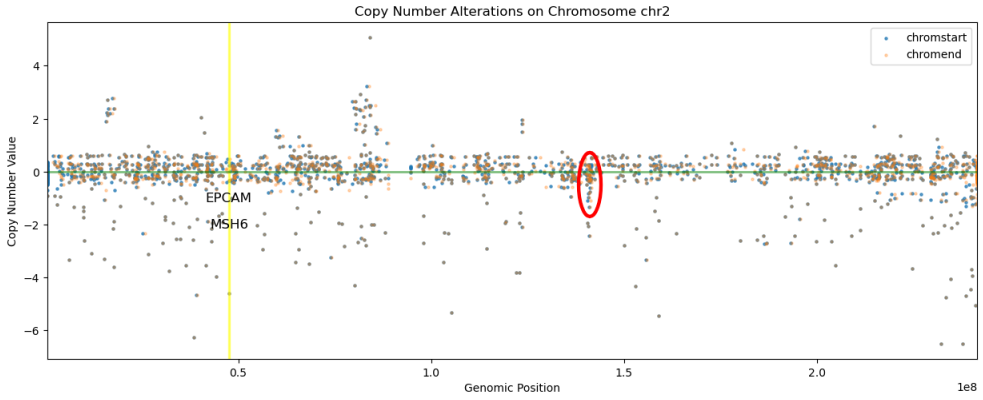
Appendix B.
Copy number heat map

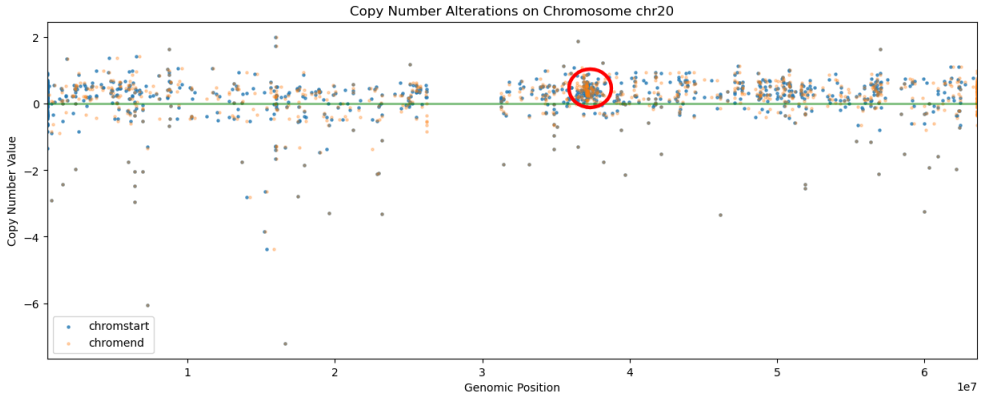


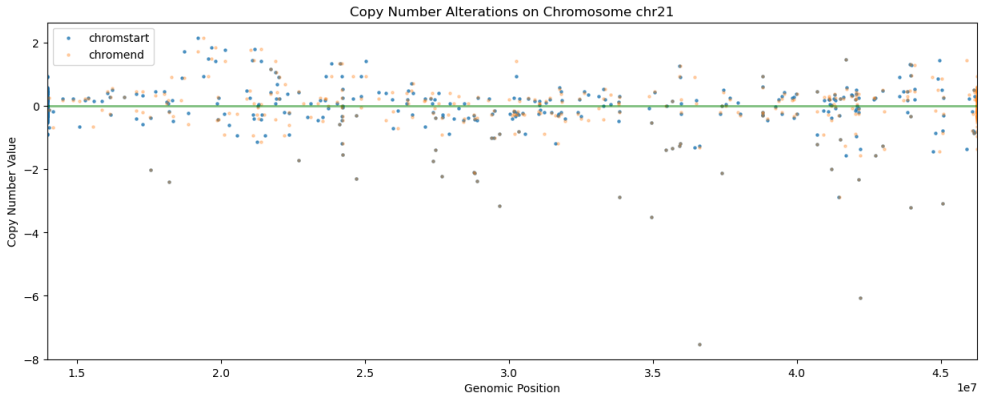
Appendix C.
Plotted copy number values for chromosomes of interest

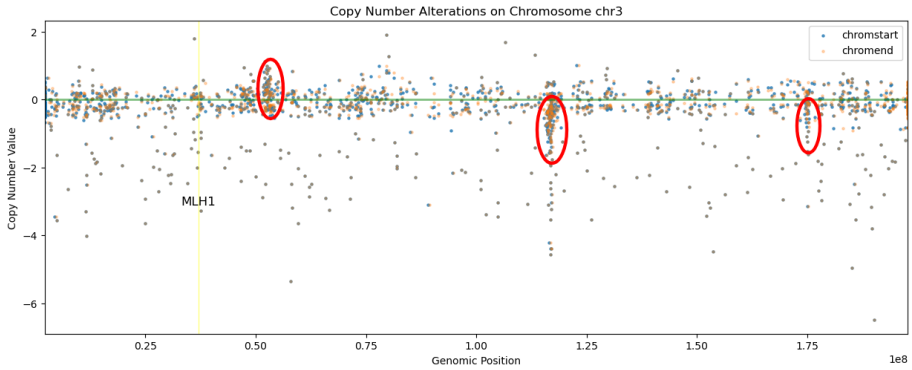


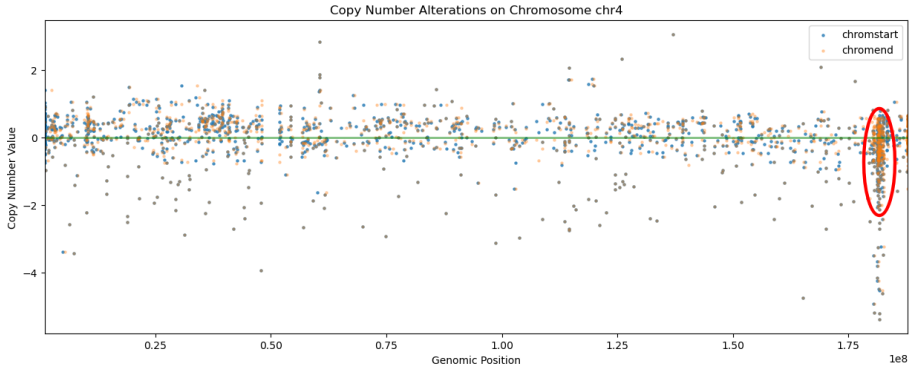

Copy Number Alterations on Chromosome chr22
chromstart
chromend
KREMEN1
ZNRF3
Copy Number Value
Genomic Position
1e7

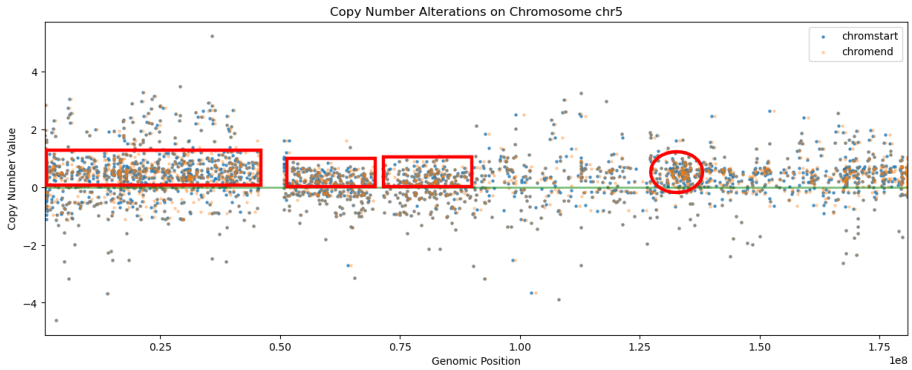

Copy Number Alterations on Chromosome chr2
chromstart
chromend
EPCAM
MSH6
Copy Number Value
Genomic Position
1e8

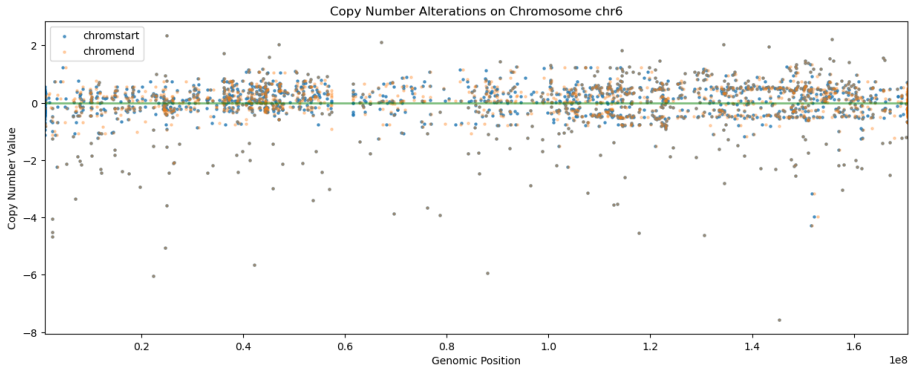

Copy Number Alterations on Chromosome chr20
chromstart
chromend
Copy Number Value
Genomic Position
1e7

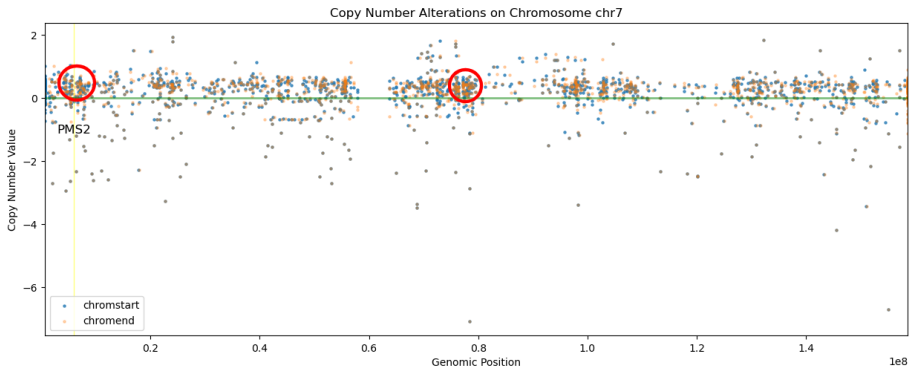

Copy Number Alterations on Chromosome chr21
chromstart
chromend
Copy Number Value
Genomic Position
1e7

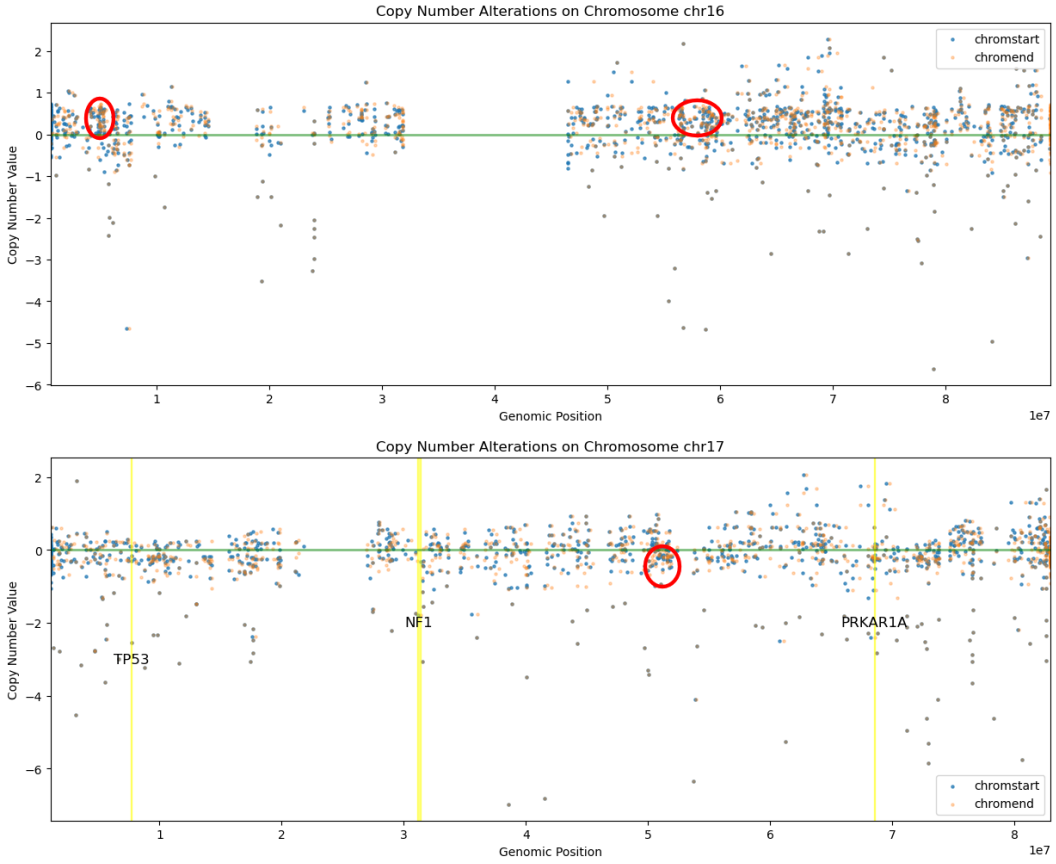

Copy Number Alterations on Chromosome chr3
chromstart
chromend
MLH1
Copy Number Value
Genomic Position
1e8


Copy Number Alterations on Chromosome chr4
chromstart
chromend
Copy Number Value
Genomic Position
1e8


Copy Number Alterations on Chromosome chr5
chromstart
chromend
Copy Number Value
Genomic Position
1e8


Copy Number Alterations on Chromosome chr6
chromstart
chromend
Copy Number Value
Genomic Position
1e8


Copy Number Alterations on Chromosome chr7
PMS2
chromstart
chromend
Copy Number Value
Genomic Position
1e8


Copy Number Alterations on Chromosome chr16
chromstart
chromend
Copy Number Value
Genomic Position
1e7
Copy Number Alterations on Chromosome chr17
TP53
NF1
PRKAR1A
chromstart
chromend
Copy Number Value
Genomic Position
1e7

Appendix D.
Classiefiers Feature Weights