

**Developing statistical and bioinformatic analysis of genomic
data from tumours**

Rohit Thakur

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Faculty of Medicine and Health

October 2018

This candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy had been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement

© 2018 The University of Leeds Rohit Thakur

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant (MELGEN) agreement No 641458. Being a part of the MELGEN program provided me with several training opportunities to develop my statistical, bioinformatic, and complementary skills. Additionally, because of this program, I was able to pursue internships at two academic institutions and at a private company. This program helped in my career development immensely by enabling me to attend and present my research at several scientific meetings.

The collection of samples in the Melanoma Cohort Study was funded by Cancer Research UK (project grant C8216/A6129, Programme awards C588/A4994 and C588/A10589) and Centre Award (C37059/A11941) and by the NIH (R01 CA83115).

I would like to thank all my supervisors for providing me with incredible mentorship. Without their supervision this dissertation would not have been possible. I would also like to express my sincerest gratitude to everyone who has helped me in pursuing this thesis, especially I would like to thank:

Prof. Jenny Barrett, my main supervisor, for introducing me to the field of biomedical statistics and for always guiding me. Prof. Barrett helped me in developing statistical thinking skills by providing constructive feedback on my monthly progress reports and statistical analyses. Prof. Barrett always kept track of my progress due to which I was able to finish this project in a timely manner. When I started my PhD, I had an inherent fear about scientific writing, Prof. Barrett's mentorship helped me overcome this fear by developing good scientific writing skills.

Prof. Julia Newton-Bishop, my co-supervisor, for introducing me to the field of translational melanoma research. Prof. Newton-Bishop helped me improve my thinking from a clinical perspective and her clinical insights have immensely helped in the development of this thesis. Prof. Newton-Bishop also helped in developing a strong network with principal investigators from several other institutions.

Dr Jérémie Nsengimana, my co-supervisor, for helping me develop supervised and unsupervised analyses pipelines and providing a rigorous review on my writing and presentation skills. Dr Nsengimana also provided me with the pre-processed microarray data generated from the LMC study cohort.

Prof. Tim Bishop, for being a great mentor, creating a positive atmosphere, and giving me feedback on my thesis project at weekly group meetings. Despite his busy schedule, Prof. Bishop took out time for discussions, the spectrum of which varied from

academic research to ongoing cricket series. Prof. Bishop helped me move my project forward by taking a lead on requesting access to several independent melanoma datasets.

Prof. Göran Jönsson, my external supervisor, for giving valuable feedback in the supervisory meetings and collaborating with our group on this project. Prof. Jonsson also gave me an opportunity to pursue a short internship with his group at Lund University, Sweden. I also want to thank Dr Martin Laüss (postdoctoral fellow with Prof. Jönsson) for granting me access to their gene expression dataset. Dr Laüss also taught me differential gene expression analysis during my internship.

Dr Jon Laye, for his brilliant suggestions and insightful thinking on the biological aspects of this thesis. Dr. Laye, trained me in drawing figures in a logical format. I have always cherished our conversations on understating some of the results in this thesis.

Dr Juliette Randerson-Moor, for helping me with administrative tasks, preparing presentations and giving her feedback during regular group meetings. Dr. Randerson-Moor provided unique guidance and support to write this thesis in the correct format.

Dr Mark Harland, for helping me with the analysis of outlying samples in the dataset and explaining me the technical protocols behind generation of the gene expression dataset.

Dr John Davies, for providing feedback on statistical analysis during the regular lab meetings and improving my understanding on the Cox proportional hazard models.

Dr Sally O'Shea, for sharing words of wisdom on writing the PhD thesis. Dr. O'Shea calculated the mitotic rate from the tumour cores which was used in the univariable and multivariable analyses.

May Chan, for always motivating me and giving access to the datasets analysed in this study.

Faye Elliott, for positively encouraging me to complete this thesis on time.

Dr Mark Iles, for providing feedback on the statistical analysis in aims 1 and 2.

Katie Cairns, for kindly helping me to keep a track of my annual leave, and arranging the logistics when I went for international conferences and internships.

Dr Will Spooner and Eleanor Stanley (members of the Eagle Genomics company), for giving an opportunity to pursue an internship at the Eagle Genomic company. Dr Bram Vandekerckhove (a member of the company) helped me with applications of machine learning technique support vector machine to our dataset.

Prof. Manolis Kellis for giving me the opportunity to pursue an internship in his group at Massachusetts Institute of Technology, USA. Alvin Shi (PhD student with Prof. Kellis) helped me with the feature selection in machine learning.

Joey Diaz, for introducing me to the field of copy number data analysis and helping me during my knee injury. Joey and I have always motivated each other on writing the PhD thesis.

Sathya Muralidhar, for giving insightful feedback during group meetings.

Joanna Pozniak for helping me with the Cytoscape and ReactomeFiViz based network enrichment analysis and improving my understanding of immune cell subtypes.

Theofanis Tsismentzoglou, for sharing words of encouragement during lunch breaks.

I would also like to thank my friend Payal Ganguly for always being supportive and encouraging and especially for going out of her way to help me during my knee injury period.

Finally, I would like to thank my mother (Revti Devi) and father (Duni Chand), who made several sacrifices to get me to this point today, for their love, support, and encouragement. I want to thank my sister (Yogita Kumari) for always having my back and being a constant source of motivation. I also want to thank my uncle (Gauri Dutt Thakur) who taught me to aim for the stars. I am also thankful to my grandmother (Soma Devi), my partner (Saumya Sisoudiya), and her family (Yogeshwar Singh Sisoudiya, Dushyant Sisoudiya, Rita Sisoudiya and Swati Sisoudiya) for their strong support during the course of this PhD.

Abstract

Previous prognostic signatures for melanoma based on tumour transcriptomic data were developed predominantly on cohorts of AJCC (American Joint Committee on Cancer) stages III and IV melanoma. Since 92% of melanoma patients are diagnosed at AJCC stages I and II, there is an urgent need for better prognostic biomarkers to allow patient stratification for receiving early adjuvant therapies.

This study uses genome-wide tumour gene expression levels and clinico-histopathological characteristics of patients from the Leeds Melanoma Cohort (LMC). Several unsupervised and supervised classification approaches were applied to the transcriptomic data, to identify biological classes of melanoma, and to develop prognostic classification models respectively.

Unsupervised clustering identified six biologically distinct primary melanoma classes (LMC classes). Unlike previous molecular classes of melanoma, the LMC classes were prognostic in both the whole LMC dataset and in stage I tumours. The prognostic value of the LMC classes was replicated in an independent dataset, but insufficient data were available to replicate in an AJCC stage I subset.

Supervised classification using the Random Forest (RF) approach provided improved performances when adjustments were made to deal with class imbalance, while this did not improve performance of the Support Vector Machine (SVM). However, RF and SVM had similar results overall, with RF only marginally better. Combining clinical and transcriptomic information in the RF further improved the performance of the prediction model in comparison to using clinical information alone. Finally, the agnostically derived LMC classes and the supervised RF model showed convergence in their association with outcome in some groups of patients, but not in others.

In conclusion, this study reports six molecular classes of primary melanoma with prognostic value in stage I disease and overall, and a prognostic classification model that predicts outcome in primary melanoma.

Table of Contents

Acknowledgements	iii
Abstract	vi
Table of Contents	vii
List of Tables	xiii
List of Figures	xvi
Abbreviations	xviii
Chapter 1 Introduction	1
1.1 Human skin and melanoma	1
1.1.1 Melanocytes.....	2
1.1.2 Melanin, melanogenesis and response to UVR.....	2
1.1.3 Melanocytes to melanoma.....	3
1.2 Cutaneous melanoma.....	3
1.3 Melanoma epidemiology.....	4
1.4 Melanoma genetics.....	5
1.5 Melanoma histopathological classification, prognostic factors and survival.....	6
1.5.1 AJCC staging system.....	6
1.5.2 Prognostic factors in primary melanoma.....	9
1.5.3 Melanoma survival based on the AJCC stage	10
1.6 Genomic classification of melanoma.....	11
1.6.1 Mutational subtypes of melanoma	11
1.6.2 Somatic copy number association with mutational subtypes	12
1.7 Gene expression profiling technologies	12
1.7.1 DNA microarrays.....	13
1.7.2 RNA-sequencing	13
1.8 Gene expression based classification of melanoma.....	15
1.8.1 Supervised classification using gene expression.....	15
1.8.2 Unsupervised classification using gene expression.....	20
1.8.3 Overlap between the signatures.....	22

1.9	Outline of this study	23
1.9.1	Aims and objectives	23
1.9.2	Outline of the chapters	24
Chapter 2	General Methods.....	26
2.1	Leeds Melanoma Cohort (LMC).....	26
2.2	Gene expression data.....	27
2.2.1	Tumour core generation	27
2.2.2	RNA extraction and Expression data generation	27
2.2.3	Data pre-processing.....	28
2.3	Clinico-pathological characteristics	29
2.4	Survival analysis.....	30
2.4.1	Survival Outcome.....	31
2.4.2	Kaplan-Meier survival estimate	31
2.4.3	Log-rank test	32
2.4.4	Cox Proportional Hazards model (CPH).....	32
2.5	Statistical tests.....	33
2.5.1	Mann-Whitney/Wilcoxon test.....	33
2.5.2	Kruskal-Wallis test.....	34
2.5.3	Pearson's chi-squared test.....	34
Chapter 3	Analysing existing melanoma gene signatures and devising a new one by clustering of LMC tumours.....	36
3.1	Introduction.....	36
3.1.1	Gene expression based cluster analyses in melanoma	36
3.1.2	Replication of Lund and TCGA signatures in the whole dataset	41
3.2	Methods.....	41
3.2.1	Replicating Lund 4-class and Lund 2-grade signatures in LMC	41
3.2.2	The Lund and TCGA signatures association with clinico-histopathological variables	42
3.2.3	Lund and TCGA signatures association with melanoma-specific survival	42

3.2.4	Statistical interaction test between the Lund grade and AJCC stage	42
3.2.5	Clustering gene expression data of LMC tumours	43
3.2.6	Comparing the HC, KM and PAM based clusters	48
3.3	Results	48
3.3.1	Applying the existing signatures to the whole LMC dataset	48
3.3.2	Clinico-histopathological association with the Lund and TCGA signatures 50	
3.3.3	Lund and TCGA signatures association with survival in the LMC primary tumours.....	54
3.3.4	Signature's prognostic value when stratified on AJCC stage	58
3.3.5	Statistical Interaction between the Lund signature and AJCC stage	64
3.3.6	Devising a new signature by consensus-based clustering	66
3.3.7	Comparing HC, KM, and PAM based clusters	69
3.3.8	Agreement with Lund and TCGA signatures.....	71
3.4	Discussion	72
3.4.1	The Lund and TCGA signatures were reproducible in the Leeds Melanoma cohort	73
3.4.2	The Lund and TCGA signatures' lack of association with AJCC stage ...	74
3.4.3	Unsupervised clustering of the LMC dataset	74
	Chapter 4 Properties of the new tumour classes in primary melanoma: prognostic significance and biological characterisation	76
4.1	Introduction.....	76
4.2	Methods.....	77
4.2.1	Identifying LMC classes using a cluster separation measure.....	77
4.2.2	Association of the LMC classes with clinico-histopathological characteristics.....	77
4.2.3	Association of LMC classes with MSS.....	78
4.2.4	Refining the LMC class signature	78
4.2.5	Replicating the prognostic value of the LMC signature in the Lund cohort 79	
4.2.6	Area under Receiver Operating Characteristic Curve	80

4.2.7	Biological significance of LMC classes	82
4.2.8	Assessing the Lund module activity in LMC classes.....	85
4.3	Results	86
4.3.1	Identifying the number of classes in the LMC dataset.....	86
4.3.2	The LMC class association with clinico-histopathological features	91
4.3.3	Prognostic value of the LMC signature.....	93
4.3.4	Comparing LMC signatures with Lund 4-classes and TCGA 3-classes ..	99
4.3.5	Replicating prognostic value of LMC signature.....	100
4.3.6	Biological significance of the LMC classes	107
4.4	Discussion	116
4.4.1	Clinical-pathological characteristics of the LMC classes.....	116
4.4.2	Prognostic significance of the LMC classes in stage I melanoma.....	117
4.4.3	Biological interpretation of the LMC classes	118
	Chapter 5 Machine learning applications to predict melanoma prognosis using gene expression.....	120
5.1	Introduction.....	120
5.2	Methods.....	121
5.2.1	Devising a supervised classification framework.....	121
5.2.2	Defining a dichotomous outcome in the LMC	121
5.2.3	Creating the training and test sets in LMC.....	122
5.2.4	Prediction performance measures.....	123
5.2.5	Developing classification models using Random Forest (RF).....	124
5.2.6	Developing classification models using SVM.....	128
5.2.7	Calculating agreement between RF and SVM models.....	134
5.2.8	Generating permutation-based baseline models	135
5.3	Results	135
5.3.1	Training and test sets in LMC.....	135
5.3.2	RF applications to predict outcome	135
5.3.3	SVM applications to predict outcome	140
5.3.4	SVM using balanced class design	141

5.3.5	Comparing RF and SVM predictions	142
5.3.6	Permutation-based RF model.....	145
5.4	Discussion	147
Chapter 6 Combining clinical information and gene expression for predicting outcome and independent validation		150
6.1	Introduction.....	150
6.2	Methods.....	151
6.2.1	Using clinical information to generate RF model.....	151
6.2.2	Removing clinical information from the gene expression	152
6.2.3	Combining clinical information and gene expression	152
6.2.4	Refining the RF model via variable selection	153
6.2.5	ROC analysis for the refined RF model	155
6.2.6	Application of the refined RF model on independent dataset.....	155
6.2.7	Pathway enrichment of predictor genes in the refined RF model	156
6.2.8	Comparison between the LMC classes and the refined RF model predictions	156
6.3	Results	157
6.3.1	Baseline RF model using clinical information alone.....	157
6.3.2	Prediction after removing clinical information from gene expression.....	157
6.3.3	RF model from combining clinical information and gene expression.....	161
6.3.4	Variable selection to generate the final refined RF model.....	162
6.3.5	ROC analysis of the final refined RF model.....	164
6.3.6	Validating prognostic significance of the refined RF model on Lund dataset	166
6.3.7	Biological interpretation of the refined RF model	168
6.3.8	The LMC class association with the refined RF model predictions	168
6.4	Discussion	170
6.4.1	Prognostic value of clinical information and gene expression based RF models	170
6.4.2	Refining combination of clinical information and gene expression based RF model	171

6.4.3	Prognostic value of refined RF model and validation.....	171
6.4.4	Biological interpretation of the refined RF model	172
Chapter 7 Final summary and discussion.....		173
7.1	Summary of the two main aims of this study.....	173
7.2	Context and discussion of findings.....	174
7.2.1	Class discovery using unsupervised clustering.....	175
7.2.2	Class prediction using machine learning	178
7.3	Strengths and limitations.....	180
7.4	Future perspectives	182
Chapter 8 Appendix I		184

List of Tables

Table 1.1 TNM staging of cutaneous melanoma.....	8
Table 1.2 Clinical staging of cutaneous melanoma	9
Table 1.3 Summary of studies investigating prognostic signatures of melanoma	18
Table 2.1 Summary of clinico-histopathological characteristics of the LMC cohort	30
Table 3.1 Summary statistics of the LMC tumour classification using the Lund and TCGA signatures	49
Table 3.2 Lund signatures association with clinico-pathological characteristics of LMC primary tumours.....	52
Table 3.3 TCGA signatures association with clinico-pathological characteristics of LMC primary tumours.....	53
Table 3.4 Unadjusted Cox proportional hazard analysis of the Lund 4-classes.. ..	54
Table 3.5 Unadjusted Cox proportional hazard analysis of the Lund 2-grades... ..	56
Table 3.6 Unadjusted Cox proportional hazard analysis of the TCGA 3-classes.....	57
Table 3.7 Unadjusted Cox proportional hazard models for the Lund 4-class, Lund 2-grades and TCGA 3-classes in the AJCC stage I group.....	59
Table 3.8 Unadjusted Cox proportional hazard models for the Lund 4-class, Lund 2-grades and TCGA 3-classes in the AJCC stages II & III	62
Table 3.9 Test of interaction between AJCC stage and Lund 2-grade signature in Cox proportional hazards model of MSS	65
Table 3.10 Cramer's V statistic comparing agreement between HC, KM, and PAM derived clusters and the Lund and TCGA signatures	72
Table 4.1 Summary of cluster separation measure at various k values.....	87
Table 4.2 Summary of LMC class tumours across the three batches sent for gene expression profiling.....	88
Table 4.3 Summary of cluster separation measure at various k values after removing 15 samples.....	90
Table 4.4 Comparing new classification with initial classification after removing 15 samples	90
Table 4.5 LMC classes association with clinico-histopathological variables.....	92
Table 4.6 Summary of the univariable and multivariable analysis of the LMC signature with melanoma-specific survival in the whole LMC dataset	94

Table 4.7 Summary of the univariable and multivariable analysis of the LMC signature with melanoma-specific survival in the AJCC stage I group.....	97
Table 4.8 Summary of the univariable analysis of the LMC signature with relapse-free survival and overall survival in the Lund cohort.....	103
Table 4.9 Summary of the univariable analysis of the LMC signature with relapse-free survival and overall survival in the AJCC stage I group of Lund cohort...	104
Table 5.1 Life table of patients in LMC for 16-year time interval.....	122
Table 5.2 Comparing histopathological differences between the training and test set	136
Table 5.3 Summary of the RF model performance on training and test set.....	138
Table 5.4 Summary of SVM model performance on training and test set.....	140
Table 5.5 Summary of SVM non-linear kernel functions on training and test set	141
Table 5.6 Summary of SVM linear kernel model after under sampling the majority class.....	142
Table 5.7 Comparing overlap between RF model and SVM model prediction on test set.....	144
Table 6.1 Refining the RF model by selecting variable number of genes, <i>mTry</i> , and number of trees.....	154
Table 6.2 Summary of clinical variable based RF model performance	157
Table 6.3 Summary of residual based RF model performance.....	161
Table 6.4 Summary of combined RF model performance	161
Table 6.5 Summary performance of RF models generated after selection of genes based on Gini index measure	163
Table 6.6 Summary of refined RF model performance on the Lund dataset	166
Table 8.1 The 150-gene based LMC 6 class signature.....	184
Table 8.2 Summary of upregulated biological pathways in LMC class 1, FDR is the false discovery rate	189
Table 8.3 Summary of upregulated biological pathways in LMC class 2, FDR is the false discovery rate	191
Table 8.4 Summary of upregulated biological pathways in LMC class 3, FDR is the false discovery rate	192
Table 8.5 Summary of upregulated biological pathways in LMC class 4, FDR is the false discovery rate	193
Table 8.6 Summary of upregulated biological pathways in LMC class 5, FDR is the false discovery rate	193
Table 8.7 Summary of upregulated biological pathways in LMC class 6, FDR is the false discovery rate	194

Table 8.8 Summary of downregulated biological pathways in LMC class 1, FDR is the false discovery rate	195
Table 8.9 Summary of downregulated biological pathways in LMC class 2, FDR is the false discovery rate	196
Table 8.10 Summary of downregulated biological pathways in LMC class 3, FDR is the false discovery rate	197
Table 8.11 Summary of downregulated biological pathways in LMC class 4, FDR is the false discovery rate	199
Table 8.12 Summary of downregulated biological pathways in LMC class 5, FDR is the false discovery rate	201
Table 8.13 Summary of downregulated biological pathways in LMC class 6, FDR is the false discovery rate	202
Table 8.14 Gini index value for 200 predictor genes of the refined RF model..	203

List of Figures

Figure 1.1 Anatomy of human skin.....	1
Figure 1.2 Measurement of tumour thickness across the T stages in the AJCC staging system.....	7
Figure 1.3 Summary steps of a microarray experiment	14
Figure 1.4 Melanoma tumour subtypes identified by the Lund group	21
Figure 3.1 Comparison of Lund and TCGA melanoma subtypes.	40
Figure 3.2 Consensus clustering workflow.....	46
Figure 3.3 Melanoma-specific survival for the Lund 4-classes.....	55
Figure 3.4 Melanoma-specific survival for the Lund 2-grades.....	56
Figure 3.5 Melanoma-specific survival for the TCGA 3-classes	58
Figure 3.6 Melanoma-specific survival for the Lund 4-classes, Lund 2-grades and TCGA 3-classes in the AJCC stage I group.....	61
Figure 3.7 Melanoma-specific survival for the Lund 4-classes, Lund 2-grades and TCGA 3-classes in the AJCC stages II&III	64
Figure 3.8 Sample size and power calculation.....	65
Figure 3.9 Selecting k by examining the relative change in area under the CDF curve	67
Figure 3.10 Consensus matrix heatmaps for HC, KM and PAM.....	68
Figure 3.11 Comparing stability of HC, KM and PAM derived clusters	71
Figure 4.1 Summary of SAM workflow	84
Figure 4.2 Re-clustering of LMC tumours after excluding 15 samples, using consensus clustering PAM method.....	89
Figure 4.3 Melanoma-specific survival for the LMC classes on the whole LMC dataset	93
Figure 4.4 Melanoma-specific survival for the LMC classes on the AJCC stage I group	96
Figure 4.5 Overlap between the LMC signatures and the Lund 4-classes and TCGA 3-classes.....	99
Figure 4.6 Refining the LMC class gene signature	100
Figure 4.7 Relapse-free survival and overall survival for the LMC classes on the Lund dataset.....	102
Figure 4.8 Independent prognostic value of the LMC signature.....	106
Figure 4.9 Summary of biological pathways associated with the LMC class 1..	108
Figure 4.10 Summary of biological pathways associated with the LMC class 2	109

Figure 4.11 Summary of biological pathways associated with the LMC class 3	110
Figure 4.12 Summary of biological pathways associated with the LMC class 4	111
Figure 4.13 Summary of biological pathways associated with the LMC class 5	112
Figure 4.14 Summary of biological pathways associated with LMC class 6	113
Figure 4.15 Characterising LMC classes using the Lund modules	115
Figure 5.1 Summary of supervised classification framework.....	121
Figure 5.2 Summary of reference terms when comparing actual status with predicted status	123
Figure 5.3 Workflow of RF algorithm.....	125
Figure 5.4 Example of a RF decision tree.....	126
Figure 5.5 Maximum margin classifier.....	129
Figure 5.6 Polynomial kernel SVM and radial kernel SVM	133
Figure 5.7 RF model using unbalanced class design.....	137
Figure 5.8 RF model using balanced class design.....	139
Figure 5.9 Comparing RF and SVM predictions.....	144
Figure 5.10 Performance summary of the permuted model.....	146
Figure 6.1 Gini importance index.....	154
Figure 6.2 Clinical variable based RF model.....	158
Figure 6.3 Gene expression association with clinical predictors of melanoma.	159
Figure 6.4 RF model from adjusted gene expression.....	160
Figure 6.5 Comparison of the refined RF model with baseline RF models.....	165
Figure 6.6 ROC analysis of the RF models in the Lund dataset.....	167
Figure 6.7 Biological interpretation of the refined RF model.....	168
Figure 6.8 Comparing the LMC classes with refined RF model.....	169

Abbreviations

ABCDE	Asymmetry Border Colour Diameter Evolution
AUC	Area Under the Curve
CDF	Cumulative Distribution Function
cDNA	complementary Deoxy-ribonucleic acid
CPH	Cox Proportional Hazards
CSD	Chronically Sun Damaged skin
CV	Cross Validation
DASL	cDNA-mediated Annealing, Selection, extension and Ligation
FDR	False Discovery Rate
FFPE	Formalin-Fixed Paraffin-Embedded
FN	False Negative
FP	False Positive
FPF	False Positive Fraction
GWAS	Genome Wide Association Studies
HC	Hierarchical Clustering
KM	Kaplan-Meier
LMC	Leeds Melanoma Cohort
mRNA	messenger Ribonucleic Acid
MSS	Melanoma Specific Survival
NCC	Nearest Centroid Classification
NGS	Next Generation Sequencing
NPV	Negative Predictive Value
ONS	Office of National Statistics
OOB	Out Of Bag
PAM	Partitioning Around Medoids
PCR	Polymerase Chain Reaction
PPV	Positive Predictive Value

RF	Random Forest
RGP	Radical Growth Phase
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
RSN	Robust Spline Normalisation
SVM	Support Vector Machine
TCGA	The Cancer Genome Atlas
TILS	Tumour Infiltrating Lymphocytes
TMA	Tissue Microarray Needle
TN	True Negative
TP	True Positive
TPF	True Positive Fraction
UVR	Ultra Violet Radiations
VGP	Vertical Growth Phase

Chapter 1

Introduction

This chapter provides an introduction to melanoma research by presenting an overview of melanoma formation, epidemiology and genetics. This is followed by different methods of classification: histopathological, genomic and transcriptomic.

1.1 Human skin and melanoma

Human skin, the largest organ of the human body, is not directly thought of as a functional organ like the heart or the liver, but is simply believed to just protect us from external stimuli, like extreme hot or cold temperatures [1]. However it plays a unique and complex role in maintaining a barrier between internal and external environment, and in maintaining the steady state of internal body (homeostasis) in terms of heat and hydration [1, 2].

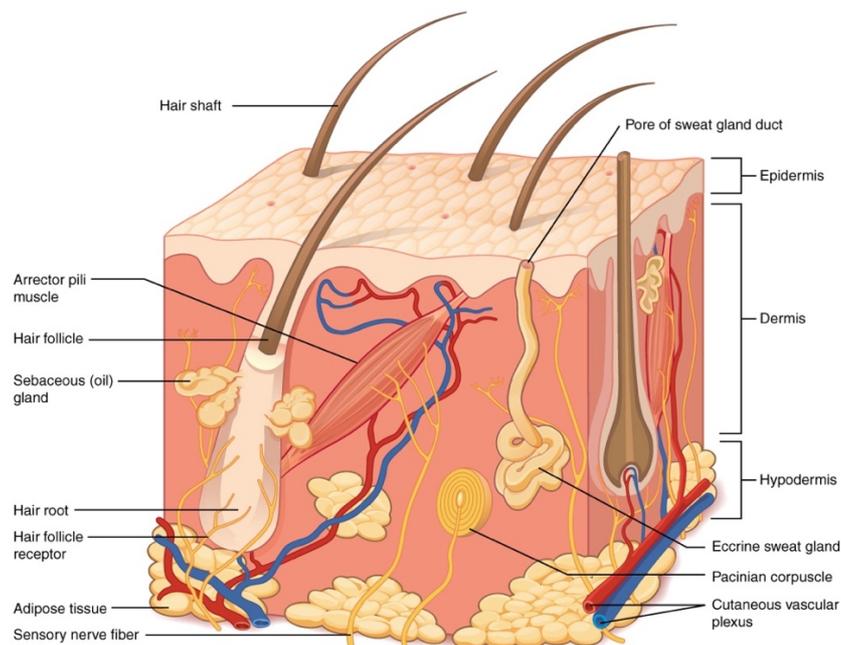


Figure 1.1 Anatomy of human skin

Adapted from [2].

Human skin is arranged into three layers: epidermis (outer layer), dermis (inner layer) and hypodermis (inner most layer) (Figure 1.1). The epidermis can be further divided into an outer 'dead' layer and an inner 'living' cell layer [2]. The outer layer contains all the dead and peeling cells, and its major function is to maintain a barrier against

microbes and other environmental factors. The inner layer contains a higher proportion of keratinocytes (up to 90%) and lower proportion (up to 10%) of other cell types known as melanocytes, Merkel cells and Langerhans cells [2]. The main function of the inner layer is to help in the formation of the main barrier, which protects against invasion of chemical substances and microbes. The majority of skin cancers originate in the epidermis [3]. Melanoma, the most aggressive form of skin cancer, originates from melanocytes or their precursor cells, melanoblasts, which are mainly present in the epidermis.

1.1.1 Melanocytes

In 1889, S. Meyerson first introduced the term 'melanocyte' to represent a dendritic cell which originates from the neural crest [4]. As described above, melanocytes are small subpopulations of cells located in the inner layer of the epidermis. A melanoblast is a precursor cell of melanocytes, and it is derived from the neural crest cells [4]. The melanoblasts differentiate from the neural crest cells and migrate extensively during embryonic development [1, 2]. Once melanoblasts have reached their terminal locations, most of these cells then differentiate into melanocytes. The process of differentiation from neural crest cell to becoming a melanocyte completes within the first 6 months of the gestation period [1, 2]. Most of the melanocytes migrate from dermis to epidermis and produce a pigment known as melanin.

1.1.2 Melanin, melanogenesis and response to UVR

After a melanocyte has been formed, it starts producing an organelle known as the melanosome [4]. Melanosomes are the organelles that produce a pigment known as melanin. The variation in human skin colour is primarily due to variations in melanin. Melanin is classified into two types: eumelanin and pheomelanin [5, 6]. Eumelanin, a major source of pigmentation, is a high density dark pigment which is contained in eumelanosomes. Pheomelanin is a cysteine-derived compound which is reddish/yellow in colour that is largely responsible for the colour of red hair.

The biological process of melanin synthesis is known as melanogenesis. After melanoblasts have differentiated into melanocytes, the melanosome formation is initiated. During melanogenesis, tyrosinase and tyrosinase-related proteins 1 and 2 (TRP1 and TRP2) catalyse the synthesis of melanin [6]. After producing melanin, it is transferred to keratinocytes, where melanin plays an important role in protecting the skin from harmful ultraviolet radiation (UVR) [6]. Generally, in humans, one melanocyte is in contact with ~35 neighbouring keratinocytes [7]. Melanin is

transferred to keratinocytes located on the inner layer of the epidermis, and it is degraded before keratinocytes migrate to the outer dead layer [6].

UVR exposure increases melanin production, which causes tanning of the skin [4-6]. A sudden increase in pigmentation in response to UVR, also referred to as immediate tanning, is achieved by darkening and shifting pre-existing melanosomes from the nuclear region to the dendritic region. Delayed tanning is caused by a gradual increase in production of melanin (i.e. eumelanin) over 2-3 weeks after UVR exposure. Overall, in both immediate and delayed tanning, a pigment is produced that protects the skin from UVR. In pale skin, melanin levels increase by 500 to 1000 fold, in comparison to dark skin where the levels only increase by 10 to 15 fold [8, 9].

1.1.3 Melanocytes to melanoma

The transition from melanocyte to the most readily recognisable form of melanoma, i.e. superficial spreading melanoma, occurs in five different phases as follows: benign naevi; dysplastic naevi; radial growth phase (RGP); vertical growth phase (VGP); metastatic melanoma [3]. This superficial spreading melanoma is typified by change in shape, size and colour and is the commonest form of melanoma in pale-skinned populations. As described above, many superficial spreading melanomas originate in naevi, although some melanoma may arise *de novo* from normal skin.

Benign naevi are the normal moles present on the skin: they are benign proliferations of melanocytes such that the proliferation step is self-limiting. A small proportion of these benign naevi undergo continued cellular proliferation which leads to formation of an asymmetric dysplastic (or atypical) naevus. Even though the majority of such naevi eventually cease, a proportion of them emerge into RGP melanoma. In RGP, the cells contained in the epidermis invade the outer dermis layer. In VGP, the malignant cells invade further into the inner dermis and subcutaneous fat layers. The melanoma at this stage has a high metastatic potential and can metastasize to distant organs. In the metastatic phase, the melanoma metastasizes both by lymphatics (probably the most frequent) but also via the blood vessels and then subsequently to different organs in the human body, such as the lungs, liver or brain [3].

1.2 Cutaneous melanoma

Melanoma is a cancer that arises from genetic changes in melanocytes and melanoblasts leading to an uncontrolled growth of these cells. Melanoma occurs most frequently on the skin, also referred as cutaneous melanoma, but also on body sites such as the uvea of the eye, mucosa, and leptomeninges. Although cutaneous

melanoma represents only 5% of skin cancer cases but it accounts for 80% of skin cancer related mortality [3].

Cutaneous melanoma most commonly presents as a mole in clinical settings, and initial diagnosis is based on the ABCDE criteria, where A stands for asymmetrical nevi, B for irregular border, C for multiple colours, D for diameter >5mm, and E for evolving shape and size [10]. Melanoma is further divided based on these criteria into four distinct subtypes: superficial spreading melanoma, nodular melanoma, lentigo maligna melanoma, and acral lentiginous melanoma [11]. Superficial spreading melanoma is the most common melanoma subtype and accounts for ~70-75% of melanoma cases; it occurs mostly on the limbs and trunk of the body [11]. The ABCDE criteria best describe this sort of melanoma. Nodular melanoma accounts for ~20-25% of melanoma cases and can appear on any body site, including the sites mentioned for superficial spreading melanoma [11]. Lentigo maligna melanoma is the least frequent melanoma and accounts only for ~5-10% of melanoma cases [11]. It occurs mostly at older age and, unlike previous subtypes, it does not originate from an existing naevus but occurs as a result of prolonged sun exposure. The previous three melanoma subtypes are mostly observed in the Caucasian population. However, the fourth subtype, acral lentiginous melanoma, is rarely observed in this population, and the majority of cases are individuals with darker skin; it occurs mostly on non-sun exposed sites, such as feet, finger nail beds, and toes [11].

1.3 Melanoma epidemiology

Cutaneous melanoma is the 5th most common form of cancer in the UK and is ranked 19th worldwide [12, 13]. In the last 50 years, the world wide incidence of melanoma has risen sharply, with greatest incidence observed in pale-skinned populations. In the UK, the incidence rates have increased by 50% over the last decade [13], and males have had a higher increase (64%) in the incidence rate in comparison to females (39%). Across Europe, approximately 100,000 new cases of melanoma were diagnosed in 2012, and the incidence rate in the UK was ranked 9th in Europe [14]. In the UK, every year (2012-2014) ~15,000 patients are diagnosed with melanoma [13]. Melanoma is predominantly an adult cancer with very few cases prior to puberty.

The risk factors for melanoma are a history of severe sunburn, a higher number of dysplastic naevi, older age, family history of melanoma, pale skin, and light hair colour [15-19]. A pooled analysis of 15 case-control studies showed that sun exposure increases melanoma risk [18]. This study provides stronger evidence for intermittent sun exposure than cumulative sun exposure [18], but a so-called UVR

signature, i.e. C>T somatic mutations in the tumour [16], implicates sun exposure as the major environmental factor in pathogenesis. Along with sun exposure, greater numbers of melanocytic naevi, also known as “moles”, increases melanoma risk, and individuals with ≥ 100 moles have 7 fold increased risk of melanoma in comparison to individuals with <100 moles. Along with melanocytic naevi, the presence of dysplastic naevi further increases melanoma risk [17]. Age is an important risk factor for melanoma, and growing old increases the relative risk of melanoma [16, 19]. Family history of melanoma increases the risk 2 fold [16, 17]. A systematic review of 10 case-control studies showed that hair colour predicts melanoma risk, as blond hair individuals had 1.8 fold increased risk, red hair individuals had 2.4 fold increased risk in comparison to black and dark brown hair individuals [20]. A recent study of a Norwegian-Swedish cohort reported consistent findings, as red haired, blond haired and brown haired individuals had increased risk of melanoma in comparison to dark brown and black haired individuals [21].

1.4 Melanoma genetics

A family history of melanoma is reported in approximately 8% of melanoma cases [16]. In families with 3 or more cases, almost 40% of cases carry a germline mutation in the cyclin-dependent kinase inhibitor 2A (*CDKN2A*) gene [16, 22-25]. Linkage studies of melanoma have also identified other high penetrance genes such as *CDK4*, *BAP1*, *POT1*, and *TERT* [26-31]. GenoMEL, a consortium of familial melanoma research, led one of the biggest studies to examine these mutations and reported that these mutations characterise only a proportion of melanoma families, and mutations in more than 50% of families remain unexplained [25].

Completion of the Human Genome Project has enabled development of cost effective genome-wide genotyping technologies. In addition to the previous linkage-based studies used to identify highly penetrant susceptibility genes, genome-wide association analysis studies (GWAS) have further characterised the genetic architecture of melanoma by identifying genes with intermediate and low penetrance. Several GWAS have identified a total of 20 loci associated with different melanoma phenotypes (physical characteristics associated with melanoma), such as pigmentation phenotypes, naevi in the skin and telomere length [32-39]. The variants mapping to the *CDKN2A/MTAP*, *PLA2G6* and *IRF4* regions were associated with development of naevi. The variants mapping to the *MC1R*, *ASIP*, *OCA2*, *SLC45A2* and *TYR* regions were observed to be associated with pigmentation phenotypes such as hair colour and eye colour. Variants in the *TERT*, *PARP1*, *ATM* and *OBFC1* regions were observed to be associated with telomere length. Several other variants

mapping to the *CCND1*, *RMDN2*, *CDKAL1*, *ARG3*, *TMEM38B*, *ARNT1*, *MX2*, and *CASP8* regions were not associated with any specific melanoma phenotype despite being strongly associated with increased melanoma risk.

1.5 Melanoma histopathological classification, prognostic factors and survival

Melanoma has the highest mortality rate among the skin cancers, but if detected early, the majority of the patients survive for more than 10 years [40]. In clinical settings the initial diagnosis of melanoma is made using visual examination of the mole with the naked eye and then using a magnifying system (dermoscopy). Suspicious moles are then removed by performing a surgical excision [10, 11]. Several histopathological factors of melanoma and host factors (site, sex and age) have been shown to be determinants of melanoma prognosis [41-43]. After clinical diagnosis of melanoma, the extent (stage) of cancer is decided based on histopathological classification described in the AJCC (American Joint Committee on Cancer) staging system [44].

1.5.4 AJCC staging system

The final version of the 7th edition of the AJCC staging system classifies melanoma tumours using TNM staging, where T is primary tumour characteristics, N is the number of regional lymph nodes and M is metastasis of tumours to distinct organs [44]. The T staging is based on three main histopathological factors as follows: Breslow thickness, mitotic rate and ulceration status of tumours.

Mitotic rate provides an estimate of the proliferation rate of cells [45] and is defined as the number of tumour cell divisions per square millimetres (mm) of the tissue slide. An increase in mitotic rate significantly reduces survival time [44]. In primary melanoma, mitotic rate has been identified as an independent predictor of poor prognosis in thin melanomas (stage I, Table 1.1, Table 1.2). The proposed threshold for mitotic rate is $<1 /\text{mm}^2$ or $\geq 1 /\text{mm}^2$, classifying patients into T1a and T1b stages respectively (Table 1.1, Table 1.2). Several studies have shown an association of mitotic rate with poor prognosis in thicker melanomas as well, but when jointly analysed with other clinical predictors, mitotic rate did not reach the significance threshold. Therefore, mitotic rate was not included in the AJCC classification of thicker melanomas [44, 46, 47]. In fact, mitotic rate has been dropped in the 8th edition [48] but this system is only now being applied, and therefore in this study the 7th Edition was used [44].

Breslow thickness is a well-established, independent prognostic factor for melanoma which is used in TNM staging of primary tumours [44, 45, 47, 49, 50]. Breslow thickness measures the thickness of the tumour in mm from the surface of the skin (granular cell layer) to the deepest part of the tumour (Figure 1.2). As shown in Table 1.1 Breslow thickness is divided into four categories in the AJCC staging system, as: $\leq 1\text{mm}$, 1.01-2.0mm, 2.01-4.0mm, $>4.0\text{mm}$, and these categories classify primary tumours in T1, T2, T3, and T4 stages respectively. The recently published 8th edition of the AJCC staging system [51] uses $\leq 0.8\text{ mm}$ as a cut-off for T1 stage, in comparison to the previous $\leq 1.0\text{mm}$ cut-off [48].

Ulceration status of the tumour is another prognostic indicator used in TNM classification and has been identified as an independent predictor of prognosis when jointly analysed with other clinical predictors of melanoma [44, 52]. The presence of ulceration increases the T stage (Table 1.1), and ulceration of the tumour also confers an increased risk of metastasis in comparison to non-ulceration. However, the detection of ulceration is associated with some interobserver variation, and several attempts has been made to standardise this variable [53].

The advanced stage classification (AJCC stage IV) utilises additional information from N and M stages which are based on detection of regional lymph node metastases, distant organ metastases and lactate dehydrogenase levels respectively (Table 1.1, Table 1.2) [44].

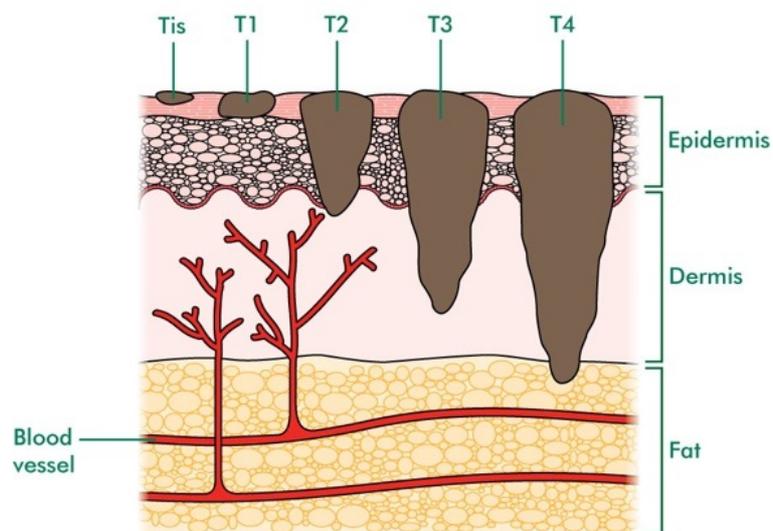


Figure 1.2 Measurement of tumour thickness across the T stages in the AJCC staging system

Adapted from Macmillan Cancer support [54]

Table 1.1 TNM staging of cutaneous melanomaAdapted from Balch *et al.* [44].

Primary Tumour Characteristics		
T stage	T1	T1a: ≤ 1.0 mm in thickness without ulceration, mitoses $< 1/\text{mm}^2$
		T1b: ≤ 1.0 mm in thickness with ulceration or mitoses $\geq 1/\text{mm}^2$
	T2	T2a: 1.01-2.0 mm in thickness without ulceration
		T2b: 1.01-2.0 mm in thickness with ulceration
	T3	T3a: 2.01-4.0 mm in thickness without ulceration
		T3b: 2.01-4.0 mm in thickness with ulceration
	T4	T4a: >4.0 mm in thickness without ulceration
		T4b: >4.0 mm in thickness with ulceration
Regional Lymph Nodes (N)		
N stage	N0	No regional metastases detected
	N1	N1a: Micrometastases in one lymph node
		N2b: Micrometastases in one lymph node
	N2	N2a: Micrometastases in 2-3 lymph nodes
		N2b: Micrometastases in 2-3 lymph nodes
		N2c: In-transit metastases/satellites without metastatic lymph nodes
	N3	≥ 4 metastatic lymph nodes, or matted lymph nodes, or in-transit metastases/ satellites with metastatic lymph node (s)
Distant metastases (M)		
M stage	M0	No evidence of distant metastases
	M1a	Metastases to the skin, subcutaneous tissue, or distant lymph nodes, normal serum lactate dehydrogenase (LDH) level
	M1b	Lung metastases, normal LDH level
	M1c	Metastases to all other visceral sites or distant metastases to any site combined with an elevated serum LDH level

Table 1.2 Clinical staging of cutaneous melanoma

Adapted from Balch *et al.* [44]. See Table 1.1. for definition of the TNM stages.

Staging		T	N	M
Stage I	IA	T1a	N0	M0
	IB	T1b T2a	N0 N0	M0 M0
Stage II	IIA	T2b T3a	N0 N0	M0 M0
	IIB	T3b T4a	N0 N0	M0 M0
	IIC	T4b	N0	M0
Stage III		Any T	N>N0	M0
Stage IV		Any T	Any N	M1

1.5.5 Prognostic factors in primary melanoma

As described before, histopathological factors such as tumour thickness, ulceration and mitotic rate were determined to be strong independent prognostic factors in melanoma and have been included in the calculation of the AJCC staging system [44]. Host factors such as sex, age at diagnosis and site of melanoma have also shown strong association with melanoma prognosis after adjusting for the histopathological factors of melanoma.

Sex is a well-known predictor of melanoma prognosis with males being consistently identified as a poor prognostic group in comparison to females [55, 56]. However, the underlying biological reasons to explain the survival advantage for women are still not clear. It was also reported that men have increased propensity to develop metastases and have higher mutation burden in their metastases in comparison to females [57]. The 5 year survival for males is around 70% and for females it is around 82% [13].

Age at diagnosis predicts poor prognosis, with the age group >70 years having the poorest prognosis [58]; this group also has a lower rate of sentinel lymph node positivity, which may imply that the poor prognosis of this group is due to the high morbidity [59, 60], or because of blood borne metastases which are more common

in older patients. Furthermore, older patients are less likely to have a strong immune system, leading to poor ability to mount an appropriate anti-tumour response [61]. The site of melanoma is another strong prognostic factor for melanoma. The site of primary melanoma is usually classified as limbs, head and neck, trunk, and other rare sites on the body, e.g. on genital skin or the sole of the foot. Melanoma occurring on the trunk has a worse prognosis in comparison to melanoma on the limbs or head and neck [62]. Melanoma most commonly occurs on the trunk and lower limbs for males and females respectively [13].

The other relevant histopathological features of melanoma that are not included in the AJCC staging system are tumour infiltrating lymphocytes (TILs), the presence of vascular or lymphatic invasion and the presence of tumour regression. The histopathological classification of tumours based on TILs was developed by Clark *et al.*, and it quantifies presence of immune cell populations surrounding and infiltrating the tumour [63]. Clark *et al.* categorised TILs as absent (when no lymphocyte aggregation is detected), non-brisk (presence of few lymphocytes) and brisk (when the lymphocyte population is large and lymphocytes have surrounded the tumour). The TILs classification has been shown to predict melanoma prognosis, with brisk TILs predicting good prognosis and the absence of TILs predicting the worst prognosis, independent of age, sex, tumour site, and the AJCC staging system [63, 64] [65]. Vascular or lymphatic invasion is the presence of tumour cells in blood vessels, which has been shown to be associated with a higher Breslow thickness, ulceration of the tumour, mitotic rate and a nodular melanoma subtype; furthermore it predicts poor prognosis [66-68]. Tumour regression is the disappearance of tumour cells, most likely as a consequence of interactions between the tumour cells and immune cells leading to replacement of the tumour tissue with non-malignant tissue [69-71]. A few studies have shown that tumour regression predicts prognosis in thin melanomas and is associated with other clinical prognostic variables such as sex, older age at diagnosis, head and neck or trunk site of melanoma [69, 70, 72]. However other studies have found no evidence to indicate that regression influences survival [73, 74].

1.5.6 Melanoma survival based on the AJCC stage

In the UK, the 5-year survival rate for patients diagnosed with melanoma is more than 95% for AJCC stage I [40]. The survival rate decreases by ~20% for the patients diagnosed at AJCC stage II [40]. The survival for AJCC stage III patients is 50%; at this stage the tumour has reached the lymph nodes. The survival further drops down

to 15-17% for AJCC stage IV melanoma [40]. At stage IV, the tumour has already metastasized to different organs in the body.

1.6 Genomic classification of melanoma

The AJCC staging system is a powerful tool for predicting melanoma prognosis, but it only considers the pathological state of the tumour and does not include the genomic changes that contribute to driving tumour progression. Advances in cost-effective DNA microarray and next-generation sequencing technologies have paved the way to discovery of biomarkers predicting disease outcome or response to treatment. These technologies have been broadly applied to detect changes in the tumour DNA to help in characterising the genomic landscape of melanoma.

1.6.1 Mutational subtypes of melanoma

BRAF and *NRAS* oncogenes have been identified as the most commonly mutated genes in melanoma [75-78]. Mutations in these genes are effectively mutually exclusive [79], and both activate the Mitogen-Activated Protein Kinase (MAPK) pathway [80]. In 2015, a study by The Cancer Genome Atlas consortium (TCGA) identified four mutation subtypes in a cohort predominantly of metastatic melanoma tumours [81]. The four mutation subtypes were named as *BRAF*, *RAS*, *NF1* and *triple wild type* mutant groups. In this study, among 318 tumours, 52% (n=166) harboured a *BRAF* mutation, with 144 tumours specifically having a *BRAF*^{V600E} mutation. The *RAS* subtype tumours harboured mutations mainly in *NRAS* (28%), and only a few tumours had mutation in other *RAS*-family member genes such *KRAS* and *HRAS*. Fourteen percent of tumours harboured *NF1* mutation, classified as the *NF1* subtype. The *triple-wild type* group contained a heterogeneous group of tumours harbouring mutations in different genes and was characterised based on the lack of mutation in *BRAF*, *NRAS* and *NF1* [81].

Melanoma when classified based on body site showed different mutation patterns on chronically sun-damaged skin (CSD) in comparison to non-chronically sun damaged skin (non-CSD) sites [82]. Early stage tumours of the CSD class harboured *NF1*, *BRAF*, *NRAS* or *KIT* mutations in the tumour, and the non-CSD class harboured only *BRAF* mutations [83-85]. In both groups, metastatic stage melanomas harboured somatic mutations in genes associated with key signalling pathways relating to cell proliferation (*NRAS*, *BRAF*, *NF1*), cell growth (*PTEN* and *KIT*), cell identity (*ARID2*), resistance to apoptosis (TP53) and cell cycle control (*CDKN2A*) [86, 87].

A study led by Shain *et al.* showed that mutation in the *BRAF* gene is acquired at an early stage when the tumour is benign [88]. Mutation in the *NRAS* gene is a characteristic feature of intermediate stage, along with mutations in other genes such as *TERT* and *PPP6C*. Overall, the sequence of mutations that lead to melanoma initiation are still unclear because different individuals harbour different mutations which lead to activation of different pathways. This supports the view that there are several biological routes to melanoma progression, and only a few of them are highlighted by these analyses [86]. The research on melanoma genomes continues with an expectation that understanding of disease biology and routes of melanoma progression can be improved by examining other data types such as copy number alterations, gene expression, and methylation status of the tumours.

1.6.2 Somatic copy number association with mutational subtypes

In melanoma, genomic regions associated with genes such as *CCND1*, *KIT*, *CDK4*, *TERT* and *MITF* are frequently amplified, and regions associated with *CDKN2A*, *PTEN* are frequently deleted [86]. The level of amplifications and deletions are a measure to estimate whether regions of the genome are duplicated or deleted during cell replication. In the TCGA study, analyses testing the association between the mutational subtypes and copy number alterations revealed that the *BRAF* mutational subtype had significantly higher copy number gains in *BRAF*, *MITF* and *PD-1*, *PDL-1* genomic regions of the tumour DNA than other mutational subtypes. The *NRAS* subtype had significantly more copy number gains in the *NRAS* genomic region in comparison to other mutational subtypes. The *NF1* subtype had significantly higher deletions in the *PTPRD* region. The *triple-wild type* mutational subtype had amplifications in *KIT*, *PDGFRA*, *KDR*, *CDK4*, *CCND1* and *MDM2* regions. Overall, the amplifications in *CCND1* and *TERT*, and deletions in *CDKN2A* and *PTEN* genomic regions were observed across all the mutational subtypes, suggesting that these copy number alterations are an intrinsic characteristic of metastatic melanoma tumours [81].

1.7 Gene expression profiling technologies

Gene expression is a process by which the genetic code or the nucleotide sequence of a gene is used in the synthesis of a functional gene product during transcription. Expression profiling at whole genome level is achieved using DNA microarray or RNA-sequencing technologies.

1.7.1 DNA microarrays

Microarray based gene expression profiling captures the molecular state of the cell by quantifying expression of thousands of genes simultaneously (Figure 1.3) [89-91]. In this high throughput technique, mRNA extracted from tissues samples is further amplified using PCR (Polymerase Chain Reaction), and then reverse transcribed into cDNA using reverse transcriptase enzyme (Figure 1.3). The cDNA sequences are stained using a fluorescent dye. As a next step, the stained cDNA sequences are hybridised to the unique short fragments of single stranded DNA bound to the plates, known as probes. After the hybridisation step, the plates are washed to retain the strong probe-cDNA binding sequences. The fluorescently labelled cDNA bound to a probe generates a signal, and the intensity of signal provides an estimate of the gene expression (Figure 1.3). There exist two main techniques for DNA microarray: oligonucleotide microarray and cDNA microarray. The oligonucleotide microarray uses short probe sequences which are 25 to 70 bases in length, and the cDNA microarray used probes of 200 to 2000 bases in length [89].

Gene expression profiling of disease using tissue specimens (such as blood, fresh tumour or archived Formalin Fixed Paraffin Embedded (FFPE) tumour) can lead to the discovery of biomarkers with prognostic potential that go beyond the histopathological classification [92]. Extracting RNA from FFPE samples stored over a long time yields low-quantity RNA, and analysing it is difficult because formalin fixation leads to crosslinking of RNA with proteins and causes nucleic acid to fragment [93]. To overcome these problems Fan *et al.* developed a gene expression assay known as DASL[®] (cDNA mediated Annealing, Selection, extension and Ligation) [94], which had locus specific probes designed for probe-cDNA hybridisation. The limitation of this assay was the limited number of probes, but introduction of a whole-genome DASL[®] assay has allowed profiling of approx. 29000 transcripts [92]. The RNA from the frozen FFPE samples is partially degraded but studies have shown that transcriptome signature based disease subtyping is still feasible using this technique [92, 95].

1.7.2 RNA-sequencing

RNA-sequencing (RNA-seq) is a another widely used technique for gene expression profiling which is based on next-generation sequencing (NGS) [96, 97]. In this technique, the mRNA extracted from the tissue sample is annealed to oligo-dT magnetic beads, and then a fragmentation agent is added to generate multiple fragments of mRNA. The fragments are reverse transcribed into single stranded

cDNA sequences. The opposite strands of cDNA are synthesized, and the resultant product is a double stranded cDNA. The double stranded cDNA sequences are end-repaired, ligated to adaptors, and PCR amplified to generate a library that is ready for sequencing. The sequences in the library are then sequenced using a NGS platform, and reads of ~30-500 bases long are generated. The reads are then aligned to a reference genome to produce a genome-wide transcriptional state which determines the level of expression for each gene [96, 97].

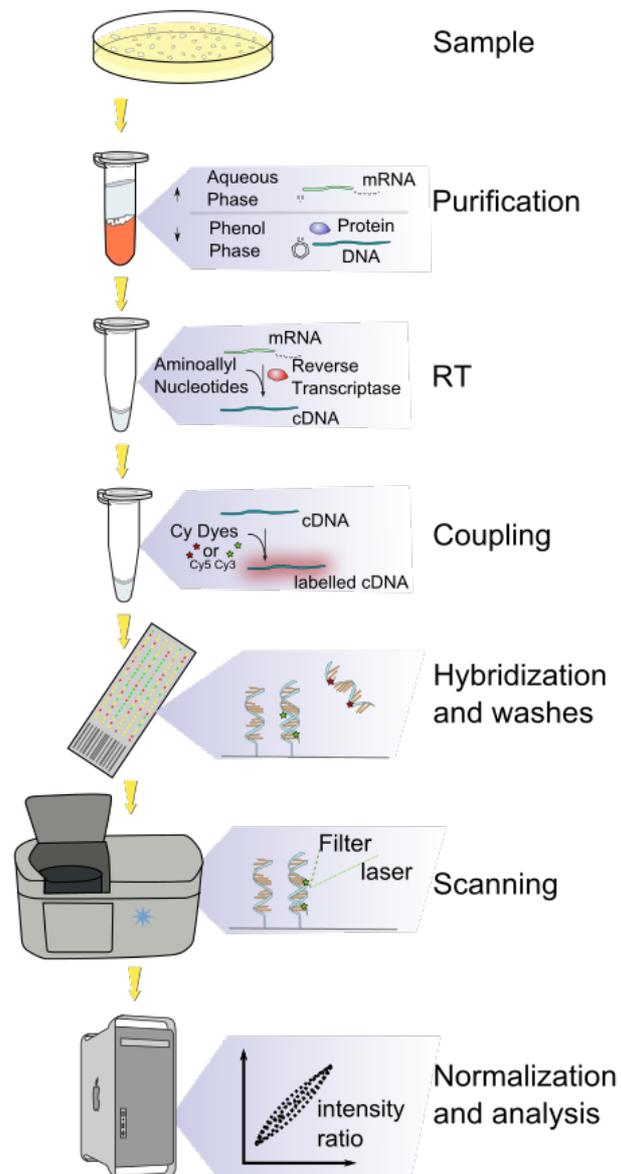


Figure 1.3 Summary steps of a microarray experiment

Adapted from [98].

Malone *et al.* compared DNA microarray and RNA-sequencing technologies and reported that both the technologies had a comparable performance and provided a reliable estimate of gene expression [99]. Both the technologies showed poor performance in detecting genes with low expression. RNA-seq has an advantage over DNA microarrays as one can study a lot of other characteristics (e.g. splicing, gene/isoforms, gene fusions, structural variations) apart from studying gene expression. The limitation of RNA-seq is that it is an expensive and more time-consuming technology than DNA microarrays [99].

Along with the advantages of high-throughput profiling techniques come various challenges: how to deal with the variability due to the use of different platforms, variability due to heterogeneous sources of material, and different ways of analysing the data.

1.8 Gene expression based classification of melanoma

Previously, numerous studies have used supervised and unsupervised classification approaches to generate gene expression-based signatures which predict melanoma tumour outcome (Table 1.3) [84, 100-109]. The unsupervised classification approach identifies novel subgroups by exploring distinct patterns in the dataset. The supervised classification, including machine learning, on the other hand mine the knowledge from known subgroups or outcomes and develop prediction models to classify new observations into these groups [110-112].

1.8.1 Supervised classification using gene expression

In 2006, Winnepenninckx *et al.* published the first study (Table 1.3) that analysed the gene expression data from primary melanoma tumours and generated a 254 gene-based signature predictive of metastasis free survival for 4 years (Agilent oligonucleotide microarray) [100]. When comparing the prediction performance of this signature with histopathological factors like tumour stage based on TNM staging, the signature misclassified 29% of samples in comparison to 28% misclassification when using histopathological factors alone. Overall the gene signature had similar prognostic value in comparison to the histopathological factors [100].

In 2007, Alonso *et al.* analysed the gene expression data from 34 vertical growth melanomas with every patient followed for at least 36 months (OncoChip DNA microarray) (Table 1.3) [101]. The study compared the patients who developed nodal metastases (n=21) with those who did not (n=13) and generated a 243 gene-based signature that predicted risk of metastasis. The biological pathway enrichment

analyses revealed that genes in the signature were associated with the epithelial-mesenchymal transition pathway, and these findings were further confirmed in an independent dataset [101].

In 2008, John *et al.* performed molecular profiling of 29 patients diagnosed at AJCC stages IIIB and IIIC with clinical outcome defined as melanoma progression in 2 years (oligonucleotide microarray) (Table 1.3) [102]. Thirteen patients did not experience disease progression in 2 years whereas 16 patients experienced disease progression. Comparing gene expression differences between the two groups identified 21 genes, which were experimentally validated and used to generate a predictive risk score. This score was applied on two independent datasets, and it accurately predicted outcome in 90% and 85% of patients respectively [102].

In 2009, Bogunovic *et al.* analysed gene expression data, mitotic rate and TILs from 38 melanomas to explore the molecular basis for metastasis and to generate biomarkers of melanoma survival (Affymetrix Human Genome microarray) (Table 1.3) [103]. Comparing gene expression differences between patients with prolonged survival and patients with short survival led to the generation of a 266 gene-based signature. For the 266 genes, pathway enrichment analysis revealed that immune response related genes were enriched in the good prognosis groups and cell proliferation related genes were enriched in the poor prognosis group. The prognostic value of the gene signature was validated in an independent dataset of comparatively similar size [103].

Conway *et al.* (former PhD student in the Leeds research group) performed gene expression profiling using archived FFPE tumour blocks of 254 melanoma patients (part of Leeds Melanoma Cohort described in chapter 2) diagnosed at AJCC stages I, II, and III (DASL microarray cancer chip based on 502 genes) (Table 1.3) [104]. Increased expression of the Osteopontin gene (SPP1) was identified as a prognostic biomarker predicting relapse-free survival in the training set. When jointly analysed with other clinico-histopathological factors of melanoma in a multivariable analysis, this gene remained a significant predictor of relapse-free survival. The prognostic significance of this gene was validated in an independent dataset (n=218). A follow up study in 2010 by Jewell *et al.* (another PhD student in the group) jointly analysed the two datasets analysed in Conway *et al.* (n=472) and identified that genes associated with DNA repair mechanisms as significant predictors of relapse-free survival [113]. The increased expression of DNA repair mechanism genes in progressive tumours supported the view that melanoma progression requires genetic stability.

In 2013, Mann *et al.* analysed the mutational and gene expression data generated from 73 melanoma patients diagnosed at AJCC stage III and identified a 46 gene-based signature that predicted outcome (Expression BeadChips microarray) (Table 1.3) [84]. Pathway enrichment analysis revealed overrepresentation of immune response mechanisms. The gene signature showed independent prognostic value when jointly analysed with clinico-pathological variables. The prognostic value of the signatures was validated in two previously published cohorts of AJCC stage III melanoma. The study concluded that *BRAF* and *NRAF* mutation, along with absence of immune related gene expression, is associated with poor prognosis in stage III melanoma [84].

In 2015, Gerami *et al.* analysed the differences between primary and metastatic melanoma tumours using available gene expression datasets and selected the 28 most discriminatory genes (Table 1.3) [109]. These 28 genes were used to develop a classification model to predict risk of metastasis using training set observations (n=164). When applying the classification model to the validation set (n=104), the model robustly predicted risk of metastasis (area under Receiver Operating Characteristic (ROC) curve =0.91) [109].

In 2018, Brunner *et al.* developed a classification model that predicted Melanoma-Specific Survival (MSS) based on expression values of 8 genes (Table 1.3) [108]. The classification model was developed using tumours in the training set (n=125) and was validated on the tumours in the test set (n=211). Both the sets were a mixture of primary tumours from AJCC stages I, II and III. The generated classification model significantly predicted MSS in both the training set and the test set observations. When the gene signature was jointly analysed with the AJCC staging system, the area under the ROC predicting MSS increased by 4% in the training set and 6% in the test set when compared to the AJCC staging system alone [108].

Table 1.3 Summary of studies investigating prognostic signatures of melanoma

Study	Cohort	Outcome	Gene signature and prognostic performance
Winnepenninckx <i>et al.</i> (2006) [100]	38 AJCC stage I-IV melanoma tumours	≥ 4 years distant metastasis-free survival versus < 4 year distant metastasis free survival	259 gene-based signature, signature misclassified 29% cases and histopathological variables misclassified 28% cases
Alonso <i>et al.</i> (2007) [101]	34 vertical phase melanoma tumours (21 with metastasis and 13 without)	Metastasis versus without metastasis	243 gene-based signature, only 3 genes were validated and showed association with survival
John <i>et al.</i> (2008) [102]	29 AJCC stage III and stage IV melanoma tumours	≥2 years to tumour progression versus <2 years to tumour progression	21 gene-based signature, no prognostic impact shown
Bogunovic <i>et al.</i> (2009) [103]	38 AJCC stage III and stage IV melanoma tumours	≥ 1.5 years of survival versus <1.5 years of survival	266 gene-based signature, independent prognostic indicator of survival when jointly analysed with mitotic rate and TNM stage
Conway <i>et al.</i> (2009) [104]	254 primary melanoma AJCC stage I-III	Relapse free survival and overall survival	1 gene, Osteopontin (<i>SPP1</i>) expression, gene expression did not maintain significance in the validation set
Jonsson <i>et al.</i> (2010) [106]	57 AJCC stage IV melanoma tumours	Overall survival difference in four groups	503 gene-based signature, strong association with overall survival in four groups and association of increased immune response with good prognosis.

Study	Cohort	Outcome	Gene signature and prognostic performance
Harbst <i>et al.</i> (2012) [105]	223 AJCC stage I and stage II melanoma tumours	Overall survival (OS) and relapse-free survival (RFS) difference in two groups	1864 gene-based signature, strong association with OS and RFS in two groups. Association of increased immune response with good prognosis
Mann <i>et al.</i> (2013) [84]	79 AJCC stage III melanoma tumours	<1 year survival versus survival > 4 years post-surgery	46 gene-based signature, strong over representation of immune response gene with good prognosis
The Cancer Genome Atlas Network (2015) [81]	331 primary and metastatic melanomas, AJCC stage I-IV	Overall survival difference between three groups	1500 gene-based signature, increased immune gene expression predicted good prognosis
Gerami <i>et al.</i> (2015) [109]	268 AJCC stage I-III melanoma tumours	Risk of metastasis	28 gene-based signature predicted risk of metastasis in the training and validation cohorts
Brunner <i>et al.</i> (2018) [108]	336 AJCC I-III melanoma tumours	Melanoma-specific survival (MSS)	8 gene-based signature predicted MSS in the training and validation dataset

1.8.2 Unsupervised classification using gene expression

The other studies have taken an unsupervised classification (clustering) approach to identify gene signatures of melanoma (Table 1.3). In 2010, Jonsson *et al.* (a group from the University of Lund with whom we collaborate) performed unsupervised hierarchical clustering on gene expression of 57 metastatic melanoma tumour samples; data was generated using an Illumina bead microarray [106]. The clustering identified four distinct classes: *High-Immune*, *Normal-like*, *Proliferative* and *Pigmentation*. This 4-class signature based on 503 genes will be referred to as the Lund 4-class signature in this report. The Lund 4-classes significantly predicted survival in stage IV melanoma as *high-immune* and *normal-like* class tumours had a good prognosis, and *pigmentation* and *proliferative* class tumours had significantly poorer prognosis. Histopathological factors such as Breslow thickness, mitotic rate, ulceration and AJCC stage were not observed to be associated with the Lund 4-classes, which could have been due to the small sample size (n=57) and the advanced stage of those tumours [106].

In 2012, the same group published another study where the Lund 4-class signature was validated in 223 melanoma primaries [105], and the signature collapsed into a 2-grade signature. The survival curves of Lund 4-classes showed a convergence into just two significantly different survival curves; furthermore this led to generation of a 2-grade signature, *high-grade* and *low-grade* (Figure 1.4A-B), from now on referred to as the Lund 2-grade signature [105]. The *high-grade* reflected metastatic tumours from *proliferative* and *pigmentation* classes, and the *low-grade* contained localised tumours from *high-immune* and *normal-like* classes [105].

In 2015, our group validated the Lund 4-classes and 2-grades on a subset of the Leeds dataset observing the previously reported association with survival [114].

In 2015, another study by the TCGA group clustered a mixture of metastatic and primary melanomas into three distinct classes [81]. The tumour classes were named as: *immune*, *keratin* and *MITF-low*. Survival analysis illustrated that patients from the *immune* class had better survival, while patients from the *MITF-low* and *keratin* class had intermediate and worse survival (Figure 1.4C) [81].

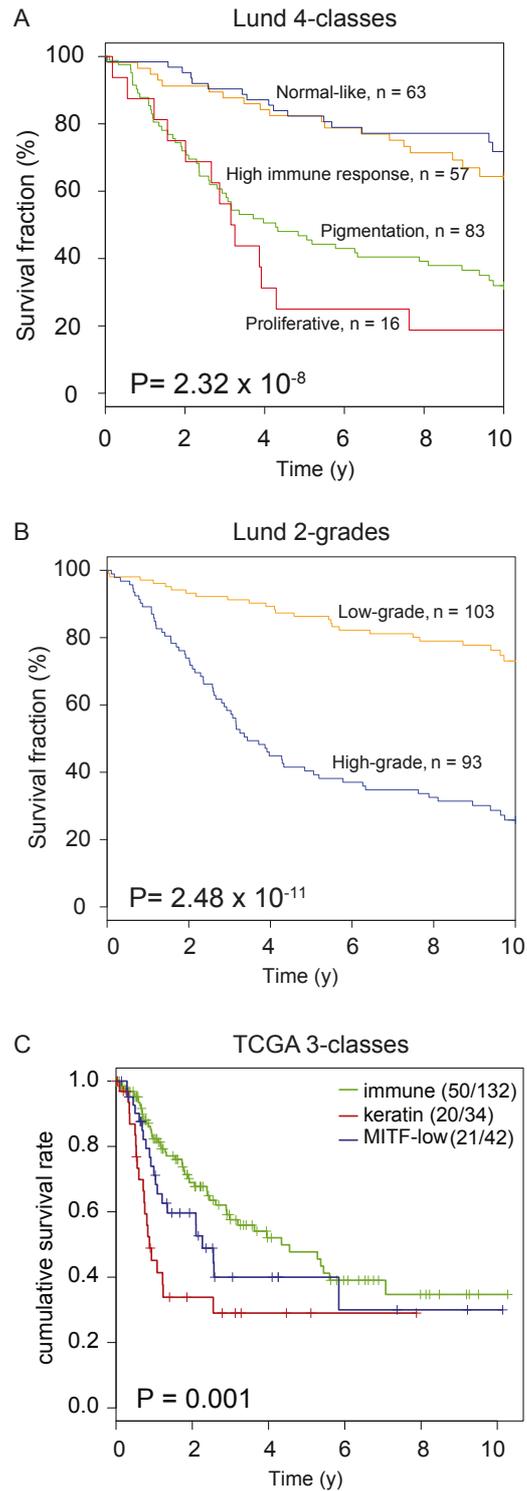


Figure 1.4 Melanoma tumour subtypes identified by the Lund group

Adapted from Harbst *et al.* [105], Kaplan–Meier curves showing overall survival of patients in (A) the Lund 4-classes and (B) the Lund 2-grades. (C) Adapted from the TCGA study [81], Kaplan-Meier curve showing survival of the TCGA 3-classes. The survival comparison was done using the log-rank test.

1.8.3 Overlap between the signatures

In 2011 Schramm *et al.*, conducted a systematic review of some of the previously published gene signatures of melanoma (between 2006-2010) and performed cross-validation of these signatures on different datasets from the respective studies [107]. A direct comparison of gene lists of these signatures identified only a few genes in common, but pathway enrichment analysis revealed consistent enrichment of immune response related mechanisms. Although some differences in classification by gene signatures was observed, the gene signatures consistently predicted prognosis in the majority of the datasets [107]. The analysed gene signatures did not outperform the clinico-histopathological characteristics in predicting prognosis, and therefore it was noted that the prognostic value of new gene signatures in future should continue to be assessed in a multivariable model including existing prognostic factors of melanoma [107].

In 2013, Liu *et al.*, analysed four previously published microarray based studies and made similar observations, as they also found that only a few genes were common across the published gene signatures [115]. The meta-analysis of the four datasets revealed a 200 gene-based signature that distinguished melanoma cells from normal skin cells. Comparison with other studies and gene enrichment analysis led to selection of a 12 gene-based signature which was validated experimentally. This study did not test the association of the signature with melanoma prognosis [115].

In 2016, Lauss *et al.* compared the molecular signatures of melanoma from unsupervised clustering [116]. It was reported that only 34 genes overlapped from the Lund and TCGA signatures, but pathway enrichment analysis revealed a good overlap of the biological pathways [116]. The Lund and TCGA signatures were developed agnostically using a similar clustering algorithm but were based on different types of analysis platform (array-based platform in Lund, RNA-sequencing in TCGA). Their convergence supports the view that transcriptomic data can indeed produce stable results. This indicated that even though gene signatures were generated from different platforms, they identify similar biological differences across tumours.

All the three studies mentioned above highlighted that different classification or clustering methods may select different genes, but an overlap of the biological pathways associated with these genes suggests that the methods might be selecting genes which are highly correlated and also belong to the same pathway. Hence a direct comparison of gene lists may not be useful in analysing the overlap between the signatures.

1.9 Outline of this study

Previous studies on unsupervised and supervised classification of melanoma have analysed gene expression datasets of comparatively small size, generated predominantly from advanced stage primary tumours, i.e. AJCC stages II and III. In 2015, our group had published a validation of the Lund 4-class and 2-grade signatures on a subset of LMC (Leeds Melanoma Cohort) tumours [114]. This study was based on 224 tumours of the LMC that were transcriptome-profiled at the time.

Currently the gene expression profiling has been done for 702 tumours of the LMC. In this study, various unsupervised and supervised classification approaches will be applied to the LMC dataset with an aim to identify molecular and prognostic signatures of melanoma.

1.9.1 Aims and objectives

Aim 1: Previously, melanoma tumours have been classified into 2-grades and 4-classes by the Lund group and into 3-classes by the TCGA group. The tumour groups have been shown to predict prognosis in metastatic and primary melanoma cohorts. To extend this further the current research has the following objectives,

Objective 1: To assess the prognostic value of the Lund and TCGA gene signatures in the whole LMC dataset (702 tumours)

Hypothesis: The gene signatures with prognostic value in the published literature will also predict survival in the LMC

Objective 2: To assess the prognostic value of the Lund and TCGA gene signatures across the AJCC stages

Objective 3: To cluster tumours of the LMC dataset using different methods to those used in deriving the existing signatures

Hypothesis: *De novo* cluster analysis of a comparatively larger dataset (up to 3 times larger than previous studies) may reveal novel tumour classes

Objective 4: To use an independent measure of cluster separation to assess stability of the newly identified tumour classes referred to as the LMC classes (Leeds Melanoma Cohort classes)

Objective 5: To test association of the LMC classes with clinico-histopathological features of melanoma

Objective 6: To test prognostic significance of the LMC classes in the whole LMC dataset and in the stage I subset

Objective 7: To generate a reduced signature of the LMC classes and replicate prognostic significance of the LMC classes in an independent dataset from Lund, Sweden

Objective 8: To explore biological differences between LMC classes using pathway enrichment analysis and using melanoma-specific biological modules

Aim 2: Supervised classification can be used to develop models to predict outcome in future patients using analyses from existing data where the outcome is known. Classification models will be developed using Random Forest (RF) and Support Vector Machine (SVM),

Objective 9: To apply RF to develop a classification model for predicting outcome in the LMC. The performance of the model will be assessed using sensitivity, specificity, and the kappa index

Objective 10: To apply SVM using linear and non-linear kernel functions to develop classification models for predicting outcome in the LMC. The performance of SVM model will be compared with the RF model

Objective 11: To generate a RF model after combining clinical information and gene expression data and compare its performance with a clinical information based RF model

Objective 12: To generate a refined RF model by performing variable selection
Hypothesis: The refined RF model using top predictor genes will predict prognosis similarly to the model using all genes

Objective 13: To validate the prognostic value of the refined RF on an independent dataset of primary melanoma from Lund, Sweden

Objective 14: Biological interpretation of the refined model using pathway enrichment analysis

1.9.2 Outline of the chapters

Chapter 2 provides a general methodological overview of the LMC cohort, generation and pre-processing of gene expression dataset and definition and analysis of MSS.

In Chapter 3, the Lund and TCGA signatures were applied to the LMC gene expression dataset, and association with MSS was tested across the AJCC stages especially in stage I tumours. It was hypothesized that clustering of LMC tumours

may identify novel molecular signatures of melanoma. Therefore, LMC tumours were clustered using three different clustering algorithms and the stability of clusters and agreement with the Lund and TCGA signatures was calculated. The clusters identified by an algorithm that demonstrated highest cluster stability were further explored.

In Chapter 4, the identified clusters, referred to as the LMC classes, were tested for association with clinico-histopathological characteristics and MSS stratified by the AJCC stages. To validate the prognostic value of the LMC classes on an independent dataset of primary melanoma (Lund dataset) generated by the Lund group, a refined gene signature was developed using the LMC data. This gene signature was applied to the Lund dataset to classify tumours into the LMC classes and association with prognosis was tested. The independent prognostic value of the LMC classes was assessed in both the datasets (LMC and Lund) using the AJCC staging system as a baseline.

In Chapter 5, two supervised classification approaches (Random Forest and Support Vector Machine) were applied to the gene expression dataset to develop classification models to predict outcome. The outcome was defined as MSS up to a selected time point. The classification models were generated using unbalanced and balanced training datasets, and performance was compared on a separate test dataset. To appraise the performance of the best performing classification model, several permutation-based classification models were generated and performance was compared with the best performing gene expression-based model.

In Chapter 6, performance of the selected gene expression-based model was compared with the classification model generated using clinical variables alone. It was hypothesized that combining the gene expression and clinical variables datasets may further improve prediction performance. Hence, the gene expression dataset and clinical variables were combined, and prediction of the combined model was compared to a clinical variable based model. Variable selection was performed to generate a final refined model which was further applied on the Lund dataset to predict outcome. Finally, the LMC classes identified in chapters 3 and 4 were compared with the predicted classes of the final refined model in this chapter.

Chapter 7 summarizes the main findings, sets them in context, presents a final discussion of the results, and highlights the strengths, limitations, and future perspectives of this study

Chapter 2

General Methods

2.1 Leeds Melanoma Cohort (LMC)

The Leeds Melanoma Cohort (LMC) is a population-based sample cohort of 2184 primary melanoma patients recruited from Yorkshire and the northern region of the United Kingdom [114, 117]. The melanoma cases in the LMC study were identified by clinicians, pathologists and from the cancer registry. The recruitment took place between 2000 and 2012.

The study was designed to understand the role of genetic and environmental factors on melanoma risk and on patient survival. After consenting to participate in the study, participants were asked to complete postal and telephone questionnaires. This generated detailed information about lifestyle factors and lifetime environmental exposures. Blood samples of the participants were collected during recruitment, for extraction of plasma, serum and DNA.

Participants in the LMC study were followed annually for first 5 years (active follow-up). General practitioner records and linkage to Office for National Statistics (ONS) death data was also used (passive follow-up). At the annual follow-up, research nurses contacted the consented participants to complete an extended questionnaire and sought their consent for donating another blood sample for studying blood constituents that are likely to be relevant for melanoma prognosis. Participants were consented for assessing formalin-fixed paraffin-embedded (FFPE) primary tumour blocks stored in the NHS (UK National Health Service) department where they had surgery for tissue sampling.

The approval for this study was granted by the Multi-centre Research Ethics Committee (MREC) and the Patient Information Advisory Group (PIAG 3-09(d)/2003). An amendment was sought from the Northern and Yorkshire Research Ethics Committee (01/3/057) for using the plasma samples for biomarker discovery [114, 117].

The variables used in this research were the genome-wide gene expressions extracted from a subset of FFPE tumours of the LMC and the clinical characteristics of these primary melanomas.

2.2 Gene expression data

Gene expression data was generated from the FFPE tumour blocks of 702 patients of the 2184 total cohort [117]. The data was generated from a selected group of participants in the cohort based on two criteria: 1) Breslow thickness greater than 0.75mm, and 2) patients with the longest follow-up time, to increase statistical power in survival analyses. Tumour less than 0.75mm thick were not selected due to difficulties in sampling. As a result, not all incident cases were sampled from the LMC. Among 702 patients, only 16 were treated with targeted therapies or immunotherapies, so the LMC data effectively reflects the outcome in treatment-naïve patients.

2.2.3 Tumour core generation

The tumour cores were extracted from FFPE blocks using a 0.6mm diameter tissue microarray (TMA) needle. A standard protocol has been developed by our group for managing and sampling tissues from FFPE blocks. The Human Tissue Act manager, Ms. Sandra Tovey led the process of tracing the FFPE blocks. The tissue sectioning, staining and sampling was performed by Dr. Filomena Esteves and Dr. Jonathan Laye. The sectioned tissues were stained using Mayer's Haematoxylin and 1% Eosin (H&E) to allow identification of areas for sampling after review. Prof. Julia Newton-Bishop and Dr. Jonathan Laye reviewed the H&E slides under a microscope and selected the areas for sampling. Up to two cores were sampled from each block, and to increase the comparability between tumours, the samples were consistently drawn from the regions with highest tumour content and least inflammation. All the tumour blocks were reviewed prior to sampling, and if there was a small amount of tumour left in the block then the block was not sampled, lest a clinically very important block be destroyed.

2.2.4 RNA extraction and Expression data generation

RNA was extracted from the tumour cores for the whole-genome gene expression assay by the lab technicians, following a previously established protocol (1,2). Whole-genome mRNA expression was profiled on Illumina's DASL (cDNA-mediated Annealing, Selection, extension and Ligation) HT12-v4 array [92]. For quality control, the mRNA was extracted from multiple cores for a number of patients (117 duplicates in total). However, gene expression data was analysed from only one extraction per patient in subsequent analyses, selecting the sample showing the best performances in terms of the number of genes detected.

2.2.5 Data pre-processing

Data pre-processing is a crucial step for removing technical variation or bias in the data [118-120]. This obscuring variation arises from different sources, for example differences in sample preparation, processing or scanning of the array. As described previously, the pre-processing of the gene expression data was conducted by Dr Jeremie Nsengimana (senior statistician) [117]. The pre-processing steps involved 1) image extraction using GenomeStudio (Illumina, Inc., San Francisco), 2) background correction to adjust out non-specific hybridisation (as measured by the negative control probes that are present on the array), 3) quantile-normalisation (steps 2 and 3 were accomplished using the R-package *Lumi* [121]), and 4) batch correction.

2.2.5.1 Sample quality control

Quality control was performed by examining the density plots of log-intensity distribution for each sample before and after pre-processing of data. This allowed identification of samples that had an abnormal distribution, which were subsequently removed. Data were re-normalised after removing the samples with an abnormal distribution. In previous studies conducted in our group [104, 122], the number of probes detected at $P < 0.05$ was shown to be correlated with sample quality and was a good measure of array performance. The detection P-value is the probability that the signal for a given probe is greater than the signal from the background noise, i.e. the average of negative controls. When sample duplicates existed, the sample with largest number of probes detected was chosen.

2.2.5.2 Batch Correction

Batch correction requires particular attention in microarray studies, in particular as not all samples were processed at the same time. They were run in 3 separate batches, each containing multiple plates which were loaded on multiple chips before scanning. As part of the normalisation process, Single Value Decomposition (SVD) was applied to the expression data to check for any biases. The top 25 principal components were tested for association with technical variables like chip, plate, batch number, age of the FFPE block, RNA concentration, and extraction assay. The principal components were found to be strongly associated with the chip, batch number and plate. The adjustment of expression data for batch number using a linear

model (R-package *SWAMP* [118]) removed the association with other technical variables. The data were converted to the log₂ scale for variance stabilisation.

2.2.5.3 Gene quality control

The array included 29,262 probes corresponding to 20,715 unique genes. For genes with multiple probes, the probe detected in the higher proportion of tumours was retained. Genes were further selected on the basis of two additional filters: 1) genes which were detected at P-value <0.05 in at least 40% of tumours and 2) had standard deviation (SD) > 0.40 (this SD threshold was chosen based on the overall distribution across the 20,715 genes on the log₂ scale, median 0.68) were selected for downstream analysis. Finally, the dataset contained 13,688 genes across 702 tumour samples. The expression values were standardised so that for each gene the mean was 0 and SD was 1.

2.3 Clinico-pathological characteristics

The following clinico-pathological characteristics of melanoma patients from the LMC (Table 2.1) were used:

1. Sex (male or female)
2. Tumour site (limbs, trunk, head and neck, other rare sites). Rare sites are those rarely sun-exposed such as subungual, anal, penile, vulvar, etc.
3. Presence of ulceration (yes or no)
4. Tumour stage (American Joint Committee on Cancer, or AJCC staging system)
5. Age at diagnosis (years)
6. Breslow thickness (mm)
7. Mitotic rate (per mm²) was stratified into two categories (<1/ mm² and ≥1/ mm²) based on the 7th edition AJCC staging system guidelines [44].
8. BRAF and NRAS mutation status (yes or no – V600 for BRAF, codons 12, 13 and 61 for NRAS)
9. Tumour Infiltrating Lymphocytes (TILs) (absent, non-brisk, brisk)

Table 2.1 Summary of clinico-histopathological characteristics of the LMC cohort

Variable		LMC cohort
Sex	<i>females n(%)</i>	384(55)
	<i>males n(%)</i>	318(45)
Site	<i>limbs n(%)</i>	298(42)
	<i>head and neck n (%)</i>	80(11)
	<i>trunk n(%)</i>	233(33)
	<i>other sites n(%)</i>	91(13)
Ulceration status	<i>no n(%)</i>	468(67)
	<i>yes n(%)</i>	234(33)
AJCC stage	<i>I n(%)</i>	233(34)
	<i>II n(%)</i>	355(51)
	<i>III n(%)</i>	106(15)
Mitotic rate	<i><1 n(%)</i>	88(14)
	<i>≥1 n(%)</i>	521(86)
BRAF mutant	<i>no n(%)</i>	309(53)
	<i>yes n(%)</i>	273(47)
NRAS mutant	<i>no n(%)</i>	432(75)
	<i>yes n(%)</i>	142(25)
TILS	<i>absent n(%)</i>	76(15)
	<i>non-brisk n(%)</i>	340(68)
	<i>brisk n(%)</i>	82(16)
<i>Age at diagnosis, median(range)</i>		58.4(18.3,81.3)
<i>Breslow Thickness, median(range)</i>		2.3(0.3,20)

2.4 Survival analysis

The aim of survival analysis is to test whether independent variables are associated with an event of interest. In our case, the event of interest was death from melanoma, and the predictors were the clinico-pathological and gene expression features. The survival time was the time from diagnosis to death from melanoma. Observations are censored when information about their survival time are incomplete [123]. Censoring can be further classified into left and right censoring. Left censoring occurs when the event of interest occurs before observations begin. Right censoring occurs for the

following reasons: a) a patient has not experienced the event of interest by the end of the study, b) a patient is lost to follow-up during the study, c) a patient died due to some other reason [123]. Patients who did not experience death from melanoma within the period of the study were censored at the time of last follow-up or of death from another cause. Kaplan-Meier curves were used for visualising survival curves, and Cox proportional hazards models were used for estimating the hazard ratios. In this study, it was assumed that all survival times are independent of one another, left censoring is not an issue because all the patients were recruited soon after diagnosis, and censoring solely occurs as right censoring.

2.4.1 Survival Outcome

Melanoma-specific survival (MSS) was used as the survival outcome. The MSS time was calculated from the date of diagnosis to the date of death caused by melanoma. For censored patients, survival time was calculated from the date of diagnosis to the date of the last follow-up or death from causes other than melanoma. Survival status was determined by looking at the definitive cause of death from the Office of National Statistics (ONS) and from the death certificate

2.4.2 Kaplan-Meier survival estimate

The Kaplan-Meier method is a non-parametric approach to estimate and visualize survival probabilities as a function of time [123, 124]. The survival probability $S(t)$ is the probability that an individual survives from the initial time (t_0) (e.g. date of cancer diagnosis) to a specified future time t . Survival analysis models $S(t)$ as a function of a set of predictor variables.

The survival probability is estimated as follows:

Suppose X patients have experienced the event of interest during follow-up at distinct times as $t_1 < t_2 < t_3 < t_4 \dots < t_n$.

Assuming that the events occurred independently of one another, the survival probability $S(t_j)$ at time t_j (indexed j) is calculated from $S(t_{j-1})$, the probability of survival at t_{j-1} , by:

$$S(t_j) = S(t_{j-1}) \left(1 - \frac{d_j}{n_j} \right) \quad (2.1)$$

where d_j is the number of deaths at t_j and, n_j is the number of patients alive before t_j . The value of S is constant along the time interval between events and changes only at the time of each event. At t_0 the value of S is equal to 1. Therefore, $S(t)$ is a step

function and this estimator includes each patient in the denominator for as long as they are known to be event-free. The graphical visualization of Kaplan-Meier survival probability against time provides a useful summary that can be used to estimate the median survival time.

2.4.3 Log-rank test

The Kaplan-Meier survival probabilities were compared across different groups using the log-rank test. This is a non-parametric test used for comparing survival times between two or more groups. The null hypothesis is that there is no difference between survival curves of the groups, and the alternative hypothesis is that there is. At each event time, this test calculates the expected number of events for each group since the previous event, under the null hypothesis. The expected values are summed over all the event times to give the total number of expected events (E_i) for each group (indexed i). The observed number of events (O_i) for that group are compared with their expectation under the null using the test statistic

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (2.2)$$

where k is the total number of groups compared.

The P value for this test statistic is derived from χ^2 distribution with $(k-1)$ degrees of freedom. The log-rank test provides a P value for the difference between the groups but it offers no estimate of the actual effect size. Several other survival modelling strategies have been proposed to overcome these limitations.

2.4.4 Cox Proportional Hazards model (CPH)

The Cox Proportional Hazards (CPH) model is a semi-parametric approach for modelling survival data [124, 125]. It is a regression-based method which describes the relationship between an event's occurrence and a set of covariates using a hazard function. All individuals who have not yet experienced the event of interest at any given time are considered to be at risk of experiencing it at a later time. The hazard function, usually denoted by $h(t)$ or $\lambda(t)$, is the probability that an individual will experience the event at time t , conditional on them having survived to t . Mathematically, the CPH model is written as

$$h(t) = h_0(t) \times \exp\{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p\} \quad (2.3)$$

where the hazard function $h(t)$ is dependent on the vector of p covariates (x_1, x_2, \dots, x_p) whose effect sizes are measured by the coefficients $(\beta_1, \beta_2, \dots, \beta_p)$. The coefficients are unknown and are estimated by maximum likelihood or, more precisely using partial likelihood. The partial likelihood approach considers the individuals with an event and not the ones which are censored and does not estimate the baseline $h_0(t)$:

$$L(\beta) = \prod_{i=1}^D \frac{\exp(\beta x_i)}{\sum_{j \in R_i} \exp(\beta x_j)} \quad (2.4)$$

Where D is the vector of failure times and R_i is the number of patients at risk at ordered time intervals $t_1 < t_2 < \dots < t_D$.

The baseline hazard function is estimated non-parametrically i.e. the survival times are not assumed to follow a particular statistical distribution. The function can vary with time but remains independent of the covariates, whence the assumption of proportional hazards. The quantity $\exp(\beta_i)$ is known as the hazard ratio (HR). $HR > 1$ corresponds to increased probability of the event of interest, while $HR < 1$ implies a decreased probability, for presence (or increased level) of that covariate. The proportional hazards assumption of the Cox model can be tested using Schoenfeld residuals. The null hypothesis is that the slope of scaled residuals on time is zero and the alternate hypothesis is that slope is not zero. The Schoenfeld residual value for each predictor variable is the difference between the expected value and observed value of the predictor variable. The residual values are regressed against time to test independence between residuals and time.

2.5 Statistical tests

2.5.1 Mann-Whitney/Wilcoxon test

The Mann-Whitney U test, also called Wilcoxon or rank sum test, is a non-parametric test equivalent of Student's t-test, used to test whether two statistical samples comes from the same distribution [126]. The test assumes that observations in both groups are independent and have no relationship between and within each group. The null hypothesis is that the distributions of two populations are the same and the alternate hypothesis is that they are not. The test calculates the U statistic whose distribution is known under the null hypothesis. The observations from both groups are merged

in one data set and ordered from the smallest (rank 1) to the largest (rank n , where n =total number of observations). The sum of ranks, R , is calculated for each group. The U statistic is given by:

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2} \quad (2.5)$$

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2} \quad (2.6)$$

$$U = \min(U_1, U_2) \quad (2.7)$$

where n_1 and n_2 is the number of observations in groups 1 and 2 respectively. R_1 and R_2 are the sums of ranks in groups 1 and 2. The U statistic follows a normal distribution, which is thus used to derive the significance level [126]. This test was used to compare groups of patients for dichotomous tumour and patient characteristics.

2.5.2 Kruskal-Wallis test

The Kruskal-Wallis test is an extension of the Mann-Whitney test for categorical variables with more than two categories. This is a non-parametric test used for comparing differences in more than two independent groups (k) [127]. The null hypothesis is the median of all groups are equal and the alternative hypothesis is that the median is different between at least two groups [127]. The observations from all groups are combined and ranked from lowest to highest with 1 being the lowest rank. The test statistic, H , is calculated as:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{r_i^2}{n_i} - 3(n+1) \quad (2.8)$$

where n is the total number of observations, k is the total number of groups, r_i is the sum of ranks in group i and n_i is the number of observations in group i . This statistic follows a χ^2 distribution with $k-1$ degrees of freedom under the null hypothesis [127].

2.5.3 Pearson's chi-squared test

The Pearson's chi-squared test is a non-parametric test for testing the association between two categorical variables [128]. The null hypothesis is that the two variables are independent and there is no association between them and the alternative

hypothesis is that there is an association. The test statistic utilises a contingency table for calculation and is calculated as:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2.9)$$

where r is the total number of rows, c is the total number of columns, O_{ij} is the observed value in the i^{th} row and j^{th} column of the table, E_{ij} is the expected value in the i^{th} row and j^{th} column of the table, computed as:

$$E_{ij} = \frac{\sum a_i * \sum a_j}{\sum_{i=1}^r \sum_{j=1}^c a_{ij}} \quad (2.10)$$

where a_i is the sum of i^{th} row and a_j is the sum of j^{th} row, a_{ij} is the sum of all the values in the table. Under the null hypothesis the statistic follows a chi-squared distribution with degrees of freedom equal to product of $r-1$ and $c-1$. [128].

Chapter 3

Analysing existing melanoma gene signatures and devising a new one by clustering of LMC tumours

The objectives of this chapter are:

Objective 1: To assess the prognostic value of the Lund and TCGA gene signatures in the whole LMC dataset (702 tumours)

Hypothesis: The gene signatures with prognostic value in the published literature will also predict survival in the LMC

Objective 2: To assess the prognostic value of the Lund and TCGA gene signatures across the AJCC stages

Objective 3: To cluster tumours of the LMC dataset using different methods to those used in deriving the existing signatures

Hypothesis: *De novo* cluster analysis of a comparatively larger dataset (up to 3 times larger than previous studies) may reveal novel tumour classes

3.1 Introduction

3.1.1 Gene expression based cluster analyses in melanoma

Previously, gene expression based cluster analysis of cutaneous melanoma has generated gene signatures with demonstrated prognostic value. Two studies led by the Lund [106] and the TCGA group [81] have identified biologically distinct classes of melanoma (Lund 4-classes and TCGA 3-classes) using similar clustering algorithms, i.e. hierarchical clustering. Both studies were conducted using the gene expression data derived from predominantly metastatic melanoma tumours. In 2012, Harbst *et al.* showed the prognostic value of the Lund signatures in primary tumours from the Lund cohort [114]. In 2015, our group replicated the prognostic value of the Lund signatures in the LMC. In 2016, a study led by the Lund group reported the overlap of biological pathways between the Lund and TCGA signatures [116].

3.1.1.1 Lund 4-class signature

As described previously (Section 1.8.2), the cluster analysis performed by the Lund group identified four tumour classes from the Swedish metastatic melanoma cohort (n=57 FFPE tumours), referred to as the Lund 4-class signature [106]. Hierarchical cluster analysis was used to generate this signature on the basis of the ~9,000 genes with the best detection performance on the Illumina DASL HT8 array [106]. The four classes were named *high-immune*, *normal-like*, *pigmentation* and *proliferative* [106]. The *high-immune* class was associated with higher expression of genes involved in different immunologic processes. The *normal-like* class expressed genes involved in epidermis and ectodermal development. The *pigmentation* class had higher expression of genes associated with melanin synthesis and melanocyte differentiation. The *proliferative* class had higher expression of genes involved in cell cycle related mechanisms. The four classes were also associated with *BRAF* and *NRAS* mutation status in metastatic tumours [106]. The *proliferative* and *pigmentation* class had a higher proportion of cases with *BRAF* and *NRAS* mutations in comparison to *high-immune* and *normal-like* classes [106]. In terms of prognosis, the *high-immune* class predicted good prognosis and *proliferative* class predicted poor prognosis, whereas the *pigmentation* and *normal-like* classes were predictive of intermediate prognosis [106].

3.1.1.2 Replication of Lund 4-class signature

The Lund 4-class signature was replicated in a cohort of primary melanoma tumours, referred to as the Lund primary melanoma cohort, composed of 223 patients diagnosed at AJCC stages I and II [105]. The signature robustly predicted prognosis in these primary tumours, with *high-immune* and *normal-like* classes associated with good prognosis, while *proliferative* and *pigmentation* classes predicted poor prognosis [105]. The Lund 4-classes were associated with clinico-pathological variables like Breslow thickness, ulceration status, mitotic rate, tumour site, AJCC stage, and tumour infiltrating lymphocytes (TILs) [105]. The *high-immune* and *normal-like* class tumours were thin, non-ulcerated and were predominantly diagnosed at AJCC stage IIA [105]. The *proliferative* and *pigmentation* class tumours were thick, ulcerated and were diagnosed predominantly at AJCC stage IIB [105]. The *high-immune* and *normal-like* tumours had lower mitotic rate in comparison to *proliferative* and *pigmentation* class tumours. The *high-immune* class had higher incidence of brisk TILs and the *pigmentation* class had absence of TILs [105]. The *normal-like* and *proliferative* class had intermediate levels of non-brisk TILs. Brisk

TILs are those infiltrating the tumour core to attack it, while non-brisk tend to remain at the tumour periphery and are low in number. The Lund 4-classes were not associated with *NRAS* and *BRAF* mutation status in primary melanoma [105].

3.1.1.3 Lund 2-grade signature

The survival curves of the Lund 4-classes converged into two groups in the Lund primary melanoma cohort (Section 1.8.2) [105]. Further examination of histological parameters, overlap between upregulated genes and survival across the four classes led to refining the 4-class signature into a new two-grade signature, referred to here and throughout as the Lund 2-grades. The Lund 2-grades were *high-grade* and *low-grade*. *High-grade* tumours were an aggressive class of tumours which overlapped with tumours from the *proliferative* and *pigmentation* classes [105]. *Low-grade* class tumours overlapped with *high-immune* and *normal-like* class tumours. In contrast to Lund 4-classes, the Lund 2-grades showed significant association with *BRAF* mutation status and borderline association with *NRAS* mutation status [105]. Overall, the Lund 2-grades signature is an extension of Lund 4-classes, with higher statistical power to detect associations with survival and clinical variables. This new signature was not obtained by merely merging pairs of groupings from the Lund 4-classes; rather it was generated from 1864 genes, although many of these were also among the 503 defining the Lund 4-classes.

3.1.1.4 Independent replication of Lund signatures

In 2015, our group applied the Lund 4-class and Lund 2-grade signatures to a subset of LMC tumours (n=204 FFPE primary tumours profiled at the time) and an additional set of 76 melanomas (predominantly metastases from the Leeds Chemotherapy Study, LCS) using the nearest centroid classification (NCC) approach [114]. The expression data was generated on the DASL HT12.4 array. Both signatures were well replicated in these independent cohorts. In keeping with previous reports, *high-immune* and *normal-like* classes were predicted good prognosis, while *proliferative* and *pigmentation* groups predicted worse prognosis. In Lund 2-grades, *high-grade* predicted worse prognosis and *low-grade* predicted good prognosis. The Lund 4-classes and Lund 2-grades were associated with clinico-pathological features like ulceration status, mitotic count, age at diagnosis, Breslow thickness, and AJCC stage. These observations were consistent with previous studies [105, 106]. The Lund 4-classes and Lund 2-grades were not associated with *BRAF* and *NRAS* mutations in the LMC tumours [114].

3.1.1.5 TCGA 3-class signature

In 2015, The Cancer Genome Atlas (TCGA) network published a study which identified three classes of melanoma in a predominantly metastatic cohort (262 metastases and 58 primaries from fresh frozen tumours) (Section 1.8.2) [81]. The three TCGA classes were named as *immune*, *keratin* and *MITF low*. The *immune* class was associated with good prognosis and the *keratin* class was associated with poor prognosis, whereas the *MITF low* class had intermediate prognosis. The *immune* class tumours had increased expression of genes involved in *immune* response mechanisms. The *keratin* class had increased expression of genes involved in *keratin*, epithelial and neuronal development mechanisms. The *MITF low* class tumours had increased expression of genes associated with embryonic development, nervous development, cell adhesion, and extracellular matrix protein organisation [81].

3.1.1.6 Overlap between Lund and TCGA signatures

In 2016, a collaboration between the Leeds and Lund groups compared the Lund 4-class signature with the TCGA 3-class signature (Figure 3.1) [116]. Although the 2 signatures have only 34 genes in common (among the 503 of Lund 4-classes and 1500 of TCGA), the gene ontology of both signatures showed a good overlap of the associated biological pathways (Figure 3.1) [116]. The Lund 4-class and TCGA 3-class signatures were applied to the TCGA (n=254, metastases), Lund (n=124, stage III primaries), Leeds (n=204, primaries) and Bergen (n=54, metastases) melanoma datasets using the supervised NCC approach [116] and it was observed that Lund 4-classes highly overlapped with the TCGA 3-classes in each of these datasets (Figure 3.1B). The tumours classified in the *high-immune* class of the Lund 4-classes overlapped with the *immune* class of TCGA 3-classes. Similarly, the *normal-like* class overlapped with the *keratin* class. The *proliferative* class overlapped with *MITF low* class. The *pigmentation* class overlapped with *immune* and *keratin* classes of the TCGA 3-classes. Overall, the study suggested that even though Lund and TCGA signatures were developed using different sample types (archived tissue in Lund and fresh-frozen tissue in TCGA), and different technologies (microarray-based platform in Lund, RNA-seq in TCGA) (see Section 1.8.3), the convergence of signatures supports the view that transcriptomic data do produce stable results [116]. These data also confirmed that similar mechanisms can be observed in primary tumours and in more advanced disease.

3.1.2 Replication of Lund and TCGA signatures in the whole dataset

The above-mentioned studies using the Lund and TCGA signatures have been conducted on melanoma cohorts containing mainly AJCC stages II-IV tumours. Therefore, the prognostic significance of these signatures has not been assessed in patients diagnosed at AJCC stage I. At the time of replication in LMC [114], expression data was available for only 204 tumours, including 58 stage I tumours. Now, with a dataset three times larger (n=702, including at 233 stage I), the first objective of this research was to test the utility of these signatures in the whole LMC dataset, especially in stage I tumours. As a second objective, since the whole LMC dataset had more power, it was hypothesised that re-clustering tumour transcriptomes of LMC could identify novel prognostic signatures. As described in this chapter, a consensus-based approach was applied to cluster gene expression data of primary melanomas from LMC using three different unsupervised classification algorithms. The results from these algorithms were compared for cluster stability and for their agreement with the Lund and TCGA signatures.

3.2 Methods

3.2.1 Replicating Lund 4-class and Lund 2-grade signatures in LMC

Supervised Nearest Centroid Classification (NCC) approach was used to classify the 702 primary tumours from the LMC into the Lund 4-classes [114]. The Lund 4-class signature contains the centroid values for each class. The centroid values are an average expression value of 503 genes in the signature. Among the 503 genes of the signature, 449 genes were present in LMC (incomplete overlap due to using different versions of DASL array).

In the NCC approach, the tumour expression values of the matching 449 genes were extracted from the LMC gene expression dataset. Each tumour sample was correlated with the Lund 4-class centroids, and the sample was put into the class corresponding to the largest correlation. The samples for which the largest Spearman's rank correlation coefficient value was less than 0.10 were considered unclassified.

Similarly, LMC tumours were classified into the Lund 2-grades using the supervised NCC approach. The Lund 2-grades centroids are average expression value of 1864 genes, among which 1586 were present in the LMC dataset. The classification of LMC tumours using the TCGA 3-class signature had already been performed by Dr Jérémie Nsengimana using the same NCC approach.

3.2.2 The Lund and TCGA signatures association with clinico-histopathological variables

The association of LMC tumours classified into the Lund 4-classes, Lund 2-grades, and TCGA 3-classes with clinico-histopathological variables was tested. The clinico-histopathological variables used for analyses were age at diagnosis, sex, AJCC stage, Breslow thickness, site of tumour, tumour ulceration status, mitotic rate, TILs, and *BRAF* and *NRAS* mutation status. The mitotic rate calculation for the LMC tumours was done by Dr Sally O'Shea (former PhD student in the group). Pearson's chi-squared test was used for testing categorical variables. Non-parametric Mann-Whitney test and Kruskal-Wallis rank sum test were used for testing continuous variables (refer to 2.5). Association analyses excluded the unclassified tumours.

3.2.3 Lund and TCGA signatures association with melanoma-specific survival

CPH models (refer to 2.4.4) were used to test the association of the Lund and TCGA signatures with melanoma-specific survival (MSS), and the survival differences between tumour classes were plotted in Kaplan-Meier curves (refer to 2.4.2). The differences in these curves were tested using a log-rank test (refer to 2.4.3). These analyses were conducted using R-packages *survival* [129], *ggplot2* [130] and *ggkm* function. The *high-immune* class from the Lund 4-classes, *low-grade* from the Lund 2-grades, and *immune* class from the TCGA 3-classes were used as the baseline comparison group. The analyses excluded the unclassified tumours.

3.2.4 Statistical interaction test between the Lund grade and AJCC stage

The Lund and TCGA signatures were generated from cohorts of AJCC stage II-IV melanomas. Since LMC dataset includes a high proportion of stage I patients (33.5% of data), the utility of the Lund and TCGA signatures was tested across stages. To check whether the signature's prognostic value was maintained at all disease stages, and having observed no evidence of prognostic values in Stage I, we conducted a

formal statistical interaction test between the AJCC stage and the Lund 2-grades in the CPH regression was tested using Wald's test. This signature was chosen due to its higher statistical power among the three signatures. Following the same line of reasoning, because the signature is known to be prognostic in the AJCC stages II and III, these two levels were merged, i.e. the AJCC stage variable had two levels: stage I or higher (there was no stage IV, so higher means II or III). The AJCC stage I and *low-grade* from Lund 2-grades were chosen as baseline comparison groups.

3.2.4.1 Power calculation in LMC stage I

Power and sample size calculations were performed to ascertain whether the stage I sample size is large enough to draw valid conclusions. The equation used for calculating power is [131]:

$$d = \frac{\left(\frac{z_{\alpha}}{2} + z_{\beta}\right)^2}{\pi_1 \pi_2 \theta_R^2} \quad (3.1)$$

where $z_{\alpha/2}$ and z_{β} are values taken from standard normal distribution at a given level of significance (or type I error), α ; β is the type II error and $1-\beta$ is the statistical power of the test. The π_1 , π_2 are the proportion of patients in the two groups (i.e. *low-* or *high-grade*). Parameter θ_R is the log-hazard ratio (natural base) and d is the total number of melanoma-specific deaths in AJCC stage I patients.

3.2.5 Clustering gene expression data of LMC tumours

The clustering of gene expression data was performed for 702 tumours of LMC. The whole transcriptome contained expression of 20,715 unique genes. The gene were further filtered prior to cluster analysis, based on two criteria: genes had to be detected with $P < 0.05$ in at least 40% of tumours and had to have a standard deviation (SD) > 0.40 . This SD threshold was chosen based on the overall distribution across all the genes on the log₂ scale. The median SD was 0.68. The data were standardized to give each gene a mean of 0 and SD of 1. After filtering, the final dataset contained expression of 13,688 genes, larger than the 7,200 genes used in the Lund study [106] and the 1,500 genes used in the TCGA study [81]. We chose a larger number in order to cover more biological variation. The LMC tumours were clustered using a consensus-based approach with implementation of Hierarchical clustering (HC), *k*-means (KM) and Partitioning Around Medoids (PAM) algorithms.

3.2.5.1 Hierarchical clustering (HC)

HC is a data partitioning method that generates hierarchical series of nested clusters using a distance metric e.g. correlation coefficient, dissimilarity (1-correlation coefficient), Euclidean distance etc. [132]. The tree-based graphical representation of these nested clusters is called a dendrogram. The dendrogram records the information about similarity between the clusters. The k clusters are generated by cutting the dendrogram at a specified threshold of distance/similarity. HC can be done using a divisive or agglomerative approach [132]. The divisive approach, also known as top-down, starts with placing all the samples in one big cluster. It then iteratively divides the big cluster into smaller clusters until the final clusters contains only one sample per cluster. The agglomerative approach, also known as the bottom-up, starts with placing all samples into individual clusters, which are then merged in successive steps based on the similarity between them. This process is repeated until all clusters are merged together to form one big cluster. In my thesis, an agglomerative approach was used. The similarity D between the clusters is calculated using UPGMA (Unweighted Pair Group Method with Arithmetic mean) as follows:

$$D(A, B) = \frac{1}{|A| \cdot |B|} \sum_{x \in A} \sum_{y \in B} d(x, y) \quad (3.2)$$

where A and B are the two clusters, $|A|$ is the total number of samples in cluster A and $|B|$ is the total number of samples in cluster B . The symbol $d(x, y)$ is the distance (here I used dissimilarity) between sample x from cluster A and sample y from cluster B . Although HC provides good visualisation of similarity between samples, it is computer-intensive. The limitation of HC is that it does not provide the number of clusters: it progressively groups or splits samples based on similarity, and users, often subjectively, decide the number of clusters based on the dendrogram. Another limitation is that a small perturbation in the dataset can greatly change the end results of the clustering, and it is sensitive to outliers [132, 133].

3.2.5.2 k -means clustering (KM)

The KM clustering algorithm aims to partition data into a predetermined (i.e. user input) number of clusters, k , into which the samples are placed based on their similarities [132]. The mean value of samples in a cluster is known as the cluster centroid. The clustering process has two main steps: 1) initialisation of centroid values; 2) iterations to find optimal centroid values. In the initialisation step, k samples are randomly selected and assigned to different clusters (i.e. one sample per cluster).

In the iteration step, each of the remaining $n-k$ samples is classified into the cluster with nearest centroid. Then for each cluster, the mean value is re-calculated and assigned as the new centroid of the cluster. The samples are re-classified using the new centroids and re-assigned to the cluster with the nearest centroid. These steps are repeated until the centroid values do not change. The iteration step minimises the sum of squared distances (S) between samples within a cluster and its centroid as

$$S = \sum_{i=1}^k \sum_{x \in C_i} |x - u_i|^2 \quad (3.3)$$

where k is the number of clusters, x are samples in cluster C_i , u_i is the centroid of cluster C_i . The KM method is computationally efficient, but like HC it is sensitive to outliers and a small change in starting position can change the clustering results. Also, the number of clusters, k , has to be provided as an input when the user rarely has much information on which to base this choice [132, 133].

3.2.5.3 Partitioning Around Medoids (PAM)

Partitioning around medoids (PAM), also known as k -medoid clustering, aims to minimise the within cluster distance W

$$W = \sum_{i=1}^k \sum_{x \in C_i} d(x, m_i) \quad (3.4)$$

where k is the number for clusters, x are the samples in cluster C_i , m_i is the medoid (central element) of cluster C_i , and $d(\cdot)$ is the distance function (here I used Euclidean distance) [134, 135]. It is constructed in a similar way to k -means but uses the cluster medoid (i.e. the most central element) instead of the average of all samples in a cluster, thereby avoiding the influence of outliers. The clustering solution of the data is at the minimum value of W . The algorithm performs clustering in two phases, 1) build phase; 2) swap phase. In the build phase, PAM randomly assigns k samples as the medoids (i.e. 1 sample per cluster). The rest of the samples are classified into a cluster based on their nearest medoid, and W is calculated. In the swap phase, a new medoid is determined for each cluster by finding a sample that minimizes the dissimilarity with other samples of the cluster. The samples are reclassified into the clusters based on the new medoids. The swapping phase is repeated until the W value stabilizes. The usage of medoids in PAM makes it more robust to outliers in

the dataset compared to KM [136], although like KM, PAM requires a priori information on the number of clusters, k .

3.2.5.4 Consensus-based clustering

All the methods described above, HC, KM and PAM are sensitive to starting positions. Hence, a method was needed to help inform the choice of k . Consensus-based clustering is an iterative technique for identifying stable clusters in high-dimensional datasets [137]. It sub-samples data to generate multiple datasets over N iterations (Figure 3.2). At each iteration, the sampled dataset is clustered into a varying number of clusters k using the specified clustering algorithm and distance metric. In the sampled dataset, each pair of samples is assigned a consensus value of 1 if they are classified in the same cluster or 0 otherwise. Across N iterations (re-samples) and for each k value, the sample pair consensus is the proportion of iterations in which the pair is put in the same cluster by the chosen clustering algorithm. This consensus information is stored in an $n \times n$ consensus matrix where n is the total number of samples (Figure 3.2). Finally, the consensus matrix values are clustered using hierarchical clustering and the resulting dendrogram is split at the appropriate threshold to give k clusters. In this study, at each iteration, we sampled 80% of genes and samples and varied k values from 2 to 12.

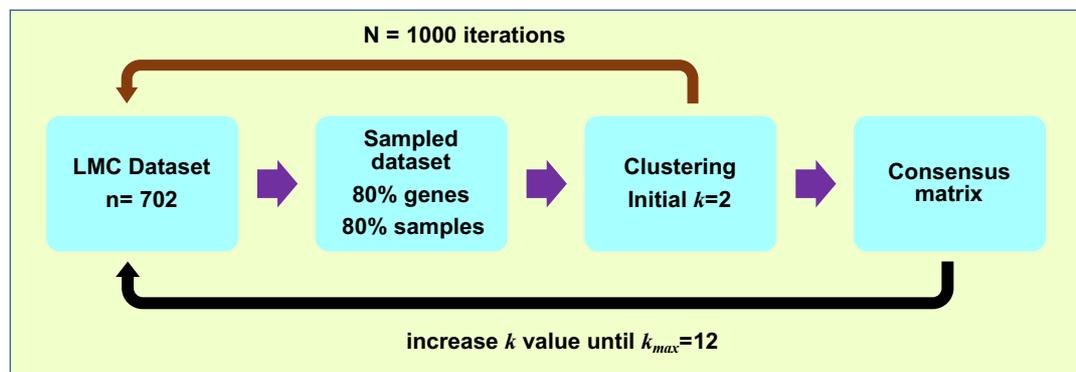


Figure 3.2 Consensus clustering workflow

The LMC data contained expression values for 702 tumours. A random subset of 80% of genes in 80% of tumours were sampled with 1000 repetitions. These data were clustered using different algorithms for $k=2$ to 12.

Consensus clustering of LMC tumours was performed with HC, KM and PAM algorithms in the consensus clustering workflow. The correlation coefficient-based

metric has been reported to perform well with HC and Euclidean distance has been observed to have higher performance when applied with KM and PAM algorithm [138]. The analysis was performed using the R-package *ConsensusClusterPlus* [137, 139] with the distance metric 1-Spearman's correlation coefficient for HC and Euclidean distance for KM and PAM.

3.2.5.4.1 Selecting the number of clusters k

As stated above, the consensus clustering was performed by varying k from 2 to 12 and a consensus matrix obtained at each k was stored for downstream analysis. The consensus matrix contains the calculated consensus score for each pair of samples. The consensus score, $M(i,j)$, is the proportion of times that samples i and j are observed in the same cluster out of all the times these 2 samples were both observed in the sampled subset, over N iterations. The choice of k was determined by visualising a number of graphics, 1) the consensus matrix heatmap; 2) the cumulative density function (CDF) of the consensus score values in consensus matrix; 3) relative change (delta) in the area under the CDF curves. The consensus matrix heatmap is generated for each k value. For a consensus matrix, the $CDF(c)$ can be defined as the proportion of pairs whose consensus score is less than or equal to c ($0 \leq c \leq 1$). The term $A(k)$ is the area under CDF curve of a consensus matrix at k clusters. The delta graph contains relative changes in $A(k)$ in comparison to $A(k-1)$.

The CDF and area under the CDF was calculated as:

$$CDF(c) = \frac{\sum_{i < j} 1\{M(i,j) \leq c\}}{\frac{n(n-1)}{2}} \quad (3.5)$$

$$A(k) = \sum_{i=2}^m [x_i - x_{i-1}] CDF(x_i) \quad (3.6)$$

where $1\{ \}$ denotes the indicator function, n is the total number of samples, k is the number of clusters being tested, M is the consensus matrix obtained at that k and (x_1, x_2, \dots, x_m) is the sorted set of entries of the consensus matrix M (with $m = n(n-1)/2$, i.e. the total number of possible sample pairings). The relative change in the CDF, $\Delta(k)$, was computed as

$$\Delta(k) = \begin{cases} A(k), & \text{if } k = 2 \\ \frac{[A(k) - A(k-1)]}{A(k-1)}, & \text{if } k > 2 \end{cases} \quad (3.7)$$

Finally, the k value is selected on the basis of the $\Delta(k)$ value by choosing the k when there is a minimal increase in area from $A(k-1)$ to $A(k)$, and hence little improvement in cluster stability as the number of clusters decreases.

3.2.6 Comparing the HC, KM and PAM based clusters

The tumour clusters derived with consensus HC, KM and PAM were compared for stability and agreement with previous signatures. The stability of the clusters was calculated as the average consensus score of the samples in the clusters. The agreement between these new tumour clusters and those obtained by applying previously published Lund and TCGA gene signatures in a supervised NCC manner (see section 3.2.1) was calculated using Cramer's V statistic [140]. Cramer's V, an extension of Cramer's Phi, is a statistic used to measure the strength of association between two nominal variables. Cramer Phi is used when a contingency table has size 2 X 2. For larger tables, Cramer's V is used and it is calculated as:

$$Cramer's\ V = \sqrt{\frac{\chi^2}{c(m-1)}} \quad (3.8)$$

where, c is the total cell count from the contingency table, m is the minimum value between the number of rows and the number of columns and χ^2 is the chi-squared statistic. Cramer's V varies between 0 and 1 regardless of the dimension of the nominal variables. A value of 1 represents the maximum agreement, while 0 represents no agreement. The calculations were performed using R-package *VCD* [141].

3.3 Results

3.3.1 Applying the existing signatures to the whole LMC dataset

The Lund (4-class and 2-grade) and TCGA signatures (3-class) were applied to the whole LMC dataset ($n=702$) using the NCC approach. The LMC tumours which could not be classified were labelled unclassified.

With the Lund 4-class signature, the NCC method classified 25% of tumours into the *high-immune* class, 28% into the *normal-like* class, 32% into the *pigmentation* class, and 12% into the *proliferative* class (Table 3.1), while 3% of tumours could not be classified. The *high-immune* class (median $r=0.40$), *normal-like* class (median

$r=0.40$) and *pigmentation* class tumours (median $r=0.38$) had similar level of correlation with their class centroids (Table 3.1), while the *proliferative* class tumours had a much lower correlation with the *proliferative* class centroid (median $r=0.26$).

With the Lund 2-grade signature, 38% of LMC tumours were classified into the *low-grade* (Table 3.1) and 42% into the *high-grade*, whereas 20% of tumours were unclassified. The *low-grade* (median $r=0.31$) and *high-grade* (median $r=0.31$) tumour has similar correlation with their respective class centroids (Table 3.1).

The classification of LMC tumours based on TCGA 3-class signature was performed by Dr Jeremie Nsengimana. Twenty seven percent of tumours were classified in the *immune* class, 35% in the *keratin* class and 21% in the *MITF low* class, while 16% could not be classified into any of the classes. The *immune* class (median $r=0.25$) and *keratin* class (median $r=0.25$) had comparatively higher correlation with their respective class centroids and *MITF low* class (median $r=0.22$) had slightly lower correlation with its class centroid (Table 3.1).

Table 3.1 Summary statistics of the LMC tumour classification using the Lund and TCGA signatures

Three percent, twenty percent and sixteen percent of LMC tumours could not be classified using the Lund 4-class, Lund 2-grade and TCGA 3-class signatures respectively. The r value is the median correlation in each of the classes.

Signature	Class	n(%)	r
Lund 4-class	<i>High-immune</i>	175(25)	0.40
	<i>Normal-like</i>	198(28)	0.40
	<i>Proliferative</i>	222(32)	0.38
	<i>Pigmentation</i>	84(12)	0.26
	Unclassified	23(3)	0.07
Lund 2-grade	<i>Low-grade</i>	266(38)	0.31
	<i>High-grade</i>	296(42)	0.31
	Unclassified	140(20)	0.05
TCGA 3-class	<i>Immune</i>	192(27)	0.25
	<i>Keratin</i>	247(35)	0.25
	<i>MITF low</i>	150(21)	0.22
	Unclassified	113(16)	0.06

3.3.2 Clinico-histopathological association with the Lund and TCGA signatures

Previous studies have shown the association of the existing signatures with melanoma prognostic factors such as Breslow thickness, AJCC stage and ulceration status [105, 106, 114]. It was tested whether these associations can be reproduced in the LMC dataset.

In LMC dataset, the Lund 4-class signature was significantly associated with patient sex, tumour site, age at diagnosis, Breslow thickness, AJCC stage, ulceration status, mitotic rate, TILs, and *NRAS* mutation status (Table 3.2). The *high-immune* and *normal-like* tumours were thin, non-ulcerated, early stage tumours, whereas the *proliferative* and *pigmentation* tumours were thicker, ulcerated and diagnosed at an advanced stage. The *high-immune* and *normal-like* tumours occurred more frequently on limbs while the *proliferative* and *pigmentation* tumours were observed more frequently on other body sites. The *high-immune* tumours had a higher incidence of brisk TILs than the *proliferative* tumours, while *pigmentation* and *normal-like* tumours had intermediate levels of brisk TILs. *BRAF* mutation status showed weak evidence of association with the Lund 4-classes ($P=0.07$), but *NRAS* mutation status was significantly associated ($P=0.01$), with the *proliferative* tumours having more frequent *NRAS* mutations (Table 3.2). The *normal-like* and *pigmentation* tumours showed a relatively lower fraction of patients with *NRAS* mutations in comparison to the *proliferative* class, while the *high-immune* tumours contained the lowest proportion of *NRAS* mutated tumours (Table 3.2).

Similar to the Lund 4-classes, and unsurprisingly, the Lund 2-grades also showed significant association with tumour site, age at diagnosis, Breslow thickness, AJCC stages, ulceration status, mitotic rate, TILs, *BRAF* and *NRAS* mutation status (Table 3.2). The *low-grade* tumours were thin, non-ulcerated and diagnosed at an early stage compared to the *high-grade* tumours. The *low-grade* tumours also occurred more frequently on limbs and were more frequently *BRAF* mutated. By contrast, the *high-grade* tumours had a higher mitotic rate, absence of TILs and more *NRAS* mutations (Table 3.2).

The TCGA 3-classes were also significantly associated with patient's sex, age at diagnosis, Breslow thickness, AJCC stage, ulceration status, mitotic rate and TILs (Table 3.3). The *immune* and *keratin* types of tumours were thin in comparison to the *MITF low* type. The *keratin* tumours were more frequently diagnosed at AJCC stage I. *MITF low* had a higher mitotic rate in comparison to other types (Table 3.3). The *immune* class had higher incidence of brisk TILs, and *MITF low* class had the lowest.

The TCGA classes showed no association with the *BRAF* and *NRAS* mutation status (Table 3.3).

Table 3.2 Lund signatures association with clinico-pathological characteristics of LMC primary tumours

The associations were tested using Pearson's chi-squared test for categorical variables and the Mann-Whitney Kruskal-Wallis test for continuous variables. The symbol *n* is the number of samples, *m* is the median and *r* is the range. Unclassified tumours were excluded from the analysis.

Clinico-pathological variables	Lund 4-classes				P	Lund 2-grades		P
	HI n(%)	NL n(%)	Pigm. n(%)	Prolif. n(%)		High-grade n(%)	Low-grade n(%)	
Sex: male	85(49)	72(36)	107(48)	41(49)	0.04	141(48)	118(44)	0.5
Tumour site: limbs	85(49)	98(50)	81(37)	27(32)	0.004	112(38)	126(47)	0.03
Age at diagnosis, (years) m(r)	59(21, 76)	57(20, 75)	59(28, 82)	58(19, 82)	0.001	60(20, 81)	58(19,76)	0.04
Breslow thickness(mm), m(r)	1.9(0.7, 18)	1.8(0.3, 10)	3(0.8, 20)	3(1.1, 15)	8×10^{-24}	3(0.5, 20)	1.8(0.7, 18)	5×10^{-19}
AJCC stage(%): I	70(40)	103(52)	34(16)	18(23)	8×10^{-14}	56(19)	129(49)	8×10^{-13}
II	86(49)	76(38)	137(63)	45(56)		172(60)	112(42)	
III	19(11)	19(10)	48(22)	17(21)		61(21)	25(9)	
Ulceration(yes)	57(33)	30(15)	104(47)	37(44)	3×10^{-11}	132(45)	56(21)	6×10^{-9}
Mitotic rate, =>1(%)	124(81)	131(78)	82(92)	67(94)	5×10^{-5}	240(93)	175(76)	2×10^{-7}
TILs(%): Absent	8(6)	22(16)	30(19)	11(21)	0.01	40(19)	18(9)	0.005
Non-brisk	97(72)	93(77)	106(67)	37(71)		144(69)	140(72)	
Brisk	29(22)	24(17)	23(15)	4(7)		24(12)	37(19)	
BRAF mutant, yes(%)	74(50)	79(51)	86(45)	23(33)	0.07	99(40)	110(50)	0.04
NRAS mutant, yes(%)	23(16)	38(24)	50(27)	24(35)	0.01	75(32)	35(16)	2×10^{-4}

Table 3.3 TCGA signatures association with clinico-pathological characteristics of LMC primary tumours.

The associations were tested using Pearson's chi-squared test for categorical variables and the Kruskal-Wallis test for continuous variables. The symbol *n* is the number of samples, *m* is the median and *r* is the range. Unclassified tumours were excluded from the analysis.

Clinico-pathological variables	TCGA 3-classes			P
	<i>Immune</i> n(%)	<i>Keratin</i> n(%)	<i>MITF low</i> n(%)	
Sex: male	102(53)	98(40)	67(45)	0.02
Tumour site: limbs	79(41)	121(49)	57(38)	0.07
Age at diagnosis, (years) m(r)	61(21, 77)	56(18, 75)	60(20, 81)	1×10^{-4}
Breslow thickness(mm), m(r)	2.5(0.7, 20)	2.0(0.3, 10)	3.2(0.8, 18)	7×10^{-11}
AJCC stage(%): I	52(27)	107(44)	32(22)	3×10^{-5}
II	114(59)	105(43)	84(58)	
III	26(14)	33(13)	29(20)	
Ulceration(yes)	71(37)	64(26)	65(43)	0.001
Mitotic rate, =>1(%)	147(86)	173(82)	119(95)	0.003
TILs(%): Absent	9(7)	33(18)	23(22)	7×10^{-4}
Non-brisk	97(71)	113(63)	73(70)	
Brisk	30(22)	33(18)	8(8)	
<i>BRAF</i> mutant, yes(%)	73(44)	94(49)	63(49)	0.6
<i>NRAS</i> mutant, yes(%)	38(24)	53(28)	31(24)	0.7

3.3.3 Lund and TCGA signatures association with survival in the LMC primary tumours

In keeping with previous studies [81, 105, 106, 114], the association of Lund and TCGA signatures was tested with melanoma-specific survival (MSS) in this larger dataset. The association with MSS was assessed using CPH and survival differences were visualised using Kaplan-Meier plots.

3.3.3.1 Lund 4-classes

Previously, *high-immune* and *normal-like* tumours were shown to predict better survival than *proliferative* and *pigmentation* tumours. In this analysis, 18% of *high-immune* class and 19% of *normal-like* class patients died from melanoma, and the corresponding rates were at least twice as high in *pigmentation* (40%) and *proliferative* (38%) classes (hazard ratio 2.5 in each case), (Table 3.4, Figure 3.3). These results replicate earlier reports from smaller datasets.

Table 3.4 Unadjusted Cox proportional hazard analysis of the Lund 4-classes

The *pigmentation* and *proliferative* class had increased hazard of melanoma-specific deaths in comparison to the baseline *high-immune* class. All deaths were caused by melanoma. CI is confidence interval. HR is the hazard ratio.

Lund 4-classes	HR	95% CI	P
<i>High-immune</i> (n=175, deaths= 32)	1.0	-	-
<i>Normal-like</i> (n=198, deaths= 38)	1.0	0.6-1.6	0.96
<i>Pigmentation</i> (n=222, deaths= 88)	2.5	1.7-3.8	8×10^{-6}
<i>Proliferative</i> (n=84, deaths=32)	2.5	1.5-4.0	2.9×10^{-4}

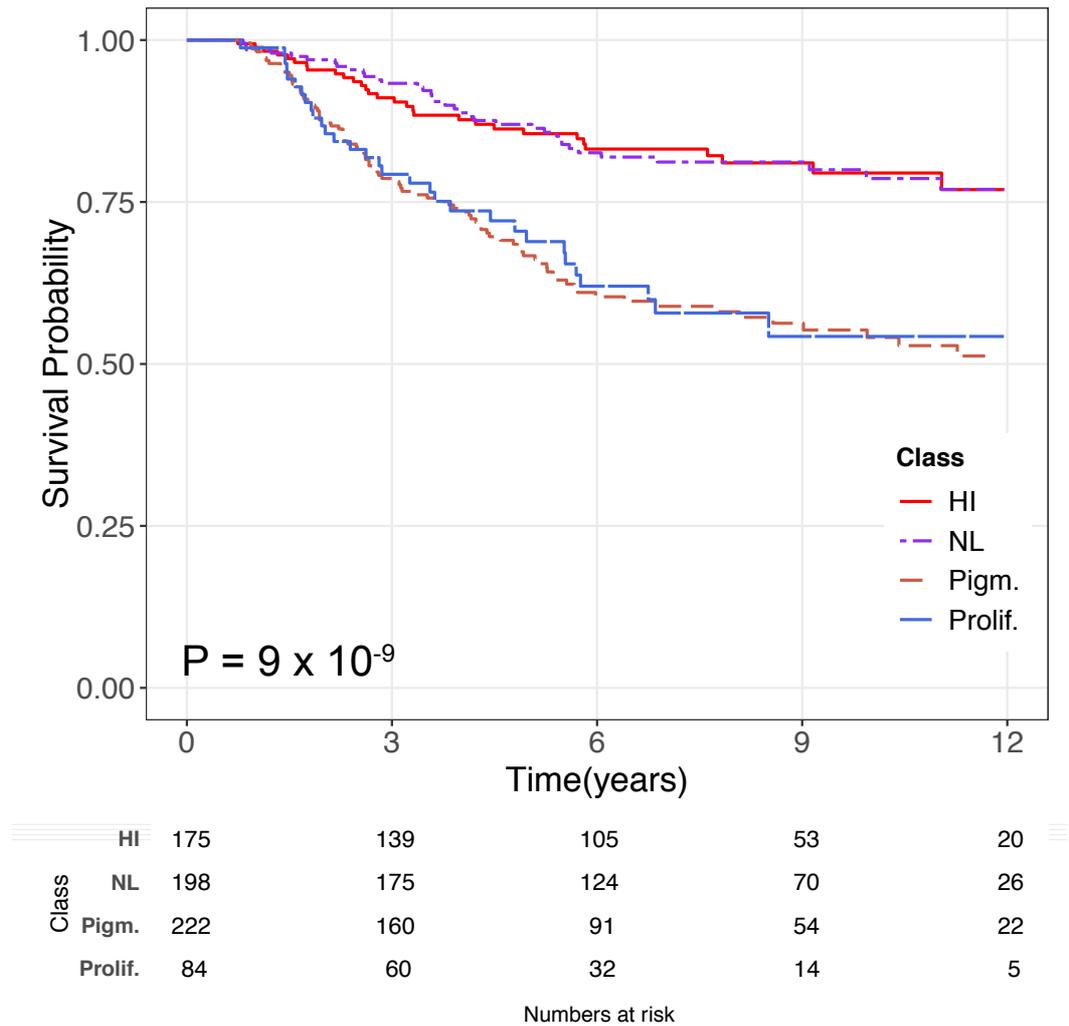


Figure 3.3 Melanoma-specific survival for the Lund 4-classes

(HI- *high-immune*, NL- *normal-like*, Pigm- *pigmentation*, Prolif- *proliferative*). The *high-immune* class and *normal-like* class predicted a better survival than the *pigmentation* and *proliferative* classes (P value from log-rank test). The risk table below the plot shows the number of patients at risk at a given time.

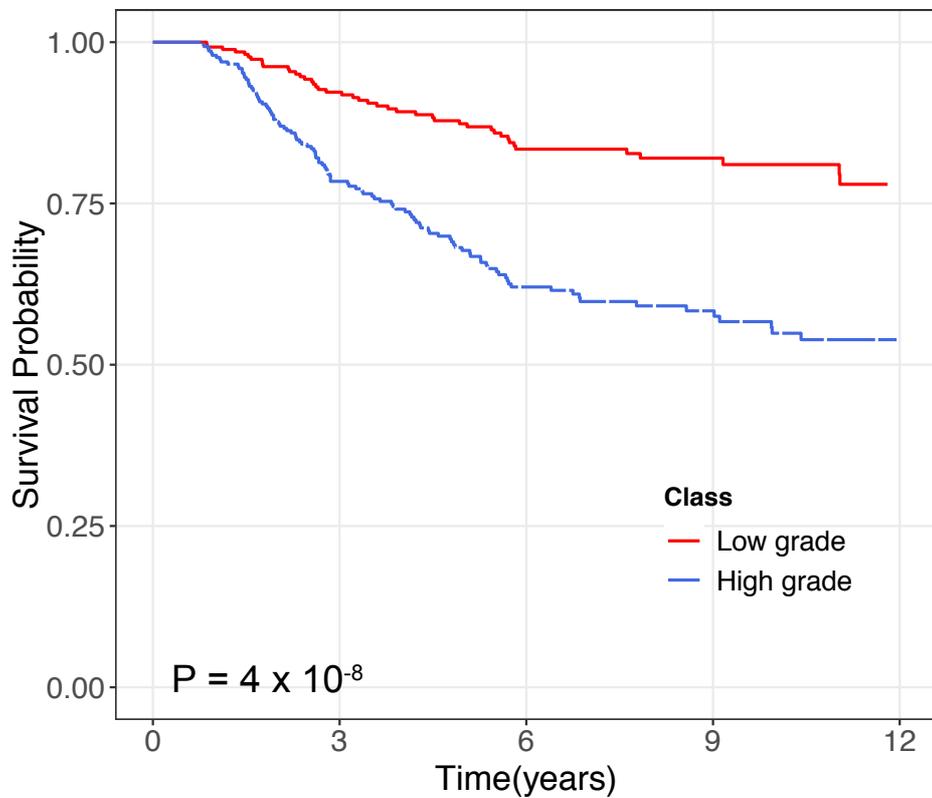
3.3.3.2 Lund 2-grades

Previously *high-grade* of the Lund 2-grades has been shown to be associated with poor survival in melanoma compared to *low-grade* [105, 114]. These observations were recapitulated in this analysis: 38% of patients in *high-grade* and 18% of patients in *low-grade* have died from melanoma, with a hazard risk ratio of 2.5 in unadjusted CPH model (Table 3.5, Figure 3.4).

Table 3.5 Unadjusted Cox proportional hazard analysis of the Lund 2-grades

Low-grade was chosen as baseline, all deaths were caused by melanoma. CI is confidence interval. HR is the hazard ratio.

Lund 2-grades	HR	95% CI	P
Low-grade (n=266, deaths= 48)	1.0	-	-
High-grade (n=296, deaths= 113)	2.5	1.8-3.5	1×10^{-7}



Low grade	266	220	163	86	33
High grade	296	211	124	70	25

Numbers at risk

Figure 3.4 Melanoma-specific survival for the Lund 2-grades

Low-grade was associated with better survival than the *high-grade* (P value from log-rank test). The risk table below the plot shows the number of patients at risk at a given time.

3.3.3.3 TCGA 3-classes

The TCGA 3-classes had been shown to predict survival in metastatic melanomas [81, 116]. In our dataset, 23% of patients in the *immune* class, 25% in *keratin* class, and 42% of patients in *MITF low* class have died from melanoma. *MITF low* class predicted therefore worse survival (HR=2.0) while the *immune* and *keratin* class predicted good survival (Figure 3.5). Unlike in metastatic tumours where the *keratin* class tumours had a survival close to *MITF low* [81], the *keratin* class had similar survival profile to the *immune* class in this primary melanoma dataset (Table 3.6 and Figure 3.5).

Table 3.6 Unadjusted Cox proportional hazard analysis of the TCGA 3-classes

Immune class was chosen as baseline, all deaths were caused by melanoma. CI is confidence interval. HR is the hazard ratio.

TCGA 3-classes	HR	95% CI	P
<i>Immune</i> (n=192, deaths= 44)	1.0	-	-
<i>Keratin</i> (n=247, deaths= 61)	1.0	0.7-1.5	0.8
<i>MITF low</i> (n=150, deaths=63)	2.0	1.4-3.0	4 x 10 ⁻⁴

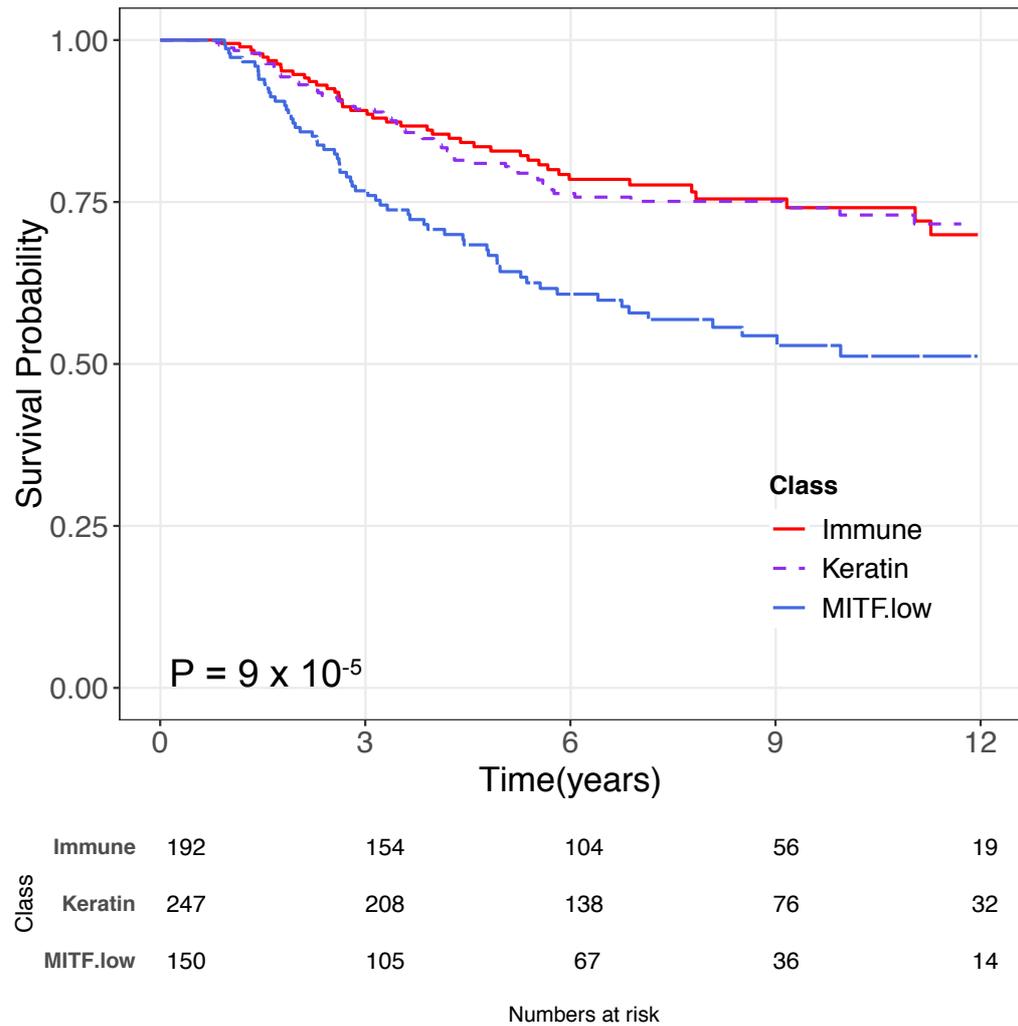


Figure 3.5 Melanoma-specific survival for the TCGA 3-classes

The *MITF low* class was associated with poor survival compared to the *immune* and *keratin* class (P value from log-rank test). The risk table below the plot shows the number of patients at risk at a given time.

3.3.4 Signature's prognostic value when stratified on AJCC stage

The LMC dataset contains 34% stage I tumours, 51% stage II tumours and 15% stage III tumours. The patients were stratified into two groups: AJCC stage I group and AJCC stage II & III pooled group. The prognostic values of the Lund and TCGA signatures was assessed across these two groups. In stage I, 14% of patients died from melanoma, while in stages II & III 35% of patients died. The Lund and TCGA signatures showed distinct prognostic properties across these two groups.

3.3.4.1 Prognostic value of the signatures in AJCC stage I

In the AJCC stage I group, none of the 3 signatures (Lund-4 class, Lund 2-grade and TCGA-3class) showed a significant association with MSS (Table 3.7 and Figure 3.6). In Lund 4-classes, 11% of patients in *high-immune* class diagnosed at stage I died from melanoma, the *normal-like* class had 14% melanoma-specific deaths, *pigmentation* class had 24% and *proliferative* class had 5%. As shown on the Kaplan-Meier curves, the *proliferative* class appeared to have the best survival profile, comparable to *high immune* (Figure 3.6A), which is the opposite of the result from whole data analysis (Figure 3.3). However this is only based on 18 cases in the *proliferative* class.

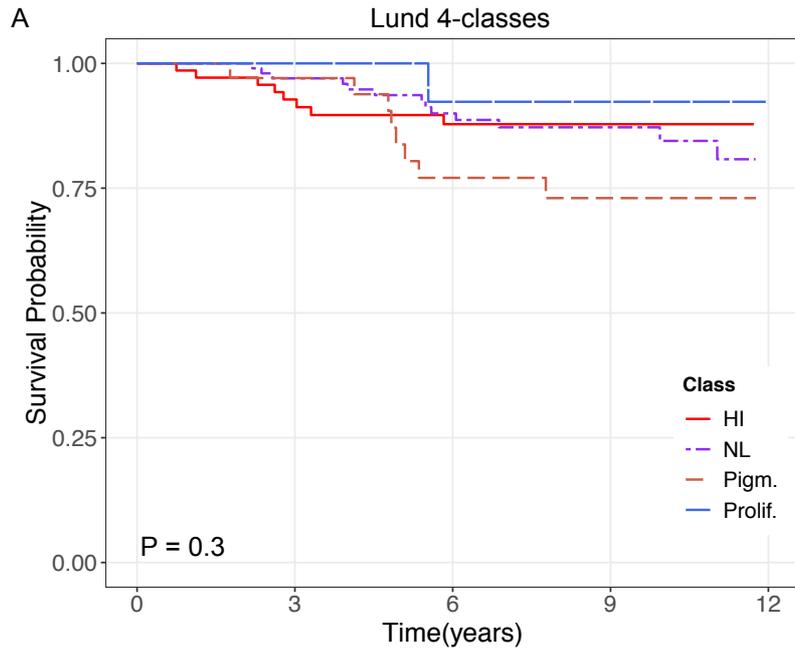
In Lund 2-grades, 13% of AJCC stage I diagnosed patients classified in *low-grade* and 12% of patients in *high-grade* died from melanoma. These numbers are similar, and the Lund 2-grades did not show any association with MSS in the stage I group (Table 3.7 and Figure 3.6B).

Similarly, the number of melanoma-caused deaths in patients diagnosed at stage I in TCGA 3-classes were: 15% in *immune* class, 12% in *keratin* class, and 21% in *MITF low* class, and there was no significant association with MSS in CPH model and the log-rank test (Table 3.7 and Figure 3.6C).

Table 3.7 Unadjusted Cox proportional hazard models for the Lund 4-class, Lund 2-grades and TCGA 3-classes in the AJCC stage I group

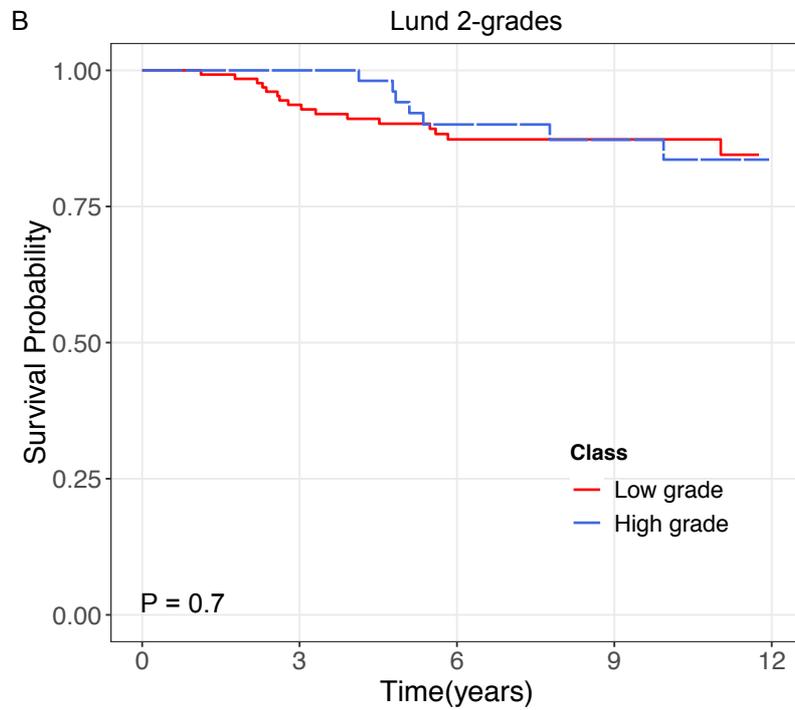
High-immune, *low-grade*, and *immune* classes were chosen as baseline, all deaths were caused by melanoma. CI is confidence interval. HR is the hazard ratio.

Signature	Class (n)	HR	95% CI	P
Lund 4 classes (n=225, deaths=31)	<i>High-immune</i> (70)	1.0	-	-
	<i>Normal-like</i> (103)	1.2	0.5-2.9	0.6
	<i>Pigmentation</i> (34)	2.1	0.8-5.5	0.1
	<i>Proliferative</i> (18)	0.5	0.1-3.8	0.5
Lund 2-grades (n= 185, deaths=24)	Low grade (129)	1.0	-	-
	High grade (56)	0.9	0.4-2.1	0.7
TCGA 3-classes (n=191, death=28)	<i>Immune</i> (52)	1.0	-	-
	<i>Keratin</i> (107)	0.8	0.3-1.9	0.6
	<i>MITF low</i> (32)	1.3	0.5-3.7	0.6



Class	HI	70	60	48	28	10
	NL	103	94	69	37	10
	Pigm.	34	33	23	13	6
	Prolif.	18	17	10	8	2

Numbers at risk



Class	Low grade	129	111	87	49	17
	High grade	56	55	41	26	10

Numbers at risk

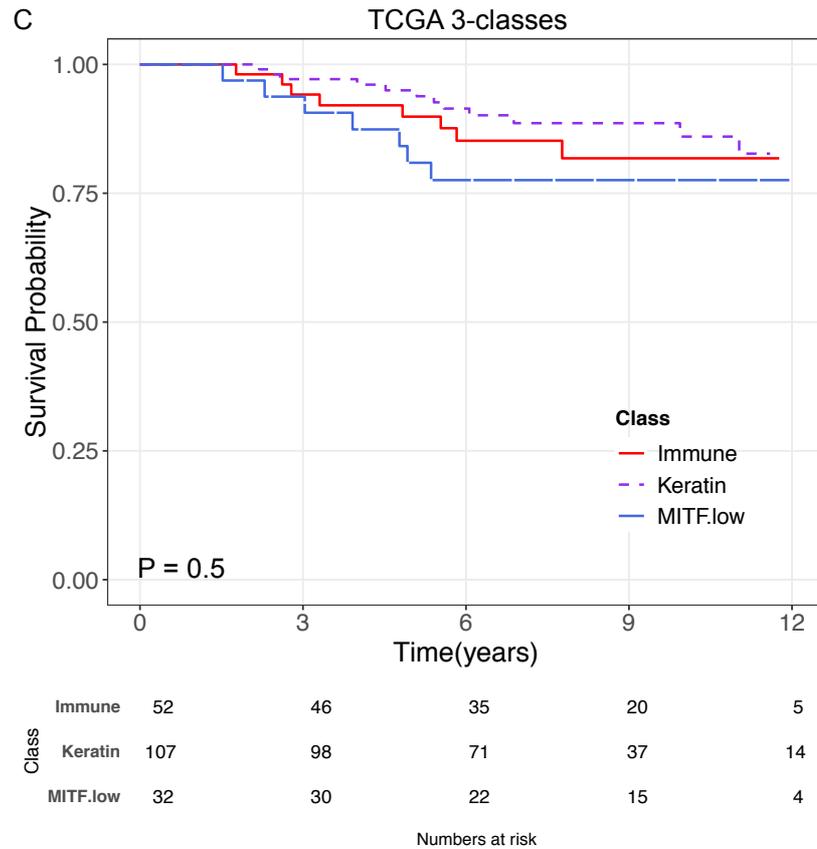


Figure 3.6 Melanoma-specific survival for the Lund 4-classes, Lund 2-grades and TCGA 3-classes in the AJCC stage I group.

P values were calculated from the log-rank test. The risk table below the plot shows the number of patients at risk at a given time.

3.3.4.2 Prognostic value of the signatures in AJCC stage II & III group

Similar to the AJCC stage I group, the Lund and TCGA signatures were tested for association with MSS in the pooled AJCC stage II & III group. In Lund 4-classes, 23% of patients in *high-immune*, 25% of *normal-like* class died from melanoma. In the *pigmentation* (43%) and *proliferative* (47%) class, a higher proportion of people died from melanoma. Therefore, in advanced stage tumours, the Lund 4-classes significantly predicted MSS (Figure 3.7A) with the *pigmentation* and *proliferative* class having the worse survival (HR=2.1 and HR=2.4 respectively, Table 3.8).

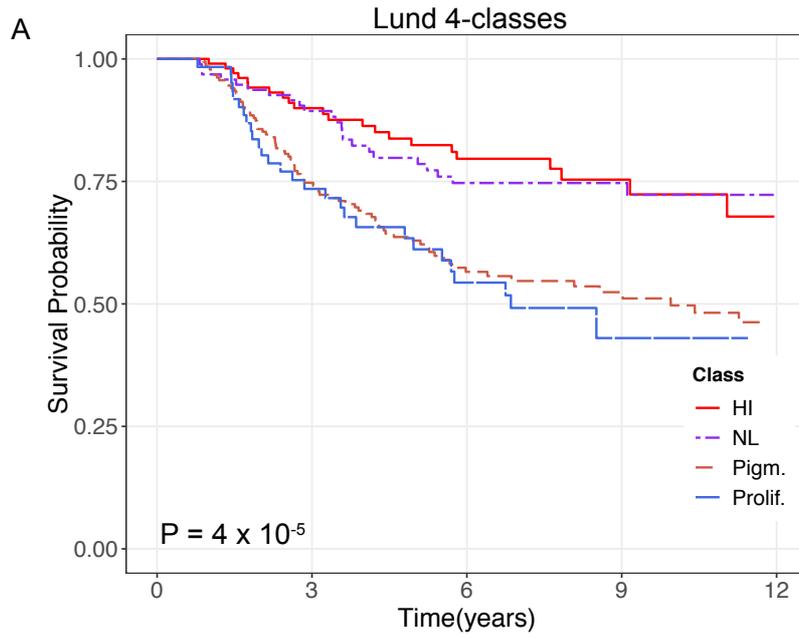
In Lund 2-grades, 23% of patients in *low-grade* and 45% of patients in *high-grade* died from melanoma; the Lund 2-grades were significantly associated with MSS (Figure 3.7B). The CPH model suggested that *high-grade* (HR=2.4) had increased hazard of death from melanoma in comparison to the baseline *low-grade* (Table 3.8).

Similar to Lund 4-classes and 2-grades, the TCGA 3-classes were also significantly associated with MSS in the stage II & III group (Figure 3.7C). The Cox model indicated that *MITF low* class (HR=2.1) had increased hazard of death from melanoma in comparison to the baseline *immune* class (Table 3.8). The *keratin* class (HR=1.3) had a hazard ratio comparable to that of the *immune* class.

Table 3.8 Unadjusted Cox proportional hazard models for the Lund 4-class, Lund 2-grades and TCGA 3-classes in the AJCC stages II & III

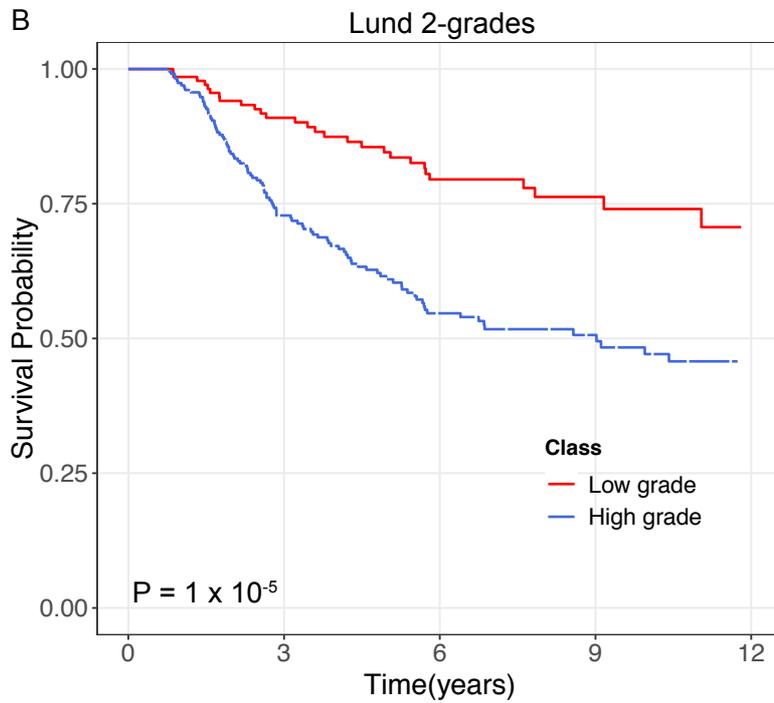
CI is confidence interval. HR is the hazard ratio.

Signature	Class (n)	HR	95% CI	P
Lund 4 classes (n=447, deaths=157)	<i>High-immune</i> (105)	1.0	-	-
	<i>Normal-like</i> (95)	1.0	0.6-1.7	0.9
	<i>Pigmentation</i> (185)	2.1	1.4-3.4	0.001
	<i>Proliferative</i> (63)	2.4	1.4-4.2	0.001
Lund 2-grades (n= 370, deaths=135)	<i>Low-grade</i> (137)	1.0	-	-
	<i>High-grade</i> (233)	2.4	1.6-3.6	1.9 x 10 ⁻⁵
TCGA 3-classes (n=391, death=138)	<i>Immune</i> (140)	1.0	-	-
	<i>Keratin</i> (138)	1.3	0.8-2.0	0.2
	<i>MITF low</i> (113)	2.1	1.4-3.2	4 x 10 ⁻⁴



	HI	105	79	57	25	10
Class	NL	95	81	55	33	16
	Pigm.	185	124	67	41	16
	Prolif.	62	40	22	6	3

Numbers at risk



Class	Low grade	137	109	76	37	16
	High grade	233	150	82	44	15

Numbers at risk

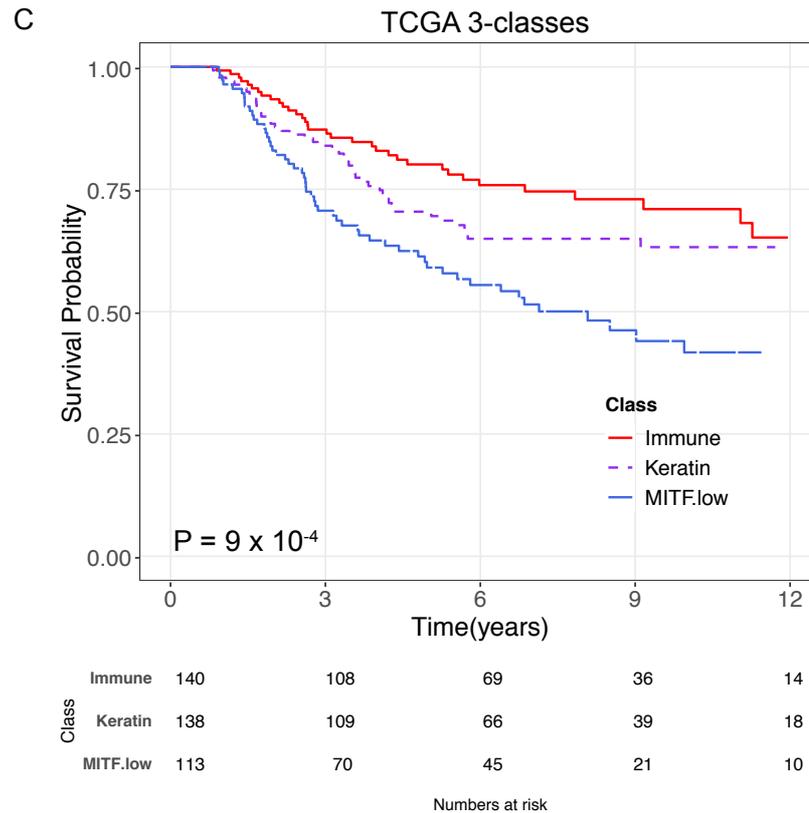


Figure 3.7 Melanoma-specific survival for the Lund 4-classes, Lund 2-grades and TCGA 3-classes in the AJCC stages II&III

P values were calculated from the log-rank test. The risk table below the plot shows the number of patients at risk at a given time.

3.3.5 Statistical Interaction between the Lund signature and AJCC stage

As shown above, the Lund and TCGA signatures strongly predicted MSS in stage II & III tumours but not in the stage I group. The AJCC stage I group tumours are thin, have lower mitotic rate and are mainly non-ulcerated [44], while AJCC stage II & III tumours are thicker, have higher mitotic rate and are mainly ulcerated.

To rule out the possibility that the difference in prognostic value of the signatures might be solely due to limited power in the stage I group, a formal test of statistical interaction was conducted between the AJCC stage and the Lund 2-grades. This analysis (Table 3.9) confirmed AJCC stage as an independent prognostic factor (HR=1.9, P=0.03), but also showed an evidence of interaction with the Lund 2-grades (P=0.04), suggesting a heterogeneity of effect of the Lund 2-grades on the AJCC stage.

Table 3.9 Test of interaction between AJCC stage and Lund 2-grade signature in Cox proportional hazards model of MSS

Low-grade and AJCC stage I was chosen as baseline, all deaths were caused from melanoma, HR is the hazard ratio.

Variable (n=555, deaths=159)		HR	95% CI	P
Lund 2-grades	<i>Low-grade</i>	1.0	-	-
	<i>High-grade</i>	0.9	0.4- 2.1	0.7
AJCC stage	I	1.0	-	-
	II&III	1.9	1.1-3.5	0.03
<i>High-grade: AJCC stage II&III</i>		2.8	1.1-7.3	0.04

3.3.5.1 Statistical power in AJCC stage I

To test if the effect of Lund 2-grades on hazard of death is not limited by the number of deaths in AJCC stage I, a statistical power calculation analysis was performed. In AJCC stage I group, 70% of patients were classified in *low-grade* and 30% were *high-grade*. The total number of deaths in AJCC stage I group was 33. Power and sample size calculations showed that the stage I group had 70% power to detect a hazard ratio of 2.6 at a significance level 0.05 (Figure 3.8).

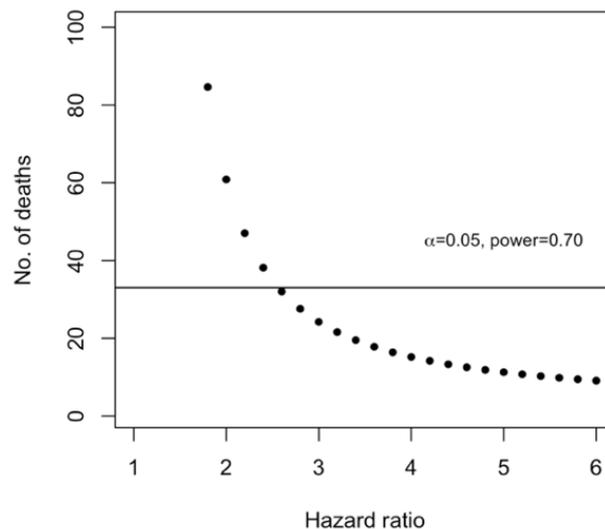


Figure 3.8 Sample size and power calculation

There were 33 deaths among patients diagnosed at AJCC stage I, providing a reasonable statistical power for $HR > 2.5$ ($\alpha=0.05$).

3.3.6 Devising a new signature by consensus-based clustering

The lack of prognostic effect in stage I tumours from the 3 published signatures was surprising since these signatures have been shown to predict prognosis in primary melanoma [105, 114]. Having shown this limitation, I sought to design a new signature based on a dataset with a good mixture of tumours at all stages. It was hypothesised that re-clustering the whole LMC dataset and exploring different clustering algorithms may reveal new melanoma molecular subtypes. Three different algorithms, HC, KM and PAM were applied in a consensus clustering approach.

The optimal k was selected based on examination of cumulative density function (CDF) graph, relative change in area under the CDF (delta) graph and heatmap of the consensus matrices.

Examining the HC CDF and delta graph suggested good k values at 4 and 5 (Figure 3.9A, B). However, the consensus matrix heatmap did not show a clear separation between the clusters at $k=4$ or 5 (Figure 3.10A, B). The KM clustering CDF and delta graphs suggested good k values at 5 and 6 (Figure 3.9C, D) and the corresponding consensus matrix heatmaps showed a neater cluster separation at $k=6$ (Figure 3.10C, D). On the other hand, PAM clustering suggested good k values at 6 and 7 from the CDF and delta CDF (Figure 3.9E, F) and from the consensus matrix heatmap (Figure 3.10E, F). For further comparisons of clustering algorithms, k was selected as 5 for HC, 6 for KM and 7 for PAM algorithms.

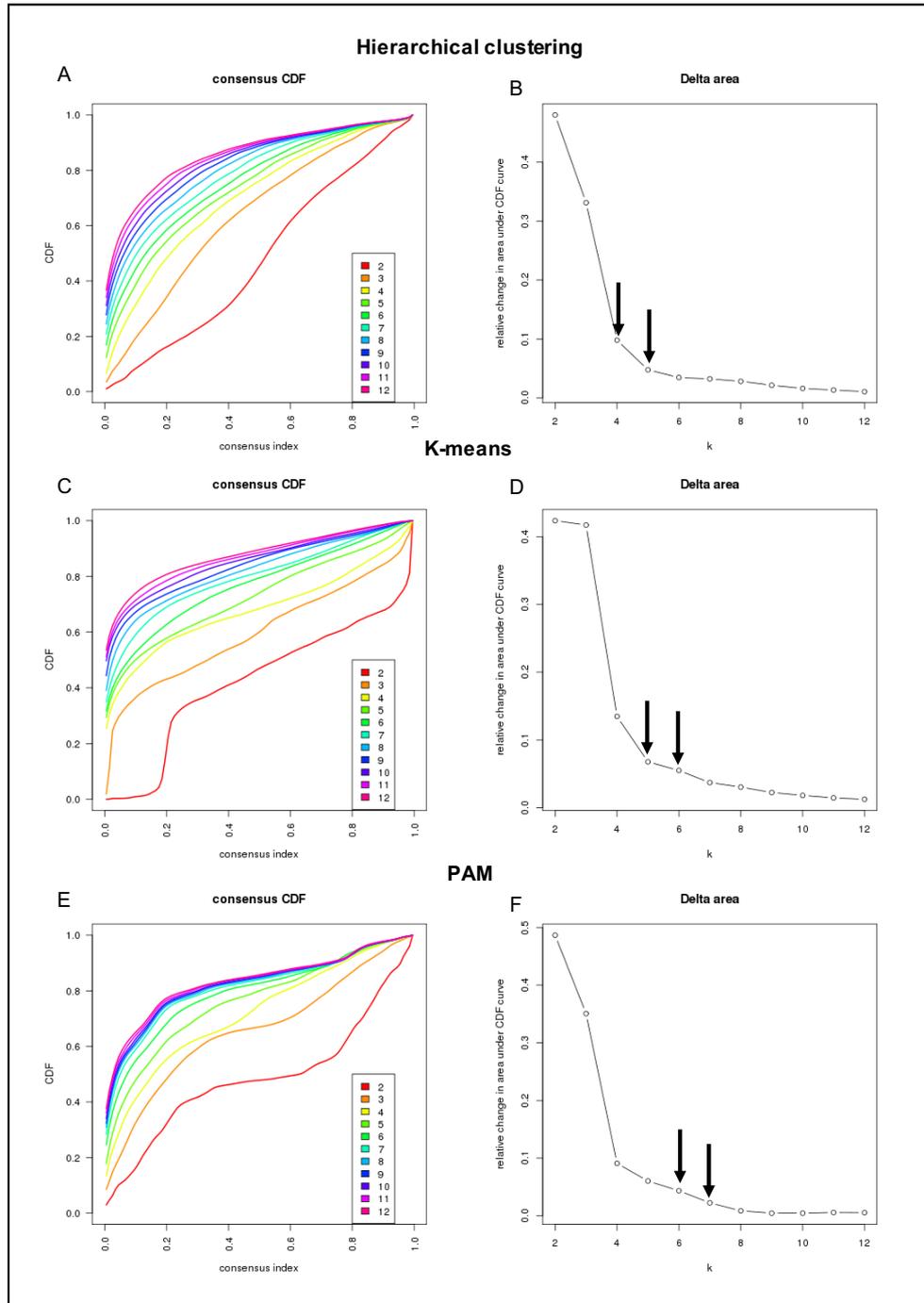


Figure 3.9 Selecting k by examining the relative change in area under the CDF curve

The LMC tumours were clustered using HC, KM and PAM based consensus clustering. The CDF plot shows the distribution of consensus indices for all possible pairings of tumours at each value of k . Typically, the majority of pairs have a low consensus index, reflecting their belonging to different clusters (panels A, C, E). The relative change in area under the CDF curves (delta graph) shows increments in the area under the CDF with increasing k (panels B, D, F). Arrows show suggested k value on the basis of flattening of the curves.

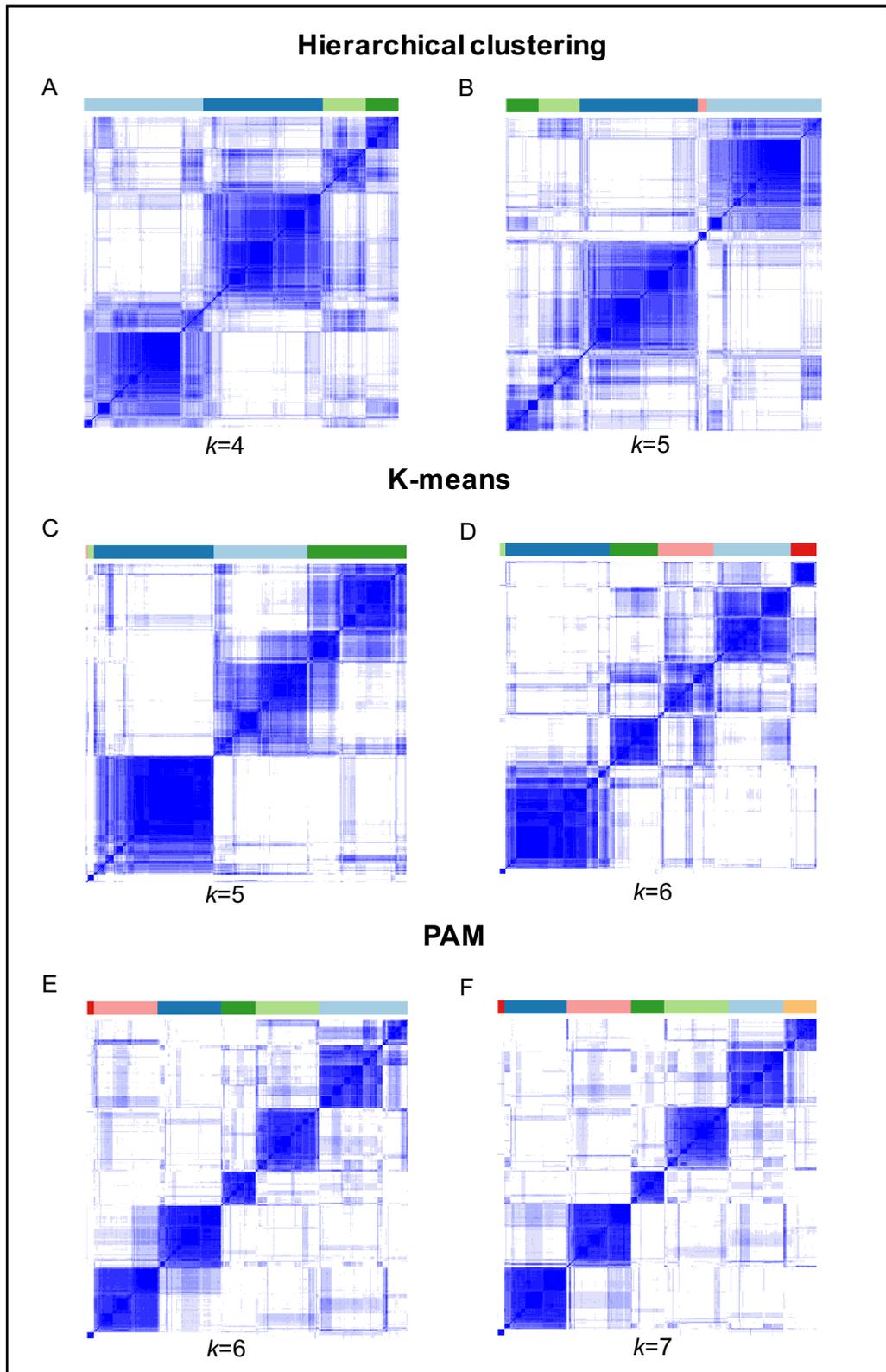


Figure 3.10 Consensus matrix heatmaps for HC, KM and PAM

Row and columns represent tumours in their respective inferred clusters; each heatmap is symmetrical. The blue represents consensus index=1 (i.e. samples frequently clustered together), white colour represents consensus index=0 (i.e. samples rarely clustered together).

3.3.7 Comparing HC, KM, and PAM based clusters

The clusters derived from the HC, KM, and PAM were compared using cluster stability measures, and their overlap with the Lund and TCGA signatures was evaluated.

3.3.7.1 Measuring cluster stability

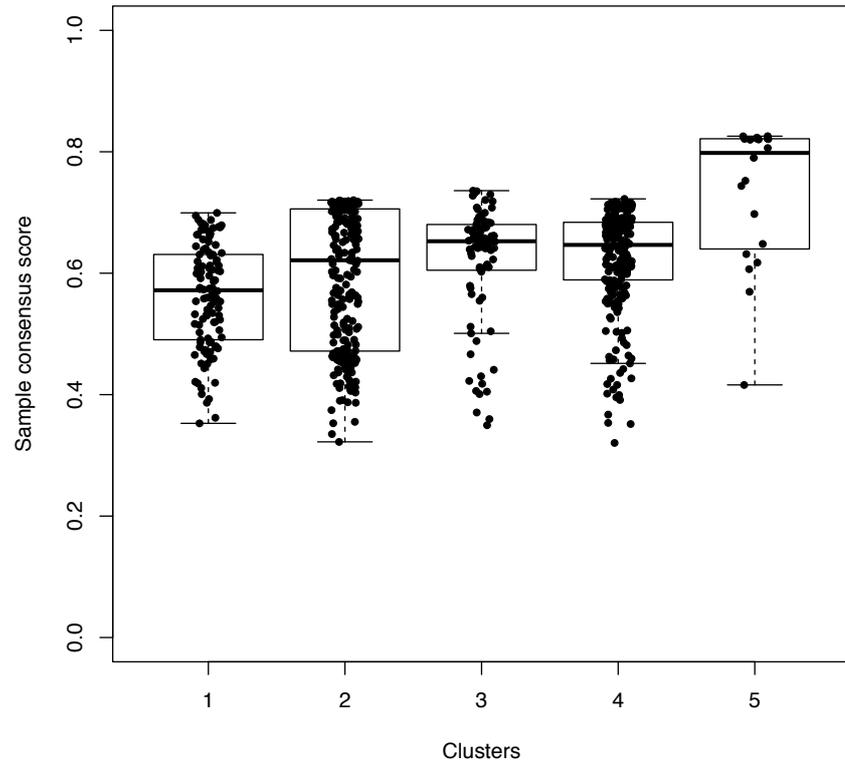
The cluster stability was calculated as the average consensus score for samples in a given cluster. In HC derived clusters, median consensus score was 0.57 for cluster 1, 0.62 for cluster 2, 0.65 for cluster 3, 0.65 for cluster 4, and 0.80 for cluster 5 (Figure 3.11A). The overall median consensus score across the HC clusters was 0.64.

Among KM derived clusters, median consensus score was 0.71 for cluster 1, 0.79 for cluster 2, 0.79 for cluster 3, 0.60 for cluster 4, 0.83 for cluster 5, and 0.95 for cluster 6 (Figure 3.11B). The overall median consensus score across the KM derived clusters was 0.73.

Among PAM derived clusters, median consensus score was 0.80 for cluster 1, 0.84 for cluster 2, 0.7 for cluster 3, 0.79 for cluster 4, 0.75 for cluster 5, 0.71 for cluster 6 and 0.94 for cluster 7 (Figure 3.11). The overall median consensus score was highest for the PAM derived clusters (0.76).

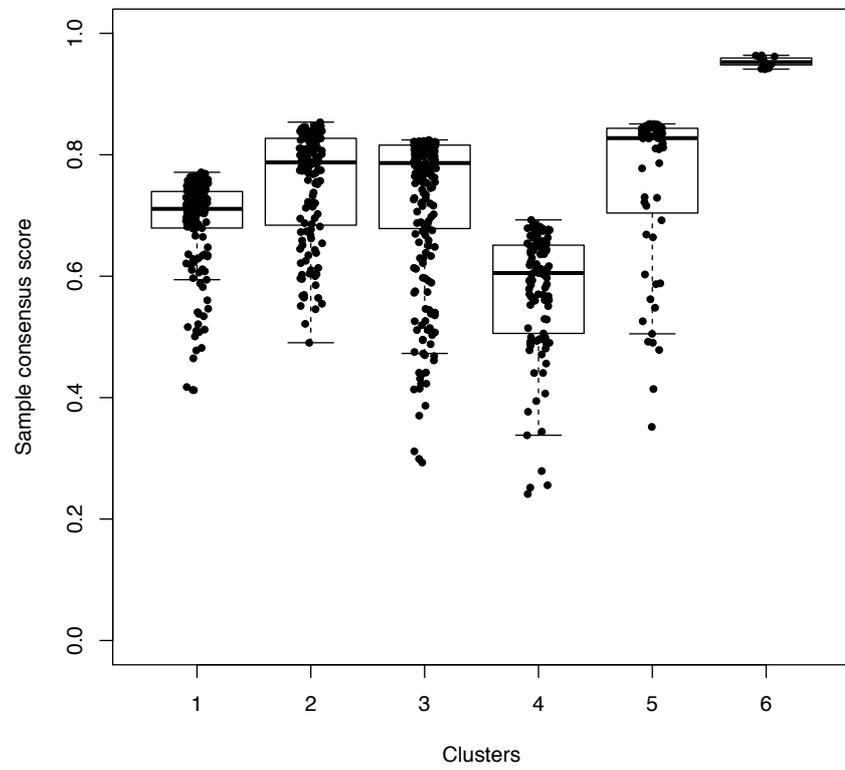
A

Hierarchical clustering



B

K-means



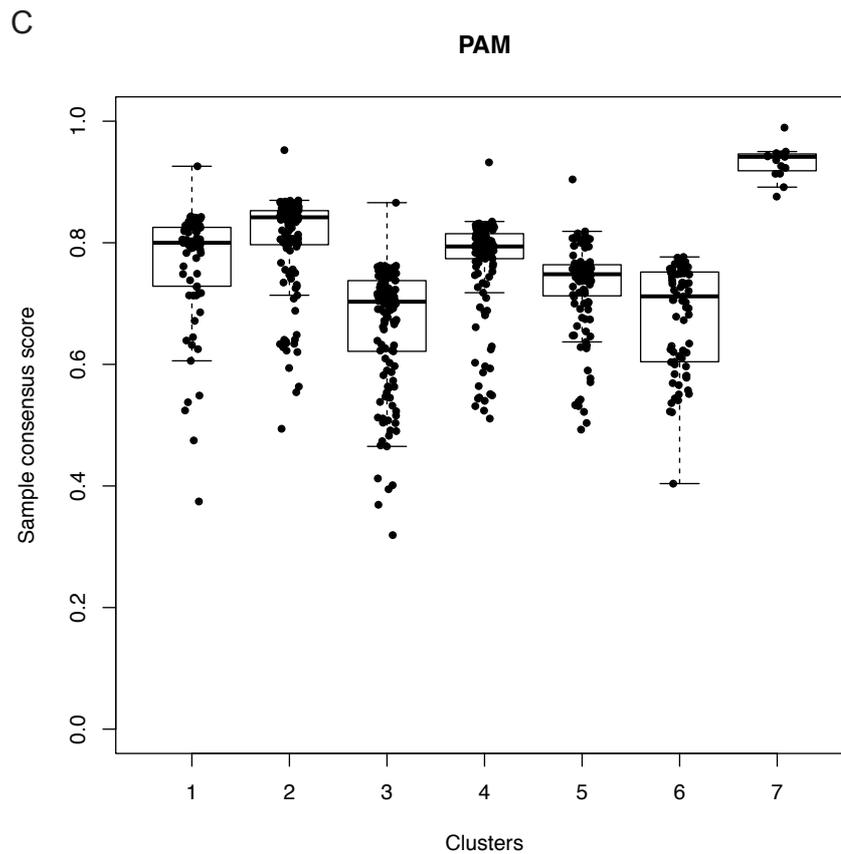


Figure 3.11 Comparing stability of HC, KM and PAM derived clusters

(A-C) The boxplot shows the consensus score for samples in each cluster.

3.3.8 Agreement with Lund and TCGA signatures

The HC, KM, and PAM derived clusters were assessed for agreement with the Lund and TCGA signatures in the LMC dataset (Table 3.10). Cramer's V statistic was used for measuring the agreement between the signatures. As described earlier, the Lund 2-grades largely overlap with Lund 4-classes [105]. We found that Cramer's V statistic between them was 0.90, the highest of any comparison made (Table 3.10). The TCGA signatures were developed from a different platform (RNA seq) but they were shown to overlap with the Lund 4-classes in terms of biological significance, which was also reflected in a higher (0.63) Cramer's V statistic (Table 3.10).

Among our three clustering methods, HC (0.76) and KM (0.78) derived clusters had similar agreement with the Lund 2-grades. However, both had a clearly lower overlap with Lund 4-classes (Cramer's $V = 0.46$ and 0.57 respectively) and with the TCGA 3-classes ($V = 0.37$ and 0.50 respectively). PAM derived clusters had weakest overlap with the Lund 2-grades ($V = 0.59$) but agreement with TCGA 3-classes ($V = 0.54$) was

slightly stronger. Among the three clustering methods, HC derived clusters had stronger agreement ($V=0.67$) with KM derived clusters in comparison to PAM ($V=0.38$) derived clusters (Table 3.10).

Table 3.10 Cramer's V statistic comparing agreement between HC, KM, and PAM derived clusters and the Lund and TCGA signatures

	Lund 2-grades	Lund 4-classes	TCGA 3-classes	HC	KM	PAM
Lund 2-grades	1	0.90	0.39	0.76	0.78	0.59
Lund 4-classes		1	0.63	0.46	0.57	0.49
TCGA 3-classes			1	0.37	0.50	0.54
HC				1	0.67	0.38
KM					1	0.45
PAM						1

3.4 Discussion

Previous studies have demonstrated the prognostic value of the Lund and TCGA signatures in smaller primary melanoma datasets [105, 114]. The Lund 4-classes initially developed from metastatic tumours have been applied to the primary melanoma datasets from Lund and Leeds [105, 114]. The Lund 2-grade is a derivative signature of Lund 4-classes which was applied on the primary cohort from Leeds and Lund [105, 114]. The TCGA signature had been developed from a mixture of primary and metastatic melanoma tumours [81].

Previously, Dr Jeremie Nsengimana (senior statistician in the group) applied the Lund signatures to the subset of LMC tumours which were profiled at the time [114]. Now with a bigger LMC dataset, the first objective was to test the prognostic value of Lund and TCGA signatures in the whole LMC dataset. Secondly it was hypothesized that re-clustering the primary tumour transcriptomes of LMC may improve the existing signatures and could lead to the discovery of new signatures.

3.4.1 The Lund and TCGA signatures were reproducible in the Leeds Melanoma cohort

The Lund 4-class and Lund 2-grade signatures were applied to the whole LMC dataset using the supervised NCC approach. Dr. Jeremie Nsengimana applied the TCGA signatures to the classify the LMC tumours using the same approach. Twenty and sixteen percent of LMC tumours could not be classified into the Lund 2-grades and TCGA 3-classes respectively, whereas only 3% of tumours could not be classified into the Lund 4-classes. The higher proportion of unclassified tumours in the Lund 2-grades and TCGA 3-classes may reflect the fact that these signatures were developed from more advanced stage tumours and have not captured all tumour subtypes in primary melanoma.

In this larger and independent dataset, the Lund signatures showed similar associations with clinico-histopathological variables and MSS as previously reported [114]: the *high-immune* and *normal-like* class of Lund 4-classes predicted good prognosis in LMC, while the *proliferative* and *pigmentation* class predicted poor prognosis. The *high-immune* and *normal-like* classes contained thin, non-ulcerated early stage melanomas, whereas *proliferative* and *pigmentation* classes contained thicker, mainly ulcerated advanced stage melanomas. The pathological reports further indicated that good prognosis classes had a higher incidence of brisk TILs, and poor prognosis groups had fewer TILs. Also, consistent with previous reports, another pathological indicator, number of mitoses, was lower in the good prognosis groups compared to poor prognosis groups. In keeping with previous reports, *low-grade* of Lund 2-grades predicted good prognosis and *high-grade* predicted poor prognosis. The *low-grade* tumours were thin, non-ulcerated early stage melanomas and *high-grade* tumours were thick, mainly ulcerated advanced stage melanomas. In TCGA 3-classes, the *MITF low* class predicted poor prognosis and *keratin* and *immune* class predicted good prognosis. This is consistent with the previous replication report of the TCGA signature [116]. However, in the TCGA study, *keratin* class was associated with poor prognosis. The differences in survival of *keratin* class may be driven by the differences in the nature of samples used in these studies (mostly metastatic tumours in TCGA dataset and exclusively primary tumours in the LMC study). Overall, the prognostic value of the Lund and TCGA signatures was successfully replicated in LMC dataset.

3.4.2 The Lund and TCGA signatures' lack of association with AJCC stage

Despite an apparent reproducibility when the LMC is analysed as one large dataset, the Lund and TCGA signatures showed different prognostic properties when the dataset was stratified on the basis of AJCC stage. None of the signatures predicted prognosis in AJCC stage I, but they all showed strong prognostic value in AJCC stages II & III. This was surprising because the Lund 4-class and Lund 2-grade signatures have been shown to predict prognosis in primary melanoma but these studies based on relatively smaller cohorts with only few stage I cases [105, 114]. The significant interaction between the Lund 2-grades and MSS provided further evidence that effect of Lund 2-grade signature is effectively different across the AJCC stages. The power calculation suggested that, although not very large, the data size for AJCC stage I patients is sufficiently powered (power=0.70) for detecting a hazard ratio greater than 2.5 ($\alpha=0.05$). This magnitude of effect is reasonable since Lund signatures have a comparable effect in the full data. Hence, it was hypothesised that re-clustering the LMC dataset may identify novel signatures with prognostic value across all AJCC stages, including stage I. Identifying a prognostic signature in stage I melanoma would be of particular interest because there is no other biomarker known so far to be prognostic within these early-stage melanomas. Since most of melanoma research is conducted on advanced stage tumours, it was challenging to find a validation dataset that comprises stage I melanomas.

3.4.3 Unsupervised clustering of the LMC dataset

The Lund 4-class and TCGA signatures were generated using similar a clustering algorithm, hierarchical clustering [81, 106]. HC is a popular method for identifying disease subtypes in clinical research. The significant increase in the dataset size (more than 3-fold increase) compared to previous studies, provided an opportunity to derive a new signature by exploring several clustering algorithms. The integrated framework of clustering algorithms with a consensus based resampling approach was applied to the LMC dataset [137]. Multiple resamples and unsupervised clustering using HC, KM and PAM generated a consensus matrix which was further analysed and examined to identify the k value. Previous comparisons of clustering algorithms showed that none of the algorithms consistently outperformed the other algorithms in identifying known subgroups from the analysed microarray datasets [142, 143]. In this study, the three applied algorithms identified different number of clusters and the cluster stability examination allowed selection of 7 clusters identified

by the PAM algorithm. These classes (clusters) showed a reasonable agreement with Lund and TCGA signatures (Table 1.19) although they did not entirely overlap, which is an indication that the published signatures did not explain all the variability in our dataset.

In conclusion, previous melanoma transcriptomic signatures predicted prognosis in the LMC dataset as a whole. However, when patients were stratified on the basis of AJCC stage, these signatures showed distinct prognostic properties and did not predict prognosis in patients diagnosed at AJCC stage I. Re-clustering the LMC dataset using three different clustering algorithms (HC, KM and PAM) identified new classes. The PAM classes had higher stability in comparison to the HC and KM classes and will be explored further in the next chapter.

Chapter 4

Properties of the new tumour classes in primary melanoma: prognostic significance and biological characterisation

The objectives of this chapter are:

Objective 4: To use an independent measure of cluster separation to assess stability of the newly identified tumour classes referred to as the LMC classes (Leeds Melanoma Cohort classes)

Objective 5: To test association of the LMC classes with clinico-histopathological features of melanoma

Objective 6: To test prognostic significance of the LMC classes in the whole LMC dataset and in the stage I subset

Objective 7: To generate a reduced signature of the LMC classes and replicate prognostic significance of the LMC classes in an independent dataset from Lund, Sweden

Objective 8: To explore biological differences between LMC classes using pathway enrichment analysis and using melanoma-specific biological modules

4.1 Introduction

Previously, it has been shown that the Lund signatures have independent prognostic value in comparison to the AJCC stage [114], but this analysis was based on a subset of 200 tumours from the LMC. The area under the Receiver Operating Characteristic (ROC) curve increased by 4% when Lund signatures were added to the AJCC staging system [114]. It was shown in the previous chapter that, when applied to the whole LMC dataset, the Lund signatures were only prognostic at advanced stages of primary melanoma (refer to 3.3.4). Therefore, it was hypothesized that re-clustering LMC tumours may identify novel classes which are prognostic at all stages of melanoma. As described in Chapter 3, LMC tumours were clustered using a consensus clustering approach using three different clustering algorithms (HC, KM and PAM). The PAM clustering algorithm identified seven LMC classes which had

higher stability in comparison to HC and KM based classes, using a stability metric from the consensus clustering algorithm (refer to 3.3.7).

In this chapter the PAM-derived classes (from now on referred to as the LMC classes) were further explored. An objective measure of cluster separation was used to select the number of classes. The identified LMC classes were tested for association with clinico-histopathological features of melanoma patients. The prognostic value of the LMC classes was assessed in the whole LMC dataset and in an independent dataset of primary melanoma from Lund, Sweden [105]. The tumour classes were also biologically characterised using pathway enrichment analysis and using melanoma-specific biological modules [144].

4.2 Methods

4.2.1 Identifying LMC classes using a cluster separation measure

The approach to identifying the optimal number of classes, k , was further explored using an objective criterion, cluster separation index. The cluster separation index was calculated as the ratio of intra- to inter-cluster similarity at each k value (in this case from $k = 2$ to $k = 12$). The intra-cluster similarity is the average consensus score of all pairs of sample observations in the respective clusters. The intra-cluster similarity has been previously described as the cluster stability in Chapter 3 (refer to 3.2.6). The inter-cluster similarity is calculated as the consensus score of all pairs of sample observations in different clusters. The ratio of intra-cluster to inter-cluster similarity was estimated for various k values. The intra-cluster and inter-cluster similarity calculations were done using R-package *FPC* [145]. The cluster separation index generally increases with k , and the k value where the cluster separation index had minimal increase at $k+1$ and thereafter was selected as the number of clusters in the LMC dataset.

4.2.2 Association of the LMC classes with clinico-histopathological characteristics

The LMC classes were tested for association with clinico-histopathological characteristics. The clinico-histopathological variables used in the analysis were age at diagnosis, patient's sex, AJCC stage, Breslow thickness, site of tumour, ulceration status of tumour, number of mitoses, TILs, *BRAF*, and *NRAS* mutation status. Pearson's chi-squared test (refer to 2.5.3) was used for testing categorical variables

and Mann-Whitney (refer to 2.5.1) or Kruskal-Wallis tests (refer to 2.5.2) were used for testing continuous variables.

4.2.3 Association of LMC classes with MSS

The LMC class association with melanoma-specific survival (MSS) was tested using CPH models (refer to 2.4.4). The survival differences between the classes were graphically visualised using Kaplan-Meier plots (refer to 2.4.2). The survival differences between the LMC classes were tested using the log-rank test (refer to 2.4.3). The analyses were done using R-packages *survival* [129], *ggplot2* [130] and *ggkm* [146]. In all the analyses, the LMC class 1 was chosen as the baseline. The independent prognostic value of LMC classes in the whole LMC data was tested by including age at diagnosis, sex, AJCC stage and number of mitoses in the multivariable CPH models. The TILs variable had 27% of missing values and was excluded from the survival analyses.

4.2.3.1 Association of LMC classes with MSS in stage I tumours

As described previously, mitotic rate, Breslow thickness and Ulceration status are a significant predictor of prognosis in the AJCC stage I melanoma (refer to 1.5.4). The LMC patients were stratified on the basis of AJCC stage. The association of LMC classes with MSS was tested in the AJCC stage I group. To test the independent prognostic value of LMC classes, mitotic rate, Breslow thickness, and ulceration status were included in the multivariable CPH models. The analyses were performed using R-package *survival* [129].

4.2.4 Refining the LMC class signature

The consensus-based clustering of LMC tumours was conducted using 13,688 genes which passed the previously described filters (refer to 2.2.5.3). However, such a large gene signature would not be economically/technically feasible in clinical practice. To mitigate this, a reduced gene signature for LMC classes was generated. The signature reduction was done such that the reduced signature reclassified LMC tumours in their respective classes with minimal loss in accuracy. The reduced signature was derived as follows:

- 1 The genes were tested for differential expression across the LMC classes using a Kruskal-Wallis test (refer to 2.5.2).

- 2 The P-values were adjusted for multiple-testing using a Bonferroni correction and genes with adjusted $P < 0.0001$ were selected.
- 3 The top $n=1, 5, 10, 25$ or 100 most significantly upregulated genes of each LMC class were selected to generate 6 reduced signatures.
- 4 The average gene expressions for each gene in each of the 6 reduced signatures were calculated for each LMC class to generate the class centroids.

The reduced signatures were applied to the LMC dataset to classify tumours into the LMC classes using the NCC approach (refer to 3.2.1). As a comparison baseline, the LMC tumours were also classified by including all genes in the signature. The classification using all genes in the signature may yield slightly different results in comparison to the initial consensus clustering results (refer to 3.2.5). This is because consensus clustering algorithm, when applied to the LMC dataset, identified the LMC classes by merging clustering results from the 1000 resampled subsets (by selecting only 80% of genes and samples in each subset) of the LMC dataset (refer to 3.2.5).

The classification accuracy for each reduced gene signature was calculated by comparing new classification labels with original classification labels in a contingency table. The proportion of diagonal values in the table represents classification accuracy (degree of overlap between signatures). The overlap was plotted to appraise the performance of each reduced signature, and the best performing among the 6 signatures was chosen.

4.2.5 Replicating the prognostic value of the LMC signature in the Lund cohort

4.2.5.1 *Lund primary melanoma cohort*

The cohort used for validation was generated from the melanoma study conducted in Lund University, Lund, Sweden [105]. The melanoma patients recruited into the study were diagnosed between 1995 and 2002 ($n=223$) [105]. The FFPE tumour blocks of these patients were retrieved, and up to 3 tumour core sections were generated for the extraction of mRNA. Expression profiling was done using the Illumina WG-DASL HT8 array [92], a similar array to the one used for the LMC study, but an earlier version. The generated gene expression data underwent background correction and cubic spline normalisation using GenomeStudio software from Illumina. After initial pre-processing steps, the effect of other technical variables like batch, chip, plate, etc., were removed using principal components analysis [105].

Since this version of the array showed poor detection, the genes were filtered and only those detected at $p < 0.05$ in at least 80% of samples were retained. The final gene expression data from the Lund cohort comprised expression values of 8932 genes across 223 samples. In this cohort, 29% of patients were diagnosed at AJCC stage I, 28% patients were diagnosed at AJCC stage II, and for the remainder of the patients AJCC stage was not specified [105].

4.2.5.2 Applying reduced LMC signature to the Lund cohort

The reduced LMC gene signature was applied to the gene expression data from the Lund primary melanoma cohort using the previously described NCC approach (refer to 3.2.1). Similar to LMC dataset, the gene expression data from the Lund cohort was standardised to give each gene mean 0 and standard deviation 1. The tumours which had correlation less than 0.1 were labelled as unclassified.

Since cause of death was not reported in the Lund cohort, overall survival and relapse-free survival were used as outcomes. The relapse-free survival time is the time difference between age at diagnosis and age at recurrence of melanoma. The overall survival time is the time difference between age at diagnosis and age at death. The LMC class association with overall and relapse-free survival was tested using CPH models (refer to 2.4.4). The survival differences between the LMC classes were graphically visualised using Kaplan-Meier plots (refer to 2.4.2). The survival differences between the LMC classes were tested using the log-rank test (refer to 2.4.3). The analyses were performed using R-package *survival* [129].

4.2.5.3 Stratification of analysis by the AJCC stage in the Lund cohort

The patients in the Lund primary melanoma dataset were stratified on the basis of the recorded AJCC stage. The prognostic value of LMC classes was tested across the AJCC stages using CPH models (refer to 2.4.4). The survival differences between groups were tested using the log-rank test and visualised using Kaplan-Meier plots (refer to 2.4.2). Overall survival was used as the outcome measure. The analyses were performed using R-package *survival* [129].

4.2.6 Area under Receiver Operating Characteristic Curve

The prognostic value of the LMC signature in comparison to the AJCC stage was analysed using ROC analysis [147]. The conventional way of testing classifier performance is by calculating the accuracy value using a 2×2 contingency table

containing the disease status in columns and test prediction status in rows. However, this approach for calculating the accuracy of a classifier has limitations, as the prediction status values are influenced by the cut-off criterion which is often chosen arbitrarily. ROC analysis is not influenced by the arbitrary cut-off [147]. ROC analysis chooses a set of operating points that divides the data into groups [147]. The ROC graph is obtained by plotting the sensitivity values on the y-axis and 1- specificity values on the x-axis [147, 148]. The sensitivity is also referred to as the True Positive Fraction (TPF) and 1-specificity is referred to as the False Positive Fraction (FPF). The x and y axis values vary between 0 and 1 [147, 148]. The extreme values (0,0) & (1,1) in the graph are extremes of TPF and FPF. The points on the graph are connected to the extremes to complete the graph. A ROC curve lying on the diagonal line indicates that the performance of the classifier is not better than chance. A ROC curve lying towards the upper left corner indicates that performance of the classifier is much better than expected by chance. The area under the ROC curve (AUC) measures the accuracy of the classifier. This interpretation is based on the Mann-Whitney U test statistic which is used in calculating the AUC. AUC values range between 0.50 and 1, where 1 indicates perfect separation between the diseased and non-diseased cases, and 0.50 is equivalent to expected by chance.

The performance of two classifiers can be compared using the Z-score statistic. The null hypothesis is that there is no difference between the AUC's of two classifiers and the alternate hypothesis is that the AUC of the two classifiers differ [147, 148]. The Z-score statistic is calculated as:

$$Z = \frac{\widehat{AUC}_1 - \widehat{AUC}_2}{SE(\widehat{AUC}_1 - \widehat{AUC}_2)} \quad (4.1)$$

where, \widehat{AUC}_1 is area under ROC curve for first classifier and \widehat{AUC}_2 is area under ROC curve for second classifier,

$$\begin{aligned} SE(\widehat{AUC}_1 - \widehat{AUC}_2) \\ = \sqrt{Var(\widehat{AUC}_1) + Var(\widehat{AUC}_2) - 2Cov(\widehat{AUC}_1, \widehat{AUC}_2)} \end{aligned} \quad (4.2)$$

The performance of the AJCC staging system and LMC signature in the LMC and Lund datasets were compared using the ROC analysis. The predicted probabilities for the ROC analysis was calculated from a logistic regression model predicting melanoma-specific deaths in the LMC dataset, and melanoma relapse and overall deaths in the Lund dataset. The time point cut-off was chosen at 6 years and patients censored before 6 years in both the datasets were excluded from the analysis. The univariable model was generated for the LMC classes and the AJCC stage. The

multivariable included both the LMC classes and the AJCC stage. The AUC calculation and plots were generated and compared using R-packages *ROCR*, *plotROC* and *ggplot2* [130, 149, 150].

4.2.7 Biological significance of LMC classes

To understand the biological differences between the LMC classes, differentially expressed genes in each tumour class were identified. The differentially expressed genes in LMC classes were identified by comparing one class to all other classes combined. The comparison was done using the Significance Analysis of Microarrays (SAM) tool [151, 152]. Pathway enrichment analysis was performed for over- and under-expressed genes in each LMC class.

4.2.7.1 Significance Analysis of Microarrays (SAM)

Tusher *et al.* proposed the SAM method to identify genes with a significant difference in their expression under different experimental conditions (Figure 4.1) [152]. This method is widely applied in microarray-based studies. It performs a regularised t-test for each gene separately and calculates a score based on changes in gene expression value relative to standard deviation of repeated measurements for that gene [151, 152]. The False Discovery Rate (FDR) is calculated to estimate the percentage of genes identified as significant by chance, based on random data permutations. The permutations account for the correlation between the genes and avoid parametric assumptions about the distribution of genes [152]. After setting a threshold value (referred to as delta) for the amount of difference to consider as relevant, which is often done subjectively, the set of significant genes are identified, and the FDR (or q-value) is calculated for this gene list (Figure 4.1).

For each gene i , the SAM statistic is calculated between 2 groups as

$$d_i = \frac{\bar{x}_{i2} - \bar{x}_{i1}}{s_i + s_0} \quad (4.3)$$

Where x_{i2} and x_{i1} are the expression values of gene x_i in group 1 and group 2. The term s_0 denotes a constant value called the exchangeability factor or regularisation factor. Without this factor, equation (4.3) becomes a standard t-test. The addition of term s_0 stabilises s_i which tends to be smaller at lower gene expression levels, making d_i less sensitive to small changes in the expression levels. Ideally, when making a comparison across genes, the d_i should be independent of expression levels. Therefore, the value of s_0 is set such that the dependence of d_i on s_i should be

minimal. The ideal value of s_0 is such that the coefficient of variation of d_i is approximately constant as function of s_i .

The parameters \bar{x}_{i2} and \bar{x}_{i1} are calculated as:

$$\bar{x}_{i1} = \frac{\sum_{j \in C1} x_{ij}}{n_1} \quad (4.4)$$

$$\bar{x}_{i2} = \frac{\sum_{j \in C2} x_{ij}}{n_2} \quad (4.5)$$

where n_1 and n_2 are the total number of samples in class 1 ($C1$) and class 2 ($C2$), x_{ij} are samples observations for gene x_i in each class. The variable s_i is calculated as:

$$s_i = \sqrt{\frac{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \left\{ \sum_{j \in C1} (x_{ij} - \bar{x}_{i1})^2 + \sum_{j \in C2} (x_{ij} - \bar{x}_{i2})^2 \right\}}{(n_1 + n_2 - 2)}} \quad (4.6)$$

The analysis was done using R-package *samr* [151]. Two-class comparisons were performed and 500 permutations were used. The significantly over-expressed and under-expressed genes with q value=0 were selected for pathway enrichment analysis.

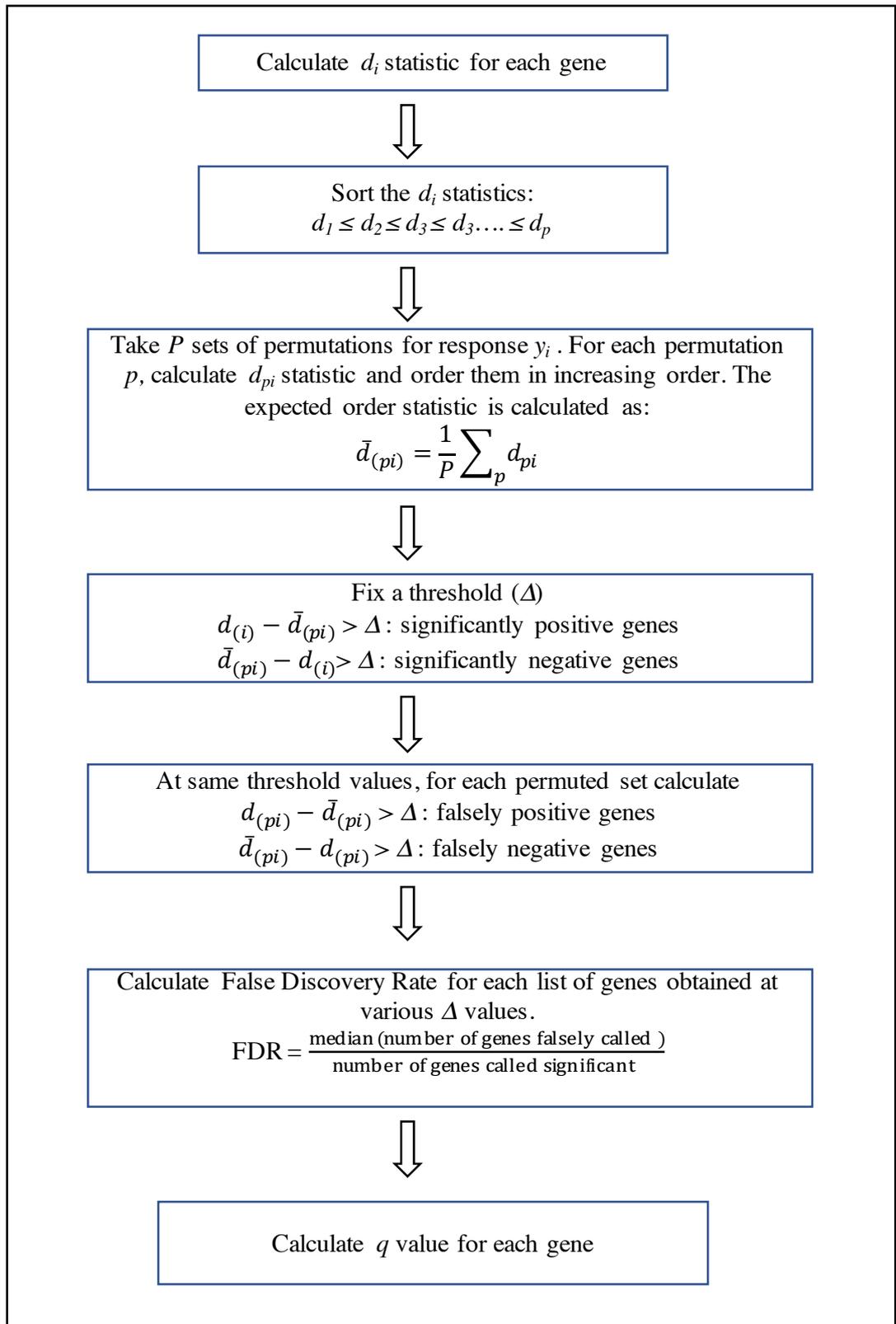


Figure 4.1 Summary of SAM workflow

Adapted from Chu *et al.* [153].

4.2.7.2 Pathway enrichment analysis

The significantly over-expressed and under-expressed genes identified from the SAM analysis were used for pathway enrichment analysis within ReactomeFIViz [154], a Cytoscape plugin [154, 155].

4.2.7.2.1 Reactome FIViz tool

Reactome FI is a Cytoscape based plugin used to perform pathway and network analysis of high dimensional datasets [154]. This tool matches the input set of genes against the Reactome functional interaction network to generate the list of enriched pathways [154]. Reactome functional interaction network is a reliable, manually curated protein functional interaction network which covers more than 60% of human proteins [154]. This tool allows construction of a functional interaction network for a set of query genes and analyses the sources of underlying evidence for physical interaction between the genes. The tool performs a binomial test comparing the list of input genes to the number of genes associated with a specific pathway. The pathways are queried from KEGG, Reactome and other open source databases [154]. The FDR of the P values is calculated using the Benjamini-Hochberg method of multiple testing correction [156]. The over- and under-expressed gene sets were input into the Reactome tool to perform pathway enrichment. The pathways with FDR < 0.01 were selected and further analysed.

4.2.8 Assessing the Lund module activity in LMC classes

4.2.8.1 Lund biological modules

Cirenajwis *et al.* performed transcriptional network analysis of highly correlated genes in an another cohort from Lund University, consisting of metastatic melanomas [144]. The pathways associated with highly correlated genes were summarised into five distinct modules named *immune*, *stroma*, *MITF*, *cell cycle* and *interferon*, from now on referred to as the Lund modules. These modules were found to be associated with the Lund 4-classes [144]. The *high-immune* class of Lund 4-classes had higher expression of the *immune* and *stroma* modules and lower expression of the *MITF* and *cell cycle* modules. The *pigmentation* class had higher expression of the *MITF* and *cell cycle* modules. The *proliferative* class had higher expression of the *cell cycle* module and lower expression of the *immune* and *stroma* modules. The *normal-like* class had increased expression of the *stroma* module and lower expression of the *cell cycle* module [144].

4.2.8.2 Application of the Lund modules to the LMC dataset

The LMC classes were characterised using the Lund modules. The *immune* module contained 231 genes, of which 215 were present in the LMC while 25/26 genes matched from the *MITF* module. All genes in the *stroma* (119 genes), *cell cycle* (11 genes) and *interferon* (7 genes) modules were present. The module score was calculated for each LMC tumour sample as:

$$s_i^m = \sum_{g \in G} x_{ig} \quad (4.7)$$

where m is the module (*immune*, *stroma*, *MITF*, *cell cycle* and *interferon*) and s_i^m is the module score for i^{th} sample, g is the gene in module gene set G and x_{ig} is the expression value of i^{th} sample for gene g . The correlation between the modules was assessed using Spearman's rank correlation.

4.3 Results

4.3.1 Identifying the number of classes in the LMC dataset

As described in the previous chapter, seven classes were selected using a subjective criterion based on the consensus clustering matrix graphics (refer to 3.2.5.4). To further confirm k using an objective criterion, a cluster separation measure was defined: the ratio of intra-cluster similarity to inter-cluster similarity. The cluster separation generally increased with the k value (Table 4.1). However, it was lower for $k=4$ than at $k=3$. In keeping with previous observations from the consensus matrix graphics, $k=7$ had the maximum increase (from $k=6$) in cluster separation in comparison to other k values. This observation confirmed selection of $k=7$ as the number of classes.

Table 4.1 Summary of cluster separation measure at various k values

The cluster separation generally increased with k values, with a maximum increase at $k=7$.

k	Cluster separation
2	5.1
3	5.5
4	5.1
5	5.9
6	7.2
7	9.1
8	9.6
9	9.9
10	10.2

4.3.1.1 Association of the LMC classes with Batch

To test whether clustering of tumours was influenced by differences due to the batch in which the samples were processed, the association of the LMC classes with batch number was tested. The LMC tumours were sent for profiling in three batches to a service provider (Service XS, Leiden, The Netherlands). Apart from the LMC class 7, all other LMC class samples were distributed across the three batches (Table 4.2). The LMC class 7 had lower number of samples ($n=15$), and 14 out of 15 samples were processed in the same batch. The statistically significant ($P=0.01$) association between batch and LMC classes suggested that clustering of some LMC classes may have been affected by artefacts arising from the batch in which samples were processed. Further exploration of LMC class 7 tumour samples suggested that these tumours were mainly present on the edges of a plate. After excluding class 7 samples the LMC classes were not associated with the batch ($P=0.2$).

Table 4.2 Summary of LMC class tumours across the three batches sent for gene expression profiling

LMC classes	Batch 101998	Batch 102232	Batch 102503
1	18	32	21
2	31	60	31
3	23	29	21
4	36	78	29
5	33	56	47
6	38	75	29
7	1	14	0

4.3.1.2 Re-clustering the LMC tumours after removing 15 samples

The above analysis suggested that the class with small sample size may be driven by artefactual features. The low sample sized class would also have limited power for downstream statistical analysis. For these reasons LMC class 7 samples were excluded from the LMC dataset, and the remaining samples were re-clustered using the same method as described previously (refer to 3.2.5.4) (consensus-based PAM clustering with the same parameters as before).

Re-clustering the LMC tumours indicated $k=6$ as a good number of classes. The CDF of the consensus matrix and delta area graph suggested minimal increase in area at $k=6$ (Figure 4.2). The cluster separation measure also had maximum increase at $k=6$ (Table 4.3). These observations clearly indicated that $k=6$ was the new optimal number of classes in the LMC dataset. These six classes largely overlapped with previously identified LMC classes with a very high agreement (Cramer' V= 0.97) (Table 4.4).

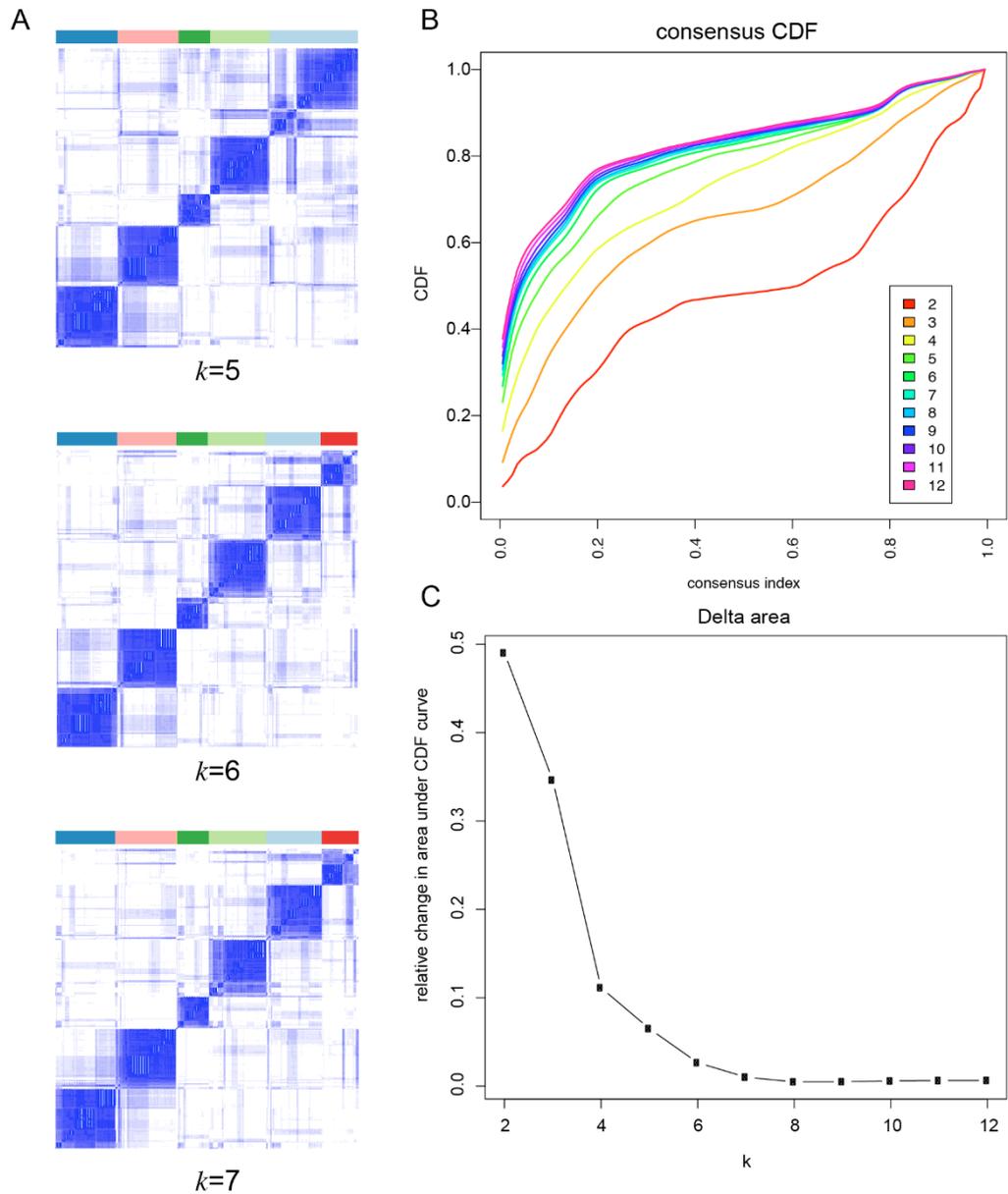


Figure 4.2 Re-clustering of LMC tumours after excluding 15 samples, using consensus clustering PAM method

(A) The heatmap of consensus matrix at $k=5, 6$ and 7 . At $k=7$ only sample was classified into the new cluster. (B) Cumulative Distribution Function (CDF) of the consensus scores at various values of k . (C) Relative change (delta) in the area under the CDF curve comparing k with $k-1$.

Table 4.3 Summary of cluster separation measure at various k values after removing 15 samples

k	Cluster separation measure
2	4.92
3	5.73
4	4.82
5	6.62
6	8.55
7	9.18
8	9.20
9	9.48
10	9.76

Table 4.4 Comparing new classification with initial classification after removing 15 samples

Class		Initial clustering					
		1	2	3	4	5	6
Re-clustering after removing 15 samples	1	71	0	0	0	0	1
	2	0	122	0	0	1	3
	3	0	0	72	5	0	6
	4	0	0	1	135	0	0
	5	0	0	0	3	135	0
	6	0	0	0	0	0	132

4.3.2 The LMC class association with clinico-histopathological features

As shown in Table 4.5, the LMC classes were significantly associated with tumour site, age at diagnosis, Breslow thickness, AJCC stage, ulceration status, TILs, *NRAS*, and *BRAF* mutation status. The LMC class 1, class 2 and class 5 tumours occurred more frequently on *limbs*. The LMC class 1 and class 5 tumours were most frequently non-ulcerated and thin while class 2 and class 4 tumours were thicker more likely to be ulcerated. The LMC class 3 and class 6 tumours were the thickest and mainly ulcerated. In term of AJCC stage, class 1 and class 5 tumours had a higher proportion of early stage tumours whereas class 3 and class 6 tumours had a higher proportion of advanced stage tumours (Table 4.5). LMC class 1 had a higher percentage of brisk TILs and LMC class 3 absence of TILs. LMC class 2 and class 6 had higher percentages of non-brisk TILs, and LMC classes 4 and 5 had intermediate levels of brisk and non-brisk TILs. The LMC classes also differed in *NRAS* and *BRAF* mutation status. LMC class 5 and class 6 were frequently *BRAF* mutated. LMC class 2, class 3 and class 4 contained frequently *NRAS*-mutated tumours (Table 4.5).

Table 4.5 LMC classes association with clinico-histopathological variables

Where n is the number of samples, m is the median and r is the range.

Histopathological variables	Whole dataset n=687 (%)	LMC class						P
		Class 1 (n=71)	Class 2 (n=122)	Class 3 (n= 73)	Class 4 (n=143)	Class 5 (n=136)	Class 6 (n=142)	
Sex : male n (%)	310 (45)	39 (55)	51 (42)	34 (47)	56 (39)	55 (40)	75 (52)	0.07
Tumour site: limbs n (%)	289 (42)	37 (52)	58 (48)	26 (36)	58 (41)	64 (47)	46 (32)	0.03
Age at diagnosis (years), m(r)	58 (18, 81)	59 (21,76)	59 (22,79)	60 (20,77)	58 (18,81)	53 (25,76)	59 (22,81)	0.03
Breslow thickness (mm) n=691 m(r)	2.3 (0.3, 20)	1.7 (0.7, 5.5)	2.1 (0.8, 8.9)	3.2 (0.8, 20)	2.3 (0.3, 15)	1.8 (0.7, 12)	3.0 (0.8, 18)	9.5×10^{-14}
AJCC stage n=680 (%)								5.2×10^{-9}
I	230 (34)	37 (52)	41 (34)	10 (14)	46 (33)	69 (51)	27 (19)	
II	344 (51)	29 (41)	59 (49)	46 (65)	77 (55)	47 (34)	86 (61)	
III	106 (15)	5 (7)	21 (17)	15 (21)	16 (12)	20 (15)	29 (20)	
Ulceration (present) n (%)	228 (33)	16 (23)	26 (32)	30 (41)	53 (37)	38 (28)	59 (42)	0.01
Number of mitoses (n= 577) \geq1	358 (62)	30 (50)	53 (58)	43 (70)	83 (64)	74 (64)	75 (64)	0.2
TILs n=490 (%)								5.5×10^{-4}
Absent	76 (15)	2 (4)	13 (14)	17 (32)	14 (16)	15 (16)	15 (13)	
Non-Brisk	333 (68)	30 (60)	65 (71)	32 (60)	60 (68)	63 (66)	83 (74)	
Brisk	81 (17)	18 (36)	14 (15)	4 (8)	14 (16)	17 (18)	14 (13)	
BRAF mutant (n = 568) yes (%)	266 (47)	26 (43)	38 (30)	23 (40)	44 (36)	63 (59)	72 (61)	5.6×10^{-5}
NRAS mutant (n= 561) yes (%)	138 (25)	8 (14)	35 (34)	17 (30)	41 (34)	20 (19)	17 (15)	3.2×10^{-4}

4.3.3 Prognostic value of the LMC signature

4.3.3.1 Prognostic value in the whole LMC dataset

LMC classes strongly predicted MSS in the LMC dataset (Figure 4.3, Table 4.6). The proportion of deaths from melanoma ranged from 14% in LMC class 1 patients to 49% in class 3, with 20% in class 2, 30% in class 4, 20% in class 5 and 36% in class 6.

In univariable analysis, LMC class 1 patients had good prognosis. LMC class 2 (HR=1.8 compared with class 1), class 5 (HR=1.6) and class 4 (HR=2.3) patients had intermediate prognosis (Figure 4.3). LMC class 3 (HR=5.0) and class 6 (HR=3.2) patients had the worst prognosis (Figure 4.3). Clinical variables such as the AJCC stage, age at diagnosis, tumour site, sex, and number of mitoses also significantly predicted MSS in the LMC dataset (Table 4.6).

In a multivariable model adjusting for clinical variables, LMC class 3 (HR= 3.6), class 4 (HR=2.2) and class 6 (HR= 2.2) maintained an independent association with survival (Table 4.6).

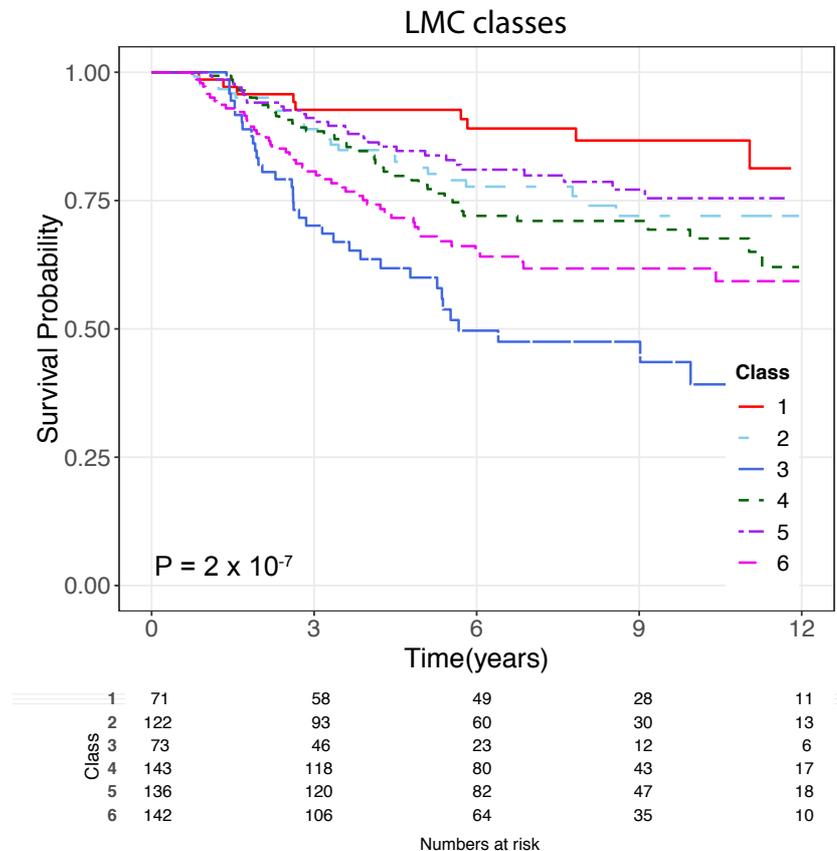


Figure 4.3 Melanoma-specific survival for the LMC classes on the whole LMC dataset

Table 4.6 Summary of the univariable and multivariable analysis of the LMC signature with melanoma-specific survival in the whole LMC dataset

The variable *n* is the total number of samples, *deaths* are the number of melanoma-specific deaths. CI is confidence interval and *HR* is the hazard ratio.

	Univariable					Multivariable (n=571, deaths=157)			
	Class (n)	HR	95% CI	P		Class (n)	HR	95% CI	P
LMC class (n=687, deaths=194)	1 (71)	1.00	-	-	LMC class	1 (61)	1.00	-	-
	2 (122)	1.8	0.9-3.7	0.1		2 (91)	0.9	0.4-2.0	0.7
	3 (73)	5.0	2.5-10.1	8×10^{-6}		3 (59)	3.6	1.7-7.6	0.001
	4 (143)	2.3	1.8-4.7	0.02		4 (126)	2.1	1.0-4.4	0.04
	5 (136)	1.6	0.8-3.2	0.2		5 (116)	1.4	0.7-3.1	0.3
	6 (142)	3.2	1.6-6.2	9×10^{-4}		6 (188)	1.9	0.9-3.7	0.08
AJCC stage (n= 680, deaths=192)	I (230)	1.00	-	-	AJCC stage	I (182)	1.0	-	-
	II (344)	2.5	1.7-3.7	5×10^{-6}		II (301)	1.6	1.0-2.6	0.04
	III (106)	6.0	3.8-9.2	2×10^{-15}		III (88)	4.7	2.8-7.9	5×10^{-9}
Sex (n=687, deaths=194)	F (377)	1.0	-	-	Sex	F (310)	1.0	-	-
	M (310)	1.5	1.1-1.9	8×10^{-3}		M (261)	1.2	0.8-1.7	0.3

	Univariable					Multivariable (n=571, deaths=157)			
	Class (n)	HR	95% CI	P		Class (n)	HR	95% CI	P
Age at diagnosis (n=687, deaths=194)	years	1.03	1.01-1.04	4×10^{-6}	Age at diagnosis	years	1.03	1.01-1.05	1×10^{-5}
Mitotic rate (n=577, deaths=158)	<1 (219)	1.0	-	-	Mitotic rate	<1 (216)	1.0	-	-
	>=1 (358)	1.9	1.3-2.7	4×10^{-4}		>=1 (355)	1.6	1.2-2.3	0.01
Site of tumour	Limbs (289)	1.0	-	-	Site of tumour	Limbs (240)	1.0	-	-
	Head (80)	1.1	0.6-1.9	0.6		Head (74)	1.0	0.5-1.7	0.9
	Trunk (230)	1.8	1.3-2.5	6×10^{-4}		Trunk (187)	1.8	1.2-2.7	0.007
	Other (88)	2.9	1.9-4.3	3×10^{-7}		Other (70)	2.0	1.2-3.2	0.002

4.3.3.2 Prognostic value of LMC signature in stage I tumours

As described in the previous chapter, the Lund and TCGA signatures did not predict survival in stage I tumours (refer to 3.3.4.1). To test the prognostic value of the LMC signature in stage I tumours, a univariable analysis and multivariable analyses (adjusting for the effect of clinical predictors) were performed.

In the univariable analysis, the LMC class 6 had relatively poor survival (Figure 4.4, Table 4.7). The overall survival difference between the six LMC classes was not significant (Figure 4.4) but the LMC class 6 had a significantly higher melanoma death hazard ratio (HR=6.6) than class 1 at P=0.02 (Table 4.7). LMC class 3 (HR=4.0) and class 4 (HR=3.1) had similar hazard ratios in AJCC stage I compared to the whole LMC dataset, although these did not reach the significance threshold of 0.05 due to the much reduced sample size. Among the clinical variables, mitotic rate (HR=2.4), Breslow thickness (HR=3.1), Sex (males: HR=2.2) and Site (limbs: HR=2.1) significantly predicted survival in the AJCC stage I group (Table 4.7).

In a multivariable model after adjusting for clinical variables (mitotic rate, Breslow thickness, Sex and Site), although not significant (P=0.07) but the LMC class 6 maintained a higher melanoma death hazard ratio (HR=7.8) (Table 4.7).

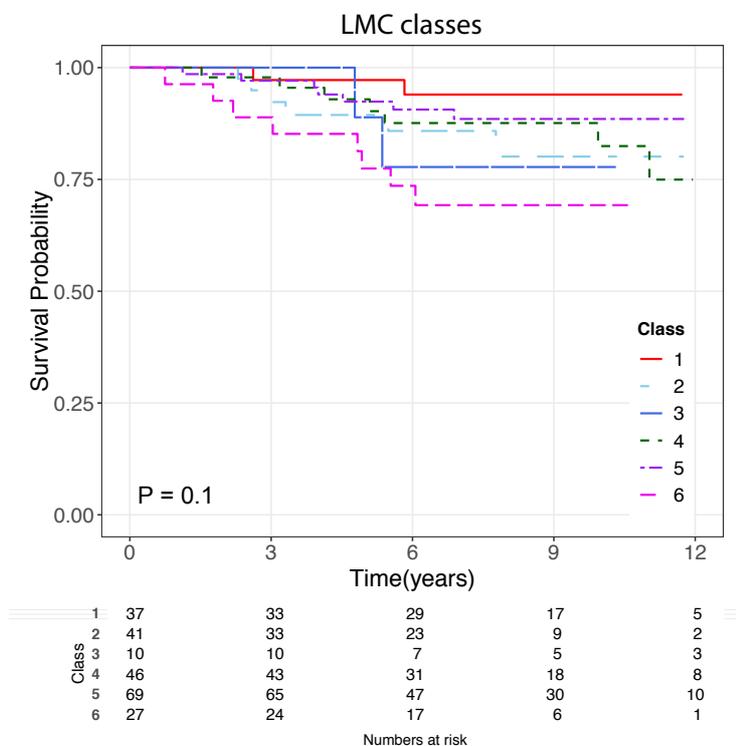


Figure 4.4 Melanoma-specific survival for the LMC classes on the AJCC stage I group

Table 4.7 Summary of the univariable and multivariable analysis of the LMC signature with melanoma-specific survival in the AJCC stage I group

The other tumour site variables like head and other rare site were excluded as only few samples were classified into these categories. The variable *n* is the total number of samples, *deaths* are the number of melanoma-specific deaths. CI is confidence interval and *HR* is the hazard ratio.

	Univariable					Multivariable (n=158, deaths=25)			
	Class (n)	HR	95% CI	P		Class (n)	HR	95% CI	P
LMC class (n=230, deaths=33)	1 (37)	1.0	-	-	LMC class	1 (26)	1.0	-	-
	2 (41)	3.5	0.7-17.3	0.1		2 (22)	5.7	0.6-54.1	0.1
	3 (10)	4.0	0.6-28.5	0.2		3 (7)	4.6	0.3-80.2	0.3
	4 (46)	3.1	0.7-15.2	0.2		4 (34)	5.5	0.7-46.4	0.1
	5 (69)	2.3	0.5-10.7	0.3		5 (50)	3.2	0.4-27.0	0.3
	6 (27)	6.6	1.4-31.0	0.02		6 (19)	7.8	0.9-70.7	0.07
Sex (n=230, deaths=33)	F (134)	1.0	-	-	Sex	F (94)	1.0	-	-
	M (96)	2.2	1.1-4.5	0.02		M (64)	2.6	1.1-6.5	0.05
Age at diagnosis (n=230, deaths=33)	Years	1.01	0.98-1.04	0.3					

	Univariable					Multivariable (n=158, deaths=25)			
	Class (n)	HR	95% CI	P		Class (n)	HR	95% CI	P
Breslow thickness (n=230, deaths=33)	mm	3.1	1.2-8.2	0.02	Breslow thickness	mm	1.32	0.4-4.2	0.6
Site (n=201, deaths=32)	Limbs (115)	1.0	-	-	Site	Limbs (90)	1.0	-	-
	Trunk (86)	2.1	1.0-4.2	0.04		Trunk (68)	1.6	0.6-4.2	0.3
Mitotic rate (n=182, deaths=25)	<1 (96)	1.0	-	-	Mitotic rate	<1 (96)	1.0	-	
	>=1 (86)	2.4	1.0-5.5	0.04		>=1 (86)	2.6	1.1-6.3	0.04
Ulceration of tumour (n=230, deaths=33)	No	1.0	-	-					
	Yes	1.1	0-Inf	0.9					

4.3.4 Comparing LMC signatures with Lund 4-classes and TCGA 3-classes

The LMC signatures overlapped with previously described Lund 4-classes and TCGA 3-classes (Figure 4.5). The LMC class 1 tumours overlapped to a large extent with *high-immune* class, class 3 overlapped with *pigmentation* class, class 5 overlapped with *normal-like* class of Lund 4-classes (Figure 4.5A). The LMC class 2, class 4 and class 6 were a mixture of the Lund 4-classes (Figure 4.5A).

Similarly, for TCGA 3-classes, the LMC class 1 overlapped with the *Immune* class, LMC class 3 overlapped with *MITF low* class and LMC class 5 overlapped with *Keratin* class (Figure 4.5B). LMC class 2, class 4 and class 6 were a mixture of the TCGA 3-classes.

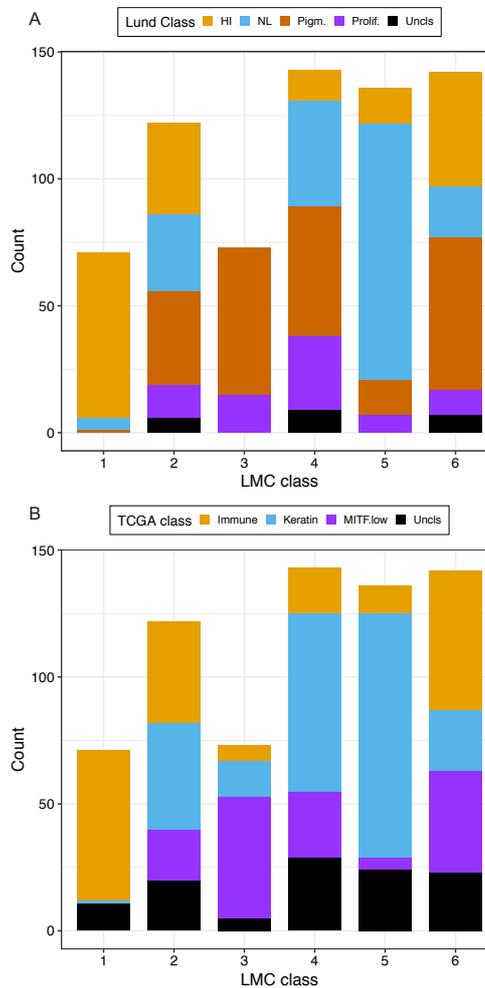


Figure 4.5 Overlap between the LMC signatures and the Lund 4-classes and TCGA 3-classes

Lund 4-classes: HI- *high immune*, NL- *normal-like*, Pigm.- *pigmentation*, Prolif.- *proliferative*; TCGA 3-classes: *immune*, *keratin* and *MITF low*. Unclassified samples are shown in black.

4.3.5 Replicating prognostic value of LMC signature

4.3.5.1 Generating a refined LMC signature

The initial LMC signature comprised ~13,000 genes. Applying this gene signature in clinical practice would not be feasible. To mitigate this, the gene signature was reduced to form multiple smaller subsets by selecting the top n (1, 5, 10, 25, 100 and 500) ranked upregulated genes in each of the LMC classes, resulting in signatures of 6, 30, 60, 150, 600 (599 unique) and 3000 (2908 unique) genes (refer to 4.2.4). For all signatures, the LMC classes 1, 3 and 5 had lower misclassification rate and LMC classes 2, 4 and 6 has higher misclassification rate (Figure 4.6). The 6 gene-based, 30 gene-based and 60-gene based signature sets had low accuracy in classifying LMC tumours into their respective classes (Figure 4.6). All other gene set signatures had a comparatively similar performance to the all gene-based signature. Among these gene signatures, the signature with fewer genes, the 150 gene-based signature, was selected for validation on an external dataset (Figure 4.6; Appendix I Table 8.1).

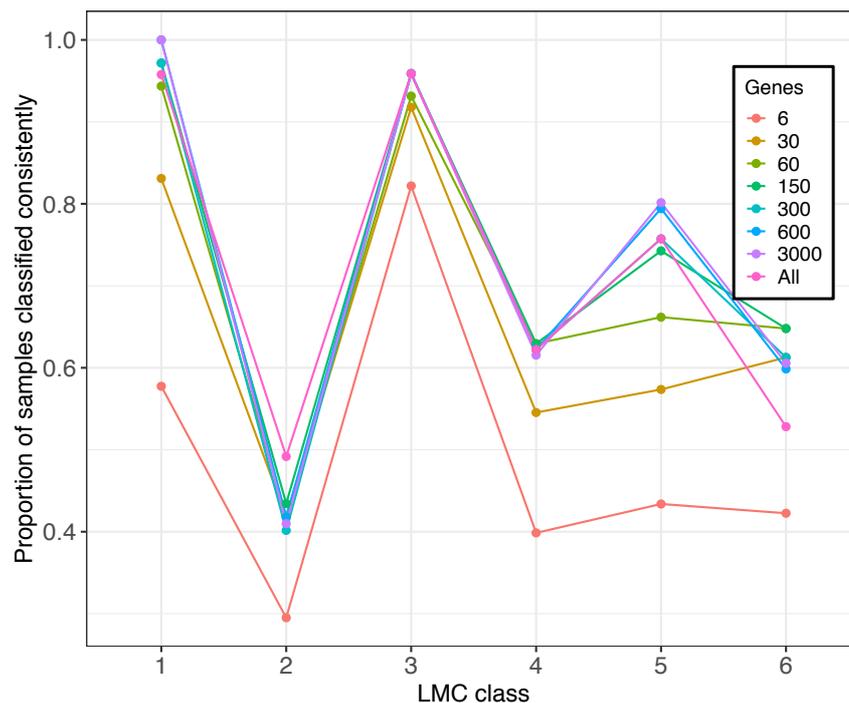


Figure 4.6 Refining the LMC class gene signature

The six gene signatures were generated by combining the top n (1, 5, 10, 25, 50, 100 and 500) highly expressed genes from each LMC class. The LMC tumours were reclassified into the LMC classes based on these signatures using the NCC approach. The plot compares the proportion of samples in each class in the original classification classified in the same way by the reduced signature.

4.3.5.2 Prognostic value of LMC signature in Lund cohort

The 150-gene based LMC signature was applied to the independent cohort of primary melanomas from Lund, Sweden. Among 150 genes in the LMC signature, 73 genes were present in the gene expression dataset from the Lund cohort. In the Lund cohort, 27% of tumours were classified into LMC class 1, 4% in class 2, 23% in class 3, 9% in class 4, 18% in class 5 and 14% tumours in class 6. The LMC class 1 tumours ($r=0.44$) had the highest median correlation with LMC class 1 centroids. The LMC class 2 ($r=0.37$) and class 3 ($r=0.40$) tumours had a similar correlation with their respective centroids. The LMC class 4 ($r=0.26$), class 5 ($r=0.32$) and class 6 ($r=0.28$) tumours had comparatively lower correlation with their respective class centroids. Three percent of tumours in the Lund cohort could not be classified.

The LMC classes significantly predicted relapse-free and overall survival in the Lund cohort (Figure 4.7, Table 4.8). In LMC class 1, 33% patients died, 27% in class 2, 75% in class 3, 71% in class 4, 35% in class 5 and 58% in class 6. The survival plot indicated that LMC class 1, class 2 and class 5 predicted good prognosis in the Lund cohort (Figure 4.7). The LMC class 3, class 4 and class 6 predicted worse prognosis with increased hazard for relapse and overall deaths in the Lund cohort (Figure 4.7, Table 4.8).

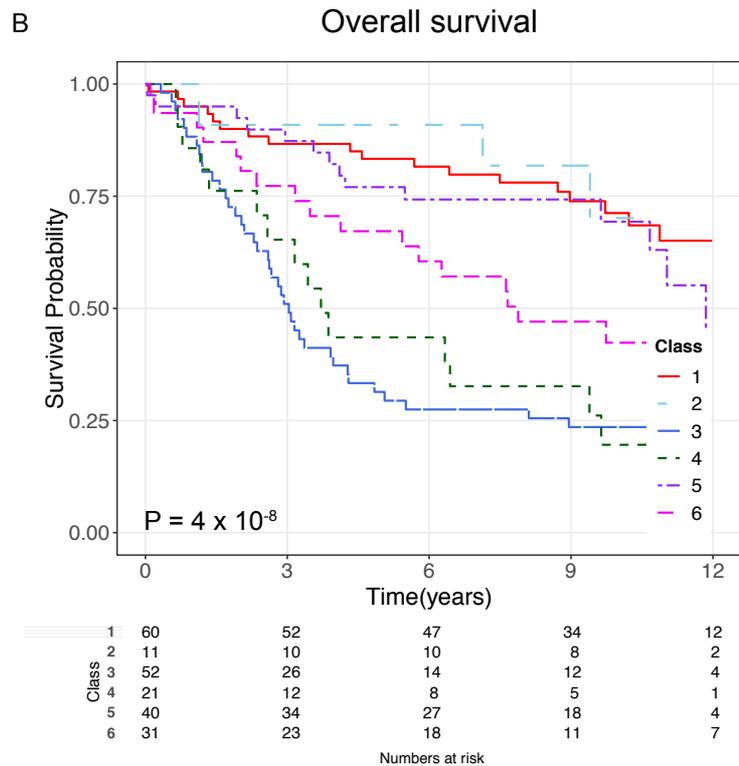
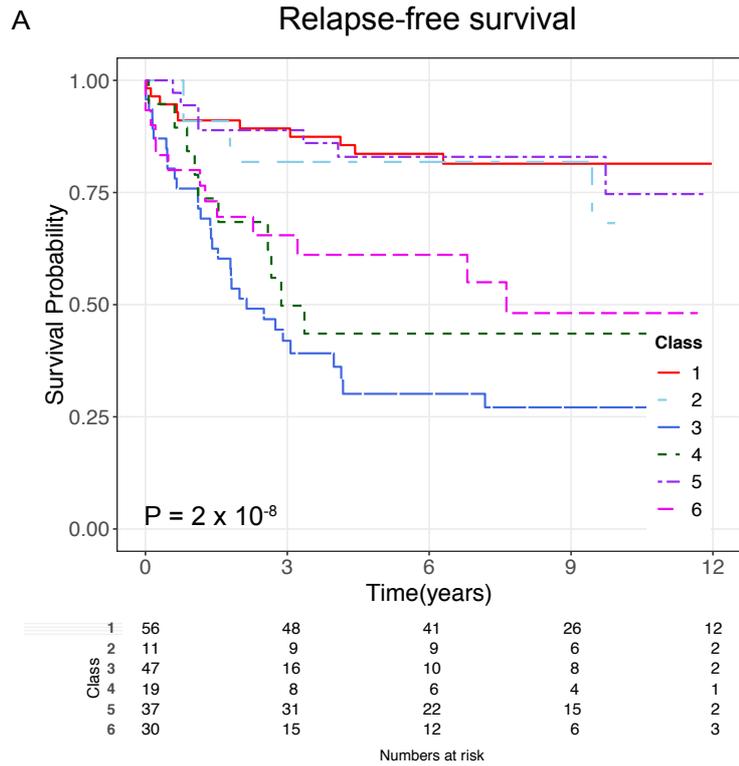


Figure 4.7 Relapse-free survival and overall survival for the LMC classes on the Lund dataset

Since melanoma-specific survival was not available, (A) relapse-free survival and (B) overall survival were used as clinical outcomes for survival analysis.

Table 4.8 Summary of the univariable analysis of the LMC signature with relapse-free survival and overall survival in the Lund cohort

The variable n is the total number of samples, e is the event (relapsed or death) and HR is the hazard ratio.

LMC class	Relapse-Free Survival (n=200, relapsed=75)			Overall Survival (n=215, deaths=109)		
	HR	95% CI	P	HR	95% CI	P
Class 1	1.0	-	-	1.0	-	-
Class 2	1.5	0.4-5.4	0.5	0.8	0.2-2.7	0.7
Class 3	6.3	3.1-12.9	6×10^{-7}	4.0	2.3-6.9	6×10^{-7}
Class 4	4.2	1.7-10.1	0.001	3.6	1.8-7.0	2×10^{-4}
Class 5	1.1	0.4-3.0	0.8	1.2	0.6-2.3	0.6
Class 6	3.7	1.6-8.4	0.001	2.1	1.1-4.0	0.02

4.3.5.3 Prognostic value of LMC signature in AJCC stage I group

The LMC classes showed no significant association with relapse-free survival or overall survival in the AJCC stage I group of the Lund cohort (Table 4.9). In AJCC stage I, 7 out of 28 people in LMC class 1, 3 out of 5 people in class 3, 4 out of 18 people in LMC class 5, and 2 out of 6 people in LMC class 6 died. The LMC classes 2 and 4 had less than 5 samples in AJCC stage I and were excluded from the survival analysis. The CPH model analysis indicated that class 3 had increased hazard for melanoma relapse ($HR=2.3$) as well as death ($HR=3.5$) in comparison to baseline LMC class 1 (Table 4.9), although this was not statistically significant. The hazard for relapse and death was also slightly higher in LMC classes 5 and 6 compared to class 1.

Table 4.9 Summary of the univariable analysis of the LMC signature with relapse-free survival and overall survival in the AJCC stage I group of Lund cohort.

The LMC classes 2 and 4 had less than 5 samples in AJCC stage I and were excluded from the analysis, *n* is the number of samples, *e* is the event (relapsed or death), CI is confidence interval, and *HR* is the hazard ratio.

LMC class	Relapse-free Survival (n=54, relapsed= 8)				Overall Survival (n=74, deaths=16)			
	n, e	HR	95% CI	P	n, e	HR	95% CI	P
Class 1	26, 4	1.0	-	-	28, 7	1.0	-	-
Class 3	5, 1	2.3	0.2-21.3	0.5	5, 3	3.5	0.9-14.0	0.07
Class 5	18, 3	1.7	0.3-8.5	0.5	18, 4	1.1	0.3-3.9	0.8
Class 6	6, 1	1.2	0.1-12.1	0.8	6, 2	1.2	0.2-6.1	0.8

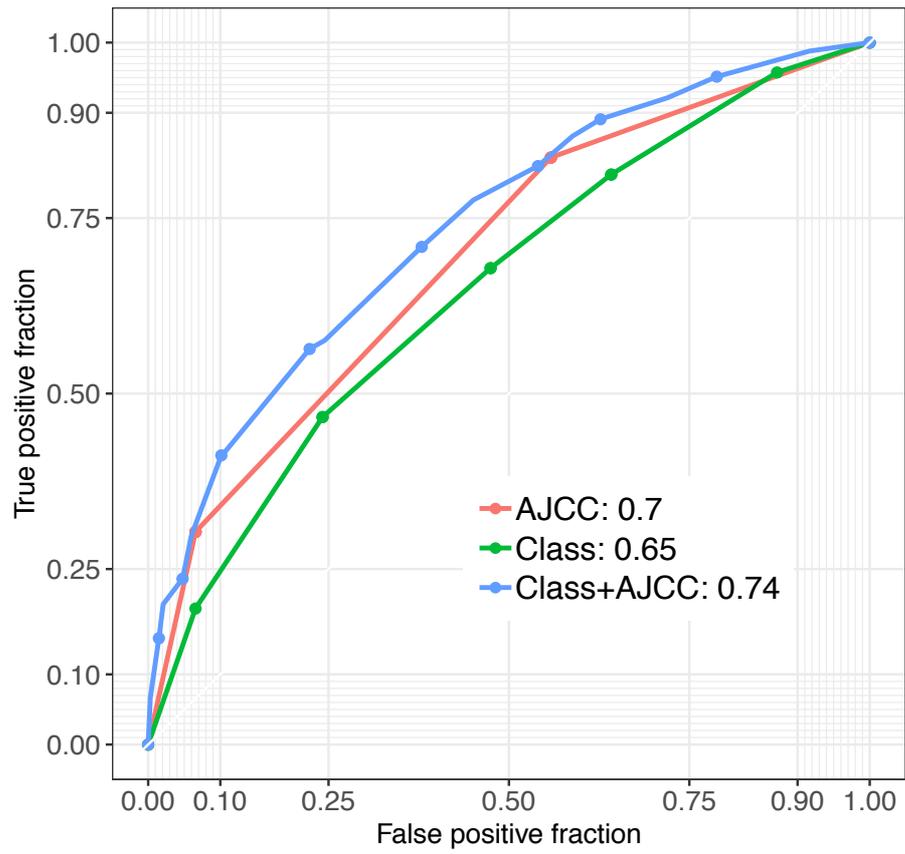
4.3.5.4 Independent prognostic value of the LMC signature

To assess the independent prognostic value of the LMC signature, ROC analysis was performed predicting melanoma-specific deaths in the LMC dataset and melanoma relapse and overall deaths in the Lund dataset (Figure 4.8). The time point cut-off was chosen at 6 years and patients censored before 6 years were excluded from the analysis.

The AJCC classification system showed an AUC of 70% in the LMC dataset (Figure 4.8A) and 81% and 78% in the Lund dataset (Figure 4.8B-C). The LMC signature had similar performance to the AJCC staging system in the LMC dataset (AUC=65%) and the Lund data (AUC= 78%-79%) (Figure 4.8). Combining the two i.e. the AJCC staging system with the LMC signature showed 4% improvement of AUC in the LMC dataset in comparison to the AJCC staging system alone (Figure 4.8A). The improvement for the combination improved by 5% and 7% in the Lund dataset for the two outcomes, melanoma relapse and overall deaths respectively (Figure 4.8B-C). The improvement in AUC for both the datasets was statistically significant (LMC MSS P=0.001, Lund relapse P=0.02, Lund overall deaths P=0.02).

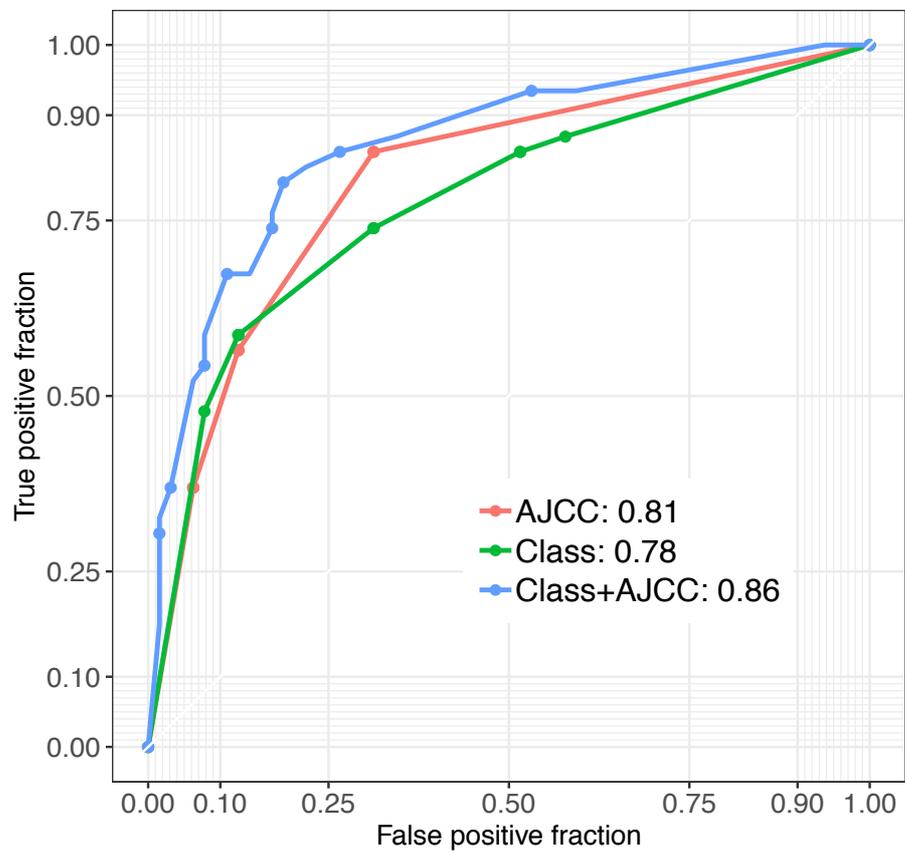
A

Leeds data



B

Lund data



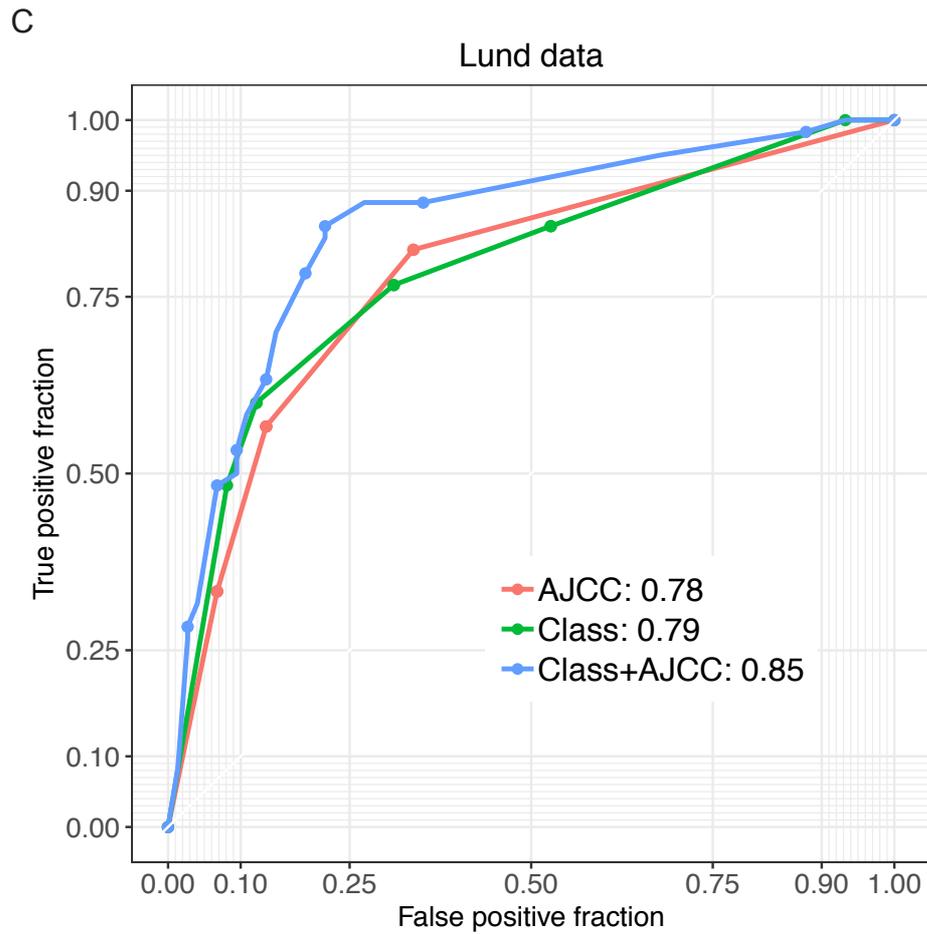


Figure 4.8 Independent prognostic value of the LMC signature

ROC curves of the LMC classes and the AJCC staging system predicting (A) melanoma-specific deaths in the LMC dataset and (B) melanoma relapse and (C) overall deaths in the Lund dataset.

4.3.6 Biological significance of the LMC classes

4.3.6.1 Pathway enrichment analyses of LMC classes

The biological differences between the LMC classes were explored using differentially expressed genes identified from SAM analysis [151, 152]. These genes were input into ReactomeFIViz tool [154] for identifying the associated biological pathways. All pathways reported here were significant at FDR <0.01.

The LMC class 1 tumours had increased expression of genes in the immune response pathways (Figure 4.9). The most significant pathways were NF-KB signalling, chemokine signalling, osteoclast differentiation, T-cell receptor signalling, etc. (Figure 4.9). The LMC class 1 tumours had decreased expression of genes in the cell cycle related pathways (Figure 4.9), such as mitotic G1-G1/S phase, cell cycle checkpoints, cell cycle, etc. The LMC class 2 tumours had increased expression of genes in the eukaryotic translation and with nonsense-mediated decay pathways and decreased expression of genes in the PLK1 signalling and FOXM1 transcription factor network pathways (Figure 4.10).

The LMC class 3 tumours mirrored class 1 tumours and had increased expression of genes in the cell cycle related pathways and decreased expression of genes in the immune response pathways (Figure 4.11). The LMC class 4 tumours had very few genes with significantly increased expression. The associated pathways were assembly of the primary cilium and RNA polymerase II transcription (Figure 4.12).

The LMC class 5 tumours had increased expression of genes in the Cell junction organisation and extra cellular matrix-receptor interaction (Figure 4.13). These tumours had decreased expression of genes in the Wnt signalling, Cadherin signalling and assembly of primary cilium pathways.

The LMC class 6 showed some characteristics of LMC class 1 and class 3 tumours. The class 6 tumours had increased expression of genes in the cell cycle related, focal adhesion and immune response pathways (Figure 4.14). These tumours had decreased expression of genes in the eukaryotic translation and nonsense-mediated decay pathways.

The complete list of the associated biological pathways is provided in Appendix I (Table 8.2 - Table 8.13).

LMC class 1

Upregulated pathways



Downregulated pathways

Figure 4.9 Summary of biological pathways associated with the LMC class 1

The upregulated pathways are represented in red and downregulated pathways are represented in blue. The font size is based on the significance level, with most significant pathways shown in larger fonts. The bracket indicates the database used for each pathway (K- KEGG, R- Reactome, N- NCBI database).

LMC class 2

Upregulated pathways

Regulation of nuclear SMAD2/3 signaling(N)
Ribosome(K)
Eukaryotic Translation Termination(R)
Eukaryotic Translation Elongation(R)
Eukaryotic Translation Initiation(R)
SRP-dependent cotranslational protein targeting to membrane(R)
Nonsense-Mediated Decay (NMD)(R)
Selenoamino acid metabolism(R)
Processing of Capped Intron-Containing Pre-mRNA(R)
SUMOylation(R)

Downregulated pathways

NoRC negatively regulates rRNA expression(R)
p73 transcription factor network(N)
Cell Cycle Checkpoints(R)
Mitotic Metaphase and Anaphase(R)
Mitotic G2-G2/M phases(R)
Signaling by Rho GTPases(R)
ATR signaling pathway(N)
Cell cycle(K) **Alcoholism(K)**
PLK1 signaling events(N)
Systemic lupus erythematosus(K)
FOXM1 transcription factor network(N)
APC/C-mediated degradation of cell cycle proteins(R)
Progesterone-mediated oocyte maturation(K)
Mitotic Prometaphase(R)
Oxidative Stress Induced Senescence(R)
Assembly of the primary cilium(R)

Figure 4.10 Summary of biological pathways associated with the LMC class 2

The upregulated pathways are represented in red and downregulated pathways are represented in blue. The font size is based on the significance level, with most significant pathways shown in larger fonts. The bracket indicates the database used for each pathway (K- KEGG, R- Reactome, N- NCBI database).

LMC class 3

Upregulated pathways

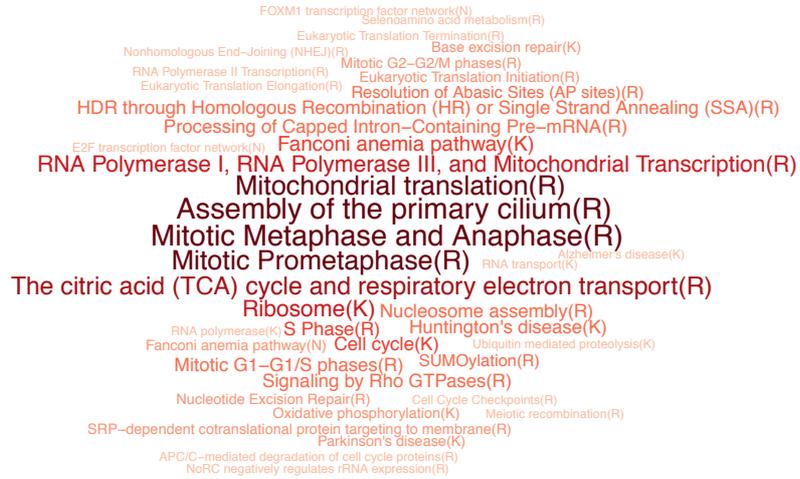


Figure 4.11 Summary of biological pathways associated with the LMC class 3

The upregulated pathways are represented in red and downregulated pathways are represented in blue. The font size is based on the significance level, with most significant pathways shown in larger fonts. The bracket indicates the database used for each pathway (K- KEGG, R- Reactome, N- NCBI database).

LMC class 4

Upregulated pathways

p75(NTR)-mediated signaling(N)
 RNA Polymerase II Transcription(R)
 Assembly of the primary cilium(R)
 Processing of Capped Intron-Containing Pre-mRNA(R)

Downregulated pathways



Figure 4.12 Summary of biological pathways associated with the LMC class 4

The upregulated pathways are represented in red and downregulated pathways are represented in blue. The font size is based on the significance level, with most significant pathways shown in larger fonts. The bracket indicates the database used for each pathway (K- KEGG, R- Reactome, N- NCBI database).

LMC class 5

Upregulated pathways

Validated targets of C-MYC transcriptional repression(N)
 Direct p53 effectors(N)
 Validated transcriptional targets of deltaNp63 isoforms(N)
 HIF-1-alpha transcription factor network(N)
 Alzheimer disease-presenilin pathway(P)
 Hippo signaling pathway(K)
 AP-1 transcription factor network(N)
 Validated transcriptional targets of TAp63 isoforms(N)
 ECM-receptor interaction(K)
 Focal adhesion(K) EPH-Ephrin signaling(R)
 Validated transcriptional targets of AP1 family members Fra1 and Fra2(N)
 Cell junction organization(R)
 Pathways in cancer(K) Sphingolipid signaling pathway(K)
 Beta1 integrin cell surface interactions(N)
 Extracellular matrix organization(R) Bladder cancer(K)
 Proteoglycans in cancer(K) Amoebiasis(K)
 Axon guidance(K) PI3K-Akt signaling pathway(K)
 Ras signaling pathway(K)
 a6b1 and a6b4 Integrin signaling(N)
 Generic Transcription Pathway(R)
 Rap1 signaling pathway(K)

Downregulated pathways

Processing of Capped Intron-Containing Pre-mRNA(R)
 Wnt signaling pathway(P)
 Assembly of the primary cilium(R)
 Cadherin signaling pathway(P)

Figure 4.13 Summary of biological pathways associated with the LMC class 5

The upregulated pathways are represented in red and downregulated pathways are represented in blue. The font size is based on the significance level, with most significant pathways shown in larger fonts. The bracket indicates the database used for each pathway (K- KEGG, R- Reactome, N- NCBI database).

LMC class 6

Upregulated pathways

amb2 Integrin signaling(N)
 Signaling events mediated by focal adhesion kinase(N)
 Extracellular matrix organization(R)
 Pertussis(K) Cell Cycle Checkpoints(R)
 Mitotic Prometaphase(R)
 Toll-Like Receptors Cascades(R)
 SUMOylation(R) Mitochondrial translation(R)
 Signaling by Insulin receptor(R)
 Cell cycle(K) Iron uptake and transport(R)
 Vibrio cholerae infection(K) Phagosome(K) Regulation of DNA replication(R)
 S Phase(R) Mitotic G1-G1/S phases(R)
 Signaling by Rho GTPases(R)
 Mitotic Metaphase and Anaphase(R)
 APC/C-mediated degradation of cell cycle proteins(R)
 Regulation of retinoblastoma protein(N)
 Rheumatoid arthritis(K) Focal adhesion(K)
 Beta1 integrin cell surface interactions(N)
 Oxidative phosphorylation(K) Synthesis of DNA(R)
 ROS, RNS production in response to bacteria(R)
 Mitotic G2-G2/M phases(R)
 Epithelial cell signaling in Helicobacter pylori infection(K)
 Urokinase-type plasminogen activator (uPA) and uPAR-mediated signaling(N)
 Degradation of beta-catenin by the destruction complex(R)
 Complement and coagulation cascades(K)

Downregulated pathways

EPHA forward signaling(N)
 Ras signaling pathway(K) GeneX Transcription Pathway(R)
 Sphingosine signaling pathway(K)
 Regulator of G-protein signaling(N)
 Phosphatidylinositol signaling system(K)
 Selenoamino acid metabolism(R)
 Eukaryotic Translation Elongation(R)
 Nonsense-Mediated Decay (NMD)(R)
 Eukaryotic Translation Termination(R)
 Eukaryotic Translation Initiation(R)
 SRP-dependent cotranslational protein targeting to membrane(R)
 Ribosome(K)
 Cell junction organization(R) Adherens junction(K)
 Rap1 signaling pathway(K) Focal adhesion(K)
 Regulation of actin cytoskeleton(K)
 Stabilization and expansion of the E-cadherin adherens junction(K)

Figure 4.14 Summary of biological pathways associated with LMC class 6

The upregulated pathways are represented in red and downregulated pathways are represented in blue. The font size is based on the significance level, with most significant pathways shown in larger fonts. The bracket indicates the database used for each pathway (K- KEGG, R- Reactome, N- NCBI database).

4.3.6.2 Characterising LMC classes using Lund modules

The LMC classes were further characterised using Lund modules identified by Cirenajwis *et al.* [144] in an independent cohort of metastatic melanoma from Lund, Sweden. In keeping with observations from pathway enrichment analysis (refer previous section), the LMC classes demonstrated distinct Lund module activity. The LMC class 1 tumours had higher *immune*, *stroma* and *interferon* module activity and lower *MITF* and *cell cycle* module activity (Figure 4.15A-E). The LMC class 2 tumours had higher *MITF* module activity and lower *cell cycle* module activity. As observed in pathway enrichment analysis, LMC class 3 tumours mirrored LMC class 1 tumours and had higher *MITF* and *cell cycle* module activity and lower *immune*, *stroma* and *interferon* module activity (Figure 4.15A-E). The LMC class 4 tumours had higher *MITF* module activity and lower *immune* and *stroma* module activity. The LMC class 5 tumours had higher *immune* and *interferon* module activity. The LMC class 6 tumours had higher *immune* and *cell cycle* module activity (Figure 4.15A-E).

When comparing modules scores with one another, the *immune* module scores were positively correlated with *stroma* and *interferon* modules and negatively correlated with *MITF* and *cell cycle* module scores (Figure 4.15F). The *MITF* module scores were positively correlated with *cell cycle* module score and negatively correlated with *immune*, *stroma* and *interferon* modules (Figure 4.15F).

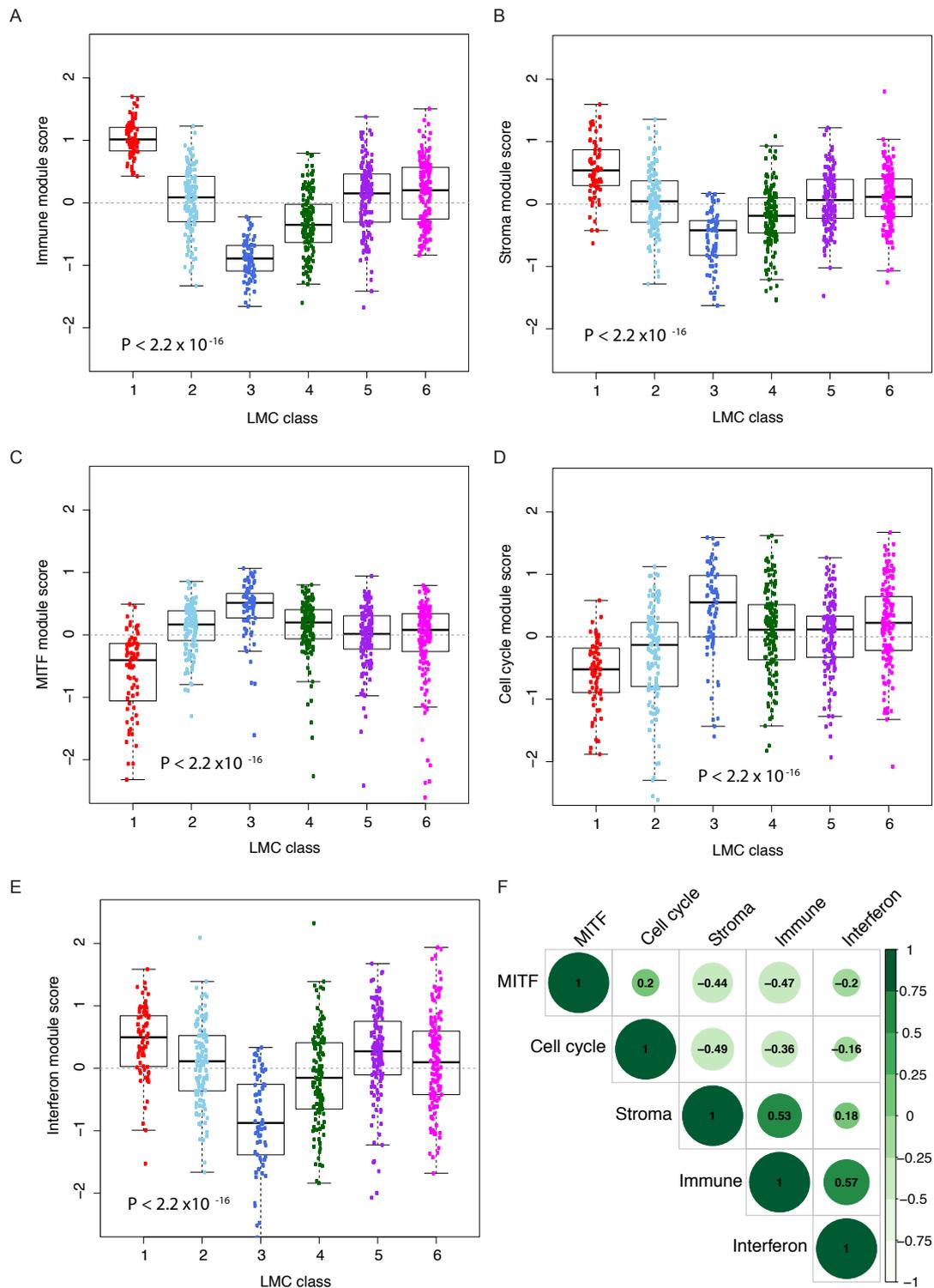


Figure 4.15 Characterising LMC classes using the Lund modules

The dot and boxplot show the distribution of scores for (A) *immune*, (B) *stroma*, (C) *MITF*, (D) *Cell cycle* and (E) *interferon* modules across the LMC classes. (F) Spearman's correlation coefficient between these modules [144].

4.4 Discussion

In this study, six melanoma tumour classes were defined using one of the largest cohorts of primary melanoma. The LMC classes were associated with distinct clinical features of melanoma and independently predicted survival. Unlike previous signatures, the LMC classes strongly predicted outcome in the whole LMC dataset including in stage I tumours. The prognostic value of LMC classes was recovered in an independent cohort of primary melanomas from Lund, Sweden. The biological differences were explored using genome-wide gene expression and candidate functional modules.

Previously, consensus clustering using the PAM algorithm identified seven classes of melanoma (refer to chapter 3). Applying a novel cluster separation method confirmed selection of seven classes in the LMC dataset. Interestingly, one of the classes contained only a few samples which were later identified to be processed in the same batch. At the time, it was not clear whether this class may represent technical bias, and these samples were excluded from the downstream analysis. Consensus clustering has been shown to be robust for obtaining a stable number of clusters [137, 157]. This was true in LMC, as re-clustering the dataset after removing the small class robustly recovered the previously observed six LMC classes with a very large agreement (Cramer V=97%).

4.4.1 Clinical-pathological characteristics of the LMC classes

Similar to Lund and TCGA signatures, the LMC classes were significantly associated with clinico-pathological features of melanoma patients like age, Breslow thickness, tumour site, AJCC stage, ulceration of tumour and TILS. The LMC class 1 and class 5 tumours were thin, non-ulcerated and presented at early stages. The LMC classes 2 and 4 had increased tumour thickness and ulceration of tumours. The LMC class 3 and class 6 were the thickest and were advanced stage tumours. In melanoma, BRAF and NRAS mutations have been reported to be mutually exclusive with strong therapeutic implications [77, 78, 83, 84, 158-162]. Interestingly, unlike previous signatures, the LMC classes were strongly associated with *BRAF* and *NRAS* mutation status [81, 106]. LMC class 5 and class 6 were frequently *BRAF* mutated and LMC class 2, class 3 and class 4 were frequently *NRAS* mutated.

4.4.2 Prognostic significance of the LMC classes in stage I melanoma

In keeping with Lund and TCGA classes, the LMC classes were also predictive of outcome in the whole LMC dataset. However, as shown previously, the Lund and TCGA signatures did not predict prognosis in the AJCC stage I tumours, and one of the objectives of this research was to devise a prognostic signature for these early stage tumours. Although the 5-year survival rate is about 95% for patients diagnosed at stage I [44], it remains of clear importance to predict those at risk of dying from their disease. In the LMC classes, LMC class 6 predicted significantly worse outcome in the AJCC stage I in the univariable model. When jointly analysed with other clinical variables the LMC class 6 still had increased melanoma death hazard for stage I tumours but it did not reach the significance threshold in the multivariate model. The lack of significance in the multivariate model could be due to the fact that this analysis was done on comparatively smaller set of samples as many samples had missing clinical information, and therefore were removed from the analysis. The comparable hazard ratios in the univariable and multivariable models suggests that the LMC class 6 predicts poor outcome in stage I group, and this class may have potential clinical relevance.

It has been shown that metastatic melanoma has distinct gene expression signatures in comparison to primary cutaneous melanoma [122, 163-165]. However, the LMC classes from primary melanoma showed an overlap with the Lund and TCGA signatures derived from metastatic melanoma. This suggests that primary and metastatic melanoma tumours do converge and share common biological characteristics which can be very well captured using the transcriptomic data. The good prognosis groups of LMC classes overlapped with good prognosis groups from Lund 4-classes and TCGA 3-classes. Similarly, the poor prognosis group LMC class 3 overlapped with the poor prognosis group from the Lund and TCGA classes. However, LMC class 2, class 4 and class 6 were novel in that they could not be distinguished using the Lund 4-classes and TCGA 3-classes.

The prognostic value of LMC classes was assessed in an independent cohort from Lund, Sweden [105]. Before this step, and to mimic real time clinical application, a reduced signature of the LMC classes was generated and trained in the LMC dataset. The reduced signature comprising 150 genes was found optimal in terms of using the smallest number of genes without loss of performance.

The reduced signature maintained similar association with outcome in the Lund cohort as it did in the Leeds dataset. The LMC classes 1 and 5 predicted good prognosis, and classes 3 and 6 predicted worse prognosis in both the cohorts. The LMC class 2 and

class 4, which had intermediate prognosis in the Leeds dataset, showed good and worse in the Lund dataset respectively. When stratifying patients on the basis of AJCC stage, the LMC classes showed no association with survival in stage I patients from the Lund dataset. As shown previously, (refer to 3.3.4.1), the Lund and TCGA signatures showed no association with prognosis in the stage I group of the Leeds dataset and had different hazard ratios in stage I when compared to the whole dataset (confirmed as a significant interaction). Although LMC classes were prognostic in the stage I group of the Leeds dataset, these classes showed no association with prognosis in the stage I group of the Lund dataset. Unlike the Lund and TCGA signatures, the hazard ratios for the LMC classes were comparable in the stage I group and the whole Lund dataset. Therefore, the lack of association for the LMC classes in the stage I group of the Lund dataset could be due to limited sample size (Leeds stage I group n=233, Lund stage I group n=74).

The gene expression data in the Lund cohort was generated using a different array platform and different pre-processing steps in comparison to Leeds data. Despite half of the genes in the LMC signature not being present in the Lund dataset, the LMC classes still demonstrated prognostic significance which supports the view that these classes are robust and capture meaningful biological information about melanoma tumours. In future, the genes that were not present on the Lund dataset will be imputed using the correlated genes selected from the LMC dataset.

AJCC stage is a strong predictor for melanoma outcome, and very few additional variables have shown an improvement in AUC by more than 2% [44, 114]. The LMC classes predicted prognosis independent of the AJCC staging system. The area under ROC value of LMC classes was comparatively similar to the AJCC classification system in the LMC and Lund datasets. However, a significant improvement in AUC by 4-7% was observed when combining the LMC classification system with AJCC staging in comparison to AJCC staging system alone.

4.4.3 Biological interpretation of the LMC classes

The LMC classes showed distinct gene expression profiles. Previously it has been shown that increased immune gene expression predicts good prognosis and good response to treatment in melanoma [81, 104-106, 117, 166, 167]. In keeping with the literature, classes associated with upregulation of immune response mechanism had the best prognosis and classes associated with increased cell cycle mechanism and low immune response had the worst prognosis in the LMC data. However, the novel non-overlapping LMC classes (LMC classes 2, 4 and 6) showed association with gradient of immune gene expression, cell cycle, cell communication, and cell

metabolism related pathways. The LMC class 4 tumours showed low immune response and had poor prognosis. In contrast, LMC class 6 tumours showed evidence of increased immune response along with cell proliferation activity but still had worse prognosis in comparison to class 4 tumours. This is interesting, as LMC class 6 highlights a class of tumours who have poor prognosis despite relatively high immune expression. Further work is required to understand the role of immune response in driving the poor prognosis of LMC class 6 tumours.

Overall, this chapter defined six molecular classes of melanoma. The classes were prognostic in the whole LMC dataset and in an independent dataset from Lund. Unlike previous signatures, LMC class signature showed prognostic value at the AJCC stage I. However, due to limited sample size in the Lund cohort this finding has not yet been replicated.

Chapter 5

Machine learning applications to predict melanoma prognosis using gene expression

The objectives of this chapter are:

Objective 9: To apply Random Forest (RF) to develop a classification model for predicting outcome in the LMC. The performance of the model will be assessed using sensitivity, specificity, and the kappa index

Objective 10: To apply Support Vector Machine (SVM) using linear and non-linear kernel functions to develop classification models for predicting outcome in the LMC. The performance of SVM model will be compared with the RF model

5.1 Introduction

The advances in cost-effective next-generation sequencing and gene expression profiling techniques have brought about an era of large biomedical datasets. The unsupervised classification methods described previously identify novel subgroups by exploring distinct patterns in these datasets. Supervised methods, including machine learning, on the other hand mine the knowledge from known subgroups or outcomes and develop prediction models to classify new observations into these groups [110-112]. Machine learning techniques are now being widely applied to learn complex relationships in these datasets and forecast health and disease outcomes [168-170]. Previously, a comprehensive performance comparison of machine learning algorithms on different datasets indicated that performance was primarily dependent on data quality, and showed that in good quality datasets the majority of these algorithms made similar predictions [170].

In this chapter two widely applied machine learning algorithms, RF and SVM, were used to predict prognosis in primary melanoma [171, 172]. The RF and SVM prediction models were generated using the training subset of LMC, and the performance of these models was assessed on the test subset of LMC.

5.2 Methods

5.2.1 Devising a supervised classification framework

The supervised classification framework presented in Figure 5.1 was applied to develop prediction models using gene expression data of LMC tumours. The gene expression dataset was divided into training (70%) and test (30%) sets based on random sampling of observations. RF and SVM algorithms were applied to the training set to develop prediction models across N number of iterations. The model with minimal classification error during training was tested on the test set (Figure 5.1).

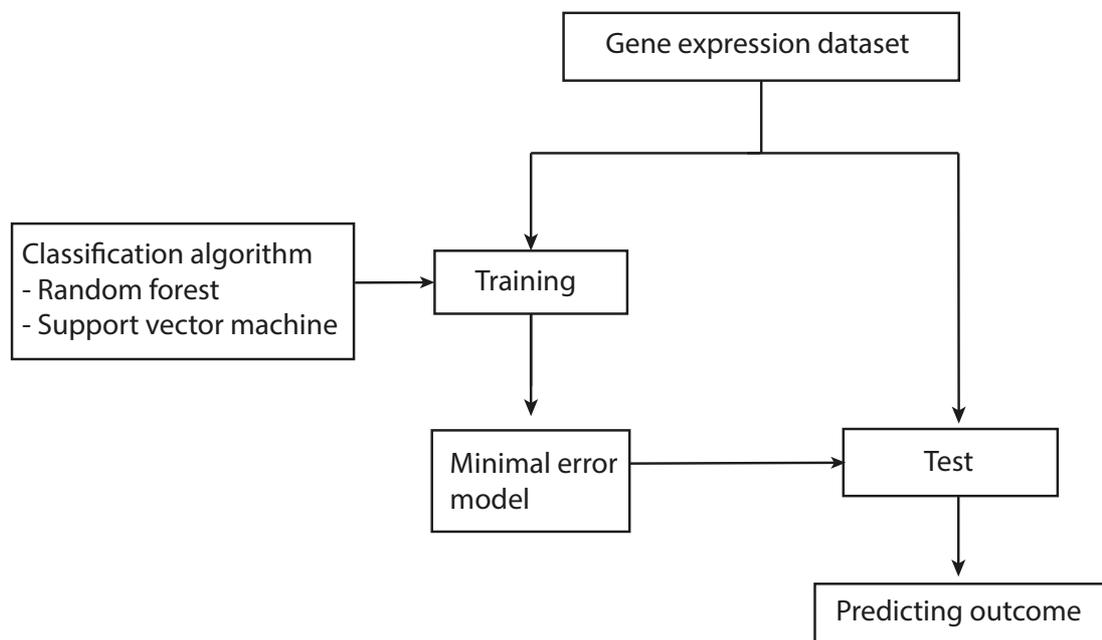


Figure 5.1 Summary of supervised classification framework

5.2.2 Defining a dichotomous outcome in the LMC

The prediction outcome was melanoma-specific survival (refer to section 2.4.1). The life table shows statistics for the numbers of deaths, people censored, people still alive, and number of people eligible for this analysis in the LMC (Table 5.1). The censored observations are patients whose cause of death was not recorded, who left the study or who have been followed up for a shorter time. In this study, to create a dichotomous outcome status, a survival time threshold was selected by aiming to maximise number of deaths and not losing many samples to censoring (Table 5.1). Therefore, melanoma-specific survival up to 6 years was chosen as a threshold. Patients who survived up to 6 years post-diagnosis were put in class 0 and those who died from their melanoma within 6 years were assigned to class 1. Patients

censored before 6 years follow-up were excluded. As Table 5.1 shows, 168 patients have died of melanoma within 6 years post-diagnosis (class 0) while 368 were still alive in that period (class 1), making a total of 536 patients.

Table 5.1 Life table of patients in LMC for 16-year time interval

Follow up time (years)	Died from melanoma	Censored	Alive	Patients eligible for analysis
1	10	3	689	699
2	55	12	635	690
3	95	54	553	648
4	122	97	483	605
5	144	128	430	574
6	168	166	368	536
7	174	217	311	485
8	178	267	257	435
9	181	322	199	380
10	186	347	169	355
11	187	384	131	318
12	190	435	77	267
13	193	483	26	219
14	196	503	3	199
15	196	506	0	196
16	196	506	0	196

5.2.3 Creating the training and test sets in LMC

Although transcriptomic data were available for all 536 patients retained for the dichotomised outcome, a small number were lost to gene expression quality-control (refer to section 4.3.1.1), leaving 525 patients as the final dataset. This dataset was randomly divided into 70% and 30% referred to as training and test sets respectively.

The genes in the training set were scaled to give each gene a mean 0 and standard deviation 1. The genes in test set were scaled using mean and standard deviation measures from the training set.

5.2.4 Prediction performance measures

The prediction performance was assessed using the sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV) and kappa index. These measures are defined below in Figure 5.2 and in the equations below.

		Reference status	
		Class 1	Class 0
Predicted status	Class 1	True Positive (TP)	False Postive (FP)
	Class 0	False Negative (FN)	True Negative (TN)

Figure 5.2 Summary of reference terms when comparing actual status with predicted status

The performance measures are calculated as follows:

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (5.1)$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \quad (5.2)$$

$$\text{Positive Predictive Value (PPV)} = \frac{TP}{(TP + FP)} \quad (5.3)$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{(TN + FN)} \quad (5.4)$$

$$\text{Accuracy} = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (5.5)$$

In unbalanced datasets, the accuracy is not a useful performance measure [173]. The Cohen's kappa index is used for measuring the agreement between the actual

and predicted outcome by accounting for agreement by chance [174, 175]. This index is calculated as:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (5.6)$$

where p_o is the observed agreement and p_e is the expected agreement by chance, derived from the Figure 5.2 as:

$$p_e = \frac{[(TP+FP)(TP+FN)] + [(FN+TN)(FP+TN)]}{(TP+FP+TN+FN)^2} \quad (5.7)$$

Kappa varies between 0 and 1, with 0 indicating no agreement, 0-0.20 a slight agreement, 0.21- 0.40 a fair agreement, 0.41-0.60 a moderate agreement, 0.61-0.80 a substantial agreement, and 0.81-1.00 almost perfect agreement between the actual and predicted classes [174, 176]. In this thesis, kappa was used to assess the prediction performance of classification models.

5.2.5 Developing classification models using Random Forest (RF)

In 2001, L Breiman formulated a classification method, known as the Random Forest, that is an aggregation of a large number of classification trees [171]. From a given set of training observations, a series of bootstrap datasets S of size X are drawn with replacement, where X is the number of observations [171, 177, 178]. For each bootstrap dataset, a decision tree is grown, and after S iterations the resultant S decision trees are collectively referred to as the Random Forest (RF). At each branch (node) of each tree, m input variables are randomly selected as potential predictors on which the bootstrap dataset is split. In each bootstrap dataset, the observations that were not included from the training set are referred to as the out-of-bag (OOB) samples.

Each OOB sample receives a classification “vote” from every decision tree that was grown independently of that sample. Finally, the RF algorithm chooses a classification label that receives the most votes, and the proportion of trees choosing that classification provides a measure of certainty about the classification. Since each training set observation has a probability of $1 - e^{-1}$ (~ 0.632) of being included in the bootstrap dataset of size X , so there will be approximately $S/3$ decision trees which will have that training observation as OOB. Similarly, a new observation can be classified in the same way using all the trees in the forest.

RF algorithm is a slight modification of the bagging algorithm [178, 179]. The bagging algorithm generates decision trees in a similar way to the RF algorithm, but in the

bagging algorithm each tree node is split using a large number of predictor variables. Hence, if a dataset has a very strong predictor variable, the bagging algorithm will make the split using this variable across most of the decision trees. This may lead to generation of decision trees which are strongly correlated [178]. The RF algorithm reduces, but does not eliminate, correlation by random selection of predictor variables at each node.

As described above, from a training set, S bootstrap datasets are generated and for each dataset the RF algorithm repeats the three following steps for constructing a decision tree (Figure 5.3):

- 1) randomly select a predefined number (m , which is usually chosen as the square root of total number of predictor variables in the dataset) of variables from the data at each node
- 2) among m variables, select the variable which gives the best data split according to the outcome status at that node
- 3) the node is split into two daughter nodes.

These steps are repeated until the terminal node size is just greater than a predefined minimum node size (Figure 5.4). The number of observations at each node is referred to as the node size. Figure 5.4 shows an example of a decision tree with 100 observations. In this example, observations are divided into smaller subsets until the daughter node contains just over 14 observations.

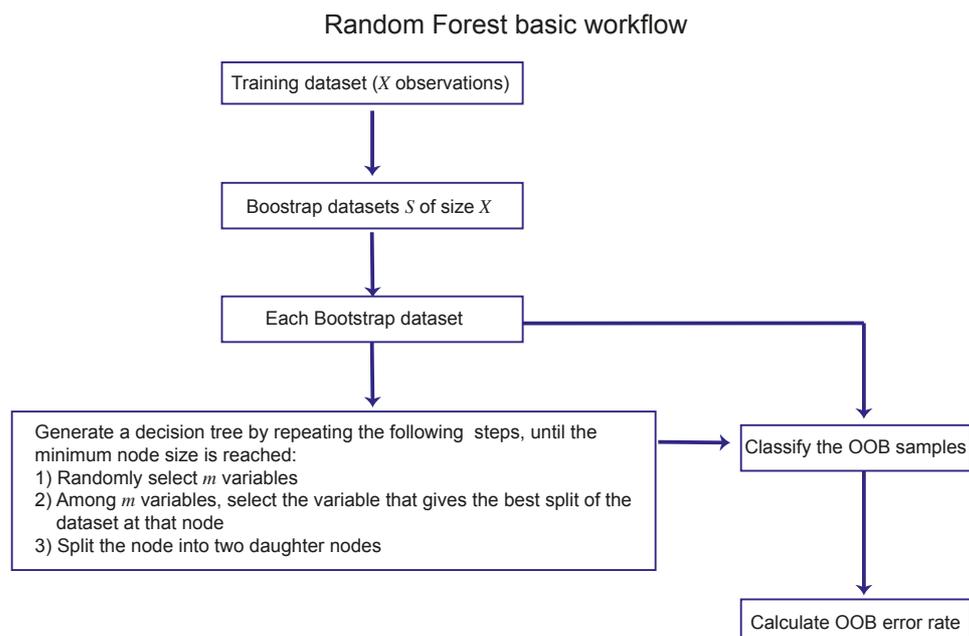


Figure 5.3 Workflow of RF algorithm

Adapted from James et al. [178]

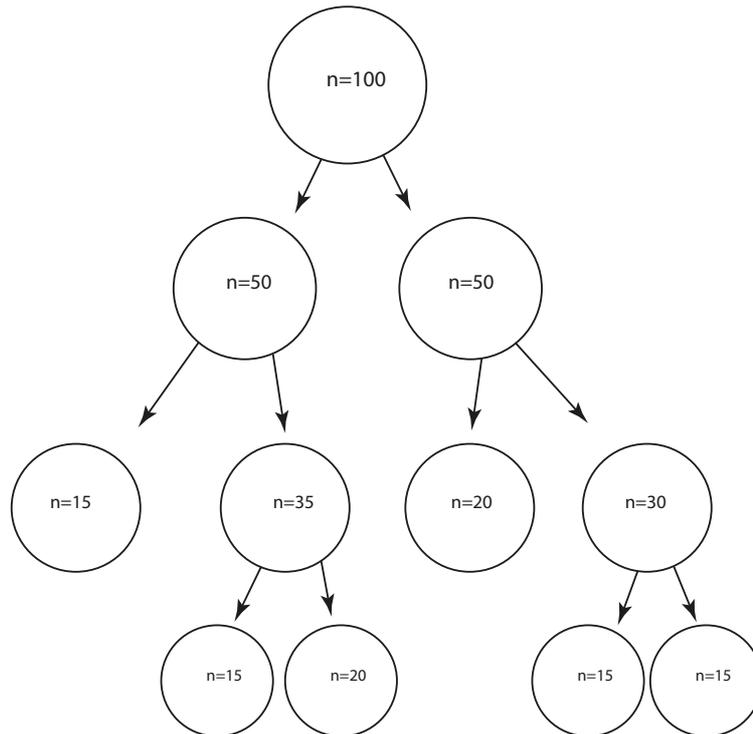


Figure 5.4 Example of a RF decision tree.

The minimum node size is 14.

After generating a decision tree, the prediction is made on OOB samples and a classification error is calculated, referred to as OOB error (Figure 5.3). The RF also calculates an importance measure for each predictor variable in the training set [177, 178]. The importance measure indicates the contribution of that variable in correctly classifying each OOB observation.

5.2.5.1 OOB error

During training, the RF algorithm does not require cross-validation or a separate test set to evaluate the error rate of a prediction model [171, 177, 178]. The classification error rate is calculated during the iteration steps as follows:

- 5 Each i^{th} decision tree is constructed using observations in the bootstrap dataset generated from random sampling of training set.
- 6 The OOB observations for the bootstrap dataset are classified using the i^{th} decision tree.
- 7 After S iterations, j^{th} observation classification label is decided based on majority vote from the decision trees that were grown independently of this observation.

- 8 Finally, after classifying all OOB observations, the proportion of OOB observations that were incorrectly classified is referred to as the OOB error of the RF model.

5.2.5.2 Variable importance measure

The RF algorithm calculates the variable importance using a permutation-based approach and Gini index [171, 178]. In the permutation-based approach, OOB samples are used to measure the prediction strength of each variable. When the i^{th} tree is grown, the OOB samples are passed down this tree and a prediction accuracy is calculated. Values for the g^{th} variable are permuted in the OOB set and the prediction accuracy is re-calculated. The difference in accuracy after permutation is averaged over all the trees and normalised using the standard deviation of differences, and this measure is used as an importance measure of the g^{th} variable [178].

Another importance measure, decrease in Gini index, is calculated every time the split is made at each node. This measure does not require permutations and is therefore less computer-intensive and gives often similar results as the permutation-based measure described above. The Gini index is calculated for the parent node and each of two daughter nodes. The Gini index is calculated as:

$$G = 1 - \sum_{k=1}^K p_k^2 \quad (5.8)$$

where p_k is the proportion of observations at each node that belong to k class and K are the total number of classes. Values of this index varies between 0 and 0.5. $G=0$ when all observations belong to the same class and $G=0.5$ when all classes contain the same number of observations. The difference in Gini index is calculated between the parent and the two daughter nodes and more informative variable will associate with larger decrease in Gini index value. The decrease in Gini index measure for g^{th} variable is calculated by averaging over all the nodes that have used that variable to make the split. Finally, this value is averaged over all the trees in the forest.

5.2.5.3 Properties of RF

RF works well with high dimensional datasets containing many predictors [178] and gives an unbiased estimate of the prediction error as the forest trees are grown [177]. It also provides an estimate of which variables are important in the classification. It can deal with unbalanced datasets by assigning different weights to the minority

classes. The algorithm also calculates a proximity matrix or “similarity” measure between the observations which can be used in clustering or detecting outliers in the dataset. Overall the random nature of selecting variables and observations for constructing decision tree makes this method less prone to overfitting. Like other methods, however, RF also has a limitation: as it uses a large number of trees for making prediction, it can be very computer-intensive, although this can be alleviated by using the Gini index instead of a permutation-based approach to assess variable importance.

5.2.5.4 Application of the RF to LMC

The RF algorithm was applied on the LMC training set to develop prediction models. The number of trees was set to 20,000, which means that 20,000 bootstrap datasets were generated. The number of predictor variables selected at each node of the decision tree was set as the square root of the total number of predictor variables in the training set ($mTry=117$). The variable importance was calculated using the mean decrease in Gini index [177]. The analysis was performed using R-packages *Caret* [180] and *randomForest* [181]. The generated RF model was applied on the test set and performance measures were calculated for both the training and test sets.

5.2.5.5 Dealing with class imbalance using undersampling

The training set was not balanced, i.e. the majority of observations were in class 0 ($n=253$) in comparison to class 1 ($n=115$). To overcome class imbalance, the under-sampling scheme was applied on the majority class [182, 183] using *samplesize* function in *randomForest* R-package [181]. This function randomly selected equal number of observations (number of observations in minority class 1 was chosen in this case) from both classes at each iteration to generate bootstrap datasets. The same parameters were used for the RF application as described in the previous section.

5.2.6 Developing classification models using SVM

The SVM is a supervised classification approach developed in the 1990s [184]. This technique is a non-linear extension of the maximum margin classifier, which is applicable when classes are linearly separable [185]. The SVM algorithm uses a training set to identify a hyperplane separating observations, which is used to make predictions on future observations.

5.2.6.1 Maximum Margin classifier

The maximum margin classifier in p dimensions identifies a hyperplane in a subspace of $p-1$ dimensions that separates and maximizes the distance from the training set observations [178]. In the special case of a two-dimensional space the hyperplane would be a straight line (Figure 5.5), which is defined as the set of points $X = (x_1, x_2)$ such that

$$f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 = 0 \quad (5.9)$$

where $\beta_0, \beta_1, \beta_2$ are the variable coefficients [178]. The test set observations are classified using the hyperplane. If $f(x)=0$, the test set observation lies on the line. If $f(x) < 0$, the observation is classified into class 0 and if $f(x) > 0$ the observation is classified into class 1. The magnitude of $f(x)$ gives an estimate about how far the observation x lies from the hyperplane.

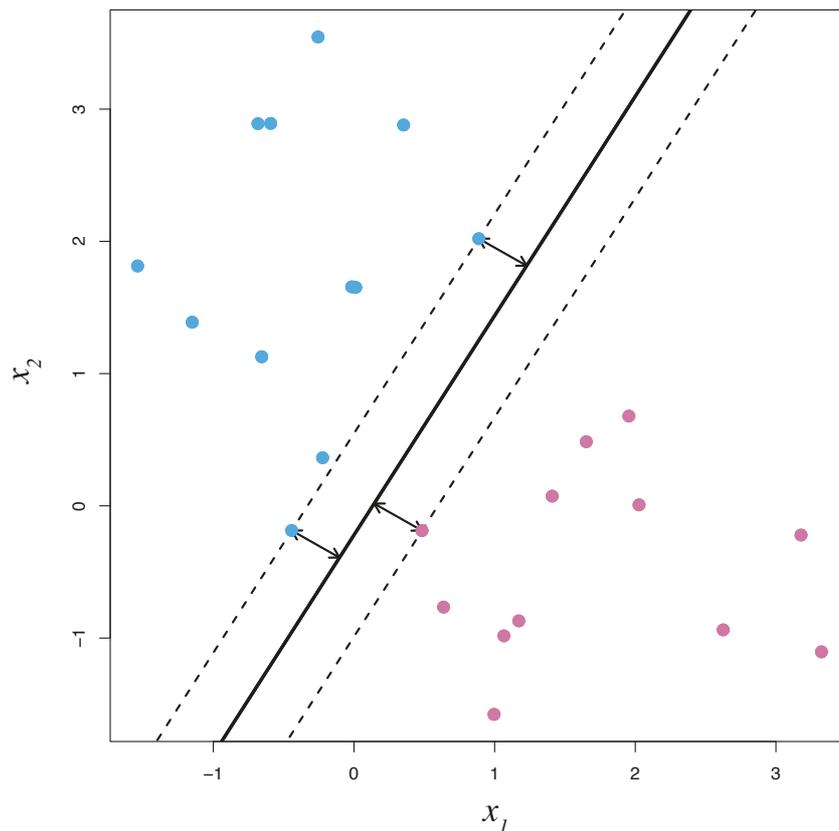


Figure 5.5 Maximum margin classifier

Adapted from James et al. [178]. Example of a maximum margin classifier identified hyperplane in two dimensions. The margin is the distance from the solid line to the dashed lines. The distance from the three points indicated by the arrows are the support vectors.

The smallest perpendicular distance from an observation to the hyperplane is referred as the margin (Figure 5.5).

Generalizing to p dimensions, with a given a set of observations $X_1, X_2, \dots, X_n \in R_p$ and associated two class labels $y_1, y_2, \dots, y_n \in (-1, 1)$, the maximum margin hyperplane solution is given by the values of β maximizing the margin, M , such that:

$$y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M, \forall i = 1, 2, 3, \dots, n \quad (5.10)$$

The maximum margin classifier has largest marginal distance on the training set observations but may not have the largest margin on the test set, which is an example of overfitting. The maximum margin classifiers are desirable when classes are clearly/linearly separable. However, the need to classify all observations accurately can often lead to overfitting on the training data. This makes the maximum margin hyperplane extremely sensitive to individual observations. The support vector classifier is an extension of this method which can tolerate some misclassifications while classifying the majority of the observations accurately [178].

5.2.6.2 Support vector classifiers

Support vector classifiers are also called soft-margin classifiers [178]. The soft-margin concept does not strictly adhere to classifying all the observations on the right side of a hyperplane; it allows some misclassifications. The solution to optimizing a soft-margin classifier is choosing β such that:

$$y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i) \quad (5.11)$$

for all $i = 1, \dots, n$, subject to the constraint:

$$\sum_{i=1}^n \epsilon_i \leq C \quad (5.12)$$

where C is a non-negative hyperparameter, M is the width of the margin and $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are the (non-negative) slack variables that allow misclassification of observations. The slack variables provide information about location of training observations with reference to the margin and the hyperplane. If $\epsilon_i = 0$, the observation is on the correct side of hyperplane. If $\epsilon_i > 0$, the observation is on the wrong side of the margin. If $\epsilon_i > 1$, the observation is on the wrong side of hyperplane.

The hyperparameter C controls the misclassification rate allowed in the training set. At large values of C , a narrow margin is chosen and no misclassification is tolerated;

the SVM algorithm tries to classify all training observations correctly [172, 186, 187]. For small values of C , the SVM algorithm tries to identify a hyperplane with wider margin and tolerates some misclassification of training set observations.

This hyperparameter also controls bias-variance trade-off: lower C values result in a model with higher bias and lower variance (underfitting), and higher C results in a model with high variance and low bias (overfitting). This hyperparameter C is tuned using a k -fold cross-validation approach. The observations that lie on margins are known as support vectors. When C is large, many support vectors are involved in determining the hyperplane, while smaller values of C correspond to fewer support vectors.

Like maximum margin classifiers, support vector classifiers work well with classes that are linearly separable by tolerating some misclassification. In case of non-linearly separable classes, support vector classifiers extend to higher order feature space functions, e.g. quadratic, cubic etc. This would require a lot computation time which is not practically feasible [178].

5.2.6.3 SVM using linear kernel function

SVM are an extension of support vector classifiers which enlarge the features space using kernel functions [172, 178, 188]. The idea of this is to find linear boundaries in the enlarged space and transform it back to non-linear boundaries in the original space. The solution to the support vector machine problem with a linear kernel involves calculation of inner product of observations, given by

$$K(X_i, X_j) = \sum_{k=1}^p x_{ik}x_{jk} \quad (5.13)$$

where K is the kernel function which quantifies similarity between two observations X_i and X_j , each having p predictor variables. The equation (5.5) is the linear kernel function that quantifies similarity between two observations using a Pearson's correlation coefficient form.

5.2.6.4 SVM using non-linear kernel function

The SVM kernel function can be extended to a higher order polynomial function as,

$$K(X_i, X_j) = \left(1 + \sum_{k=1}^p x_{ik}x_{jk} \right)^d \quad (5.14)$$

where d is the degree of the polynomial [178]. Using a higher order polynomial function provides a flexible decision boundary since it tries to fit the classifier in a higher dimensional space than the original feature space (Figure 5.6 left).

The features of observation X in higher dimensional space are calculated as:

$$f(X) = \beta_o + \sum_{i=1}^S \alpha_i K(X, X_i) \quad (5.15)$$

where $f(X)$ is the new feature vector of X . The feature vector can be input into equation 5.11 to calculate the predicted class of observation X . The parameter α is a vector of coefficients, β_o is a constant parameter and S are the total number of training set observations.

Another choice apart from polynomial kernels is the radial basis kernel (Figure 5.6 right) which takes the form,

$$K(X_i, X_j) = \exp\left(-\gamma \sum_{k=1}^p (x_{ik} - x_{jk})^2\right) \quad (5.16)$$

where γ is a positive constant. The radial basis kernel calculates the Euclidean distance between the observations, and if the distance is large then kernel term will be small.

The hyperparameter γ controls a trade-off between errors due to the bias and the variance in the model. If a test observation X^* is far from a training observation X_i (based on Euclidian distance), then the kernel term will be small and X_i will play no role in predicting the class of X^* . This suggests that radial kernels have a local behaviour, and training observations have a larger impact on predicting the class of test observations closest to them (Figure 5.6 right). A further advantage of using kernels in comparison to support vector classifiers is that it is computationally efficient.

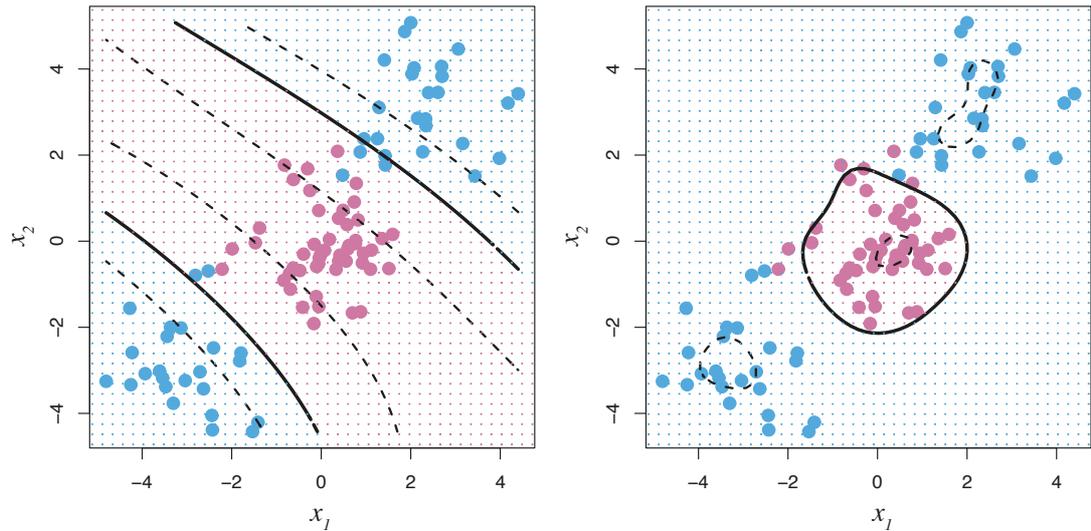


Figure 5.6 Polynomial kernel SVM and radial kernel SVM

Adapted from James et al. [178]. The figure shows an example of polynomial kernel SVM, with degree=3 (left) and radial Kernel SVM (right).

5.2.6.5 *k*-fold Cross-validation

Resampling-based strategies have been applied to estimate the error rate of prediction models [178, 189]. The two widely used resampling strategies are bootstrapping and cross-validation [178]. Bootstrapping proceeds by randomly re-sampling a subset of data multiple times, with replacement, to generate several prediction models. The performance of these models is assessed on observations that were not included in the bootstrapped subset [178]. (This is the approach taken in RF algorithm). The cross-validation approach, on the other hand, proceeds by re-sampling data subsets multiple times without replacement to generate prediction models.

The *k*-fold cross-validation approach divides training data into *k* equal subsets referred to as *k*-folds. The first fold is held out and referred to as the validation set. The prediction model is generated by training on the remaining *k*-1 folds. The performance of this prediction model is assessed on the validation set. This process is repeated *k* times and each time a different set of observations are assigned to the validation set. The overall performance estimate of the prediction model after *k* iterations is calculated as,

$$CV = \frac{1}{k} \sum_{i=1}^k Perf_i \quad (5.17)$$

where CV is the cross-validation performance estimate across k iterations and $Perf_i$ is the performance estimate on the validation set at the i^{th} iteration. The advantage of using the k -fold cross-validation approach is that it is not prone to overfit and it gives a reliable estimate of performance of the model during training [178].

5.2.6.6 Application of SVM to LMC

SVM algorithm was applied on the training set using linear and non-linear kernel (polynomial and radial basis) functions to generate prediction models. The hyperparameters of these kernel functions were tuned via grid-search using a 10-fold cross-validation approach. The cross-validation sets were generated using stratified sampling which maintained the proportion of classes across the training and validation sets. Kappa measure was used to estimate the performance of the model in 10-fold cross-validation. The hyperparameter values which showed the best performance on the cross-validation set were selected in the final model. The predictions made by the final model on the left-out observations were used to estimate model performance during training. The final model was applied on the test set observations and performance measures were calculated. The analyses were performed using R-packages *Caret* [180] and *kermlab* [190].

5.2.6.7 Dealing with class imbalance using under sampling

As for the RF, the SVM model was generated using the under-sampling scheme for the majority class. This was done at each cross-validation iteration using *down-sampling* function in *Caret* R-package [180]. The model was generated using a linear kernel, and cross-validation performance was assessed using Kappa index. The performance measures on the training and test sets were calculated as described in the previous section. The analyses were performed using R-packages *Caret* [180] and *kermlab* [190].

5.2.7 Calculating agreement between RF and SVM models

The agreement between the RF model and SVM model was calculated using the Cramer's V statistic (refer to section 3.2.6). The predictions of these two models were compared on the test set and the model with higher performance was further explored.

5.2.8 Generating permutation-based baseline models

Permutation-based classification models were developed using random datasets generated after shuffling of class labels in the training set over 100 iterations. The best performing algorithm among RF and SVM was used to generate prediction models using each of these random datasets and the model performance was assessed on the test set (non-shuffled). The aim of these analyses was to produce average classification performance metrics from 100 random models and use them as baseline to appraise the real value of similar metrics obtained from the unshuffled training dataset.

5.3 Results

5.3.1 Training and test sets in LMC

The gene expression dataset was divided into training and test sets after selecting survival cut-off at 6 years. The training set contained 70% of the observations and test set contained 30% of the observations. As expected, the training and test sets showed no significant differences in clinico-histopathological characteristics of the observations (Table 5.2).

5.3.2 RF applications to predict outcome

5.3.2.1 RF with unbalanced class design

The RF algorithm sampled 20,000 bootstrap datasets from the training set and generated 20,000 decision trees that predicted outcome. The performance was assessed on OOB samples for each bootstrap dataset. Initially, the overall OOB error decreased with the addition of more decision trees, however the error stabilised after generating approximately 2500 trees (Figure 5.7). The OOB error was much lower for majority class observations (class 0, survivors) in comparison to minority class observations (class 1, non-survivors). Since the training set was unbalanced, the prediction model based on this set will be referred to as the RF model with unbalanced class design.

Table 5.2 Comparing histopathological differences between the training and test set

Characteristics n(%)	LMC dataset		P
	Training n=368(%)	Test n=157(%)	
Sex: male , 226 (43)	159 (43)	67 (43)	0.98
Site: limbs , 222 (58)	155 (42)	67 (43)	0.98
Ulceration status: yes 176 (34)	127 (34)	49 (31)	0.53
AJCC stage: I 181 (35)	124 (34)	57 (37)	0.33
II 266 (51)	194 (53)	72 (47)	
III 74 (14)	48 (13)	26 (17)	
Breslow (mm) median =2.3 (range= (0.3,20))	2.3 (0.3, 18)	2.1 (0.5, 20)	0.20
Age (years) median=57.4 (range= (20,81))	58 (20, 81)	57 (20, 76)	0.96
TILS: Absent 59 (15)	37 (15)	17 (15)	0.83
Non-Brisk 260 (72)	181 (72)	79 (70)	
Brisk 49 (13)	32 (13)	17 (15)	

When the RF model with unbalanced class design was applied on test set observations, the model had similar prediction performance when compared with the training set (Table 5.3). The model performance parameters indicated that the RF model had high specificity (0.95) and low sensitivity (0.23) on test set (Table 5.3). There was no difference between the NPV (0.71) and PPV (0.71) of the RF model. The kappa index (0.22) for the model suggested low agreement between actual and predicted classes (Table 5.3).

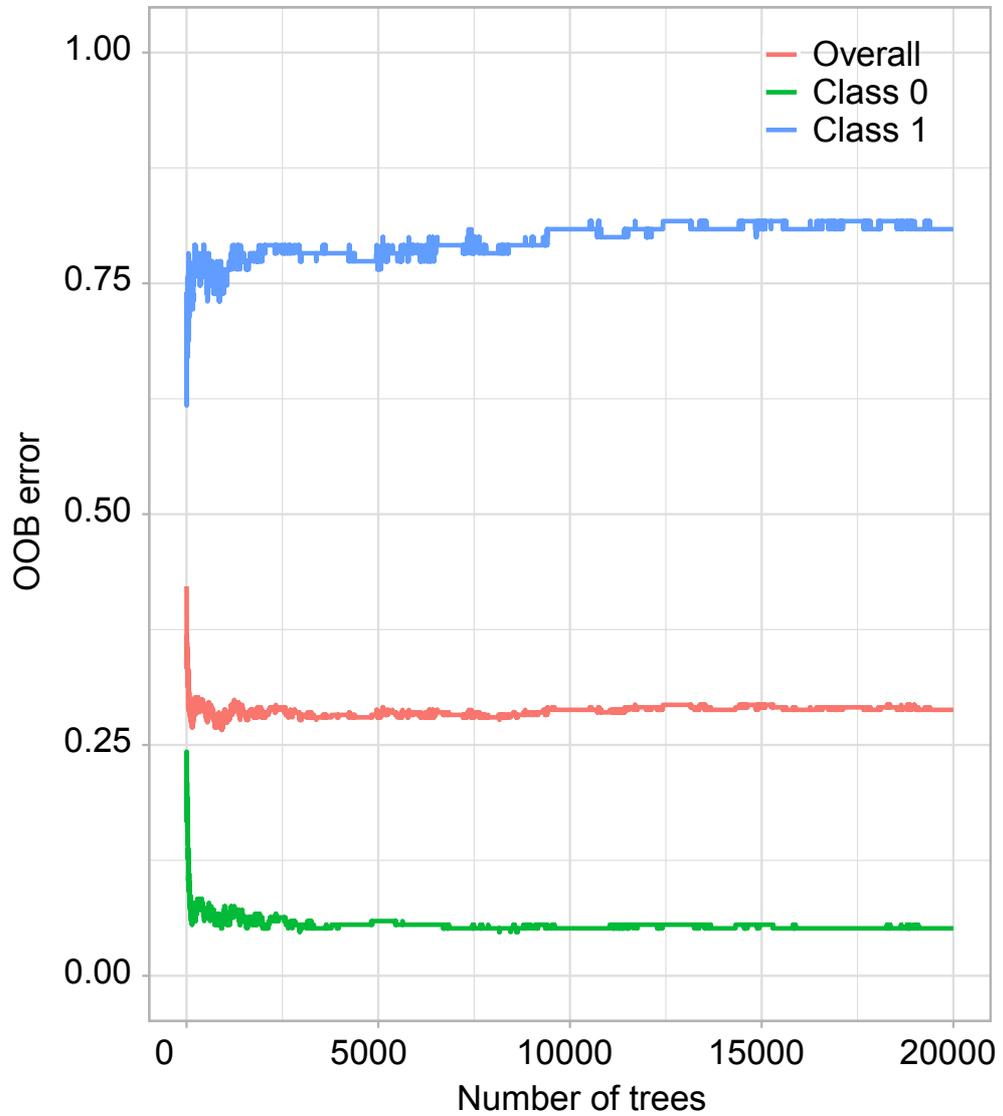


Figure 5.7 RF model using unbalanced class design

RF OOB error estimate as the number of decision trees were added to the forest. The red line is the overall OOB error, green line shows the error rate for class 0 observations, and blue line shows the error rate for class 1 observations.

Table 5.3 Summary of the RF model performance on training and test set

Sens. is the sensitivity, spec. is the specificity, PPV is the positive predictive value and NPV is the negative predictive value.

Class design	Set	Sens.	Spec.	PPV	NPV	kappa
Unbalanced	Training	0.19	0.95	0.62	0.72	0.17
	Test	0.23	0.95	0.71	0.71	0.22
Balanced	Training	0.44	0.77	0.47	0.75	0.22
	Test	0.54	0.86	0.65	0.79	0.41

5.3.2.2 RF with balanced class design

As shown above, in the unbalanced class design, the RF model predicted majority class (class 0) observations more accurately than minority class (class 1) observations. To overcome this, RF model was re-run under-sampling the majority class at each iteration to achieve a balanced class design. The new model showed a much lower OOB error for class 1 observations in comparison to the previous model with unbalanced design (Figure 5.7 and Figure 5.8). However, this improvement came at the expense of an increased error rate for class 0; the overall error rates in the two designs were comparable (Figure 5.7, Figure 5.8). As a result of the shift in error rates corresponding to class 1 and class 0, the RF model in the balanced design had an higher sensitivity (0.54) and lower specificity (0.86) in comparison to the RF model with unbalanced design (sens.= 0.23, spec.= 0.95) (Table 5.3). Compared with the unbalanced design, the kappa index was much higher for the balanced design (kappa= 0.41), and this further indicated that the balanced design RF model had better agreement between predicted class labels and actual class labels on the test set (Table 5.3).

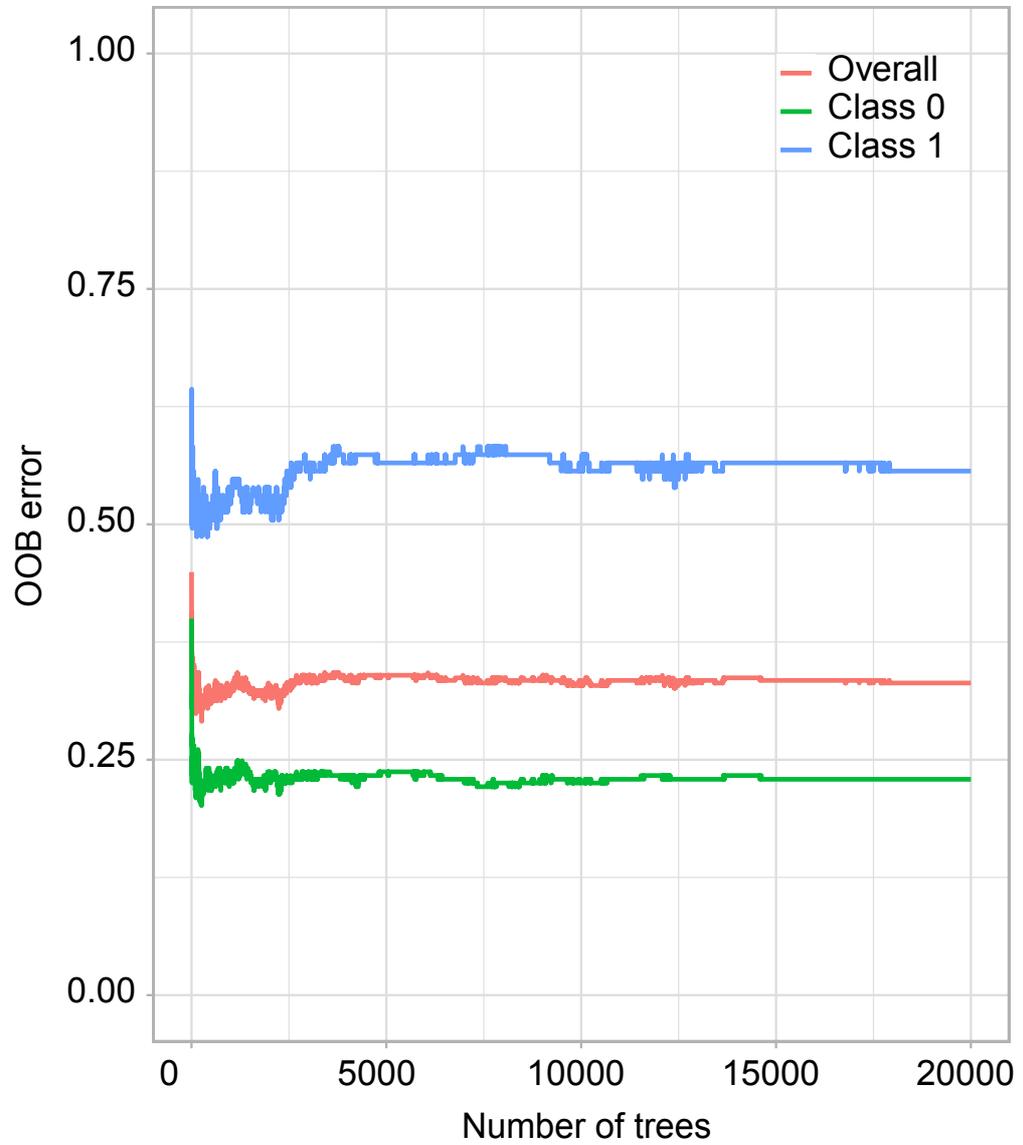


Figure 5.8 RF model using balanced class design

RF OOB error estimate using balanced class design. The red line is the overall OOB error, green line is the error rate for class 0, and blue line shows the error rate for class 1 observations.

5.3.3 SVM applications to predict outcome

The SVM prediction models were generated using linear and non-linear kernel functions.

5.3.3.1 SVM application using linear kernel

The SVM algorithm using a linear kernel generated a model that achieved similar classification performance on both the training and test sets. The kappa index increased with the hyperparameter C and stabilised after C reached 0.0005. The SVM model with the highest average kappa index across the 10-fold cross validation during training was selected and was applied to the test set.

The SVM model had comparable performance measures on the training (sensitivity=0.43, specificity=0.83, kappa=0.27) and test sets (sensitivity=0.46, specificity=0.80, kappa=0.27) (Table 5.4). Like the RF model, the SVM model had a high specificity (0.80) and NPV (0.75), and a low sensitivity (0.46) and PPV (0.53) (Table 5.4).

The initial grid search for hyperparameter C was done for a large range of values. It was hypothesized that fine-tuning the grid search around the above selected C value may further improve the model performance. Fine-tuning the hyperparameter C value did not improve the performance of SVM model as this model achieved comparatively similar performance to the previous SVM model (Table 5.4). As before, the fine-tuned SVM model had a higher specificity (0.90) and NPV (0.74), than the sensitivity (0.34) and PPV (0.60) on the test set.

Table 5.4 Summary of SVM model performance on training and test set

Sens. is the sensitivity, spec. is the specificity, PPV is the positive predictive value and NPV is the negative predictive value.

SVM	Set	Sens.	Spec.	PPV	NPV	kappa
Wide range C	Training	0.43	0.83	0.53	0.76	0.27
	Test	0.46	0.80	0.53	0.75	0.27
Fine-tuned C	Training	0.38	0.86	0.56	0.75	0.26
	Test	0.34	0.90	0.60	0.74	0.27

5.3.3.2 SVM application using non-linear kernel

Overall, the SVM algorithm with non-linear kernel functions (radial and polynomial) performed poorly in predicting outcome (Table 5.5). The SVM radial kernel function generated a majority class model with low sensitivity (0) and high specificity (1) and a kappa index of 0 in both training and test sets (Table 5.5). The SVM algorithm with a polynomial kernel function generated a model with low sensitivity (0.15) and high specificity (0.93) (Table 5.5) in the test set with a kappa index of 0.10, a PPV of 0.53 and NPV of 0.69 (Table 5.5).

Table 5.5 Summary of SVM non-linear kernel functions on training and test set

Sens. is the sensitivity, spec. is the specificity, PPV is the positive predictive value and NPV is the negative predictive value.

Kernel	Set	Sens.	Spec.	PPV	NPV	kappa
SVM radial	<i>Training</i>	0	1	NaN	0.66	0
	<i>Test</i>	0	1	NaN	0.69	0
SVM poly	<i>Training</i>	0.24	0.95	0.68	0.73	0.23
	<i>Test</i>	0.15	0.93	0.53	0.69	0.10

5.3.4 SVM using balanced class design

The SVM models described above were generated using unbalanced classes. Like RF, it was hypothesised that under-sampling the majority class to achieve a balanced class design may further improve the model performance.

Unlike RF, using balanced class design did not generate a better performing SVM model. This newly generated SVM model had similar performance to the SVM model with unbalanced class design. The SVM model with balanced design had similar kappa index on the test set (kappa=0.26) when compared with SVM model with unbalanced design (kappa=0.27) (Table 5.4, Table 5.6). However, the new model had improved sensitivity (0.61) in comparison to the model with unbalanced design (sensitivity=0.46) on the test set (Table 5.4, Table 5.6), at the expense of a decreased specificity (0.67 in the balanced design, 0.80 in the unbalanced design) (Table 5.4, Table 5.6).

Table 5.6 Summary of SVM linear kernel model after under sampling the majority class

Sens. is the sensitivity, spec. is the specificity, PPV is the positive predictive value and NPV is the negative predictive value.

Set	Sens.	Spec.	PPV	NPV	Kappa
Training	0.63	0.66	0.47	0.80	0.28
Test	0.61	0.67	0.48	0.78	0.26

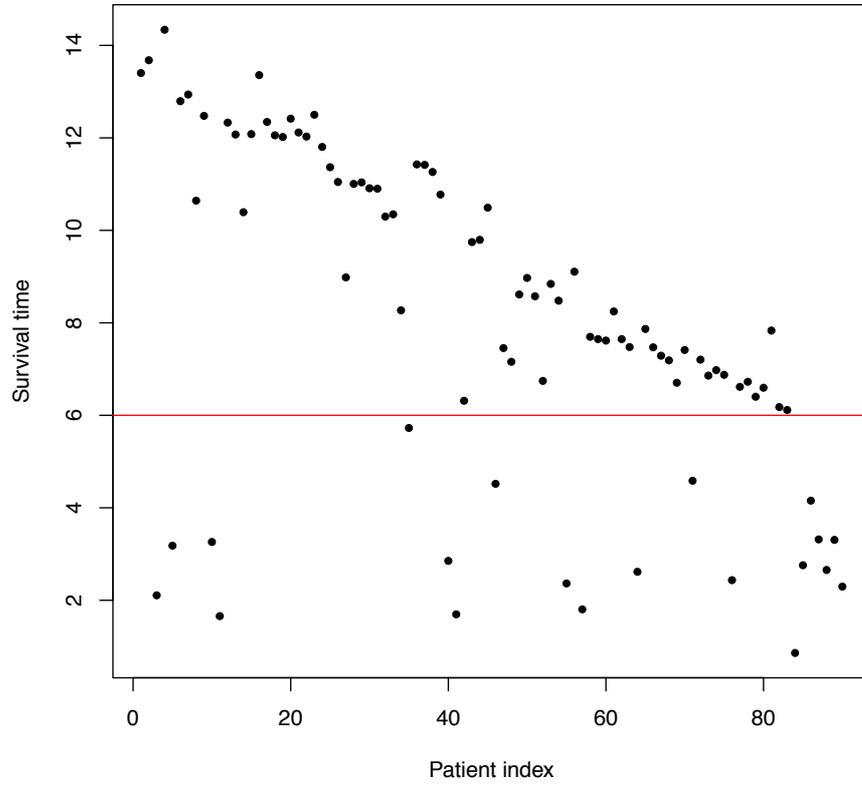
5.3.5 Comparing RF and SVM predictions

The Cramer's V statistic (0.71) indicated good agreement between predictions from the RF and SVM classification models (Table 5.7). The RF and SVM models were generated using the balanced class design approach (refer to 5.3.2.2 and 5.3.4). All the observations classified by RF in class 1 were also classified in same class by SVM model (Table 5.7). Seventy nine percent of observations classified in class 0 by RF were classified also in the same class by SVM and rest 21% of observation showed different class labels (Table 5.7). Overall both the methods made consistent predictions on test set but the RF model (kappa=0.41) had slightly better performance when compared with SVM model (kappa=0.27) (Table 5.6).

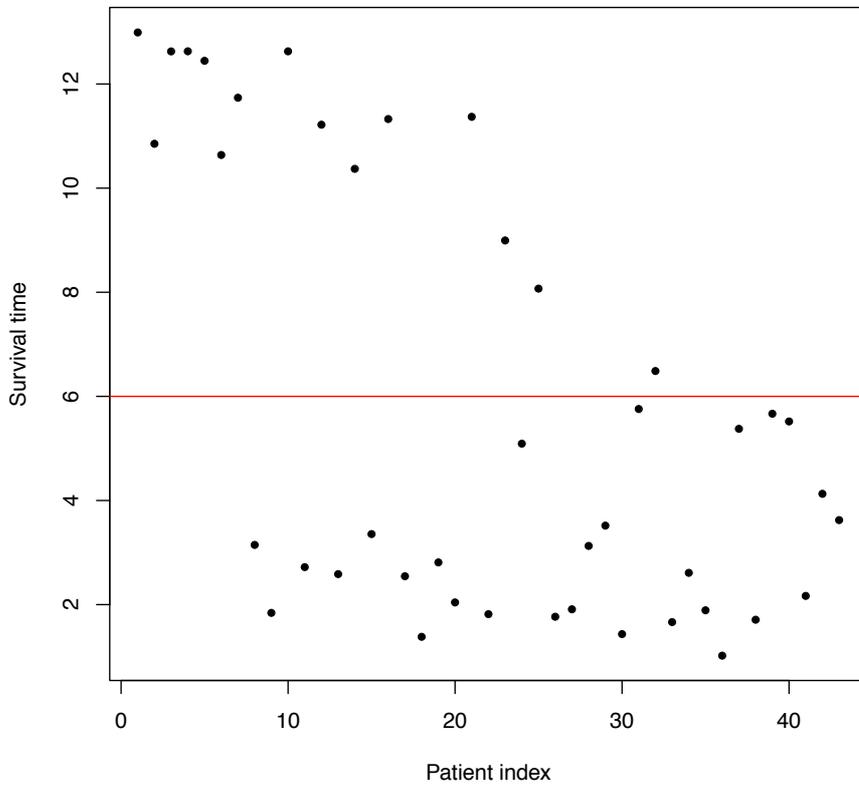
The RF and SVM models classified 90 observations as class 0 (survivors, n=90) and 78% of these observations had survival time greater than 6 years (Figure 5.9A). The observations that were classified as class 1 (non-survivors, n=43) by both the models, 65% experienced death from melanoma within 6 years (Figure 5.9B).

Among 24 disagreeing observations, 20 had survived for more than 6 years (Figure 5.9). The RF model predicted these 24 observations as survivors and the prediction was correct for 20 cases and incorrect for just 4 cases. SVM model on the other hand predicted these 24 observations as non-survivors and the prediction was correct only for 4 observations and was incorrect for 20 observations. Overall, these results highlight that RF model had higher classification performance than the SVM model on the test set.

A Class 0: agreements between RF and SVM



B Class 1: agreements between RF and SVM



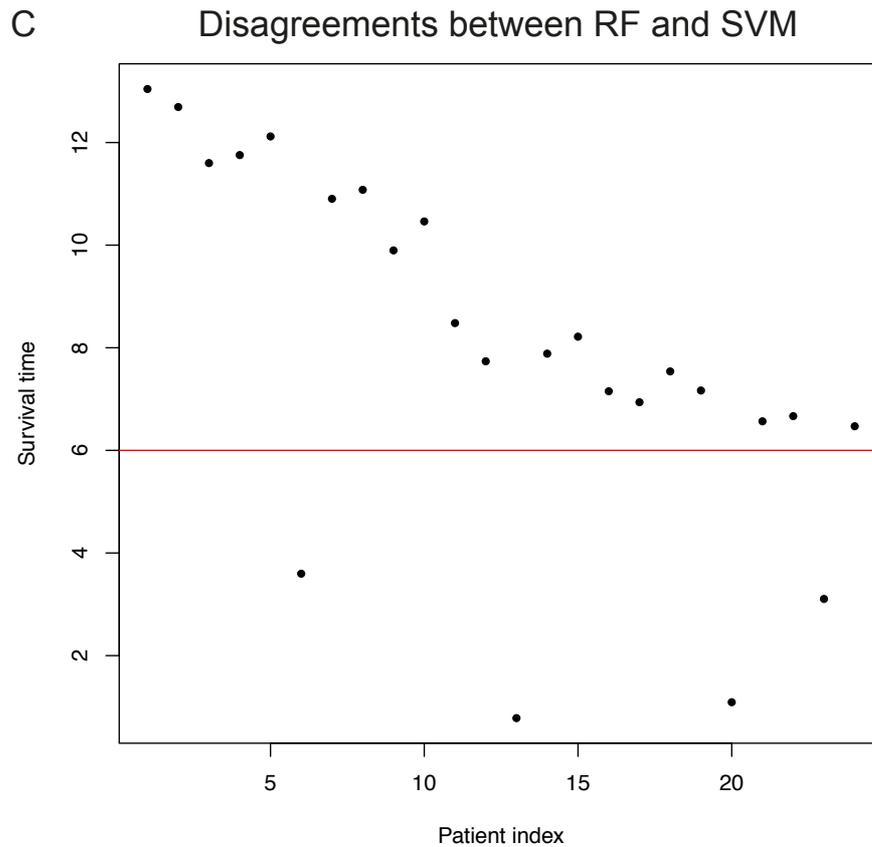


Figure 5.9 Comparing RF and SVM predictions

Survival time of the observations classified in the same classes by RF and SVM, i.e. (A) class 0, (B) class 1. (C) Survival time of the observations which had disagreement between their predictions by RF and SVM models. The red line indicates the survival cut-off at 6 years.

Table 5.7 Comparing overlap between RF model and SVM model prediction on test set

Test set		RF	
		Class 1 n(%)	Class 0 n(%)
SVM	Class 1	43 (100)	24 (21)
	Class 0	0	90 (79)

5.3.6 Permutation-based RF model

As shown in previous section, the RF algorithm had higher performance than the SVM algorithm. Therefore, RF algorithm was selected and applied to 100 random datasets generated by permuting class labels in the training set. As in the real data, the prediction in these random datasets showed a higher specificity than sensitivity (Figure 5.10). Over 100 iterations, the median specificity, sensitivity, and kappa index were respectively 0.89, 0.09 and -0.01 (Figure 5.10).

The non-permuted RF model (refer to section 5.3.2.2) had higher sensitivity and kappa in comparison to the permuted models. Across the 100 permuted datasets, none of the models showed a kappa index and sensitivity value higher than 0.41 and 0.54 respectively (Figure 5.10). However, the median specificity of the permuted model was sometimes higher than the non-permuted RF model. This analysis therefore shows that specificity is a poor performance metric for the analysis of our dataset; sensitivity and the kappa index are good metrics as they are close to 0 on average in random data.

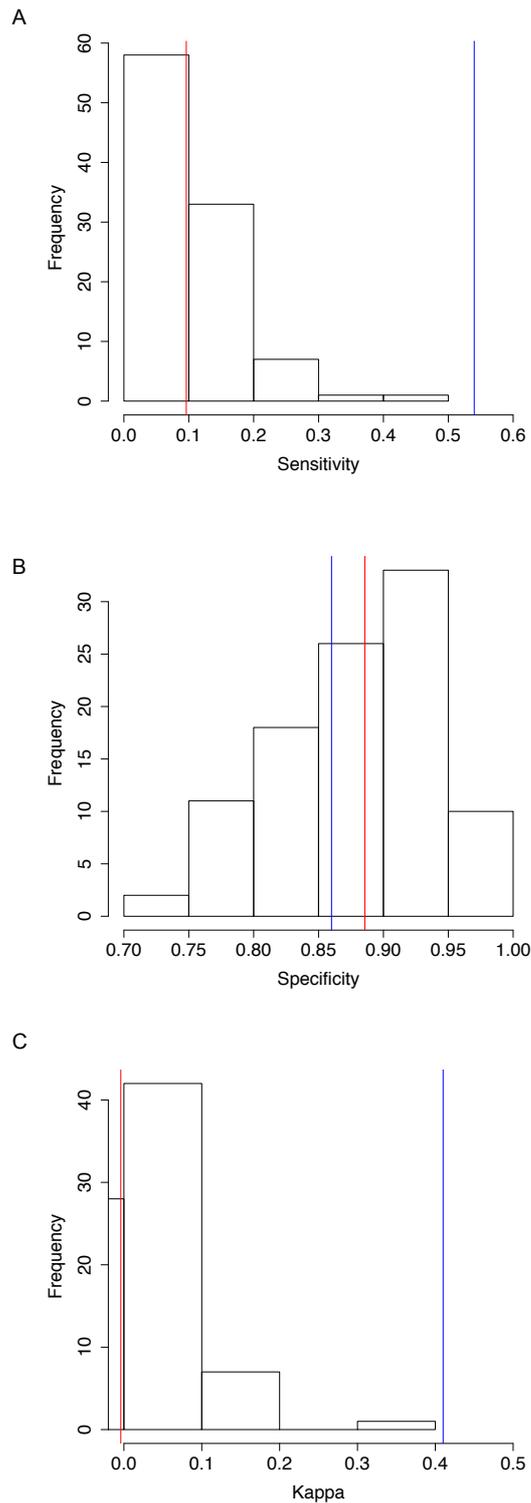


Figure 5.10 Performance summary of the permuted model

Permuted model performance parameters (A) sensitivity (B) specificity and (C) kappa on the test set over 100 iterations of permutation based RF model. The red line is the median value of performance measure for permuted models and blue line represents the observed performance measure for the non-permuted RF model.

5.4 Discussion

Machine learning applications have shown great promise in improving diagnosis and predicting prognosis of cancer patients [109, 111, 112, 168, 191, 192]. Recently, it was shown that a machine learning approach achieved better diagnostic performance than a group of trained dermatologists in detecting melanoma from dermoscopic images [169].

The main aim of this chapter was to test whether transcriptomic information from the tumours of patients from LMC can predict clinical outcome (death from melanoma within 6 years). To this end, the LMC data were partitioned into a training and test set to respectively develop and evaluate two machine learning classification methods, RF and SVM. The random data splitting meant that the two obtained subsets had no difference between them in terms of clinical characteristics. The prediction outcome was defined as surviving up to 6 years, as the majority of patients who have melanoma metastasis and die from it, normally do so very early after the initial diagnosis [193].

One major advantage of using the RF is that the use of OOB samples to evaluate performance in the training step yields an unbiased estimate of the model performance [171]. The unbiasedness means that the results observed in the training dataset are reproduced in the test set. This was confirmed in the LMC as the RF model showed similar performance measures in both training and test sets (Table 5.3). The RF model predicted the majority class observations better than the minority class observations, i.e. had a higher specificity than the sensitivity. This class-imbalance is a well-recognized problem in supervised classification [173, 183, 194]. Several methods were proposed to overcome this issue and have shown some improvement in performance of machine learning algorithms [173, 182, 183, 194-196]. Among the available methods, over-sampling the minority class and synthetic minority over-sampling technique are more prone to overfit the training data, and hence the majority class was under-sampled (less prone to overfitting) to achieve balanced class design. Applying an under-sampling approach resulted in improved sensitivity and kappa index of RF model. The new RF model with balanced design had a lower OOB error for the minority class in comparison to the unbalanced design.

The SVM approach identifies a hyperplane using linear or non-linear kernel functions [184]. The hyperparameters in SVM were tuned using 10-fold cross validation and the selected parameter value was used in the final SVM model [186]. In LMC dataset, the SVM with linear kernel outperformed SVM with non-linear kernel functions. In particular, the SVM model with a radial kernel showed no utility at all as it classified

all observations in the majority class (maximum specificity, zero sensitivity), which is equivalent to the best guess without using the available transcriptomic data. The linear SVM outperformed the non-linear and its performance on the training set was comparable to the test set. The initial grid search for hyperparameter C of the linear kernel function was done on a large interval and its fine-tuning on the grid search around the previously selected C value did not improve prediction performance of the SVM model.

Like the RF model with unbalanced design, the SVM model also had a much higher specificity than sensitivity. Undersampling the majority class to overcome class imbalance brought an improvement to the sensitivity of the SVM model but this improvement came at a greater cost to the specificity. The kappa value however remained unchanged, suggesting a similar level of agreement overall between actual and predicted class labels with and without undersampling.

Shi et al. compared several machine learning methods by generating more than 30,000 models using different microarray datasets and assessing their performances on independent test sets [170]. They found none of the methods to consistently outperform all the others. The prediction performance was observed to be mainly dependent on the quality of the microarray data (i.e. pre-processing) and prediction endpoint variable. Overall, in good quality microarray datasets, the majority of methods made similar predictions [170]. In our analysis of the LMC, the RF and SVM models made similar predictions on the test dataset. However, for the observations with conflicting predictions, the RF model showed higher classification accuracy and classified most of the observation into the actual classes.

Transcriptomic information of a patient's tumour has been demonstrated to be of prognostic significance [105, 106, 114, 116, 197]. To assess whether the generated RF model has captured truly prognostic information from the dataset, several null models were generated after randomly shuffling training set observations. As expected, the RF model had consistently a better performance (higher sensitivity and kappa index) than the permuted RF models. The permuted models had high specificity and low sensitivity and kappa, which suggests that these models are likely to misclassify non-survivors as survivors. It also hints that higher sensitivity and higher kappa are good measures to assess the performance of prediction models in LMC.

Overall, these results provide evidence that machine learning methods can predict prognosis of a patient using transcriptomic information from the patient's tumour. The limitation of these approaches is that although training and test sets were scaled

independently, some dependency still lies between the two sets as the samples were profiled using the same technique and were pre-processed together. In next chapter, these issues will be addressed by pursuing model validation on an independently generated dataset.

In summary, RF and SVM algorithms generated models that predicted the patient prognosis in the LMC. Undersampling to achieve balanced class design further improved the performance of RF but not the SVM model. Overall, the RF and SVM made comparable predictions on the test dataset. In the next chapter, only the RF will be further explored, both because this method performed better as a classifier and because it provides an estimate of the importance of each gene in the model, while SVM works as “black box” in nature and generates very limited output. In what is presented so far, the RF model was based on gene expression data alone. The next step is to combine gene expression data and clinical information to see if an improvement is possible. The performance of the combined model will be validated on an independent dataset from Lund, Sweden.

Chapter 6

Combining clinical information and gene expression for predicting outcome and independent validation

The objectives of this chapter are:

Objective 11: To generate a RF model after combining clinical information and gene expression data and compare its performance with a clinical information based RF model

Objective 12: To generate a refined RF model by performing variable selection
Hypothesis: The refined RF model using top predictor genes will predict prognosis similarly to the model using all genes

Objective 13: To validate the prognostic value of the refined RF on an independent dataset of primary melanoma from Lund, Sweden

Objective 14: Biological interpretation of the refined model using pathway enrichment analysis

6.1 Introduction

Machine learning approaches like Random Forest (RF) and Support Vector Machines (SVM) are black-box in nature; they allow the user a very limited control over the steps involved in making predictions. The RF algorithm partly overcomes this issue by calculating a Gini index for each predictor variable in the model (refer to 5.2.5.2). This importance measure can be used to assess the contribution of each variable included in the model. The prediction models previously generated using RF with a balanced class design will be further explored in this chapter (refer to 5.3.2.2).

In clinical settings, AJCC staging system is a widely used procedure for predicting prognosis and deciding treatment protocols [44]. However it predicts outcome accurately only for 68% of the population (AUC:0.68) [114]. Therefore, it is of interest to develop better prognostic biomarkers which can complement AJCC staging in clinical settings. Other clinical variables, such as age at diagnosis, sex and body site of tumour have been shown to be associated with melanoma prognosis [42]. These variables independently predicted melanoma-specific survival in the LMC (refer to

4.3.3.1). Hence, a baseline RF model was generated by jointly analysing the clinical variables described above.

In this chapter, a RF model will be constructed after combining clinical variable information and gene expression data. The performance of this model will be compared with the baseline clinical variable based RF model. To maximise clinical utility, a refined model with a limited number of predictor variables will be generated. The refined model will be validated by comparing its performance with the baseline model in an independent cohort of primary melanoma from Lund, Sweden (refer to section 4.2.5.1). To get biological insights into the decision-making process of the refined RF model, the predictor genes included in this model will be tested for their overrepresentation in curated biological pathways. Finally, to compare the unsupervised clustering signature with the supervised classification signature, the six LMC classes defined in chapter 4 will be compared with the refined RF model classification.

6.2 Methods

6.2.1 Using clinical information to generate RF model

The RF model was generated using clinical variable information of melanoma tumours in LMC. The clinical variables included in the model were AJCC stage, age at diagnosis, sex and body site of tumour. AJCC stage was treated as an ordered categorical variable, age at diagnosis was treated as a continuous variable, sex was coded as 0 for females and 1 for males, and body site of tumour (limbs, trunk, head and neck, rare) was converted into a series of dummy variables using *model.matrix* function in the R *base* package. The dummy variables from the main variable was generated by assigning its categorical levels a value of 1 if true and 0 if false. In this case, from site of tumour, four new variables were generated named as limbs, trunk, head and neck, and rare. These variables were initialised with a value of 1 if true for a patient or 0 if not.

As described previously, the RF model was generated using a balanced class design approach (refer to 5.2.5.5), the number of trees was set to 500 and the number of variables selected for consideration at each node (*mTry*) was 2. The RF analysis was performed using R-packages *randomForest* and *Caret* [180, 181]. The error for OOB samples was plotted using R-package *ggplot2* [130]. The performance of this

model was assessed on the test set using previously described performance measures (refer to 5.2.4).

Since clinical variables Breslow thickness and ulceration are included in the definition of AJCC stage [44], these variables were not included in the model. Four observations in the dataset had missing information on AJCC stage, and these samples were excluded. TILs and mitotic rate had more than 21% (n=115) and 14% (n=74) of missing values respectively. Hence, these variables were not included in the analysis.

6.2.2 Removing clinical information from the gene expression

To remove clinical variable information from gene expression data, a linear regression approach was used with the gene expression as the outcome variable (equation (6.1)). The regression approach assumes a linear relationship between the outcome variable and the predictors. The linear regression line is represented as:

$$Gene_i = \beta_{i0} + \beta_{i1}x_1 + \beta_{i2}x_2 + \dots + \beta_{ip}x_p + \varepsilon_i \quad (6.1)$$

where $Gene_i$ is a vector of expression values of i^{th} gene (outcome variable) and x is a vector of p predictor variables (e.g. x_1 for AJCC stage, x_2 for age, x_3 for sex). β s are the parameter coefficients for predictor variables and are different for different genes. β_0 is the intercept and ε_i is the error term in the model or the residual, i.e. the difference between the actual outcomes and those predicted by the model. All parameters are estimated using a least squares approach which tries to minimise the sum of squared errors. Residuals are assumed to be normally distributed with mean zero and are independent of the predictor variables.

The linear regression was performed using clinical variables as independent predictors and gene expression data as the outcome variable. The residual values for each gene from the linear regression were used as input data in the RF algorithm. A new RF analysis was conducted using the same parameters as described previously (refer to the 5.2.5.5), and the performance measures were calculated for both training and test sets (refer to 5.2.4).

6.2.3 Combining clinical information and gene expression

The RF model was generated after combining clinical variable information and gene expression residuals for LMC patients. The clinical variables used in the previous section were included in this analysis. During training, at each node the predictor

variables were selected such that the selected variable pool included all the clinical variables along with randomly selected genes. The variable from the selected pool that gave the best split at the node was chosen to make the split into two daughter nodes. The number of randomly selected genes was decided based on the value of *mTry* parameter which was set to 117 (square root of total number of gene). The number of trees was fixed at 20,000.

In keeping with previous analyses, the RF model was generated using a balanced class design by selecting an equal proportion of samples from both classes at each iteration (refer to 5.2.5.5). During training, the error rate of the RF model was calculated on OOB samples, and final performance was assessed on the test set. This analysis was done using R-packages *Caret* [180] and *ranger* [198]. The *ranger* R-package contains *ranger* function, which performs fast implementation of the RF algorithm and also forces selection of a desired set of variables at each node of the decision tree (clinical variables in this instance).

6.2.4 Refining the RF model via variable selection

The RF model developed in the previous chapter was generated using ~13,000 genes (refer to 5.2.5.5). In clinical settings however, it may not be practical to use a biomarker/signature based on such a large number of genes. To mitigate this, smaller gene sets of varying size were selected based on Gini index (Figure 6.1). The top *n* (10, 50, 100, 200, 400, 800, 1600 and all genes) ranked genes were selected and combined with the clinical variables to generate several RF models (Table 6.1). These RF models were built using the same training set, and the performances of these models were compared on the test set. The RF model with best performance was retained as the final refined model. The analysis was done using R-packages *Caret* [180] and *ranger* [198].

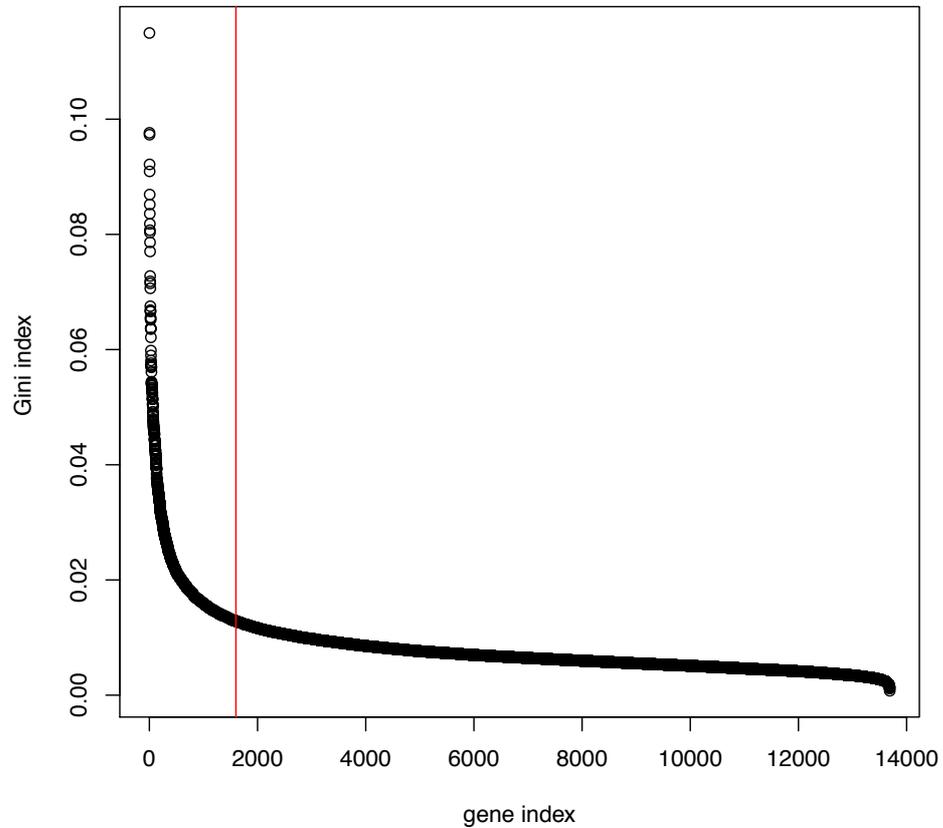


Figure 6.1 Gini importance index

Plot showing Gini importance measure for each gene in the RF model, x-axis is the gene index and y-axis is the Gini index for each gene. The red line shown the importance measure of the 1600th gene.

Table 6.1 Refining the RF model by selecting variable number of genes, $mTry$, and number of trees

Number of genes	$mTry$	Number of trees
10	3	100
50	7	700
100	10	1000
200	14	2000
400	20	4000
800	29	5000
1600	40	5000

6.2.5 ROC analysis for the refined RF model

The prognostic value of the refined RF model was compared with the baseline RF model (based on clinical variables alone) using ROC analysis. Previously, for each sample logistic regression was used for calculating probabilities of MSS up to 6 years (refer to 4.2.6). In this analysis, for each sample, the proportion of votes received that classified in class 0 in the refined RF and baseline RF models were used in place of probability of surviving up to six years. The comparison was drawn for four RF models described as follows:

- 1) Baseline RF model using clinical variable alone;
- 2) RF model based on gene expression alone (refer to 5.2.5.5);
- 3) Combined RF model;
- 4) The final refined RF model.

ROC curves were plotted for each of these models and AUCs were calculated. The test of AUC improvement in nested models was performed using DeLong's test (refer to 4.2.6). These analyses were performed using R-packages *ROCR*, *plotROC*, *pROC* and *ggplot2* [130, 149, 150, 199].

6.2.6 Application of the refined RF model on independent dataset

The prognostic value of the refined RF model was validated on an independent dataset of primary melanomas from Lund, Sweden (refer to 4.2.5.1). Clinical variables including AJCC stage, sex, age at diagnosis and tumour site were provided upon request by Prof. Göran Jonsson and Dr. Martin Lauss (collaborators at Lund University). Some of the tumour samples had missing AJCC stage value (n=76). For a subset of these (n=49), the AJCC stage was calculated manually based on the guidelines of the 7th edition of AJCC staging system [44], using the records of Breslow thickness and ulceration status of tumours. This AJCC edition had been used for all other samples in that dataset, as was the case also in the LMC.

The tumour samples for which AJCC stage could not be calculated were excluded from the analysis (n=37). Finally, the dataset had gene expression values of 186 tumours. Since cause of death was not recorded in the Lund cohort, death from any cause (deaths due to melanoma and other causes) and melanoma relapse before 6 years were used as outcomes. The cut-off at six years was consistent with the LMC, and samples were classified as class 0 or class 1, henceforth referred to as actual class labels.

In keeping with previous chapters, gene expression data from the Lund cohort was standardised to give each gene mean 0 and standard deviation 1. The refined RF model was applied to the Lund dataset similarly to how it was applied to the LMC test set. The refined RF model stored after previous analyses was re-loaded in the R environment for application to the Lund dataset, and Lund samples were classified into class 0 or class 1. The predicted class labels were compared to the actual class labels and performance measures were calculated (refer to 5.2.4). The baseline RF model (refer to 6.2.1) was also applied on the Lund dataset, and the performance was compared with the refined RF model (based on gene expression and clinical variables). As before, these analyses were done using R-package *ranger* [198].

The prognostic value of the refined RF model was compared with the baseline RF model using ROC analyses. Similar to section 6.2.5, the proportion of votes received classifying in class 0 was used as a probability measure. The outcome measures were death from any cause and melanoma relapse at 6 years. The ROC curves for both the RF models were plotted, and AUC was calculated for each ROC curve. The analysis was performed using R-packages *ROCR*, *plotROC*, *pROC* and *ggplot2* [130, 149, 150, 199]. The statistical comparison of AUC in ROC was performed using DeLong's test (refer to 4.2.6).

6.2.7 Pathway enrichment of predictor genes in the refined RF model

Pathway enrichment analysis was performed for the selected predictor genes in the refined RF model. The genes were input into ReatcomeFIViz, a tool designed to identify associated biological pathways. The pathways were visualised using R-packages *wordcloud* [200] and *RColorBrewer* [201].

6.2.8 Comparison between the LMC classes and the refined RF model predictions

In this thesis, the two final gene signatures developed are: the refined LMC class signature from the unsupervised clustering (refer to 4.3.5.1) and the refined RF model signature after combining clinical variable information and gene expression data from the supervised classification model. To compare these signatures, their classification was tabulated against the actual outcome. During supervised classification, the training and test set samples were classified into the four categories, true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The bar plot

comparing samples classified in the four categories with the LMC classes was generated using R-package *ggplot2* [130].

6.3 Results

6.3.1 Baseline RF model using clinical information alone

To evaluate the combined prognostic value of the clinical variables, an RF model was generated using AJCC stage, sex, site of tumour and age at diagnosis. During the training stage, the RF model had a lower OOB error for class 0 observations in comparison to class 1 observations (Figure 6.2). The overall OOB error stabilised after generating 200 decision trees (Figure 6.2).

The performance of the model was comparable on both training and test sets. The RF model had higher specificity in both training (0.68) and test (0.71) sets than the sensitivity (training:0.63, test:0.63) (Table 6.2). This also translated in a higher NPV than the PPV in both training and test sets (Table 6.2). The kappa value was slightly higher in the test set (0.32) in comparison to training set (0.28) indicating better agreement between actual and predicted class labels in the test set (Table 6.2).

Table 6.2 Summary of clinical variable based RF model performance

Set	Sensitivity	Specificity	PPV	NPV	kappa
Training	0.63	0.68	0.48	0.80	0.28
Test	0.63	0.71	0.50	0.81	0.32

6.3.2 Prediction after removing clinical information from gene expression

Gene expression data has shown to be associated with prognostic clinical variables such as stage of tumour, sex and age of the patient. To test whether gene expression data can independently predict prognosis after removing clinical information from the expression data, a linear regression model was fitted for each gene using clinical variables as predictors in the multivariable model. A strong correlation was found between gene expression and clinical variables as illustrated in Figure 6.3. Among the 13,688 genes tested, the clinical variables were a significant predictor of gene expression for approximately 5500 genes (FDR adjusted $p < 0.05$) in the multi-

variable model (Figure 6.3A). From the linear regression model, residual values for each gene were extracted. Figure 6.3B shows an example of a gene that has different expression value across males and females but, as expected, the residuals were no longer associated with sex.

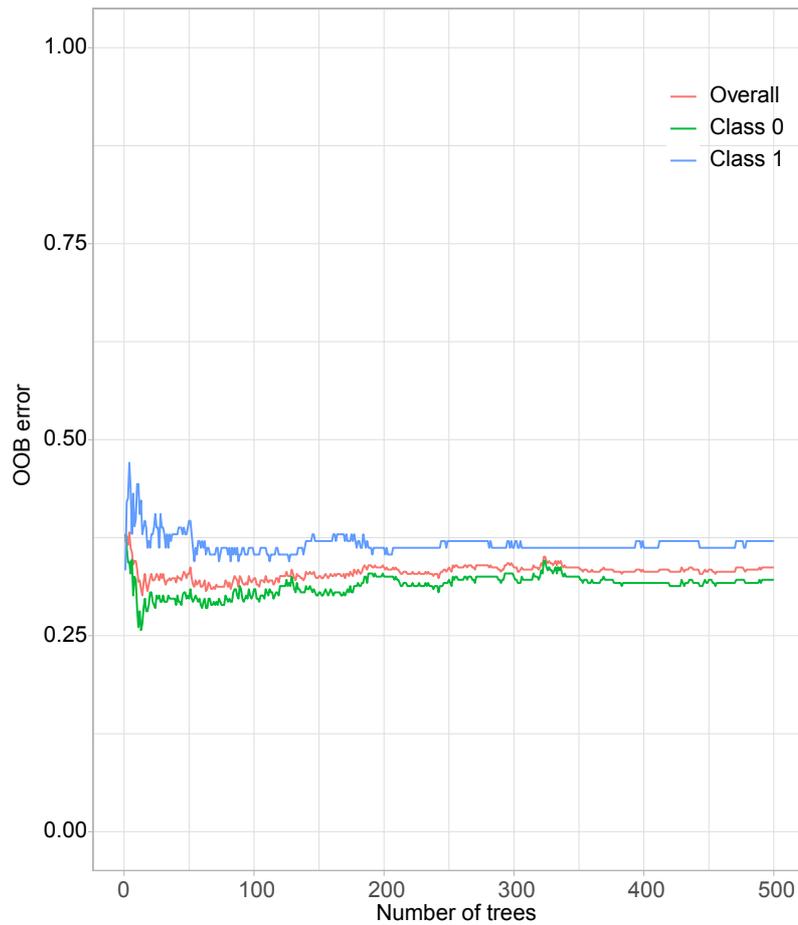


Figure 6.2 Clinical variable based RF model.

RF model OOB error estimates as the decision trees were added to the forest. The red line is the overall OOB error, the green line is the error rate for class 0, and the blue line is the error rate for class 1.

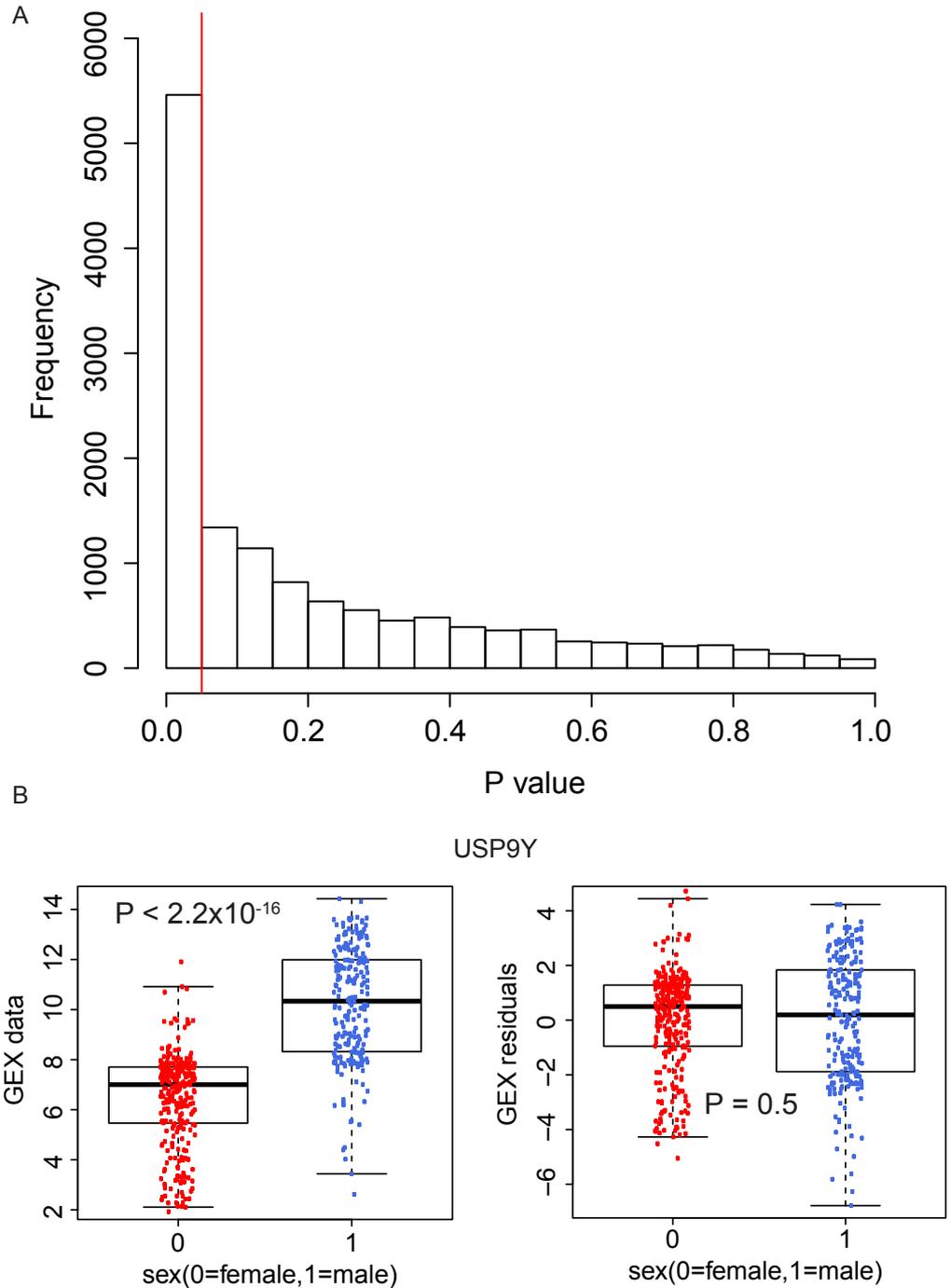


Figure 6.3 Gene expression association with clinical predictors of melanoma.

(A) Histogram plot of FDR adjusted P values from the linear regression model predicting gene expression using clinical variables. Red line shows the adj. P value (FDR) cut off at <0.05 . (B) Example boxplot showing expression values of a Y-linked gene *USP9Y* across males and females (left) and the same distribution when applied to the gene expression residuals values after adjusting for clinical variables in linear regression. P value was calculated using Mann-Whitney/Wilcoxon test.

6.3.2.1 RF model from using adjusted gene expression

To test whether gene expression can predict outcome after adjusting for clinical information, a RF model was generated using residual values as described above. To be consistent with previous analyses (refer to 5.2.5.5), the RF model was generated using a balanced class design. Similar to previous results, the RF model had a higher OOB error rate in class 1 than in class 0 during training (Figure 6.4). The overall error rate of the RF model (combining the 2 classes) stabilised after adding 5000 trees to the model (Figure 6.4).

The model had a higher specificity in both training (0.81) and test sets (0.78) than sensitivity (training: 0.32, test: 0.49), and consequently, a higher NPV than PPV in both the sets (Table 6.3). When comparing training and test sets, the RF model had higher performance on the test set ($\kappa=0.27$ in the test set and 0.14 in the training set) (Table 6.3).

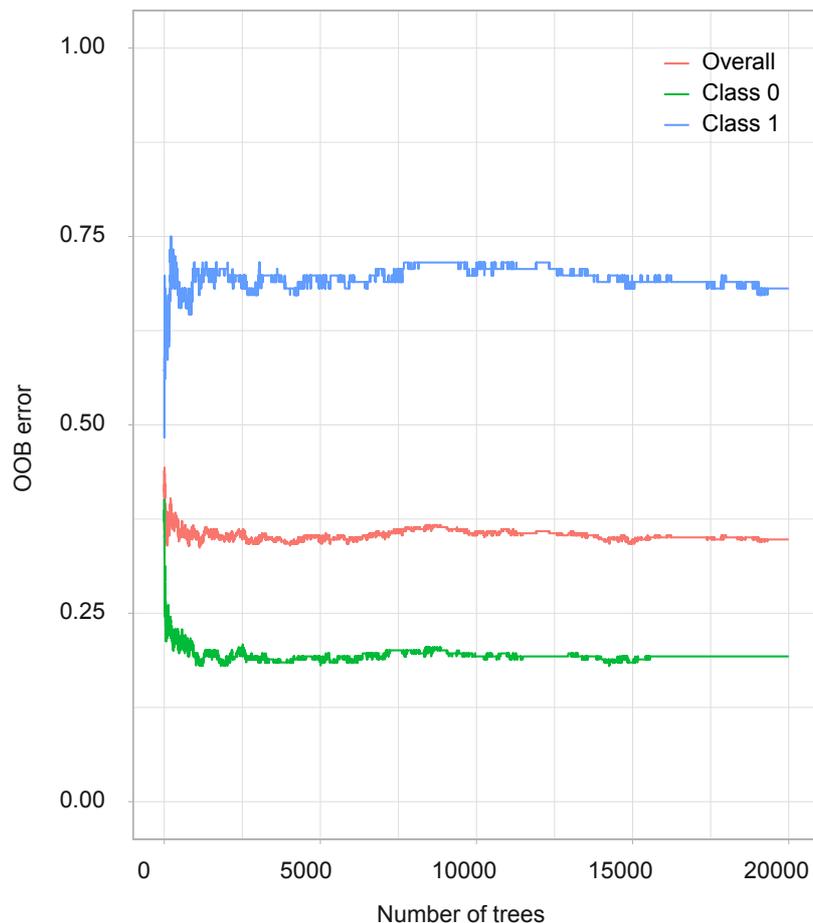


Figure 6.4 RF model from adjusted gene expression

RF model OOB error estimate as more decision trees were added to the forest. The red line is the overall OOB error, green line is the error rate for class 0 observations, and blue line is the error rate for class 1 observations.

Table 6.3 Summary of residual based RF model performance

Set	Sensitivity	Specificity	PPV	NPV	kappa
Training	0.32	0.81	0.43	0.72	0.14
Test	0.49	0.78	0.50	0.77	0.27

6.3.3 RF model from combining clinical information and gene expression

As shown earlier, even after adjusting for clinical variable information, the gene expression demonstrated prognostic significance as the resultant RF model predicted outcome on the test set. Hence it was hypothesised that combining clinical variable information and gene expression data may lead to generation of a RF model with improved performance in comparison to clinical variables alone.

As expected, the RF model from combining gene expression data and clinical variables (combined RF model) (kappa=0.37) had higher performance on the test set in comparison to the RF model from clinical variables alone (kappa=0.32). Overall the combined RF model had higher specificity in both training (0.82) and test sets (0.76) compared to the sensitivity (training:0.50, test:0.63) and a higher NPV than PPV; indicating that the model predicted class 0 observations more accurately than class 1 observations. The kappa values were similar in both training (0.33) and test sets (0.37) (Table 6.4).

The combined RF model (sens.=0.50, spec=0.82, kappa=0.33) also had higher performance on the training set in comparison to the gene expression based model developed in chapter 5 (sens.=0.44, spec.=0.77, kappa=0.22) (Table 6.4, refer to 5.2.5.5). However, performance of the combined RF model (sens.=0.63, spec.=0.76, kappa=0.37) was slightly lower than the gene expression based model (sens.=0.54, spec.=0.86, kappa=0.41) on the test set.

Table 6.4 Summary of combined RF model performance

Set	Sensitivity	Specificity	PPV	NPV	kappa
Training	0.50	0.82	0.53	0.78	0.33
Test	0.63	0.76	0.54	0.82	0.37

6.3.4 Variable selection to generate the final refined RF model

The combined RF model was based on ~13 thousand gene expression values and four clinical variables. To improve clinical utility of the RF model, variable selection was performed on the gene expression dataset. The top n genes (10, 50, 100, 200, 400, 800, 1600) were selected based on Gini index and new RF models were generated for each of the gene sets (total 7 models). Clinical data was integrated into gene expression as described above (refer to 6.2.3).

Interestingly, reducing the number of genes did not result in a reduced performance as the RF models with 10, 50, 100, 200, 400, 800 and 1600 genes along with clinical variables had a better kappa index than all genes- and clinical variable-based RF model (Table 6.5).

The RF models with reduced gene sets had comparable kappa values on the test set indicating similar levels of agreement between actual and predicted class labels (Table 6.5). All the reduced signatures (gene sets) RF models maintained a higher specificity than sensitivity, as in all previous analyses, but the difference between these two metrics was much lower in these models in comparison to RF models based on all genes or referred to as the combined RF model previously. These reduced models maintained a higher NPV than PPV, consistent with previous results. The RF model with 200 genes had the highest kappa value in both training and test sets (Table 6.5) and was selected as the “final refined RF model” for further analyses.

Table 6.5 Summary performance of RF models generated after selection of genes based on Gini index measure

ClinVar refers to the clinical variables.

Set	Gene set	Sensitivity	Specificity	PPV	NPV	kappa
Training	10 genes + ClinVar	0.61	0.81	0.60	0.81	0.42
	50 genes + ClinVar	0.59	0.76	0.53	0.80	0.34
	100 genes + ClinVar	0.59	0.78	0.56	0.80	0.36
	200 genes + ClinVar	0.59	0.82	0.61	0.81	0.42
	400 genes + ClinVar	0.58	0.82	0.60	0.81	0.40
	800 genes + ClinVar	0.59	0.82	0.60	0.81	0.41
	1600 genes + ClinVar	0.58	0.81	0.58	0.80	0.39
	All genes	0.50	0.82	0.53	0.78	0.33
Test	10 genes + ClinVar	0.71	0.75	0.56	0.85	0.43
	50 genes + ClinVar	0.69	0.78	0.59	0.85	0.45
	100 genes + ClinVar	0.71	0.76	0.57	0.85	0.44
	200 genes + ClinVar	0.69	0.79	0.60	0.85	0.46
	400 genes + ClinVar	0.67	0.77	0.57	0.84	0.42
	800 genes + ClinVar	0.67	0.75	0.55	0.83	0.40
	1600 genes + ClinVar	0.67	0.75	0.55	0.83	0.40
	All genes	0.63	0.76	0.54	0.82	0.37

6.3.5 ROC analysis of the final refined RF model

A ROC analysis was performed to compare the prognostic value of the final refined RF model with three other models on the training and test sets of LMC. Overall, the four RF models compared were:

- 1) Baseline RF model using clinical variable alone (refer to 6.3.1)
- 2) RF model based on gene expression alone (refer to 5.2.5.5)
- 3) Combined RF model (refer to 6.3.3)
- 4) The final refined RF model (refer to 6.3.4).

On the training set, the final refined RF model (0.77) had the highest AUC in comparison to the other three models (baseline clinical variable based, gene expression based, and combined RF models) (Figure 6.5A). This observation was consistent on the test set as the final refined RF model (0.83) maintained the highest AUC value in comparison to the other three models (Figure 6.5B). Among the three other models, the combined RF model (0.71) had slightly better performance than the baseline clinical variable based RF model (0.70) and had much better performance than the gene expression based RF model (0.66) on the training set (Figure 6.5A). On the test set, the combined RF model and the gene expression RF model had same AUC (0.78) and AUC for the baseline clinical variable RF model (0.73) was much lower (Figure 6.5B).

Overall, the final refined RF model consistently outperformed the other three models in both training and test sets (Figure 6.5). Moreover, the 7% (training) and 10% (test) increase in the AUC of the final refined model was statistically significant when compared with the baseline RF model using clinical variables alone (train: $P=0.03$, test: $P=0.02$) (Figure 6.5).

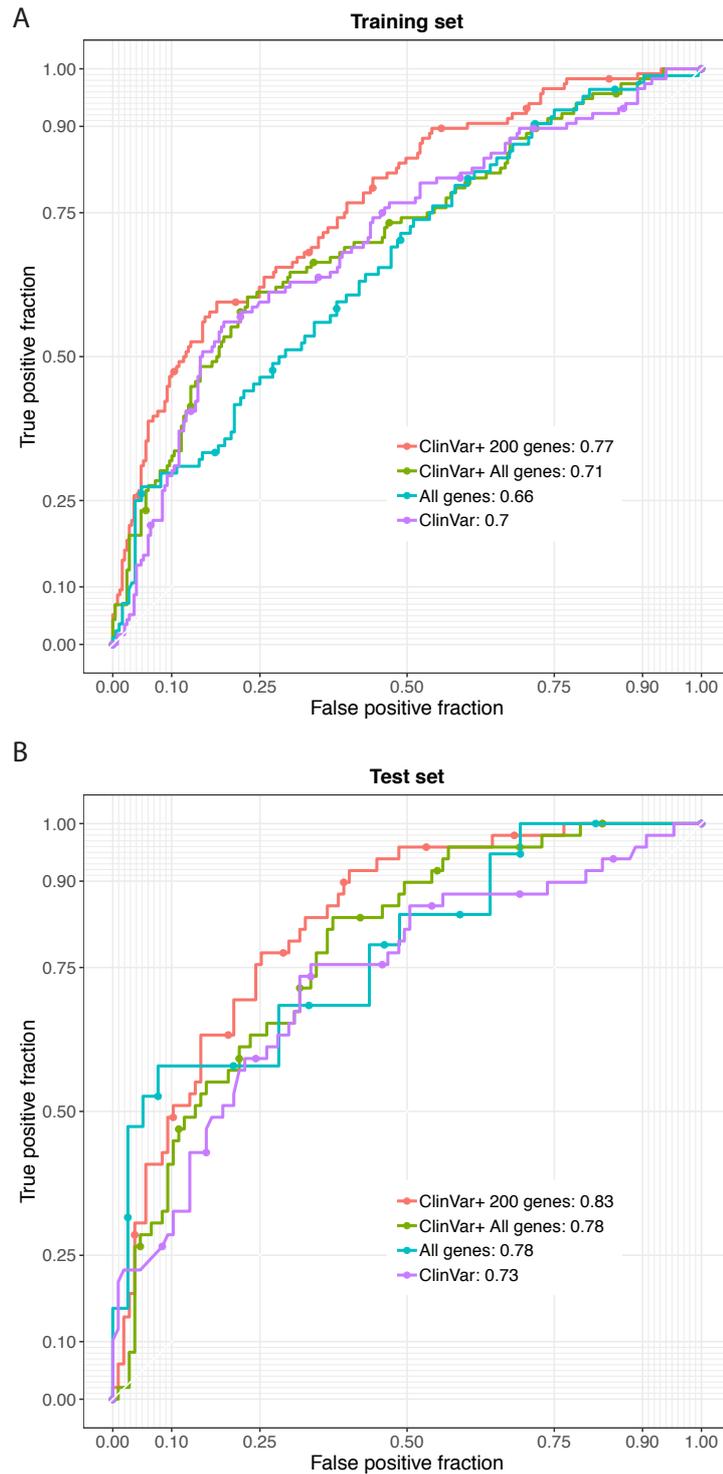


Figure 6.5 Comparison of the refined RF model with baseline RF models

ROC curves of the RF models on (A) training and (B) test sets. *ClinVar* refers to the clinical variables model, *All genes* refers to the genome-wide gene expression based model, *ClinVar+ All genes* is the model including clinical variables and genome-wide gene expression, and *ClinVar + 200 genes* is the final refined model.

6.3.6 Validating prognostic significance of the refined RF model on Lund dataset

To validate the prognostic significance of the refined RF model, it was applied to the Lund primary melanoma dataset, using the same clinical variables as used in the LMC dataset. Among 200 predictor genes of the refined RF model, 96 genes were not present in the Lund dataset. The missing genes were initialised with 0 value in the Lund data. Since cause of death was not recorded in the Lund dataset, death from any cause and melanoma relapse at 6 years were used as alternative outcomes. For both these outcomes, the refined RF model predicted outcome in the Lund dataset (Table 6.6). Similarly to the LMC dataset, this model had higher specificity (0.86-0.88) than sensitivity (0.54-0.57) and higher NPV (0.81) than PPV (0.63-0.69) in the Lund dataset (Table 6.6). The kappa index was comparable to what was observed in the LMC test set (Table 6.6).

The ROC analysis confirmed the model's prognostic value in this new dataset (Figure 6.6). For melanoma relapse before 6 years endpoint, the increase in the AUC for the refined RF model was 4% when compared with using clinical variables alone, although it failed to reach statistical significance ($P=0.2$). The AUC improvement was more modest for deaths from any cause before 6 years post diagnosis (1% increase, $P=0.7$, Figure 6.6B).

Table 6.6 Summary of refined RF model performance on the Lund dataset

Event at 6 years	Sensitivity	Specificity	PPV	NPV	Kappa
Relapse	0.57	0.88	0.69	0.81	0.47
Death from any cause	0.54	0.86	0.63	0.81	0.42

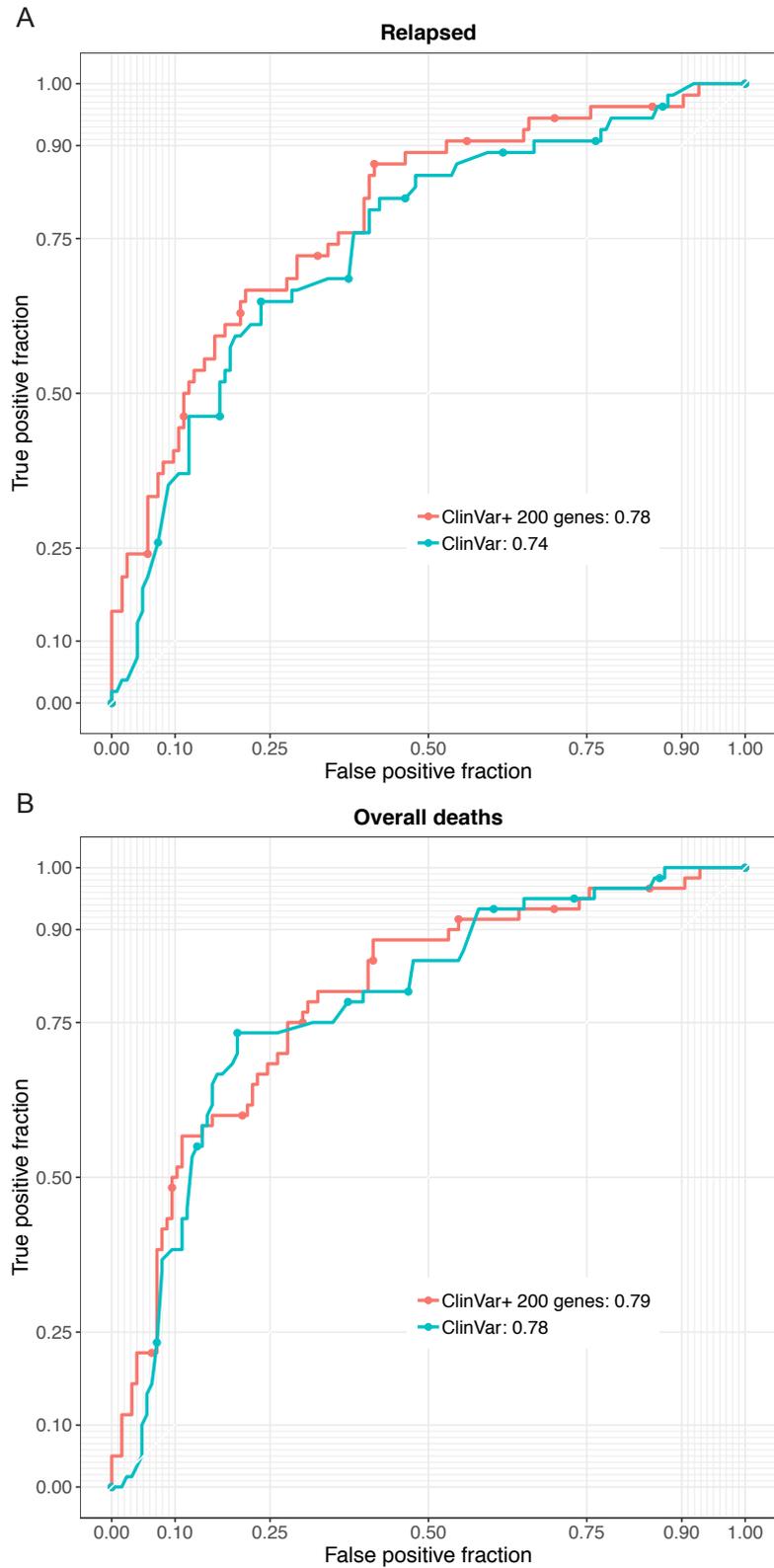


Figure 6.6 ROC analysis of the RF models in the Lund dataset

The outcome was (A) relapse and (B) deaths within 6 years post diagnosis. *ClinVar* refers to the baseline model and *ClinVar + 200 genes* is the final refined model.

6.3.7 Biological interpretation of the refined RF model

To understand the biology behind the refined RF model predictions, a pathway enrichment analysis was performed on the 200 predictor genes of the refined RF model (Appendix I, Table 8.14). The enrichment analysis identified association with pathways linked to DNA damage repair (e.g. Fanconi anemia pathway), cell cycle and cell proliferation (e.g. meiotic recombination, mitotic prometaphase, mitotic metaphase and anaphase, mitochondrial translation), PI3K-Akt signaling, and generic terms like pathways in cancer, small cell lung cancer, Parkinson's disease and Alzheimer's disease etc (Figure 6.7).

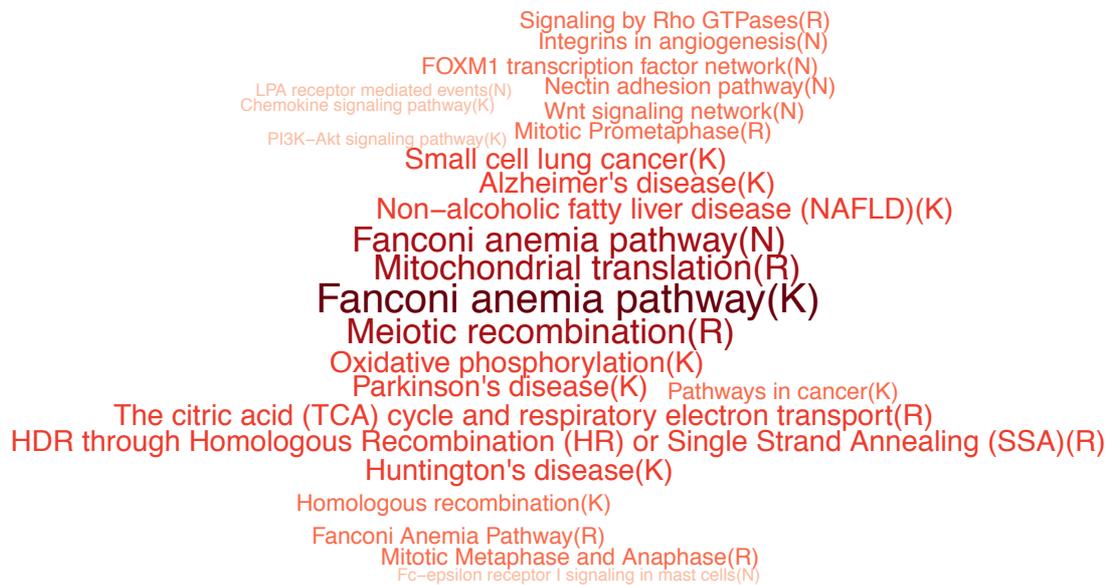


Figure 6.7 Biological interpretation of the refined RF model.

Summary of biological pathways enriched in the top 200 genes from the RF model. All the pathways with FDR < 0.1 were selected. The text size in the figure is based on FDR value associated with each pathway i.e. more significant pathways are shown in a larger font. (K)- refers to entry from KEGG database, N- NCBI database and R- Reactome database.

6.3.8 The LMC class association with the refined RF model predictions

The LMC classes defined in Chapter 4 (refer to 4.3.1) were compared to the refined RF model prediction classes (class 0 and class 1) (Figure 6.8). The LMC classes showed significant association with RF model predictions in training ($P < 2 \times 10^{-16}$) and test sets ($P = 7 \times 10^{-9}$). The LMC class 1 and class 5 samples were almost exclusively predicted as survivors (class 0) by the RF forest model (i.e. true

negatives, with some false negatives) (Figure 6.8). The LMC class 2 samples were also mainly predicted as survivors but with a few true and false positives (Figure 6.8). The LMC class 3 samples were mainly predicted as non-survivors (class 1) by the RF model (Figure 6.8). The LMC class 4 and class 6 had a mixture of samples from each of the predictor categories; these classes had the highest proportion of wrongly classified samples (Figure 6.8).

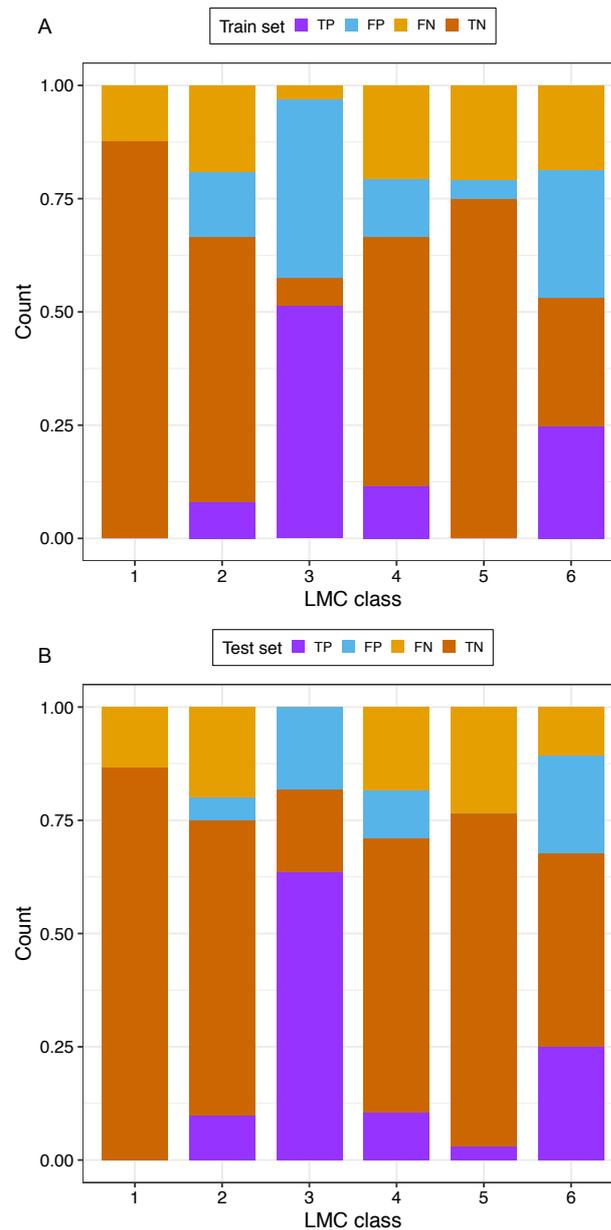


Figure 6.8 Comparing the LMC classes with refined RF model.

The overlap between the RF model predictions and LMC classes in the (A) training and (B) test dataset. TP refers to true positive samples, FP are false positive samples, FN are false negative samples, TN are true negative samples.

6.4 Discussion

In this chapter, the prognostic performance of gene expression-based RF models was compared with a clinical information-based RF model. Moreover, integration of these 2 data types demonstrated an improvement in the prediction performance. The performance further improved after refining the RF model using variable selection on gene expression. The refined RF model prediction performance was higher than the baseline clinical variable RF based model, although it showed only a little improvement in the Lund dataset.

6.4.1 Prognostic value of clinical information and gene expression based RF models

The baseline RF model using clinical variables alone predicted prognosis in training and test sets. The clinical variables included were AJCC stage, sex, body site of tumour and age at diagnosis; these variables are also most likely to be recorded during a clinical visit. The AJCC staging system is one of the most widely used clinical variable for predicting prognosis of melanoma patients and is built upon strong prognostic factors like Breslow thickness and ulceration status [44]. Other clinical variables like sex, age and tumour site have also shown an association with melanoma prognosis [42]. Therefore, as expected, the RF model generated using these clinical variables predicted prognosis in the LMC data. Previous studies have shown strong association between clinical variables and gene expression in melanoma [114, 202]. In keeping with these studies, in LMC it was observed that clinical variables significantly predicted gene expression for more than five thousand genes (Figure 6.3). Removing the clinical variable information from the gene expression data demonstrated that the resultant RF model based on gene expression residuals was still predictive of prognosis. This is interesting as it signifies that gene expression data has independent prognostic information of melanoma tumours. Hence it was hypothesised that combining clinical variables and gene expression data may improve prediction performance in comparison to using clinical variables alone. As expected, the combined RF model has improved performance in comparison to the baseline clinical variables based RF model. These results highlight the potential clinical relevance of performing gene expression profiling alongside assessing the clinical characteristics of tumours for predicting prognosis of melanoma patients.

6.4.2 Refining combination of clinical information and gene expression based RF model

Application of the combined RF model would require profiling a large number of genes, and this would limit the clinical utility due to the associated higher economical cost or practicality. Variable selection is crucial in supervised classification as selecting irrelevant features can sometimes leads to low accuracy and overfitting [203-206]. Therefore, to mitigate this, the model was refined by performing variable selection using the Gini index measure for each gene in the RF model (refer to 5.2.5.5). Multiple subsets were evaluated, and the final refined RF model with 200 genes had a higher performance than the original model. The difference in performance of the final refined RF model and original RF model may well be due the difference in the number of predictor variables used for generating the RF model at each iteration. The RF randomly selects genes to be included in the model at each iteration and selecting them from the whole genome means there is a high likelihood of using uninformative genes (i.e. those not associated with the outcome) in a significant number of iterations. The selection of which genes to consider at each node is random, and does not use information on which variables are most relevant. Therefore, incorporating variable importance measures into the model building process is a two-staged approach, and our results indicate that this approach may be more efficient than a one-staged approach which does not use variable importance information.

6.4.3 Prognostic value of refined RF model and validation

The refined RF model showed an improved prognostic value when compared with the baseline RF model based on clinical variables. Interestingly, the increase in AUC from ROC analysis of the refined RF model increased by 10% in the test set when compared with the baseline model ($p=0.02$). This highlights the potential translational value of the refined RF model in clinical settings for predicting melanoma prognosis. However, in order to confirm this, an independent validation was required.

The refined RF model was tested on an independently generated dataset of primaries from Lund. In keeping with the LMC test set, the refined RF model showed consistency in predicting prognosis and improved the AUC by 4% (relapsed) and by 1% (death from any cause) when compared with the baseline model. The improvement in AUC was modest and did not reach the significance threshold of $P<0.05$. The AUC was higher for melanoma relapse than using death from any cause as an outcome because not all deaths were caused by melanoma in the Lund cohort

(median age 67), as the cohort had high morbidity given the advanced age of patients in comparison to LMC (median age 57). As described previously, Lund data was generated in a different population and using a different microarray platform and quality control steps [105]. It is noteworthy that among the 200 genes selected in our refined model, the Lund dataset only contained 104 genes. The others were excluded in quality control steps (in fact, only ~8900 probes, representing 7753 unique genes were in the Lund data, compared to ~29000 probes and 20807 genes in the LMC). Hence, a further validation on larger primary melanoma datasets is required to fully confirm the robustness of this model and demonstrate its superiority to one based on clinical variables alone. The current limitation is that only one primary melanoma dataset is so far publicly available. Therefore, the robustness of refined RF model could not be assessed on additional datasets.

6.4.4 Biological interpretation of the refined RF model

Biological pathway enrichment analysis of predictor genes in the refined model showed association with pathways such as the Fanconi anemia pathway, PI3K-AKT signalling pathway and cell cycle related pathways. These pathways have been implicated in various cancers. The Fanconi anemia pathway has been shown to play crucial role in DNA repair mechanisms [207-209]. Disruption of this pathway has been associated with acquiring tumour resistance in various cancers [210-214]. The PI3K-AKT signalling pathway has been reported as one of the most frequently disrupted pathways in melanoma, which has now paved the way for therapeutic drug discovery [215-219]. The predictor genes were also associated with cell cycle related pathways. These pathways provide an overview of biological information being utilised by the RF model in the decision-making process. However, furthermore work is required to investigate the individual impact of the predictor genes on melanoma survival which is beyond the scope of this thesis.

In summary, integrating clinical variables into the gene expression-based RF model improved the prediction performance. Applying variable selection further improved the performance and generated a refined model with better prognostic value in the LMC and the Lund datasets. The refined model was shown to be of biological significance. Overall, this model after further validation on larger cohorts could be useful in clinical settings for predicting prognosis of melanoma patients.

Chapter 7

Final summary and discussion

7.1 Summary of the two main aims of this study

In this study, statistical and bioinformatic analysis were developed to classify primary melanoma tumours of the LMC. The analyses addressed the two main aims: 1) generating molecular classes of melanoma using unsupervised clustering and 2) developing a prognostic classification model using supervised classification. Chapters 3 and 4 reported the results of unsupervised clustering while chapters 5 and 6 dealt with supervised classification.

In the first aim, applying unsupervised clustering to gene expression data of LMC tumours led to discovery of the six LMC classes signature. This LMC class signature showed prognostic value in the whole dataset, including in the stage I tumours. A validation of this signature was conducted in primary melanomas from the Lund cohort, although the paucity of stage I samples in the Lund cohort did not allow replication in that particular subgroup.

In the second aim, supervised classification models were generated to classify the LMC patients into two classes (survivors and non-survivors) based on survival up to 6 years. The classification models were developed using Random Forest (RF) and Support Vector Machine (SVM) algorithms, after splitting the LMC dataset in a training and a test set. The training set was imbalanced (the majority being survivors), and an attempt to overcome the class imbalance problem still generated prediction models with higher accuracy in predicting the majority class (survivors) than the minority class (non-survivors).

In a comparison, the supervised classification model (refined RF model) showed good agreement with some of the LMC classes. LMC classes 1 and 5, with strong upregulation of immune and stroma related genes, were consistently predicted as the survivors by the supervised classification model. Tumours of LMC class 3, with downregulation of immune genes and upregulation of cell cycle mechanisms, were consistently predicted as non-survivors by the supervised classification model. The strong overlap between some of the LMC classes and supervised classification model prediction is not surprising because the association between the immune and cell cycle related gene expression and survival is well-established in the field [103, 117, 197, 220-224]. Interestingly, the LMC classes which showed intermediate levels

of immune and cell cycle related gene expression (LMC classes 2, 4, 6) were predicted to include both survivors and non-survivors by the supervised classification model. The LMC molecular classes are a good foundation for understanding the underlying biological pathways driving melanoma prognosis, and the LMC class signature, after successful validation on another dataset, could be of direct clinical relevance to AJCC stage I melanoma. The machine learning (refined RF model) model on the other hand would be more appealing in clinical practice because it is based on only two classes (survivors or non-survivors), and therefore has lower complexity in interpretation of the results generated from this model.

7.2 Context and discussion of findings

The AJCC staging system is one of the most powerful and widely used classification tools in clinical settings for predicting melanoma prognosis [44]. In keeping with previous reports, the AJCC staging system performed only moderately well at predicting prognosis (AUC= 0.69) (refer to 4.3.5.4) [114]. Therefore, additional biomarkers are required to complement this tool for predicting prognosis in clinical settings.

Gene expression profiling has allowed molecular classification of tumours in several cancers [85, 106, 158, 225-236]. Molecular classification has drastically improved the biological understanding of the disease and has led to development of new drug targets. Additionally, these molecular classes have also been shown to predict clinical outcome for a patient [81, 105, 106, 225, 228, 234, 235, 237-239]. In breast cancer, for example, previous retrospective studies of gene expression profiling have identified five molecular classes with prognostic significance, and furthermore these classes predicted response to adjuvant treatments [237, 240-245]. Paik *et al.* surveyed 250 genes reported to be associated with clinical outcome for breast cancer patients and generated a refined 21 gene-based prognostic biomarker (Oncotype-DX) which has been successfully validated in prospective clinical trials [231, 246-251]. This test predicts recurrence of breast cancer, and recently its ability to stratify patients for adjuvant chemotherapy has been demonstrated [251]. Overall, these studies highlight the crucial role of gene expression profiling in developing biomarkers with prognostic and predictive value.

Similarly, in melanoma, several studies have performed gene expression profiling of tumours from retrospective cohorts to generate prognostic gene signatures [81, 100, 102, 104-106, 108, 113, 114, 144, 252-254]. Unlike breast cancer, the prognostic

significance of these gene signatures has not yet been tested in prospective clinical trials. Most of these studies have been based on small sample size, and the gene signatures were generated from cohorts predominantly of advanced stage tumours [81, 84, 100, 101, 104-106, 113].

Most of the biomarkers developed from retrospective studies fail to translate into clinical settings due to low prognostic and predictive value on other validation cohorts [255, 256]. Therefore, it is highly recommended that studies are conducted with low ascertainment bias, with appropriate sample size, and with validation on other independently generated datasets to avoid overfitting [255, 256]. It has also been suggested that a biomarker based on gene expression must be jointly analysed in a multivariate model with the histopathological factors of a disease [107].

7.2.1 Class discovery using unsupervised clustering

Previously, the Lund and TCGA groups have developed molecular classes of melanoma using a similar unsupervised clustering algorithm [81, 105, 106]. These molecular classes (Lund and TCGA classes) were applied to the LMC dataset. As shown in Chapter 3, the Lund and TCGA classes were predictive of MSS in the whole LMC dataset. When the LMC patients were stratified on the basis of AJCC stage, the signatures predicted outcome in stage II & III tumours but showed no association with MSS in stage I tumours. This was confirmed as a significant statistical interaction, illustrating the fact that the lack of association in stage I was not merely due to a power issue.

The Lund and TCGA class signatures were derived from cohorts predominantly of metastatic tumours (i.e. AJCC stages III and IV) [81, 106], but subsequent replication studies have validated their prognostic value in primary tumours with only few AJCC stage I cases (n=61 in Lund, n=58 in Leeds subset) [105, 114]. The lack of association of these signatures with prognosis in stage I tumours in this larger dataset was disappointing, since currently the majority of melanomas (91%) are diagnosed at AJCC stages I and II [257]. Although these patients have a good prognosis overall, there is a significant number of deaths among these patients [257, 258]. Therefore, there is an urgent need to develop both prognostic and predictive biomarkers for those early stage melanoma patients.

The LMC dataset has a large number of stage 1 cases, and its make-up in terms of disease stages offered the potential to generate a prognostic signature across all AJCC stages in a new unsupervised analysis. As shown in Chapter 3, the LMC tumours were clustered using robust consensus based HC, KM and PAM clustering

algorithms. Previous comparisons of clustering algorithms spanning over multiple gene expression and protein sequence information datasets suggested that none of the algorithms clearly outperformed the others [142, 143, 259]. In keeping with that, although different clustering algorithms (HC, KM and PAM) identified a different number of tumour classes in the LMC, i.e. HC algorithm identified 5 classes, KM identified 6 classes, and PAM identified 7 classes, the algorithms showed agreement with each other and with the existing Lund and TCGA classes. Since the PAM classes had higher stability in comparison to HC and KM classes, the 7 PAM classes were further explored.

It has been reported that some noise may remain in the dataset even after adjusting for technical variation in gene expression quality control [260]. However, over-normalisation can remove some biological variation, thereby reducing the utility of these experiments. Although batch correction was applied during normalisation, the PAM algorithm still identified one small class whose mRNA samples were processed in the same batch [117]. On further investigation, these samples were identified to be in similar positions on the plates, i.e. in first or last rows of the plates. Previously, the LMC samples were plated and sent to a gene expression profiling company in Leiden (Netherlands). The plate edges may have been impacted (for example by evaporation) during transport or other manipulations for one particular batch. Investigating different clustering algorithms allowed identification of these samples. Taking a cautious approach, it was concluded that this class may represent technical noise, and hence samples from this class were excluded.

Re-clustering the remaining samples using the same consensus PAM algorithm confirmed the previously observed six classes (LMC classes). As shown in Chapter 4, the LMC classes were significantly associated with clinical prognostic factors of melanoma, such as sex, body site of tumour, ulceration status, age at diagnosis, Breslow thickness and TILs. Most of these factors were also associated with the Lund and TCGA classes when applied to the LMC data. However, the LMC classes further showed a strong association with tumour mutation status in *BRAF* and *NRAS* oncogenes, while the Lund and TCGA classes did not.

Interestingly, the LMC classes were prognostic in the whole dataset, including within the AJCC stage I group. To our knowledge this is the first and only clustering based prognostic molecular signature in stage I melanoma. This study used a much larger number of genes (~13,000) in comparison to the Lund (~7000) and TCGA (1500) studies [81, 106]. Using a larger number of genes may have allowed more biological variation in the data to be captured. Secondly, the LMC data is based on the HT12-

version 4 array, whereas the Lund study used a previous version of array and the TCGA study used an altogether different platform (RNA-seq). However, high overlap for some of the classes (LMC classes, 1, 3 and 5) with the Lund and TCGA classes may reflect the conservation of gene expression profiles in primary and metastatic melanoma.

In clinical settings, performing expression profiling with a large number of genes is associated with higher economical cost. Hence, a prognostic test with few genes is preferred over a test based on large number of genes. To make the LMC classes signature clinically feasible, the signature was refined from 13,688 genes to 150 genes. Since gene reduction was done by selecting the top genes which characterise the LMC classes, this may have also removed relevant biological information. As expected, application of the reduced LMC signature (150 genes) to the LMC dataset resulted in some misclassification (especially for LMC class 2) in comparison to using the original classification.

The biological characterisation of the LMC classes using pathway enrichment analysis and Lund biological modules gave consistent results [81, 106]. High immune gene expression has been reported to be predictive of good prognosis in melanoma and other cancers [103, 117, 197, 206, 220-223, 226, 261-267]. In keeping with this, the LMC classes with good prognosis had higher immune gene expression and lower cell cycle related gene expression; LMC classes with poor prognosis had higher cell cycle related gene expression and lower immune gene expression. A similar observation has also been made by a fellow PhD student in the group (Joanna Pozniak personal communication) by clustering the LMC tumours based on immune response related genes and observing an inverse correlation between the genes in proliferation pathways and the genes in immune response pathways. However, this dichotomy did not apply to all the classes: LMC class 6 showed evidence of increased expression of both immune and cell cycle genes and the corresponding prognosis was very poor, especially in stage I tumours. This class is therefore unusual and was poorly captured with the existing Lund and TCGA classification systems (refer to 4.3.4). Although it was not among the aims of this thesis, a preliminary biological investigation found significant increase in *JUN* expression and copy number in LMC class 6 tumours along with evidence for activation of the epithelial to mesenchymal transition pathway. Overall these results suggests that the LMC class 6 may be a biologically different class with strong clinical relevance for AJCC stage I melanoma.

7.2.2 Class prediction using machine learning

Several studies have performed comparisons between patients with various clinical endpoints of melanoma and have generated biomarkers based on clinico-histopathological and genomic characteristics [104, 109, 252, 254, 268-271]. A gene expression based prognostic test for melanoma, Decision-Dx by Castle Biosciences, identifies patients who are likely to metastasize in 5 years and is available commercially [109]. It contains 28 prognostic genes and 3 control genes and produces an accurate prediction of high and low risk of metastasis. Another 12-gene based biomarker from Liu *et al.* achieved higher diagnostic performance in distinguishing metastatic melanomas from normal skin and benign naevi. The 11 gene-based signature from Brunner *et al.* when combined with the AJCC staging system for predicting MSS showed 4-6% increase in performance in comparison to the AJCC staging system alone [108]. Overall, these signatures have been developed on comparatively small retrospective cohorts of primary melanoma, and the robustness of these gene signatures is yet to be determined on larger melanoma cohorts and especially in prospective cohorts. Recently, it was showed that a machine learning based algorithm achieved higher classification accuracy in detecting melanoma in comparison to several trained dermatologists [169]. This study suggested that dermatologists may benefit from using this machine learning model in clinical settings for detecting melanoma at early stages.

Previously, several other studies have developed gene signatures predicting various melanoma outcomes, but none of the studies were done using a large number of early stage patients and using machine learning algorithms [84, 100-102, 104, 108, 109, 113, 272]. The LMC dataset so far is the largest dataset available for early stage melanoma with good follow-up data. In Chapter 5, two machine learning approaches, Random Forest (RF) and Support Vector Machine (SVM), were applied to predict outcome. Although these methods have been used in different applications, they have not been widely used for class prediction based on gene expression data.

To avoid sample losses to censoring, a survival time cut-off was chosen at 6 years, but the LMC dataset still had a higher proportion of survivors than non-survivors. This is inevitable given that the majority of patients with primary melanoma survive for more than 5 years but the unbalanced dataset caused a challenge in machine learning. Consistent with the literature [182], in RF, under-sampling the majority class in the training set to achieve a balanced class design showed an improvement in prediction performance in comparison to using an unbalanced class design. Unlike RF, the SVM model performance did not change with or without balanced class

design approaches. Overall both the RF and SVM algorithms were influenced by class imbalance and generated prediction models which predicted the majority class (survivors) better than minority class (non-survivors), i.e. these models had higher specificity than sensitivity. In future, more investigations will be required to overcome the class imbalance problem.

Several studies have compared the performance of RF and SVM algorithms and have found contradictory results. The study by Diaz-Uriarte *et al.* showed that the RF algorithm achieved comparable performance to that of SVM on simulated and real microarray datasets [273]. However, Statnikov *et al.* performed comparisons across 22 diagnostic and prognostic datasets and reported that the SVM algorithm outperformed the RF algorithm [274]. In LMC, the RF model was identified to have a higher performance overall in comparison to the SVM model. A grid search to select the parameters of SVM initially improved the performance of the model, but further fine-tuning of the hyperparameter search brought no further improvement. Further comparisons indicated that the RF and SVM models made consistent predictions on test set observations which was reflected in higher Carmer's V agreement. However, when comparing the inconsistent predictions between the models, the RF model showed higher accuracy than the SVM model.

Previous studies have shown that although gene expression data has prognostic significance, when jointly analysed with other clinico-histopathological variables of melanoma it does not show much improvement in predicting outcome in comparison to clinical variable variables alone [52, 100, 107, 113, 114]. Therefore, in Chapter 6, the prediction of outcome using gene expression (analyses performed in Chapter 5) was compared to that of the known prognostic clinico-histopathological variables of primary melanoma. As expected, the RF model generated using clinical variables alone predicted outcome in the LMC test set. When clinical variables were combined with the gene expression data, the performance of the resulting model improved slightly in comparison to using clinical variables alone. A feature selection was conducted (based on their importance in the RF) which generated a reduced RF model with 200 genes and clinical variables. The predictor genes in the refined RF model had biological relevance, as many of them are implicated in pathways, e.g. fanconi anemia pathway, PI3-AKT signalling and cell cycle related mechanisms, that have previously been reported to play a role in melanoma and other cancers [207-210, 212-219].

Ferris *et al.* reported that combining Decision-DX test and AJCC stage information further improved prediction of overall deaths and risk of metastasis in comparison to

using the AJCC stage alone [253]. They found that the model based on the Decision-DX test + AJCC stage had 82% sensitivity and 62% specificity while the model based on the AJCC stage alone had 60% sensitivity and 74% specificity [253]. Brunner *et al.* reported that the combining the gene signature and the AJCC staging system (AUC=0.66) had comparable AUC to the AJCC staging system alone (AUC=0.60). In keeping with these studies, the refined RF model had 69% sensitivity and 79% specificity, while the clinical variables model had 63% sensitivity and 71% specificity. The AUC for the combined model (AUC=0.83) was 10% higher than the clinical variables alone (AUC=0.73). It is worth noting that previous studies [108, 253] used the AJCC stage as the only clinical variable, whereas in this study sex, age at diagnosis and site of primary melanoma were also included in addition to the AJCC stage.

The prognostic value of the refined RF model was validated on the dataset from Lund. Since the Lund dataset did not have MSS information, melanoma relapse and death from any cause in 6 years was used as the outcome measures. In keeping with the LMC dataset, this model predicted outcome in the Lund cohort. The AUC for the refined RF model increased by 4% (melanoma relapse as outcome) and 1% (death from any cause as outcome) when compared to clinical variables alone in the Lund cohort. The improvement in AUC reached statistical significance in the LMC, but not in the Lund dataset, probably due to the smaller sample size, exacerbated by a considerable amount of missing data. Therefore, in future, validation of the refined RF model is required on another independent dataset to fully confirm the prognostic value of this model.

7.3 Strengths and limitations

The major strengths of this study are:

1. Cohort size: This study is one of the largest studies conducted on understanding the influence of tumour transcriptomics on primary melanoma survival. The largest cohort used in previous studies (Lund cohort) was about one third of the size of the LMC.
2. Extensive genomic and phenotypic data: Detailed clinico-histopathological and survival information has been recorded for the participants of the LMC. The gene expression data were generated from the tumour samples using a genome-wide array. Altogether these data are complementary to each other

and their integration can bring additional insights in melanoma biology, patient survival and a potential basis for a stratified care.

3. Melanoma-specific survival: Unlike previous studies, this study utilised melanoma-specific survival for testing the association of signatures with survival. This was made possible by the attention to detail in the data collection, which includes records of the cause of death, rarely recorded in most other studies. This information is continually retrieved from reliable sources (Public Health England) as the cohort follow-up continues.
4. AJCC stage subsets: The LMC participants have been recruited in hospitals as they were diagnosed, so they represent more closely the population as a whole (North of England). Although population ascertained but due to difficulties in sampling thin tumours, the tumour cores were sampled predominantly from thick tumours. However, sufficient thin tumours were also sampled to ensure that the conclusions drawn are more likely to be applicable to the wider population. This is in contrast with many other transcriptomic studies which often use highly selected samples (e.g. most advanced disease), which may result in a deeper biological understanding and even drug development but may be less applicable to the whole patient group. Having a good representation of AJCC stage I patients (the most common diagnosis in the population) allowed development of a signature of relevance in this group.
5. Validation dataset: The LMC classes and supervised classification signatures developed in this thesis were validated on an independent dataset published in 2012 [105] by Prof Göran Jönsson and Dr Martin Lauss (Lund University, Sweden). My work has therefore benefitted from the long-held collaboration between the two universities.

The main limitations of this study are:

1. Availability of data at one time point: The gene expression data used in this study represent the tumour's characteristics at only one time point. Longitudinal data were not available to evaluate the consistency of tumour gene expression patterns with time. However, it was noted that previously reported survival associations with gene expression were validated in this dataset (Chapter 3), indicating the robustness of the approaches used. Moreover, there is a need to identify a prognostic biomarker relevant to patients at presentation with primary disease.

2. AJCC stage I validation set: The LMC 6 class signature was shown to be of prognostic value in AJCC stage I melanoma. Unfortunately, this finding could not be validated in an independent dataset because none of the studies have generated gene expression dataset from sufficient stage I melanomas. As shown in Chapter 4, the replication of the refined LMC signature on the Lund dataset regenerated the LMC classes in this dataset. However, this cohort had only few AJCC stage I tumours and the prognostic value of the signature could not be validated in stage I melanoma. Most of melanoma research is conducted in advanced stage melanoma because of the challenges associated with sampling thin early stage melanomas, hence our efforts to find a validation dataset comprising of AJCC stage I tumours was not successful.
3. Missing information in the Lund dataset: Half of the predictor genes of the refined RF model were missing in the Lund data. The missing genes were filtered out in quality control steps due to the poor quality of the array platform that was used, the Illumina HT8.3 (only ~7,200 genes passed the QC filters). Hence, the Lund dataset was not an ideal replication dataset but it was the only one available.
4. Clinical utility: The prognostic signatures generated in this study have been demonstrated to be of prognostic value in LMC data and on an additional cohort. However, validation in additional cohorts is required to further confirm the robustness of the signatures, especially for stage I tumours.

7.4 Future perspectives

There are several recommendations for future work following the analyses conducted in this thesis:

1. Biological characterisation of the LMC classes: In chapter 4, differentially expressed genes and an overview of biological pathways associated with LMC classes were presented. For future analysis it would be interesting to explore the key regulatory genes governing the outcome in each of the classes.
2. Predictive value of LMC classes: The LMC signature had shown to be of prognostic importance but its association with response to adjuvant therapies, especially immunotherapy, was not tested. Immunotherapy response

datasets are currently being collected to test the value of these signatures in predicting response to immunotherapy.

3. Additional validation cohorts: There are currently no large enough AJCC stage I gene expression datasets available. It would be a step forward to generate data from stage I melanoma and validate the prognostic value of the signature.
4. Comparison with other signatures: Several molecular signatures have been proposed to predict risk of metastasis in melanoma. It would be interesting to compare these signatures and those presented in this thesis for their performance and their composition and possibly to pool them into one signature if the comparison reveals a limited overlap.
5. Combining gene expressions with Copy Number Variation (CNV) data: This thesis was focused on gene expression and clinical data. Next-generation sequencing derived copy number variation data have been generated in a subset of the same tumours (N=266) and are being analysed by another PhD student in the group. Despite the limited number of tumours profiled, it would be interesting to investigate the additional power of combining the two datasets, although this would certainly not be of relevance for stage I disease where the overlap would be even smaller.

Chapter 8

Appendix I

Table 8.1 The 150-gene based LMC 6 class signature

Gene	LMC class					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
<i>RASAL3</i>	1.619	0.016	-0.996	-0.401	-0.019	0.111
<i>SH2D1A</i>	1.549	0.128	-1.047	-0.327	0.053	-0.068
<i>SMAP2</i>	1.548	0.127	-0.84	-0.584	0.073	0.067
<i>TRAT1</i>	1.542	0.139	-0.845	-0.33	-0.026	-0.099
<i>BTLA</i>	1.541	0.078	-0.801	-0.384	-0.059	0.018
<i>ZNF831</i>	1.504	0.048	-0.598	-0.408	-0.172	0.09
<i>RHOF</i>	1.498	-0.04	-0.944	-0.381	0.151	0.01
<i>NLRC3</i>	1.495	0.053	-1.04	-0.249	0.101	-0.104
<i>TIGIT</i>	1.491	0.059	-0.994	-0.349	0.017	0.05
<i>WAS</i>	1.488	0.016	-1.207	-0.359	0.003	0.222
<i>SPN</i>	1.487	-0.137	-0.866	-0.28	0.171	-0.063
<i>SPOCK2</i>	1.478	0.118	-1.273	-0.369	0.06	0.129
<i>FCRL3</i>	1.47	0.069	-0.752	-0.348	-0.045	-0.015
<i>STAP1</i>	1.47	0.113	-0.702	-0.361	-0.068	-0.042
<i>P2RY8</i>	1.468	0.084	-0.732	-0.436	-0.086	0.092
<i>CD3G</i>	1.466	0.134	-1.178	-0.366	0.116	0.015
<i>VAV1</i>	1.464	-0.082	-0.944	-0.268	0.086	0.011
<i>CD2</i>	1.463	0.109	-1.009	-0.518	0.146	0.075
<i>SLAMF6</i>	1.46	0.024	-1.003	-0.333	-0.027	0.126
<i>NOD3</i>	1.46	0.064	-0.84	-0.314	-0.021	-0.016
<i>RAB37</i>	1.459	-0.019	-0.63	-0.208	-0.187	0
<i>ELMO1</i>	1.458	0.074	-1.074	-0.383	0.106	0.043
<i>LCK</i>	1.453	0.161	-1.044	-0.492	0.061	0.109
<i>IKZF1</i>	1.449	0.109	-1.268	-0.435	0.115	0.162
<i>FAM113B</i>	1.448	0.01	-1.371	-0.325	0.106	0.198
<i>NCRNA00219</i>	-0.275	0.643	0.676	-0.132	-0.476	-0.175
<i>TIGA1</i>	-0.279	0.631	0.66	-0.096	-0.492	-0.174

Gene	LMC class					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
<i>ARID5B</i>	0.165	0.615	0.301	-0.234	-0.413	-0.135
<i>HNRNPA2B1</i>	-0.133	0.607	0.666	-0.499	-0.479	0.164
<i>C20orf199</i>	-0.435	0.603	0.746	-0.003	-0.1	-0.585
<i>RPL15</i>	-0.072	0.601	0.698	-0.009	-0.334	-0.51
<i>CCNI</i>	0.085	0.594	0.628	-0.771	-0.272	0.161
<i>RPS13</i>	-0.134	0.593	0.481	-0.273	0.004	-0.418
<i>EPS15</i>	-0.183	0.591	0.556	-0.375	-0.51	0.164
<i>ANKRD36B</i>	0.179	0.587	0.276	-0.193	-0.205	-0.346
<i>C5orf53</i>	-0.005	0.58	0.491	-0.383	-0.467	0.085
<i>RPS3A</i>	-0.162	0.558	0.708	-0.188	-0.119	-0.459
<i>APEX1</i>	-0.349	0.547	0.836	-0.351	-0.252	-0.131
<i>ENOSF1</i>	-0.416	0.547	0.237	0.196	-0.169	-0.419
<i>ARGLU1</i>	0.113	0.544	0.47	-0.534	-0.479	0.231
<i>C9orf61</i>	-0.368	0.534	0.531	-0.083	-0.392	-0.089
<i>SUPV3L1</i>	0.032	0.53	0.327	-0.481	-0.081	-0.077
<i>CCNH</i>	0.169	0.526	0.427	-0.332	-0.281	-0.152
<i>PPP3CB</i>	0.502	0.525	-0.039	-0.554	-0.331	0.193
<i>KIAA0141</i>	0.066	0.523	0.564	-0.485	-0.36	0.062
<i>RPS15A</i>	0.2	0.523	0.566	-0.418	-0.091	-0.331
<i>UNC84A</i>	-0.048	0.522	0.389	-0.235	-0.257	-0.141
<i>PAN2</i>	-0.081	0.521	0.531	-0.44	-0.356	0.104
<i>ABCA10</i>	0.042	0.518	0.292	0.108	-0.452	-0.293
<i>HNRPA1L.2</i>	-0.267	0.516	0.568	-0.267	-0.116	-0.221
<i>CCT7</i>	-0.734	-0.402	0.101	0.765	0.114	-0.219
<i>SOCS6</i>	-0.71	-0.273	0.391	0.72	-0.159	-0.184
<i>BAHCC1</i>	-0.578	-0.249	0.384	0.709	-0.073	-0.339
<i>CLN6</i>	-0.513	-0.307	0.108	0.705	0.194	-0.432
<i>RBM8A</i>	-0.365	-0.198	0.079	0.692	0.122	-0.503
<i>POLR2K</i>	-0.514	-0.435	0.032	0.68	0.074	-0.142
<i>AGPAT1</i>	-0.397	-0.395	0.032	0.672	0.052	-0.205
<i>C17orf41</i>	-0.345	-0.356	0.357	0.666	-0.048	-0.329
<i>NDNL2</i>	-0.17	-0.443	0.193	0.659	-0.11	-0.191

Gene	LMC class					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
<i>C4orf23</i>	-0.385	-0.315	0.132	0.655	0.048	-0.31
<i>ISYNA1</i>	-0.803	-0.36	0.58	0.652	-0.353	0.094
<i>SCARNA11</i>	-0.002	-0.183	-0.212	0.65	0.063	-0.447
<i>PRPF39</i>	-0.4	-0.006	0.315	0.645	-0.243	-0.375
<i>ANKRD40</i>	-0.304	-0.28	0.167	0.643	0.075	-0.412
<i>PTPN9</i>	-0.506	-0.411	0.067	0.638	0.009	-0.079
<i>LOC100128164</i>	-0.305	-0.344	0.162	0.635	-0.275	-0.012
<i>HRH4</i>	-0.35	-0.089	0.5	0.633	-0.356	-0.302
<i>EPM2A</i>	-0.343	-0.212	0.425	0.629	-0.329	-0.183
<i>CLEC17A</i>	0.109	-0.317	0.072	0.628	-0.187	-0.273
<i>ZBTB6</i>	-0.165	-0.36	0.287	0.627	-0.183	-0.211
<i>FIZ1</i>	-0.174	-0.526	-0.043	0.624	-0.184	0.109
<i>NUTF2</i>	-0.557	-0.539	0.178	0.623	0.229	-0.197
<i>LOC100125556</i>	-0.737	-0.062	0.366	0.622	-0.064	-0.332
<i>CXorf64</i>	-0.552	-0.425	0.25	0.619	-0.072	-0.041
<i>C19orf12</i>	-0.181	-0.192	0.166	0.615	-0.175	-0.281
<i>MPZL2</i>	-0.1	-0.264	-1	0.14	1.048	-0.355
<i>IL20RB</i>	-0.071	-0.007	-0.849	-0.002	1.036	-0.513
<i>FAM83A</i>	-0.195	-0.274	-0.751	0.186	1.029	-0.454
<i>AQP3</i>	-0.062	-0.136	-0.976	0.146	1.027	-0.481
<i>PPL</i>	0.005	0.001	-0.783	-0.025	1.026	-0.559
<i>PVRL1</i>	-0.183	-0.143	-0.692	0.066	1.026	-0.478
<i>S100A9</i>	0.032	-0.193	-1.237	0.141	1.019	-0.332
<i>SERPINB3</i>	-0.309	-0.183	-0.752	0.183	1.008	-0.451
<i>TRIM16</i>	-0.057	-0.18	-0.539	0.067	1.003	-0.568
<i>TMEM45A</i>	-0.128	-0.134	-0.898	0.153	1.002	-0.474
<i>GRHL1</i>	0.033	-0.031	-1.091	0.16	1.002	-0.55
<i>FAM83C</i>	-0.133	-0.096	-0.796	0.181	1	-0.582
<i>GJA1</i>	0.087	-0.041	-1.17	0.067	1	-0.432
<i>GSDMC</i>	-0.114	-0.087	-0.772	0.016	0.997	-0.442
<i>RHOV</i>	-0.099	-0.14	-0.803	0.168	0.997	-0.541
<i>PPP1R13L</i>	0.093	-0.131	-1.095	0.208	0.995	-0.533

Gene	LMC class					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
<i>KLC3</i>	-0.286	-0.104	-0.847	0.208	0.993	-0.492
<i>ABCA12</i>	-0.187	-0.063	-0.829	0.167	0.992	-0.545
<i>PLEKHN1</i>	-0.016	-0.222	-0.903	0.204	0.989	-0.49
<i>GLTP</i>	-0.202	-0.112	-0.619	-0.084	0.989	-0.347
<i>FABP5</i>	-0.379	-0.196	-0.566	0.049	0.987	-0.346
<i>CD24</i>	-0.057	-0.116	-0.682	0.061	0.985	-0.526
<i>PGLYRP3</i>	-0.155	-0.064	-0.808	0.152	0.981	-0.545
<i>FAM110C</i>	0.011	-0.16	-0.837	0.112	0.98	-0.49
<i>PKP1</i>	-0.132	-0.094	-0.831	0.193	0.977	-0.556
<i>HSPA13</i>	-0.239	0.067	0.492	-0.469	-0.491	0.752
<i>PRAF2</i>	-0.14	0.022	0.353	-0.351	-0.526	0.727
<i>GNS</i>	-0.054	0.04	-0.008	-0.464	-0.27	0.723
<i>SDSL</i>	-0.241	-0.243	0.222	-0.348	-0.16	0.719
<i>C3AR1</i>	0.447	-0.052	-0.382	-0.473	-0.206	0.69
<i>VKORC1</i>	-0.494	-0.025	0.527	-0.35	-0.353	0.688
<i>CYTSA</i>	-0.162	0.156	0.296	-0.341	-0.571	0.686
<i>RNGTT</i>	-0.139	-0.138	0.196	-0.262	-0.332	0.669
<i>SNAPIN</i>	-0.582	-0.193	0.535	-0.099	-0.397	0.662
<i>GAL3ST4</i>	-0.605	-0.11	0.322	-0.188	-0.25	0.66
<i>CDR2</i>	0.302	0.123	0.216	-0.592	-0.45	0.659
<i>MUL1</i>	-0.35	-0.054	0.292	-0.363	-0.229	0.657
<i>YIPF5</i>	-0.382	0.182	0.578	-0.503	-0.429	0.655
<i>RAB23</i>	-0.167	-0.206	0.036	0.031	-0.462	0.652
<i>ATP6V0B</i>	-0.035	-0.178	-0.05	-0.519	0.072	0.65
<i>TOMM34</i>	-0.419	0.083	0.32	-0.308	-0.382	0.649
<i>CHPF</i>	-0.683	-0.103	0.448	-0.311	-0.136	0.643
<i>C11orf17</i>	-0.222	0.096	0.375	-0.348	-0.475	0.641
<i>ATP2B1</i>	0.224	-0.08	0.347	-0.294	-0.591	0.641
<i>SNX11</i>	0.156	-0.13	0.033	-0.345	-0.287	0.639
<i>SACM1L</i>	0.008	0.067	0.436	-0.368	-0.578	0.639
<i>BIRC2</i>	0.31	0.178	0.007	-0.487	-0.478	0.637
<i>RAB8B</i>	0.619	0.192	0.134	-0.597	-0.601	0.634

Gene	LMC class					
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
<i>CNN3</i>	-0.098	0.184	0.251	-0.475	-0.412	0.634
<i>UBLCP1</i>	-0.063	0.109	0.304	-0.504	-0.352	0.627
<i>CDC73</i>	-0.361	0.071	1.03	0.113	-0.583	0.035
<i>MKI67IP</i>	-0.686	0.135	1.027	0.013	-0.392	0.061
<i>KLHL12</i>	-0.725	0.172	1.022	-0.064	-0.429	0.165
<i>FAM172A</i>	-0.242	0.282	1.014	-0.082	-0.52	-0.062
<i>ZFAND1</i>	-0.611	0.401	1.008	-0.043	-0.522	-0.014
<i>LYSMD1</i>	-0.665	0.149	1.006	-0.005	-0.493	0.165
<i>TMEM55A</i>	-0.632	0.324	0.983	0.003	-0.578	0.083
<i>SNRPE</i>	-0.398	0.277	0.972	-0.18	-0.688	0.302
<i>PHF20</i>	-0.5	0.047	0.967	0.212	-0.444	-0.075
<i>TMEM133</i>	-0.654	0.166	0.966	0.225	-0.565	0.002
<i>C12orf23</i>	-0.547	-0.013	0.96	-0.165	-0.622	0.553
<i>PLA2G12A</i>	-0.508	0.291	0.95	0.061	-0.37	-0.191
<i>ZBTB41</i>	-0.595	-0.059	0.948	0.263	-0.414	-0.008
<i>ZFP106</i>	-0.883	0.226	0.948	0.316	-0.384	-0.19
<i>WDR3</i>	-0.615	0.386	0.947	-0.091	-0.551	0.108
<i>SSR1</i>	-0.389	-0.014	0.945	-0.142	-0.373	0.222
<i>BEND5</i>	-0.237	0.167	0.935	0.123	-0.409	-0.238
<i>BTBD3</i>	-1.073	0.076	0.931	0.325	-0.306	-0.041
<i>SLC9A5</i>	-0.675	0.259	0.923	0.294	-0.444	-0.23
<i>DENND5B</i>	-0.423	0.387	0.92	-0.324	-0.356	0.073
<i>TTPA</i>	-0.537	0.298	0.914	0.135	-0.526	-0.089
<i>OSBPL9</i>	-0.699	0.36	0.908	0.049	-0.321	-0.169
<i>KANK2</i>	-0.518	0.067	0.908	0.01	-0.29	0.002
<i>LOC441743</i>	-0.402	0.411	0.903	-0.294	-0.211	-0.119
<i>LASS2</i>	-0.703	-0.026	0.899	0.073	-0.351	0.174

Table 8.2 Summary of upregulated biological pathways in LMC class 1, FDR is the false discovery rate

Pathway	P value	FDR
NF-kappa B signaling pathway(K)	1.11×10^{-16}	1.10×10^{-14}
Chemokine signaling pathway(K)	1.11×10^{-16}	1.10×10^{-14}
Natural killer cell mediated cytotoxicity(K)	1.11×10^{-16}	1.10×10^{-14}
TCR signaling in naive CD4+ T cells(N)	1.11×10^{-16}	1.10×10^{-14}
TCR signaling in naive CD8+ T cells(N)	1.11×10^{-16}	1.10×10^{-14}
Cytokine-cytokine receptor interaction(K)	1.11×10^{-16}	1.10×10^{-14}
Osteoclast differentiation(K)	1.11×10^{-16}	1.10×10^{-14}
Hematopoietic cell lineage(K)	1.11×10^{-16}	1.10×10^{-14}
IL12-mediated signaling events(N)	4.44×10^{-16}	3.91×10^{-14}
GPVI-mediated activation cascade(R)	6.66×10^{-16}	5.26×10^{-14}
T cell receptor signaling pathway(K)	1.33×10^{-15}	9.59×10^{-14}
Measles(K)	2.11×10^{-15}	1.39×10^{-13}
Primary immunodeficiency(K)	2.89×10^{-15}	1.76×10^{-13}
Interferon gamma signaling(R)	1.13×10^{-14}	6.45×10^{-13}
Signaling by Interleukins(R)	6.89×10^{-14}	3.65×10^{-12}
Tuberculosis(K)	1.21×10^{-13}	5.92×10^{-12}
TNF signaling pathway(K)	1.82×10^{-13}	8.38×10^{-12}
Pathways in cancer(K)	5.21×10^{-13}	2.29×10^{-11}
B cell receptor signaling pathway(K)	6.05×10^{-12}	2.54×10^{-10}
Jak-STAT signaling pathway(K)	7.84×10^{-12}	3.06×10^{-10}
DAP12 interactions(R)	6.75×10^{-11}	2.57×10^{-9}
T cell activation(P)	8.65×10^{-11}	3.11×10^{-9}
Transcriptional misregulation in cancer(K)	1.01×10^{-10}	3.43×10^{-9}
Toll-like receptor signaling pathway(K)	1.16×10^{-10}	3.60×10^{-9}
Staphylococcus aureus infection(K)	1.16×10^{-10}	3.60×10^{-9}
Interferon alpha/beta signaling(R)	1.39×10^{-10}	4.18×10^{-9}
Gastrin-CREB signalling pathway via PKC and MAPK(R)	1.45×10^{-10}	4.19×10^{-9}
Costimulation by the CD28 family(R)	2.90×10^{-10}	8.12×10^{-9}
Chagas disease (American trypanosomiasis)(K)	3.05×10^{-10}	8.13×10^{-9}
CXCR4-mediated signaling events(N)	3.13×10^{-10}	8.13×10^{-9}
Cell adhesion molecules (CAMs)(K)	8.53×10^{-10}	2.13×10^{-8}

Pathway	P value	FDR
Fc-epsilon receptor I signaling in mast cells(N)	1.11 x 10 ⁻⁹	2.67 x 10 ⁻⁸
Leukocyte transendothelial migration(K)	1.53 x 10 ⁻⁹	3.67 x 10 ⁻⁸
Toll-Like Receptors Cascades(R)	1.70 x 10 ⁻⁹	3.91 x 10 ⁻⁸
Cell surface interactions at the vascular wall(R)	2.08 x 10 ⁻⁹	4.57 x 10 ⁻⁸
BCR signaling pathway(N)	2.67 x 10 ⁻⁹	5.88 x 10 ⁻⁸
Interleukin signaling pathway(P)	3.57 x 10 ⁻⁹	7.50 x 10 ⁻⁸
Leishmaniasis(K)	3.79 x 10 ⁻⁹	7.95 x 10 ⁻⁸
IL12 signaling mediated by STAT4(N)	4.37 x 10 ⁻⁹	8.73 x 10 ⁻⁸
Pertussis(K)	8.16 x 10 ⁻⁹	1.55 x 10 ⁻⁷
Signaling by SCF-KIT(R)	9.21 x 10 ⁻⁹	1.75 x 10 ⁻⁷
IL4-mediated signaling events(N)	9.79 x 10 ⁻⁹	1.76 x 10 ⁻⁷
Downstream signaling in naive CD8+ T cells(N)	9.79 x 10 ⁻⁹	1.76 x 10 ⁻⁷
HTLV-I infection(K)	1.06 x 10 ⁻⁹	1.91 x 10 ⁻⁷
Inflammatory bowel disease (IBD)(K)	1.28 x 10 ⁻⁸	2.18 x 10 ⁻⁷
Antigen processing and presentation(K)	1.33 x 10 ⁻⁸	2.27 x 10 ⁻⁷
Signalling by NGF(R)	1.50 x 10 ⁻⁸	2.40 x 10 ⁻⁷
Malaria(K)	1.54 x 10 ⁻⁸	2.47 x 10 ⁻⁷

Table 8.3 Summary of upregulated biological pathways in LMC class 2, FDR is the false discovery rate

Pathway	P value	FDR
Eukaryotic Translation Initiation(R)	1.49×10^{-14}	8.48×10^{-12}
Eukaryotic Translation Elongation(R)	2.85×10^{-13}	8.11×10^{-11}
Eukaryotic Translation Termination(R)	2.13×10^{-12}	4.06×10^{-10}
SRP-dependent cotranslational protein targeting to membrane(R)	5.35×10^{-12}	7.60×10^{-10}
Nonsense-Mediated Decay (NMD)(R)	6.40×10^{-11}	6.08×10^{-9}
Selenoamino acid metabolism(R)	6.40×10^{-11}	6.08×10^{-9}
Ribosome(K)	3.09×10^{-10}	2.50×10^{-8}
Processing of Capped Intron-Containing Pre-mRNA(R)	7.98×10^{-6}	5.30×10^{-4}
Regulation of nuclear SMAD2/3 signaling(N)	8.41×10^{-6}	5.30×10^{-4}
SUMOylation(R)	6.34×10^{-5}	3.61×10^{-3}
Regulation of cytoplasmic and nuclear SMAD2/3 signaling(N)	2.80×10^{-4}	0.0143
Internalization of ErbB1(N)	3.67×10^{-4}	0.0172
Signaling events mediated by focal adhesion kinase(N)	6.83×10^{-4}	0.0294
BMP receptor signaling(N)	7.46×10^{-4}	0.0298
Nongenotropic Androgen signaling(N)	9.51×10^{-4}	0.0362

Table 8.4 Summary of upregulated biological pathways in LMC class 3, FDR is the false discovery rate

Pathway	P value	FDR
Mitotic Metaphase and Anaphase(R)	5.42×10^{-12}	3.76×10^{-9}
Assembly of the primary cilium(R)	8.95×10^{-12}	3.76×10^{-9}
Mitotic Prometaphase(R)	1.50×10^{-10}	4.21×10^{-8}
Mitochondrial translation(R)	2.14×10^{-10}	4.48×10^{-8}
The citric acid (TCA) cycle and respiratory electron transport(R)	1.15×10^{-9}	1.93×10^{-7}
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription(R)	1.47×10^{-8}	2.06×10^{-6}
Ribosome(K)	2.78×10^{-8}	3.34×10^{-6}
Fanconi anemia pathway(K)	3.60×10^{-8}	3.78×10^{-6}
S Phase(R)	1.55×10^{-7}	1.44×10^{-5}
Cell cycle(K)	3.05×10^{-7}	2.56×10^{-5}
Processing of Capped Intron-Containing Pre-mRNA(R)	4.24×10^{-7}	3.22×10^{-5}
HDR through Homologous Recombination (HR) or Single Strand Annealing (SSA)(R)	1.17×10^{-6}	7.12×10^{-5}
Mitotic G1-G1/S phases(R)	1.27×10^{-6}	7.12×10^{-5}
Signaling by Rho GTPases(R)	1.28×10^{-6}	7.12×10^{-5}
Huntington's disease(K)	1.32×10^{-6}	7.12×10^{-5}
Nucleosome assembly(R)	1.37×10^{-6}	7.12×10^{-5}
SUMOylation(R)	1.64×10^{-6}	8.04×10^{-5}
Resolution of Abasic Sites (AP sites)(R)	2.81×10^{-6}	1.29×10^{-4}
Eukaryotic Translation Initiation(R)	5.19×10^{-6}	2.28×10^{-4}
Nucleotide Excision Repair(R)	7.14×10^{-6}	3.00×10^{-4}
Fanconi anemia pathway(N)	7.73×10^{-6}	3.09×10^{-4}
Mitotic G2-G2/M phases(R)	9.63×10^{-6}	3.53×10^{-4}
Oxidative phosphorylation(K)	9.82×10^{-6}	3.53×10^{-4}
SRP-dependent cotranslational protein targeting to membrane(R)	1.14×10^{-5}	3.99×10^{-4}
Base excision repair(K)	2.68×10^{-5}	8.86×10^{-4}

Table 8.5 Summary of upregulated biological pathways in LMC class 4, FDR is the false discovery rate

Pathways	P-value	FDR
Assembly of the primary cilium(R)	5.50×10^{-7}	4.10×10^{-4}
Processing of Capped Intron-Containing Pre-mRNA(R)	1.82×10^{-6}	6.79×10^{-4}
RNA Polymerase II Transcription(R)	2.93×10^{-6}	7.26×10^{-4}
p75(NTR)-mediated signaling(N)	5.12×10^{-6}	9.53×10^{-4}
Intrinsic Pathway for Apoptosis(R)	8.11×10^{-5}	0.0105
NoRC negatively regulates rRNA expression(R)	8.48×10^{-5}	0.0105
Syndecan-3-mediated signaling events(N)	1.31×10^{-4}	0.0139
Nonsense-Mediated Decay (NMD)(R)	2.34×10^{-4}	0.0218

Table 8.6 Summary of upregulated biological pathways in LMC class 5, FDR is the false discovery rate

Pathway	P value	FDR
Cell junction organization(R)	2.01×10^{-7}	1.58×10^{-5}
Validated transcriptional targets of AP1 family members Fra1 and Fra2(N)	1.22×10^{-7}	3.81×10^{-5}
Pathways in cancer(K)	1.46×10^{-7}	3.81×10^{-5}
EPH-Ephrin signaling(R)	5.67×10^{-7}	1.11×10^{-4}
Beta1 integrin cell surface interactions(N)	1.71×10^{-6}	2.24×10^{-4}
ECM-receptor interaction(K)	1.97×10^{-6}	2.24×10^{-4}
Extracellular matrix organization(R)	2.02×10^{-6}	2.24×10^{-4}
Validated transcriptional targets of TAp63 isoforms(N)	2.73×10^{-6}	2.65×10^{-4}
AP-1 transcription factor network(N)	3.34×10^{-6}	2.91×10^{-4}
Proteoglycans in cancer(K)	4.31×10^{-6}	3.36×10^{-4}
Axon guidance(K)	1.58×10^{-5}	1.12×10^{-3}
Hippo signaling pathway(K)	1.87×10^{-5}	1.21×10^{-3}
Ras signaling pathway(K)	2.74×10^{-5}	1.55×10^{-3}

Table 8.7 Summary of upregulated biological pathways in LMC class 6, FDR is the false discovery rate

Pathway	P value	FDR
Signaling by Rho GTPases(R)	3.90×10^{-8}	3.45×10^{-5}
Mitotic Metaphase and Anaphase(R)	7.98×10^{-8}	3.52×10^{-5}
Mitotic G1-G1/S phases(R)	5.27×10^{-7}	1.43×10^{-4}
Phagosome(K)	6.48×10^{-7}	1.43×10^{-4}
APC/C-mediated degradation of cell cycle proteins(R)	9.15×10^{-7}	1.61×10^{-4}
Regulation of retinoblastoma protein(N)	1.58×10^{-6}	2.33×10^{-4}
Iron uptake and transport(R)	2.88×10^{-6}	3.62×10^{-4}
Signaling by Insulin receptor(R)	8.82×10^{-6}	8.64×10^{-4}
Rheumatoid arthritis(K)	1.10×10^{-5}	8.64×10^{-4}
Mitochondrial translation(R)	1.10×10^{-5}	8.64×10^{-4}
Toll-Like Receptors Cascades(R)	1.15×10^{-5}	8.64×10^{-4}
Beta1 integrin cell surface interactions(N)	1.18×10^{-5}	8.64×10^{-4}
Oxidative phosphorylation(K)	1.30×10^{-5}	8.72×10^{-4}
Mitotic Prometaphase(R)	1.42×10^{-5}	8.97×10^{-4}
ROS, RNS production in response to bacteria(R)	2.15×10^{-5}	1.25×10^{-3}
Mitotic G2-G2/M phases(R)	2.27×10^{-5}	1.25×10^{-3}
Cell Cycle Checkpoints(R)	2.83×10^{-5}	1.36×10^{-3}

Table 8.8 Summary of downregulated biological pathways in LMC class 1, FDR is the false discovery rate

Pathway	P value	FDR
Mitotic G1-G1/S phases(R)	1.16×10^{-11}	5.62×10^{-9}
Mitochondrial translation(R)	1.44×10^{-11}	5.62×10^{-9}
Cell Cycle Checkpoints(R)	6.52×10^{-10}	1.70×10^{-7}
Cell cycle(K)	1.35×10^{-8}	2.63×10^{-6}
Mitotic Metaphase and Anaphase(R)	1.88×10^{-8}	2.93×10^{-6}
APC/C-mediated degradation of cell cycle proteins(R)	6.04×10^{-8}	7.51×10^{-6}
Signaling by Rho GTPases(R)	6.77×10^{-8}	7.51×10^{-6}
Validated targets of C-MYC transcriptional activation(N)	3.46×10^{-8}	3.35×10^{-5}
S Phase(R)	4.33×10^{-8}	3.72×10^{-5}
FOXM1 transcription factor network(N)	7.42×10^{-7}	5.79×10^{-5}
Nucleosome assembly(R)	9.63×10^{-7}	6.83×10^{-5}
The citric acid (TCA) cycle and respiratory electron transport(R)	1.47×10^{-6}	9.55×10^{-5}
Mitotic G2-G2/M phases(R)	1.67×10^{-6}	1.00×10^{-5}
Aurora B signaling(N)	3.44×10^{-6}	1.89×10^{-5}
Mitotic Prometaphase(R)	4.22×10^{-6}	2.20×10^{-5}
M/G1 Transition(R)	5.00×10^{-6}	2.40×10^{-5}
E2F transcription factor network(N)	1.22×10^{-5}	5.60×10^{-5}
Hedgehog 'off' state(R)	1.42×10^{-5}	6.10×10^{-5}
Regulation of DNA replication(R)	1.97×10^{-5}	8.07×10^{-4}
Parkinson's disease(K)	2.10×10^{-5}	8.18×10^{-4}
Nucleotide Excision Repair(R)	2.25×10^{-5}	8.34×10^{-4}

Table 8.9 Summary of downregulated biological pathways in LMC class 2, FDR is the false discovery rate

Pathway	P value	FDR
PLK1 signaling events(N)	2.12×10^{-8}	3.63×10^{-6}
Systemic lupus erythematosus(K)	2.20×10^{-8}	3.63×10^{-6}
Alcoholism(K)	2.93×10^{-8}	3.63×10^{-6}
ATR signaling pathway(N)	2.04×10^{-7}	1.90×10^{-5}
FOXO1 transcription factor network(N)	3.70×10^{-7}	2.74×10^{-5}
APC/C-mediated degradation of cell cycle proteins(R)	1.36×10^{-6}	8.42×10^{-5}
Signaling by Rho GTPases(R)	9.17×10^{-6}	4.47×10^{-4}
Mitotic G2-G2/M phases(R)	9.72×10^{-6}	4.47×10^{-4}
Mitotic Metaphase and Anaphase(R)	1.33×10^{-6}	5.44×10^{-4}
Progesterone-mediated oocyte maturation(K)	2.66×10^{-5}	9.85×10^{-4}
Mitotic Prometaphase(R)	5.43×10^{-5}	1.85×10^{-3}
Cell Cycle Checkpoints(R)	6.26×10^{-5}	1.94×10^{-3}
Oxidative Stress Induced Senescence(R)	7.25×10^{-5}	2.03×10^{-3}
Assembly of the primary cilium(R)	1.52×10^{-4}	3.87×10^{-3}
p73 transcription factor network(N)	1.61×10^{-4}	3.87×10^{-3}
Cell cycle(K)	1.84×10^{-4}	4.23×10^{-3}
NoRC negatively regulates rRNA expression(R)	3.38×10^{-4}	7.43×10^{-3}
Oocyte meiosis(K)	9.32×10^{-4}	0.0186
General transcription by RNA polymerase I(P)	1.04×10^{-3}	0.0197
Meiotic recombination(R)	1.32×10^{-3}	0.0238
Validated targets of C-MYC transcriptional activation(N)	1.63×10^{-3}	0.0277
regulators of bone mineralization(B)	2.52×10^{-3}	0.0409
HDR through Homologous Recombination (HR) or Single Strand Annealing (SSA)(R)	2.56×10^{-3}	0.0409
Mechanism of protein import into the nucleus(B)	2.99×10^{-3}	0.0422
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription(R)	3.01×10^{-3}	0.0422

**Table 8.10 Summary of downregulated biological pathways in LMC class 3,
FDR is the false discovery rate**

Pathway	P value	FDR
IL12-mediated signaling events(N)	1.11×10^{-16}	2.22×10^{-14}
Cytokine-cytokine receptor interaction(K)	1.11×10^{-16}	2.22×10^{-14}
Osteoclast differentiation(K)	1.11×10^{-16}	2.22×10^{-14}
Hematopoietic cell lineage(K)	1.11×10^{-16}	2.22×10^{-14}
TNF signaling pathway(K)	2.22×10^{-16}	3.55×10^{-14}
NF-kappa B signaling pathway(K)	3.33×10^{-16}	4.43×10^{-14}
Extracellular matrix organization(R)	1.59×10^{-14}	1.81×10^{-14}
Chemokine signaling pathway(K)	2.02×10^{-13}	2.02×10^{-14}
Pathways in cancer(K)	9.33×10^{-13}	8.31×10^{-11}
Natural killer cell mediated cytotoxicity(K)	2.05×10^{-12}	1.64×10^{-10}
GPVI-mediated activation cascade(R)	3.61×10^{-12}	2.60×10^{-10}
Primary immunodeficiency(K)	2.56×10^{-11}	1.57×10^{-9}
TCR signaling in naive CD8+ T cells(N)	2.58×10^{-11}	1.57×10^{-9}
T cell receptor signaling pathway(K)	3.79×10^{-11}	2.16×10^{-9}
Beta1 integrin cell surface interactions(N)	5.14×10^{-11}	2.72×10^{-9}
TCR signaling in naive CD4+ T cells(N)	7.06×10^{-11}	3.53×10^{-9}
Interferon gamma signaling(R)	1.17×10^{-10}	5.48×10^{-9}
Signaling by Interleukins(R)	1.70×10^{-10}	7.47×10^{-9}
DAP12 interactions(R)	2.28×10^{-10}	9.57×10^{-9}
Interferon alpha/beta signaling(R)	2.66×10^{-10}	1.07×10^{-8}
Downstream signaling in naive CD8+ T cells(N)	7.34×10^{-10}	2.79×10^{-8}
IL12 signaling mediated by STAT4(N)	9.47×10^{-10}	3.41×10^{-8}
Cell adhesion molecules (CAMs)(K)	1.95×10^{-9}	6.64×10^{-8}
Focal adhesion(K)	2.15×10^{-9}	7.11×10^{-8}
Cell surface interactions at the vascular wall(R)	4.18×10^{-9}	1.34×10^{-7}
Staphylococcus aureus infection(K)	6.16×10^{-9}	1.85×10^{-7}
Signaling by PDGF(R)	6.42×10^{-9}	1.86×10^{-7}
Signaling by SCF-KIT(R)	9.69×10^{-9}	2.71×10^{-7}
Fc gamma R-mediated phagocytosis(K)	1.12×10^{-8}	3.02×10^{-7}
Complement and coagulation cascades(K)	1.42×10^{-8}	3.64×10^{-7}
ECM-receptor interaction(K)	1.46×10^{-8}	3.64×10^{-7}

Pathway	P value	FDR
PI3K-Akt signaling pathway(K)	1.78×10^{-8}	4.33×10^{-7}
Measles(K)	1.81×10^{-8}	4.33×10^{-7}
IL23-mediated signaling events(N)	5.17×10^{-8}	1.19×10^{-6}
Costimulation by the CD28 family(R)	5.96×10^{-8}	1.31×10^{-6}
T cell activation(P)	6.06×10^{-8}	1.33×10^{-6}
Validated transcriptional targets of AP1 family members Fra1 and Fra2(N)	7.46×10^{-8}	1.57×10^{-6}

Table 8.11 Summary of downregulated biological pathways in LMC class 4, FDR is the false discovery rate

Pathway	P value	FDR
T cell receptor signaling pathway(K)	1.98×10^{-11}	1.75×10^{-8}
Osteoclast differentiation(K)	3.82×10^{-10}	1.69×10^{-7}
Natural killer cell mediated cytotoxicity(K)	6.75×10^{-10}	1.98×10^{-7}
Toll-Like Receptors Cascades(R)	1.72×10^{-9}	3.11×10^{-7}
TCR signaling in naive CD4+ T cells(N)	1.79×10^{-9}	3.11×10^{-7}
Interferon gamma signaling(R)	2.37×10^{-9}	3.11×10^{-7}
Measles(K)	2.47×10^{-9}	3.11×10^{-7}
Signalling by NGF(R)	5.11×10^{-9}	5.62×10^{-7}
IL12-mediated signaling events(N)	8.59×10^{-9}	8.42×10^{-7}
Primary immunodeficiency(K)	1.67×10^{-8}	1.47×10^{-6}
TCR signaling in naive CD8+ T cells(N)	3.11×10^{-8}	2.49×10^{-6}
T cell activation(P)	5.51×10^{-8}	4.02×10^{-6}
Chemokine signaling pathway(K)	7.40×10^{-8}	4.96×10^{-6}
TCR signaling(R)	1.18×10^{-7}	7.37×10^{-6}
B cell receptor signaling pathway(K)	1.34×10^{-7}	7.37×10^{-6}
Leishmaniasis(K)	1.34×10^{-7}	7.37×10^{-6}
Interferon alpha/beta signaling(R)	1.46×10^{-7}	7.46×10^{-6}
DAP12 interactions(R)	1.64×10^{-7}	7.76×10^{-6}
Tuberculosis(K)	1.69×10^{-7}	7.76×10^{-6}
Signaling by SCF-KIT(R)	2.22×10^{-7}	9.78×10^{-6}
Costimulation by the CD28 family(R)	3.20×10^{-7}	1.34×10^{-6}
Signaling by Interleukins(R)	4.04×10^{-7}	1.56×10^{-6}
NF-kappa B signaling pathway(K)	4.11×10^{-7}	1.56×10^{-6}
BCR signaling pathway(N)	5.07×10^{-7}	1.82×10^{-6}
Cell surface interactions at the vascular wall(R)	5.72×10^{-7}	2.00×10^{-5}
Signaling by Rho GTPases(R)	7.02×10^{-7}	2.32×10^{-5}
MAPK signaling pathway(K)	1.21×10^{-6}	3.88×10^{-5}
Cytokine-cytokine receptor interaction(K)	1.32×10^{-6}	4.10×10^{-5}
Downstream signaling in naive CD8+ T cells(N)	1.71×10^{-6}	5.14×10^{-5}
Fc gamma R-mediated phagocytosis(K)	1.77×10^{-6}	5.14×10^{-5}
CXCR4-mediated signaling events(N)	2.70×10^{-6}	7.56×10^{-5}

Pathway	P value	FDR
HTLV-I infection(K)	3.83×10^{-6}	1.00×10^{-4}
Pertussis(K)	3.96×10^{-6}	1.00×10^{-4}
Signaling by EGFR(R)	4.01×10^{-6}	1.00×10^{-4}
IL12 signaling mediated by STAT4(N)	4.08×10^{-6}	1.02×10^{-4}
Phagosome(K)	4.25×10^{-6}	1.02×10^{-4}
GPVI-mediated activation cascade(R)	4.57×10^{-6}	1.05×10^{-4}
Fc-epsilon receptor I signaling in mast cells(N)	4.58×10^{-6}	1.05×10^{-4}
Signaling by PDGF(R)	6.60×10^{-6}	1.45×10^{-4}
TNF signaling pathway(K)	9.06×10^{-6}	1.91×10^{-4}
Hematopoietic cell lineage(K)	9.08×10^{-6}	1.91×10^{-4}
Leukocyte transendothelial migration(K)	9.13×10^{-6}	1.92×10^{-4}
Chagas disease (American trypanosomiasis)(K)	1.18×10^{-5}	2.36×10^{-4}
Beta2 integrin cell surface interactions(N)	1.23×10^{-5}	2.46×10^{-4}
Signaling by VEGF(R)	1.54×10^{-5}	2.92×10^{-4}
Signaling by FGFR3(R)	1.67×10^{-5}	3.01×10^{-4}
Signaling by FGFR4(R)	1.67×10^{-5}	3.01×10^{-4}
Signaling by FGFR1(R)	1.81×10^{-5}	3.26×10^{-4}
Signaling by FGFR2(R)	1.81×10^{-5}	3.26×10^{-4}
amb2 Integrin signaling(N)	2.25×10^{-5}	3.82×10^{-4}
Class I PI3K signaling events(N)	2.98×10^{-5}	5.07×10^{-4}
Signaling by Insulin receptor(R)	3.25×10^{-5}	5.20×10^{-4}
EPO signaling pathway(N)	3.93×10^{-5}	6.29×10^{-4}
Pathways in cancer(K)	4.07×10^{-5}	6.52×10^{-4}
Fc epsilon RI signaling pathway(K)	4.47×10^{-5}	6.83×10^{-4}
Signaling by ERBB2(R)	4.56×10^{-5}	6.83×10^{-4}
Toll-like receptor signaling pathway(K)	4.70×10^{-5}	7.04×10^{-4}
Class I MHC mediated antigen processing & presentation(R)	5.40×10^{-5}	8.10×10^{-4}

**Table 8.12 Summary of downregulated biological pathways in LMC class 5,
FDR is the false discovery rate**

Pathway	P value	FDR
Wnt signaling pathway(P)	1.99×10^{-10}	1.56×10^{-7}
Assembly of the primary cilium(R)	6.85×10^{-8}	2.69×10^{-5}
Cadherin signaling pathway(P)	1.42×10^{-7}	3.20×10^{-5}
Processing of Capped Intron-Containing Pre-mRNA(R)	1.63×10^{-7}	3.20×10^{-5}
RNA Polymerase I, RNA Polymerase III, and Mitochondrial Transcription(R)	1.00×10^{-4}	0.0126
SUMOylation(R)	1.08×10^{-4}	0.0126
E2F transcription factor network(N)	1.13×10^{-4}	0.0126
Ubiquitin mediated proteolysis(K)	1.60×10^{-4}	0.0142
Regulation of retinoblastoma protein(N)	1.63×10^{-4}	0.0142
tRNA processing in the nucleus(R)	2.33×10^{-4}	0.0182

**Table 8.13 Summary of downregulated biological pathways in LMC class 6,
FDR is the false discovery rate**

Pathway	P value	FDR
Eukaryotic Translation Termination(R)	1.11×10^{-16}	1.35×10^{-14}
Eukaryotic Translation Initiation(R)	1.11×10^{-16}	1.35×10^{-14}
Nonsense-Mediated Decay (NMD)(R)	1.11×10^{-16}	1.35×10^{-14}
SRP-dependent cotranslational protein targeting to membrane(R)	1.11×10^{-16}	1.35×10^{-14}
Eukaryotic Translation Elongation(R)	1.11×10^{-16}	1.35×10^{-14}
Selenoamino acid metabolism(R)	1.11×10^{-16}	1.35×10^{-14}
Ribosome(K)	8.88×10^{-16}	9.24×10^{-14}
Cell junction organization(R)	2.06×10^{-8}	1.87×10^{-6}
Phosphatidylinositol signaling system(K)	1.13×10^{-6}	9.18×10^{-5}
Rap1 signaling pathway(K)	4.19×10^{-5}	3.06×10^{-3}
Regulation of actin cytoskeleton(K)	5.01×10^{-5}	3.30×10^{-3}
Oxytocin signaling pathway(K)	5.71×10^{-5}	3.48×10^{-3}
Focal adhesion(K)	1.07×10^{-4}	6.00×10^{-3}
Direct p53 effectors(N)	1.15×10^{-4}	6.00×10^{-3}
Regulation of Telomerase(N)	1.66×10^{-4}	7.96×10^{-3}
Tight junction(K)	1.91×10^{-4}	8.26×10^{-3}
Sphingolipid signaling pathway(K)	1.96×10^{-4}	8.26×10^{-3}
Stabilization and expansion of the E-cadherin adherens junction(N)	2.25×10^{-4}	8.26×10^{-3}
Inositol phosphate metabolism(K)	2.27×10^{-4}	8.26×10^{-3}
Generic Transcription Pathway(R)	2.30×10^{-4}	8.26×10^{-3}
Nectin adhesion pathway(N)	2.43×10^{-4}	8.26×10^{-3}
Adherens junction(K)	2.78×10^{-4}	8.96×10^{-3}
EPHA forward signaling(N)	2.89×10^{-4}	8.96×10^{-3}
Ras signaling pathway(K)	3.12×10^{-4}	9.36×10^{-3}

Table 8.14 Gini index value for 200 predictor genes of the refined RF model

Gene	Gini
RNF31	0.11
SEC11C	0.1
LMAN2L	0.1
SNORA12	0.09
ATP5A1	0.09
PPP1R3B	0.09
C1orf43	0.09
CBLN1	0.08
MS4A1	0.08
CCL19	0.08
CUGBP2	0.08
RCN2	0.08
NDUFAB1	0.08
RMI1	0.07
HSPA13	0.07
DNAJA3	0.07
MOV10	0.07
BTBD3	0.07
PHF20	0.07
DLL3	0.07
PRICKLE4	0.07
IFI44	0.07
YTHDF3	0.07
WDSOF1	0.07
UBL5	0.06
UBA2	0.06
LAMP3	0.06
NDUFB2	0.06
COPA	0.06
FCRLB	0.06
ZNF322A	0.06
ST13	0.06
NOL11	0.06
BOC	0.06
STMN1	0.06

Gene	Gini
FOXD1	0.06
IMPA1	0.06
MLL3	0.06
C21orf63	0.05
CACNB1	0.05
CENPQ	0.05
USP50	0.05
HDGFRP3	0.05
FAM65C	0.05
GMIP	0.05
ELP4	0.05
ZNF281	0.05
ABCB6	0.05
CHKB	0.05
MRPL13	0.05
RASA4	0.05
DOK2	0.05
TNFRSF18	0.05
KLHL12	0.05
RPIA	0.05
GBAS	0.05
C8orf33	0.05
CD1A	0.05
H19	0.05
GPR19	0.05
DDX28	0.05
ZC3H12D	0.05
PECI	0.05
SLC35E3	0.05
PLEKHG1	0.05
PBK	0.05
ATP6V1G1	0.05
C12orf23	0.05
PARP12	0.05
PRRC1	0.05

Gene	Gini
<i>ZCRB1</i>	0.05
<i>TYMS</i>	0.05
<i>P2RY4</i>	0.05
<i>CYBRD1</i>	0.05
<i>PDE1B</i>	0.05
<i>ATHL1</i>	0.05
<i>LPAR1</i>	0.05
<i>ACBD3</i>	0.05
<i>TATDN1</i>	0.05
<i>FLJ42986</i>	0.05
<i>C11orf46</i>	0.05
<i>CCDC17</i>	0.05
<i>MRPS28</i>	0.05
<i>PRDX4</i>	0.05
<i>PMP22</i>	0.05
<i>ST14</i>	0.05
<i>RAB27B</i>	0.05
<i>TUBB4</i>	0.05
<i>HERC6</i>	0.05
<i>CDK4</i>	0.05
<i>C5orf20</i>	0.05
<i>RPA3</i>	0.05
<i>DHX58</i>	0.05
<i>DDX60</i>	0.05
<i>NFKB2</i>	0.04
<i>NAB1</i>	0.04
<i>ZNF385A</i>	0.04
<i>COL29A1</i>	0.04
<i>ZNF318</i>	0.04
<i>NLRP1</i>	0.04
<i>KIAA1683</i>	0.04
<i>IL28RA</i>	0.04
<i>CD37</i>	0.04
<i>CCDC7</i>	0.04
<i>RAD54B</i>	0.04
<i>ITPR2</i>	0.04

Gene	Gini
<i>LAMA2</i>	0.04
<i>LASS2</i>	0.04
<i>CLIC2</i>	0.04
<i>PTTG1IP</i>	0.04
<i>HOXC5</i>	0.04
<i>MANEAL</i>	0.04
<i>VCPIP1</i>	0.04
<i>FH</i>	0.04
<i>PGCP</i>	0.04
<i>EPRS</i>	0.04
<i>UTP14C</i>	0.04
<i>RNASE2</i>	0.04
<i>ZNF512</i>	0.04
<i>GGH</i>	0.04
<i>ERI2</i>	0.04
<i>TMEM178</i>	0.04
<i>FANCC</i>	0.04
<i>CCR7</i>	0.04
<i>SEC24A</i>	0.04
<i>ZNF146</i>	0.04
<i>N6AMT2</i>	0.04
<i>ZUFSP</i>	0.04
<i>MED20</i>	0.04
<i>MRPL30</i>	0.04
<i>NDUFS2</i>	0.04
<i>PROK2</i>	0.04
<i>DCI</i>	0.04
<i>PIGM</i>	0.04
<i>PTK2</i>	0.04
<i>CBX4</i>	0.04
<i>PLAT</i>	0.04
<i>C6orf125</i>	0.04
<i>GATA4</i>	0.04
<i>NKD2</i>	0.04
<i>CENPF</i>	0.04
<i>BLM</i>	0.04

Gene	Gini
<i>TMEM117</i>	0.04
<i>BRPF3</i>	0.04
<i>C11orf21</i>	0.04
<i>SLC25A28</i>	0.04
<i>FCER1A</i>	0.04
<i>FKBP4</i>	0.04
<i>PTAR1</i>	0.04
<i>PPM1D</i>	0.04
<i>SNORD112</i>	0.04
<i>USP37</i>	0.04
<i>CBFA2T3</i>	0.04
<i>IFI44L</i>	0.04
<i>MVP</i>	0.04
<i>C6orf111</i>	0.04
<i>C1orf53</i>	0.04
<i>DHX15</i>	0.04
<i>IL16</i>	0.04
<i>F11R</i>	0.04
<i>EPHA4</i>	0.04
<i>MKI67IP</i>	0.04
<i>KIFC1</i>	0.04
<i>JMJD2C</i>	0.04
<i>DNAJC25</i>	0.04
<i>DCTPP1</i>	0.04
<i>KITLG</i>	0.04
<i>ANXA11</i>	0.04
<i>TMEM64</i>	0.04
<i>GFRA3</i>	0.04
<i>HOXB13</i>	0.03

Gene	Gini
<i>CYTH1</i>	0.03
<i>ETV6</i>	0.03
<i>C16orf59</i>	0.03
<i>PCDH21</i>	0.03
<i>OR2W3</i>	0.03
<i>FAM78B</i>	0.03
<i>COX5A</i>	0.03
<i>KIF4B</i>	0.03
<i>PCDHB10</i>	0.03
<i>HLA.DQB1</i>	0.03
<i>ERCC6L</i>	0.03
<i>LARP6</i>	0.03
<i>HOXB8</i>	0.03
<i>SFRP2</i>	0.03
<i>SNRPA1</i>	0.03
<i>KCNH4</i>	0.03
<i>PCGF1</i>	0.03
<i>SLC25A45</i>	0.03
<i>DKK1</i>	0.03
<i>CCDC48</i>	0.03
<i>PUS7</i>	0.03
<i>MIR24.1</i>	0.03
<i>OTUD6B</i>	0.03
<i>FOSL2</i>	0.03
<i>MRPL50</i>	0.03
<i>TMEM203</i>	0.03
<i>RSPO1</i>	0.03
<i>HBB</i>	0.03
<i>POMP</i>	0.03

References

1. J. McGrath, R. Eady, and F. Pope, *Anatomy and organization of human skin*. Rook's textbook of dermatology, 2004. **3**: p. 1-15.
2. E.N. Marieb and K. Hoehn, *Human anatomy & physiology*. 2007: Pearson Education.
3. C.M. Poole and D. Guerry, *Melanoma: prevention, detection, and treatment*. 2005: Yale University Press.
4. W. Westerhof, *The discovery of the human melanocyte*. Pigment Cell Research, 2006. **19**(3): p. 183-93.
5. S.S. Sulaimon and B.E. Kitchell, *The basic biology of malignant melanoma: molecular mechanisms of disease progression and comparative aspects*. Journal of Veterinary Internal Medicine, 2003. **17**(6): p. 760-72.
6. J.Y. Lin and D.E. Fisher, *Melanocyte biology and skin pigmentation*. Nature, 2007. **445**(7130): p. 843-50.
7. M. Eisinger and O. Marko, *Selective proliferation of normal human melanocytes in vitro in the presence of phorbol ester and cholera toxin*. Proceedings of the National Academy of Sciences, 1982. **79**(6): p. 2018-2022.
8. K.H. Kaidbey, P.P. Agin, R.M. Sayre, and A.M. Kligman, *Photoprotection by melanin—a comparison of black and Caucasian skin*. Journal of the American Academy of Dermatology, 1979. **1**(3): p. 249-260.
9. K.H. Kaidbey and A.M. Kligman, *The acute effects of long-wave ultraviolet radiation on human skin*. Journal of Investigative Dermatology, 1979. **72**(5): p. 253-6.
10. C. Griffiths, J. Barker, T. Bleiker, R. Chalmers, and D. Creamer, *Rook's textbook of dermatology*. 2016: John Wiley & Sons.
11. R.L. Barnhill, M. Piepkorn, and K.J. Busam, *Pathology of melanocytic nevi and malignant melanoma*. 2004: Springer Science & Business Media.
12. *Cancer Facts and Figures 2016*. American Cancer Society. Accessed September 4, 2016.
13. Cancer Research UK, *Skin cancer incidence statistics*. 2018. [cited 2nd October]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer>.

14. A.M. Forsea, V. Del Marmol, E. de Vries, E.E. Bailey, and A.C. Geller, *Melanoma incidence and mortality in Europe: new estimates, persistent disparities*. *British Journal of Dermatology*, 2012. **167**(5): p. 1124-30.
15. V. Nikolaou and A.J. Stratigos, *Emerging trends in the epidemiology of melanoma*. *British Journal of Dermatology*, 2014. **170**(1): p. 11-9.
16. D. Schadendorf, D.E. Fisher, C. Garbe, J.E. Gershenwald, J.-J. Grob, A. Halpern, M. Herlyn, M.A. Marchetti, G. McArthur, and A. Ribas, *Melanoma*. *Nature reviews Disease primers*, 2015. **1**: p. 15003.
17. E. Cho, B.A. Rosner, D. Feskanich, and G.A. Colditz, *Risk factors and individual probabilities of melanoma for whites*. *Journal of Clinical Oncology*, 2005. **23**(12): p. 2669-75.
18. Y.-m. Chang, J.H. Barrett, D.T. Bishop, B.K. Armstrong, V. Bataille, W. Bergman, M. Berwick, P.M. Bracci, J.M. Elwood, and M.S. Ernstoff, *Sun exposure and melanoma risk at different latitudes: a pooled analysis of 5700 cases and 7216 controls*. *International Journal of Epidemiology*, 2009. **38**(3): p. 814-830.
19. Cancer Research UK, *Risks and causes of melanoma*. 2018. [cited 2nd October]. Available from: <https://www.cancerresearchuk.org/about-cancer/melanoma/risks-causes>.
20. J.M. Bliss, D. Ford, A.J. Swerdlow, B.K. Armstrong, M. Cristofolini, J.M. Elwood, A. Green, E.A. Holly, T. Mack, R.M. MacKie, and et al., *Risk of cutaneous melanoma associated with pigmentation characteristics and freckling: systematic overview of 10 case-control studies*. *The International Melanoma Analysis Group (IMAGE)*. *International Journal of Cancer*, 1995. **62**(4): p. 367-76.
21. M.B. Veierød, H.-O. Adami, E. Lund, B.K. Armstrong, and E. Weiderpass, *Sun and solarium exposure and melanoma risk: effects of age, pigmentary characteristics, and nevi*. *Cancer Epidemiology and Prevention Biomarkers*, 2010. **19**(1): p. 111-120.
22. L.A. Cannon-Albright, D.E. Goldgar, L.J. Meyer, C.M. Lewis, D.E. Anderson, J.W. Fountain, M.E. Hegi, R.W. Wiseman, E.M. Petty, A.E. Bale, and et al., *Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22*. *Science*, 1992. **258**(5085): p. 1148-52.
23. N.A. Gruis, P.A. van der Velden, L.A. Sandkuijl, D.E. Prins, J. Weaver-Feldhaus, A. Kamb, W. Bergman, and R.R. Frants, *Homozygotes for CDKN2 (p16) germline mutation in Dutch familial melanoma kindreds*. *Nature Genetics*, 1995. **10**(3): p. 351-3.

24. N.A. Gruis, L.A. Sandkuijl, P.A. van der Velden, W. Bergman, and R.R. Frants, *CDKN2 explains part of the clinical phenotype in Dutch familial atypical multiple-mole melanoma (FAMMM) syndrome families*. *Melanoma Research*, 1995. **5**(3): p. 169-77.
25. A.M. Goldstein, M. Chan, M. Harland, E.M. Gillanders, N.K. Hayward, M.F. Avril, E. Azizi, G. Bianchi-Scarra, D.T. Bishop, B. Bressac-de Paillerets, W. Bruno, D. Calista, L.A. Cannon Albright, F. Demenais, D.E. Elder, P. Ghiorzo, N.A. Gruis, J. Hansson, D. Hogg, E.A. Holland, P.A. Kanetsky, R.F. Kefford, M.T. Landi, J. Lang, S.A. Leachman, R.M. Mackie, V. Magnusson, G.J. Mann, K. Niendorf, J. Newton Bishop, J.M. Palmer, S. Puig, J.A. Puig-Butlle, F.A. de Snoo, M. Stark, H. Tsao, M.A. Tucker, L. Whitaker, E. Yakobson, and C. Melanoma Genetics, *High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL*. *Cancer Research*, 2006. **66**(20): p. 9818-28.
26. A. Molven, M.B. Grimstvedt, S.J. Steine, M. Harland, M.F. Avril, N.K. Hayward, and L.A. Akslen, *A large Norwegian family with inherited malignant melanoma, multiple atypical nevi, and CDK4 mutation*. *Genes, Chromosomes and Cancer*, 2005. **44**(1): p. 10-8.
27. M. Carbone, L.K. Ferris, F. Baumann, A. Napolitano, C.A. Lum, E.G. Flores, G. Gaudino, A. Powers, P. Bryant-Greenwood, T. Krausz, E. Hyjek, R. Tate, J. Friedberg, T. Weigel, H.I. Pass, and H. Yang, *BAP1 cancer syndrome: malignant mesothelioma, uveal and cutaneous melanoma, and MBAITs*. *Journal of Translational Medicine*, 2012. **10**(1): p. 179.
28. C.D. Robles-Espinoza, M. Harland, A.J. Ramsay, L.G. Aoude, V. Quesada, Z. Ding, K.A. Pooley, A.L. Pritchard, J.C. Tiffen, M. Petljak, J.M. Palmer, J. Symmons, P. Johansson, M.S. Stark, M.G. Gartside, H. Snowden, G.W. Montgomery, N.G. Martin, J.Z. Liu, J. Choi, M. Makowski, K.M. Brown, A.M. Dunning, T.M. Keane, C. Lopez-Otin, N.A. Gruis, N.K. Hayward, D.T. Bishop, J.A. Newton-Bishop, and D.J. Adams, *POT1 loss-of-function variants predispose to familial melanoma*. *Nature Genetics*, 2014. **46**(5): p. 478-481.
29. J. Shi, X.R. Yang, B. Ballew, M. Rotunno, D. Calista, M.C. Fargnoli, P. Ghiorzo, B. Bressac-de Paillerets, E. Nagore, M.F. Avril, N.E. Caporaso, M.L. McMaster, M. Cullen, Z. Wang, X. Zhang, N.D.C.S.W. Group, N.D.C.G.R. Laboratory, G. French Familial Melanoma Study, W. Bruno, L. Pastorino, P. Queirolo, J. Banuls-Roca, Z. Garcia-Casado, A. Vaysse, H. Mohamdi, Y. Riazalhosseini, M. Foglio, F. Jouenne, X. Hua, P.L. Hyland, J. Yin, H. Vallabhaneni, W. Chai, P. Minghetti, C. Pellegrini, S. Ravichandran, A.

- Eggermont, M. Lathrop, K. Peris, G.B. Scarra, G. Landi, S.A. Savage, J.N. Sampson, J. He, M. Yeager, L.R. Goldin, F. Demenais, S.J. Chanock, M.A. Tucker, A.M. Goldstein, Y. Liu, and M.T. Landi, *Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma*. *Nature Genetics*, 2014. **46**(5): p. 482-6.
30. S. Horn, A. Figl, P.S. Rachakonda, C. Fischer, A. Sucker, A. Gast, S. Kadel, I. Moll, E. Nagore, K. Hemminki, D. Schadendorf, and R. Kumar, *TERT promoter mutations in familial and sporadic melanoma*. *Science*, 2013. **339**(6122): p. 959-61.
31. M. Harland, M. Petljak, C.D. Robles-Espinoza, Z. Ding, N.A. Gruis, R. van Doorn, K.A. Pooley, A.M. Dunning, L.G. Aoude, K.A. Wadt, A.M. Gerdes, K.M. Brown, N.K. Hayward, J.A. Newton-Bishop, D.J. Adams, and D.T. Bishop, *Germline TERT promoter mutations are rare in familial melanoma*. *Familial Cancer*, 2016. **15**(1): p. 139-44.
32. D.T. Bishop, F. Demenais, M.M. Iles, M. Harland, J.C. Taylor, E. Corda, J. Randerson-Moor, J.F. Aitken, M.F. Avril, E. Azizi, B. Bakker, G. Bianchi-Scarra, B. Bressac-de Paillerets, D. Calista, L.A. Cannon-Albright, A.W.T. Chin, T. Debniak, G. Galore-Haskel, P. Ghiorzo, I. Gut, J. Hansson, M. Hocevar, V. Hoiom, J.L. Hopper, C. Ingvar, P.A. Kanetsky, R.F. Kefford, M.T. Landi, J. Lang, J. Lubinski, R. Mackie, J. Malvehy, G.J. Mann, N.G. Martin, G.W. Montgomery, F.A. van Nieuwpoort, S. Novakovic, H. Olsson, S. Puig, M. Weiss, W. van Workum, D. Zelenika, K.M. Brown, A.M. Goldstein, E.M. Gillanders, A. Boland, P. Galan, D.E. Elder, N.A. Gruis, N.K. Hayward, G.M. Lathrop, J.H. Barrett, and J.A. Bishop, *Genome-wide association study identifies three loci associated with melanoma risk*. *Nature Genetics*, 2009. **41**(8): p. 920-5.
33. J.H. Barrett, M.M. Iles, M. Harland, J.C. Taylor, J.F. Aitken, P.A. Andresen, L.A. Akslen, B.K. Armstrong, M.F. Avril, E. Azizi, B. Bakker, W. Bergman, G. Bianchi-Scarra, B. Bressac-de Paillerets, D. Calista, L.A. Cannon-Albright, E. Corda, A.E. Cust, T. Debniak, D. Duffy, A.M. Dunning, D.F. Easton, E. Friedman, P. Galan, P. Ghiorzo, G.G. Giles, J. Hansson, M. Hocevar, V. Hoiom, J.L. Hopper, C. Ingvar, B. Janssen, M.A. Jenkins, G. Jonsson, R.F. Kefford, G. Landi, M.T. Landi, J. Lang, J. Lubinski, R. Mackie, J. Malvehy, N.G. Martin, A. Molven, G.W. Montgomery, F.A. van Nieuwpoort, S. Novakovic, H. Olsson, L. Pastorino, S. Puig, J.A. Puig-Butille, J. Randerson-Moor, H. Snowden, R. Tuominen, P. Van Belle, N. van der Stoep, D.C. Whiteman, D. Zelenika, J. Han, S. Fang, J.E. Lee, Q. Wei, G.M. Lathrop, E.M.

- Gillanders, K.M. Brown, A.M. Goldstein, P.A. Kanetsky, G.J. Mann, S. Macgregor, D.E. Elder, C.I. Amos, N.K. Hayward, N.A. Gruis, F. Demenais, J.A. Bishop, D.T. Bishop, and M.E.L.C. Geno, *Genome-wide association study identifies three new melanoma susceptibility loci*. *Nature Genetics*, 2011. **43**(11): p. 1108-13.
34. M.H. Law, D.T. Bishop, J.E. Lee, M. Brossard, N.G. Martin, E.K. Moses, F. Song, J.H. Barrett, R. Kumar, D.F. Easton, P.D.P. Pharoah, A.J. Swerdlow, K.P. Kypreou, J.C. Taylor, M. Harland, J. Randerson-Moor, L.A. Akslen, P.A. Andresen, M.F. Avril, E. Azizi, G.B. Scarra, K.M. Brown, T. Debniak, D.L. Duffy, D.E. Elder, S. Fang, E. Friedman, P. Galan, P. Ghiorzo, E.M. Gillanders, A.M. Goldstein, N.A. Gruis, J. Hansson, P. Helsing, M. Hocevar, V. Hoiom, C. Ingvar, P.A. Kanetsky, W.V. Chen, M.E.L.C. Geno, I. Essen-Heidelberg, S.D.H.S. Group, M. Q. Q. Investigators, A. Investigators, A.M.S. Group, M.T. Landi, J. Lang, G.M. Lathrop, J. Lubinski, R.M. Mackie, G.J. Mann, A. Molven, G.W. Montgomery, S. Novakovic, H. Olsson, S. Puig, J.A. Puig-Butille, A.A. Qureshi, G.L. Radford-Smith, N. van der Stoep, R. van Doorn, D.C. Whiteman, J.E. Craig, D. Schadendorf, L.A. Simms, K.P. Burdon, D.R. Nyholt, K.A. Pooley, N. Orr, A.J. Stratigos, A.E. Cust, S.V. Ward, N.K. Hayward, J. Han, H.J. Schulze, A.M. Dunning, J.A.N. Bishop, F. Demenais, C.I. Amos, S. MacGregor, and M.M. Iles, *Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma*. *Nature Genetics*, 2015. **47**(9): p. 987-995.
35. S. MacGregor, G.W. Montgomery, J.Z. Liu, Z.Z. Zhao, A.K. Henders, M. Stark, H. Schmid, E.A. Holland, D.L. Duffy, and M. Zhang, *Genome-wide association study identifies a new melanoma susceptibility locus at 1q21. 3*. *Nature Genetics*, 2011. **43**(11): p. 1114.
36. M.M. Iles, M.H. Law, S.N. Stacey, J. Han, S. Fang, R. Pfeiffer, M. Harland, S. MacGregor, J.C. Taylor, and K.K. Aben, *A variant in FTO shows association with melanoma risk not due to BMI*. *Nature Genetics*, 2013. **45**(4): p. 428.
37. M. Falchi, V. Bataille, N.K. Hayward, D.L. Duffy, J.A.N. Bishop, T. Pastinen, A. Cervino, Z.Z. Zhao, P. Deloukas, and N. Soranzo, *Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi*. *Nature Genetics*, 2009. **41**(8): p. 915.
38. C.I. Amos, L.E. Wang, J.E. Lee, J.E. Gershenwald, W.V. Chen, S. Fang, R. Kosoy, M. Zhang, A.A. Qureshi, S. Vattathil, C.W. Schacherer, J.M. Gardner, Y. Wang, D.T. Bishop, J.H. Barrett, M.E.L.I. Geno, S. MacGregor, N.K. Hayward, N.G. Martin, D.L. Duffy, Q.M. Investigators, G.J. Mann, A. Cust, J.

- Hopper, A. Investigators, K.M. Brown, E.A. Grimm, Y. Xu, Y. Han, K. Jing, C. McHugh, C.C. Laurie, K.F. Doheny, E.W. Pugh, M.F. Seldin, J. Han, and Q. Wei, *Genome-wide association study identifies novel loci predisposing to cutaneous melanoma*. *Human Molecular Genetics*, 2011. **20**(24): p. 5012-23.
39. K.M. Brown, S. Macgregor, G.W. Montgomery, D.W. Craig, Z.Z. Zhao, K. Iyadurai, A.K. Henders, N. Homer, M.J. Campbell, M. Stark, S. Thomas, H. Schmid, E.A. Holland, E.M. Gillanders, D.L. Duffy, J.A. Maskiell, J. Jetann, M. Ferguson, D.A. Stephan, A.E. Cust, D. Whiteman, A. Green, H. Olsson, S. Puig, P. Ghiorzo, J. Hansson, F. Demenais, A.M. Goldstein, N.A. Gruis, D.E. Elder, J.N. Bishop, R.F. Kefford, G.G. Giles, B.K. Armstrong, J.F. Aitken, J.L. Hopper, N.G. Martin, J.M. Trent, G.J. Mann, and N.K. Hayward, *Common sequence variants on 20q11.22 confer melanoma susceptibility*. *Nature Genetics*, 2008. **40**(7): p. 838-40.
40. Cancer Research UK, *Melanoma skin cancer survival statistics*. [cited 6th October]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer/survival>.
41. E. Nagore, V. Oliver, R. Botella-Estrada, S. Moreno-Picot, A. Insa, and J.M. Fortea, *Prognostic factors in localized invasive cutaneous melanoma: high value of mitotic rate, vascular invasion and microscopic satellitosis*. *Melanoma Research*, 2005. **15**(3): p. 169-177.
42. U. Leiter, P.G. Buettner, T.K. Eigentler, and C. Garbe, *Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the german dermatological society*. *Journal of Clinical Oncology*, 2004. **22**(18): p. 3660-7.
43. E. Zettersten, R.W. Sagebiel, J.R. Miller, 3rd, S. Tallapureddy, S.P. Leong, and M. Kashani-Sabet, *Prognostic factors in patients with thick cutaneous melanoma (> 4 mm)*. *Cancer*, 2002. **94**(4): p. 1049-56.
44. C.M. Balch, J.E. Gershenwald, S.J. Soong, J.F. Thompson, M.B. Atkins, D.R. Byrd, A.C. Buzaid, A.J. Cochran, D.G. Coit, S. Ding, A.M. Eggermont, K.T. Flaherty, P.A. Gimotty, J.M. Kirkwood, K.M. McMasters, M.C. Mihm, Jr., D.L. Morton, M.I. Ross, A.J. Sober, and V.K. Sondak, *Final version of 2009 AJCC melanoma staging and classification*. *Journal of Clinical Oncology*, 2009. **27**(36): p. 6199-206.
45. J.F. Thompson, S.J. Soong, C.M. Balch, J.E. Gershenwald, S. Ding, D.G. Coit, K.T. Flaherty, P.A. Gimotty, T. Johnson, M.M. Johnson, S.P. Leong, M.I. Ross, D.R. Byrd, N. Cascinelli, A.J. Cochran, A.M. Eggermont, K.M.

- McMasters, M.C. Mihm, Jr., D.L. Morton, and V.K. Sondak, *Prognostic significance of mitotic rate in localized primary cutaneous melanoma: an analysis of patients in the multi-institutional American Joint Committee on Cancer melanoma staging database*. *Journal of Clinical Oncology*, 2011. **29**(16): p. 2199-205.
46. B.A. Roach, A.L. Burton, M.P. Mays, B.A. Ginter, R.C. Martin, A.J. Stromberg, L. Hagendoorn, K.M. McMasters, and C.R. Scoggins, *Does mitotic rate predict sentinel lymph node metastasis or survival in patients with intermediate and thick melanoma? The American Journal of Surgery*, 2010. **200**(6): p. 759-764.
47. M. Mandalà, F. Galli, L. Cattaneo, B. Merelli, E. Rulli, S. Ribero, P. Quaglino, V. De Giorgi, J. Pigozzo, and V.C. Sileni, *Mitotic rate correlates with sentinel lymph node status and outcome in cutaneous melanoma greater than 1 millimeter in thickness: a multi-institutional study of 1524 cases*. *Journal of the American Academy of Dermatology*, 2017. **76**(2): p. 264-273. e2.
48. J.E. Gershenwald and R.A. Scolyer, *Melanoma Staging: American Joint Committee on Cancer (AJCC) 8th Edition and Beyond*. *Annals of Surgical Oncology*, 2018. **25**(8): p. 2105-2110.
49. S.R. Freeman, B.B. Gibbs, D.G. Brodland, and J.A. Zitelli, *Prognostic Value of Sentinel Lymph Node Biopsy Compared with that of Breslow Thickness: Implications for Informed Consent in Patients with Invasive Melanoma*. *Dermatologic Surgery*, 2013. **39**(12): p. 1800-1812.
50. H. Eriksson, M. Frohm-Nilsson, J. Jaras, L. Kanter-Lewensohn, P. Kjellman, E. Mansson-Brahme, I. Vassilaki, and J. Hansson, *Prognostic factors in localized invasive primary cutaneous malignant melanoma: results of a large population-based study*. *British Journal of Dermatology*, 2015. **172**(1): p. 175-86.
51. E.Z. Keung and J.E. Gershenwald, *The eighth edition American Joint Committee on Cancer (AJCC) melanoma staging system: implications for melanoma treatment and care*. *Expert Review of Anticancer Therapy*, 2018. **18**(8): p. 775-784.
52. R. Jewell, F. Elliott, J. Laye, J. Nsengimana, J. Davies, C. Walker, C. Conway, A. Mitra, M. Harland, M.G. Cook, A. Boon, S. Storr, S. Safuan, S.G. Martin, K. Jirstrom, H. Olsson, C. Ingvar, M. Lauss, T. Bishop, G. Jonsson, and J. Newton-Bishop, *The clinicopathological and gene expression patterns associated with ulceration of primary melanoma*. *Pigment Cell Melanoma Research*, 2015. **28**(1): p. 94-104.

53. A. Spatz, M.G. Cook, D.E. Elder, M. Piepkorn, D.J. Ruitter, and R.L. Barnhill, *Interobserver reproducibility of ulceration assessment in primary cutaneous melanomas*. *European Journal of Cancer*, 2003. **39**(13): p. 1861-5.
54. Macmillan Cancer Support, *Melanoma cancer staging and grading*. 2018. [cited 2nd October 2018]. Available from: <https://www.macmillan.org.uk/information-and-support/melanoma/treating/treatment-decisions/understanding-your-diagnosis/staging-and-grading.html#9913>.
55. K. Lasithiotakis, U. Leiter, F. Meier, T. Eigentler, G. Metzler, M. Moehrle, H. Breuninger, and C. Garbe, *Age and gender are significant independent predictors of survival in primary cutaneous melanoma*. *Cancer*, 2008. **112**(8): p. 1795-1804.
56. A. Joosse, E. de Vries, R. Eckel, T. Nijsten, A.M. Eggermont, D. Hölzel, J.W.W. Coebergh, and J. Engel, *Gender differences in melanoma survival: female patients have a decreased risk of metastasis*. *Journal of Investigative Dermatology*, 2011. **131**(3): p. 719-726.
57. S. Gupta, M. Artomov, W. Goggins, M. Daly, and H. Tsao, *Gender disparity and mutation burden in metastatic melanoma*. *Journal of the National Cancer Institute*, 2015. **107**(11).
58. L. Højberg, D. Gad, N. Gyldenkerne, L. Bastholt, and R. Academy of Geriatric Cancer, *Trends in melanoma in the elderly in Denmark, 1980-2012*. *Acta Oncologica*, 2016. **55 Suppl 1**(sup1): p. 52-8.
59. C.M. Balch, J.F. Thompson, J.E. Gershenwald, S.-j. Soong, S. Ding, K.M. McMasters, D.G. Coit, A.M. Eggermont, P.A. Gimotty, and T.M. Johnson, *Age as a predictor of sentinel node metastasis among patients with localized melanoma: an inverse correlation of melanoma mortality and incidence of sentinel node metastasis among young and old patients*. *Annals of Surgical Oncology*, 2014. **21**(4): p. 1075-1081.
60. C.M. Balch, S.-j. Soong, J.E. Gershenwald, J.F. Thompson, D.G. Coit, M.B. Atkins, S. Ding, A.J. Cochran, A.M. Eggermont, and K.T. Flaherty, *Age as a prognostic factor in patients with localized melanoma and regional metastases*. *Annals of Surgical Oncology*, 2013. **20**(12): p. 3961-3968.
61. S.A. Weiss, J. Han, F. Darvishian, J. Tchack, S.W. Han, K. Malecek, M. Krogsgaard, I. Osman, and J. Zhong, *Impact of aging on host immune response and survival in melanoma: an analysis of 3 patient cohorts*. *Journal of Translational Medicine*, 2016. **14**(1): p. 299.

62. G.G. Callender, M.E. Egger, A.L. Burton, C.R. Scoggins, M.I. Ross, A.J. Stromberg, L. Hagendoorn, C.R. Martin II, and K.M. McMasters, *Prognostic implications of anatomic location of primary cutaneous melanoma of 1 mm or thicker*. The American Journal of Surgery, 2011. **202**(6): p. 659-665.
63. W.H. Clark Jr, D.E. Elder, D. Guerry IV, L.E. Braitman, B.J. Trock, D. Schultz, M. Synnestvedt, and A.C. Halpern, *Model predicting survival in stage I melanoma based on tumor progression*. Journal of the National Cancer Institute, 1989. **81**(24): p. 1893-1904.
64. N.E. Thomas, K.J. Busam, L. From, A. Krickler, B.K. Armstrong, H. Anton-Culver, S.B. Gruber, R.P. Gallagher, R. Zanetti, and S. Rosso, *Tumor-infiltrating lymphocyte grade in primary melanomas is independently associated with melanoma-specific survival in the population-based genes, environment and melanoma study*. Journal of Clinical Oncology, 2013. **31**(33): p. 4252.
65. I.S. van Houdt, B.J. Sluijter, L.M. Moesbergen, W.M. Vos, T.D. de Gruijl, B.G. Molenkamp, A.J. van den Eertwegh, E. Hooijberg, P.A. van Leeuwen, and C.J. Meijer, *Favorable outcome in clinically stage II melanoma patients is associated with the presence of activated tumor infiltrating T-lymphocytes and preserved MHC class I antigen expression*. International Journal of Cancer, 2008. **123**(3): p. 609-615.
66. F. Tas and K. Erturk, *Histological lymphovascular invasion is associated with nodal involvement, recurrence, and survival in patients with cutaneous malignant melanoma*. International Journal of Dermatology, 2017. **56**(2): p. 166-170.
67. M.E. Egger, J.E. Gilbert, A.L. Burton, K.M. McMasters, G.G. Callender, A.R. Quillo, R.E. Brown, H. St, R. Charles, and L. Hagendoorn, *Lymphovascular invasion as a prognostic factor in melanoma*. The American Surgeon, 2011. **77**(8): p. 992-997.
68. A.E. Rose, P.J. Christos, D. Lackaye, R.L. Shapiro, R. Berman, M. Mazumdar, H. Kamino, I. Osman, and F. Darvishian, *Clinical relevance of detection of lymphovascular invasion in primary melanoma using endothelial markers D2-40 and CD34*. The American journal of surgical pathology, 2011. **35**(10): p. 1441.
69. S. Ribero, E. Moscarella, G. Ferrara, S. Piana, G. Argenziano, and C. Longo, *Regression in cutaneous melanoma: a comprehensive review from diagnosis to prognosis*. Journal of the European Academy of Dermatology and Venereology, 2016. **30**(12): p. 2030-2037.

70. J.C. Rubinstein, G. Han, L. Jackson, K. Bulloch, S. Ariyan, D. Narayan, B.G. Rothberg, and D. Han, *Regression in thin melanoma is associated with nodal recurrence after a negative sentinel node biopsy*. *Cancer medicine*, 2016. **5**(10): p. 2832-2840.
71. L.V. Kalialis, K.T. Drzewiecki, and H. Klyver, *Spontaneous regression of metastases from melanoma: review of the literature*. *Melanoma Research*, 2009. **19**(5): p. 275-282.
72. S. Ribero, S. Osella-Abate, M. Sanlorenzo, P. Savoia, C. Astrua, G. Cavaliere, C. Tomasini, R. Senetta, G. Macripo, M.G. Bernengo, and P. Quaglino, *Favourable prognostic role of regression of primary melanoma in AJCC stage I-II patients*. *British Journal of Dermatology*, 2013. **169**(6): p. 1240-5.
73. R. Botella-Estrada, V. Traves, C. Requena, C. Guillen-Barona, and E. Nagore, *Correlation of histologic regression in primary melanoma with sentinel node status*. *JAMA Dermatol*, 2014. **150**(8): p. 828-35.
74. A.L. Burton, J. Gilbert, R.W. Farmer, A.J. Stromberg, L. Hagendoorn, M.I. Ross, R.C. Martin, K.M. McMasters, C.R. Scoggins, and G.G. Callender, *Regression does not predict nodal metastasis or survival in patients with cutaneous melanoma*. *The American Surgeon*, 2011. **77**(8): p. 1009-1013.
75. P.M. Pollock, U.L. Harper, K.S. Hansen, L.M. Yudt, M. Stark, C.M. Robbins, T.Y. Moses, G. Hostetter, U. Wagner, J. Kakareka, G. Salem, T. Pohida, P. Heenan, P. Duray, O. Kallioniemi, N.K. Hayward, J.M. Trent, and P.S. Meltzer, *High frequency of BRAF mutations in nevi*. *Nature Genetics*, 2003. **33**(1): p. 19-20.
76. H. Davies, G.R. Bignell, C. Cox, P. Stephens, S. Edkins, S. Clegg, J. Teague, H. Woffendin, M.J. Garnett, W. Bottomley, N. Davis, E. Dicks, R. Ewing, Y. Floyd, K. Gray, S. Hall, R. Hawes, J. Hughes, V. Kosmidou, A. Menzies, C. Mould, A. Parker, C. Stevens, S. Watt, S. Hooper, R. Wilson, H. Jayatilake, B.A. Gusterson, C. Cooper, J. Shipley, D. Hargrave, K. Pritchard-Jones, N. Maitland, G. Chenevix-Trench, G.J. Riggins, D.D. Bigner, G. Palmieri, A. Cossu, A. Flanagan, A. Nicholson, J.W. Ho, S.Y. Leung, S.T. Yuen, B.L. Weber, H.F. Seigler, T.L. Darrow, H. Paterson, R. Marais, C.J. Marshall, R. Wooster, M.R. Stratton, and P.A. Futreal, *Mutations of the BRAF gene in human cancer*. *Nature*, 2002. **417**(6892): p. 949-54.
77. M. Colombino, M. Capone, A. Lissia, A. Cossu, C. Rubino, V. De Giorgi, D. Massi, E. Fonsatti, S. Staibano, and O. Nappi, *BRAF/NRAS mutation*

- frequencies among primary tumors and metastases in patients with melanoma.* Journal of Clinical Oncology, 2012. **30**(20): p. 2522-2529.
78. K. Omholt, A. Platz, L. Kanter, U. Ringborg, and J. Hansson, *NRAS and BRAF mutations arise early during melanoma pathogenesis and are preserved throughout tumor progression.* Clinical Cancer Research, 2003. **9**(17): p. 6483-8.
 79. M. Sensi, G. Nicolini, C. Petti, I. Bersani, F. Lozupone, A. Molla, C. Vegetti, D. Nonaka, R. Mortarini, and G. Parmiani, *Mutually exclusive NRAS Q61R and BRAF V600E mutations at the single-cell level in the same human melanoma.* Oncogene, 2006. **25**(24): p. 3357.
 80. J.M. Swick and J.C. Maize, Sr., *Molecular biology of melanoma.* Journal of the American Academy of Dermatology, 2012. **67**(5): p. 1049-54.
 81. The Cancer Genome Atlas Network, *Genomic classification of cutaneous melanoma.* Cell, 2015. **161**(7): p. 1681-1696.
 82. J.A. Curtin, J. Fridlyand, T. Kageshita, H.N. Patel, K.J. Busam, H. Kutzner, K.H. Cho, S. Aiba, E.B. Brocker, P.E. LeBoit, D. Pinkel, and B.C. Bastian, *Distinct sets of genetic alterations in melanoma.* New England Journal of Medicine, 2005. **353**(20): p. 2135-47.
 83. J.H. Lee, J.W. Choi, and Y.S. Kim, *Frequencies of BRAF and NRAS mutations are different in histological types and sites of origin of cutaneous melanoma: a meta-analysis.* British Journal of Dermatology, 2011. **164**(4): p. 776-84.
 84. G.J. Mann, G.M. Pupo, A.E. Campain, C.D. Carter, S.J. Schramm, S. Pianova, S.K. Gerega, C. De Silva, K. Lai, J.S. Wilmott, M. Synnott, P. Hersey, R.F. Kefford, J.F. Thompson, Y.H. Yang, and R.A. Scolyer, *BRAF mutation, NRAS mutation, and the absence of an immune-related expressed gene profile predict poor outcome in patients with stage III melanoma.* Journal of Investigative Dermatology, 2013. **133**(2): p. 509-17.
 85. B.C. Bastian, *The molecular pathology of melanoma: an integrated taxonomy of melanocytic neoplasia.* Annual Review of Pathology, 2014. **9**: p. 239-71.
 86. E. Hodis, I.R. Watson, G.V. Kryukov, S.T. Arold, M. Imielinski, J.P. Theurillat, E. Nickerson, D. Auclair, L. Li, C. Place, D. Dicara, A.H. Ramos, M.S. Lawrence, K. Cibulskis, A. Sivachenko, D. Voet, G. Saksena, N. Stransky, R.C. Onofrio, W. Winckler, K. Ardlie, N. Wagle, J. Wargo, K. Chong, D.L. Morton, K. Stemke-Hale, G. Chen, M. Noble, M. Meyerson, J.E. Ladbury, M.A. Davies, J.E. Gershenwald, S.N. Wagner, D.S. Hoon, D. Schadendorf,

- E.S. Lander, S.B. Gabriel, G. Getz, L.A. Garraway, and L. Chin, *A landscape of driver mutations in melanoma*. *Cell*, 2012. **150**(2): p. 251-63.
87. M. Krauthammer, Y. Kong, B.H. Ha, P. Evans, A. Bacchiocchi, J.P. McCusker, E. Cheng, M.J. Davis, G. Goh, M. Choi, S. Ariyan, D. Narayan, K. Dutton-Regester, A. Capatana, E.C. Holman, M. Bosenberg, M. Sznol, H.M. Kluger, D.E. Brash, D.F. Stern, M.A. Materin, R.S. Lo, S. Mane, S. Ma, K.K. Kidd, N.K. Hayward, R.P. Lifton, J. Schlessinger, T.J. Boggon, and R. Halaban, *Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma*. *Nature Genetics*, 2012. **44**(9): p. 1006-14.
88. A.H. Shain, I. Yeh, I. Kovalyshyn, A. Sriharan, E. Talevich, A. Gagnon, R. Dummer, J. North, L. Pincus, B. Ruben, W. Rickaby, C. D'Arrigo, A. Robson, and B.C. Bastian, *The Genetic Evolution of Melanoma from Precursor Lesions*. *New England Journal of Medicine*, 2015. **373**(20): p. 1926-36.
89. V. Trevino, F. Falciani, and H.A. Barrera-Saldana, *DNA microarrays: a powerful genomic tool for biomedical and clinical research*. *Molecular Medicine*, 2007. **13**(9-10): p. 527-41.
90. M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. *Science*, 1995. **270**(5235): p. 467-70.
91. G.A. Churchill, *Fundamentals of experimental design for cDNA microarrays*. *Nature Genetics*, 2002. **32 Suppl**: p. 490-5.
92. C. April, B. Klotzle, T. Royce, E. Wickham-Garcia, T. Boyaniwsky, J. Izzo, D. Cox, W. Jones, R. Rubio, K. Holton, U. Matulonis, J. Quackenbush, and J.B. Fan, *Whole-genome gene expression profiling of formalin-fixed, paraffin-embedded tissue samples*. *PLoS One*, 2009. **4**(12): p. e8162.
93. F. Medeiros, C.T. Rigl, G.G. Anderson, S.H. Becker, and K.C. Halling, *Tissue handling for genome-wide expression analysis: a review of the issues, evidence, and opportunities*. *Archives of Pathology and Laboratory Medicine*, 2007. **131**(12): p. 1805-16.
94. J.B. Fan, J.M. Yeakley, M. Bibikova, E. Chudin, E. Wickham, J. Chen, D. Doucet, P. Rigault, B. Zhang, R. Shen, C. McBride, H.R. Li, X.D. Fu, A. Oliphant, D.L. Barker, and M.S. Chee, *A versatile assay for high-throughput gene expression profiling on universal array matrices*. *Genome Research*, 2004. **14**(5): p. 878-85.
95. K. Kojima, C. April, C. Canasto-Chibuque, X. Chen, M. Deshmukh, A. Venkatesh, P.S. Tan, M. Kobayashi, H. Kumada, J.B. Fan, and Y. Hoshida, *Transcriptome profiling of archived sectioned formalin-fixed paraffin-*

- embedded (AS-FFPE) tissue for disease classification*. PloS One, 2014. **9**(1): p. e86961.
96. Z. Wang, M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nature Reviews Genetics, 2009. **10**(1): p. 57-63.
 97. F. Ozsolak and P.M. Milos, *RNA sequencing: advances, challenges and opportunities*. Nature reviews genetics, 2011. **12**(2): p. 87.
 98. DNA Microarray experiments, Access date. Oct 2018. https://commons.wikimedia.org/wiki/File:DNA_microarray_experiment.svg.
 99. J.H. Malone and B. Oliver, *Microarrays, deep sequencing and the true measure of the transcriptome*. BMC Biology, 2011. **9**(1): p. 34.
 100. V. Winnepenninckx, V. Lazar, S. Michiels, P. Dessen, M. Stas, S.R. Alonso, M.F. Avril, P.L. Ortiz Romero, T. Robert, O. Balacescu, A.M. Eggermont, G. Lenoir, A. Sarasin, T. Tursz, J.J. van den Oord, A. Spatz, R. Melanoma Group of the European Organization for, and C. Treatment of, *Gene expression profiling of primary cutaneous melanoma and clinical outcome*. Journal of the National Cancer Institute, 2006. **98**(7): p. 472-82.
 101. S.R. Alonso, L. Tracey, P. Ortiz, B. Pérez-Gómez, J. Palacios, M. Pollán, J. Linares, S. Serrano, A.I. Sáez-Castillo, and L. Sánchez, *A high-throughput study in melanoma identifies epithelial-mesenchymal transition as a major determinant of metastasis*. Cancer Research, 2007. **67**(7): p. 3450-3460.
 102. T. John, M.A. Black, T.T. Toro, D. Leader, C.A. Gedye, I.D. Davis, P.J. Guilford, and J.S. Cebon, *Predicting clinical outcome through molecular profiling in stage III melanoma*. Clinical Cancer Research, 2008. **14**(16): p. 5173-5180.
 103. D. Bogunovic, D.W. O'Neill, I. Belitskaya-Levy, V. Vacic, Y.L. Yu, S. Adams, F. Darvishian, R. Berman, R. Shapiro, A.C. Pavlick, S. Lonardi, J. Zavadil, I. Osman, and N. Bhardwaj, *Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(48): p. 20429-34.
 104. C. Conway, A. Mitra, R. Jewell, J. Randerson-Moor, S. Lobo, J. Nsengimana, S. Edward, D.S. Sanders, M. Cook, and B. Powell, *Gene expression profiling of paraffin-embedded primary melanoma using the DASL assay identifies increased osteopontin expression as predictive of reduced relapse-free survival*. Clinical Cancer Research, 2009. **15**(22): p. 6939-6946.
 105. K. Harbst, J. Staaf, M. Lauss, A. Karlsson, A. Masback, I. Johansson, P.O. Bendahl, J. Vallon-Christersson, T. Torngren, H. Ekedahl, J. Geisler, M.

- Hoglund, M. Ringner, L. Lundgren, K. Jirstrom, H. Olsson, C. Ingvar, A. Borg, H. Tsao, and G. Jonsson, *Molecular profiling reveals low- and high-grade forms of primary melanoma*. *Clinical Cancer Research*, 2012. **18**(15): p. 4026-36.
106. G. Jonsson, C. Busch, S. Knappskog, J. Geisler, H. Miletic, M. Ringner, J.R. Lillehaug, A. Borg, and P.E. Lonning, *Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome*. *Clinical Cancer Research*, 2010. **16**(13): p. 3356-67.
107. S.J. Schramm, A.E. Campain, R.A. Scolyer, Y.H. Yang, and G.J. Mann, *Review and cross-validation of gene expression signatures and melanoma prognosis*. *Journal of Investigative Dermatology*, 2012. **132**(2): p. 274-83.
108. G. Brunner, A. Heinecke, T.M. Falk, B. Ertas, N. Blödm-Schlicht, H.-J. Schulze, L. Suter, J. Atzpodien, and C. Berking, *A Prognostic Gene Signature Expressed in Primary Cutaneous Melanoma: Synergism With Conventional Staging*. *JNCI Cancer Spectrum*, 2018. **2**(3): p. pky032.
109. P. Gerami, R.W. Cook, J. Wilkinson, M.C. Russell, N. Dhillon, R.N. Amaria, R. Gonzalez, S. Lyle, C.E. Johnson, K.M. Oelschlager, G.L. Jackson, A.J. Greisinger, D. Maetzold, K.A. Delman, D.H. Lawson, and J.F. Stone, *Development of a prognostic genetic signature to predict the metastatic risk associated with cutaneous melanoma*. *Clinical Cancer Research*, 2015. **21**(1): p. 175-83.
110. S.-B. Cho and H.-H. Won. *Machine learning in DNA microarray analysis for cancer classification*. in *Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics 2003-Volume 19*. 2003. Australian Computer Society, Inc.
111. K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, and D.I. Fotiadis, *Machine learning applications in cancer prognosis and prediction*. *Comput Struct Biotechnol J*, 2015. **13**(Supplement C): p. 8-17.
112. L. Pan, G. Liu, F. Lin, S. Zhong, H. Xia, X. Sun, and H. Liang, *Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia*. *Scientific Reports*, 2017. **7**(1): p. 7402.
113. R. Jewell, C. Conway, A. Mitra, J. Randerson-Moor, S. Lobo, J. Nsengimana, M. Harland, M. Marples, S. Edward, M. Cook, B. Powell, A. Boon, F. de Kort, K.A. Parker, I.A. Cree, J.H. Barrett, M.A. Knowles, D.T. Bishop, and J. Newton-Bishop, *Patterns of expression of DNA repair genes and relapse from melanoma*. *Clinical Cancer Research*, 2010. **16**(21): p. 5211-21.

114. J. Nsengimana, J. Laye, A. Folia, C. Walker, R. Jewell, J.J. Van den Oord, P. Wolter, P. Patel, A. Sucker, D. Schadendorf, G.B. Jonsson, D.T. Bishop, and J. Newton-Bishop, *Independent replication of a melanoma subtype gene signature and evaluation of its prognostic value and biological correlates in a population cohort*. *Oncotarget*, 2015. **6**(13): p. 11683-93.
115. W. Liu, Y. Peng, and D.J. Tobin, *A new 12-gene diagnostic biomarker signature of melanoma revealed by integrated microarray analysis*. *PeerJ*, 2013. **1**: p. e49.
116. M. Lauss, J. Nsengimana, J. Staaf, J. Newton-Bishop, and G. Jonsson, *Consensus of Melanoma Gene Expression Subtypes Converges on Biological Entities*. *Journal of Investigative Dermatology*, 2016. **136**(12): p. 2502-2505.
117. J. Nsengimana, J. Laye, A. Folia, S. O'Shea, S. Muralidhar, J. Pozniak, A. Droop, M. Chan, C. Walker, L. Parkinson, J. Gascoyne, T. Mell, M. Polso, R. Jewell, J. Randerson-Moor, G.P. Cook, D.T. Bishop, and J. Newton-Bishop, *beta-Catenin-mediated immune evasion pathway frequently operates in primary cutaneous melanomas*. *Journal of Clinical Investigation*, 2018. **128**(5): p. 2048-2063.
118. M. Lauss, I. Visne, A. Kriegner, M. Ringner, G. Jonsson, and M. Hoglund, *Monitoring of technical variation in quantitative high-throughput datasets*. *Cancer Informatics*, 2013. **12**: p. 193-201.
119. F. Cordero, M. Botta, and R.A. Calogero, *Microarray data analysis and mining approaches*. *Brief Funct Genomic Proteomic*, 2007. **6**(4): p. 265-81.
120. B.M. Bolstad, R.A. Irizarry, M. Astrand, and T.P. Speed, *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias*. *Bioinformatics*, 2003. **19**(2): p. 185-93.
121. P. Du, W.A. Kibbe, and S.M. Lin, *lumi: a pipeline for processing Illumina microarray*. *Bioinformatics*, 2008. **24**(13): p. 1547-8.
122. R. Jewell, A. Mitra, C. Conway, J. Iremonger, C. Walker, F. de Kort, M. Cook, A. Boon, V. Speirs, and J. Newton-Bishop, *Identification of differentially expressed genes in matched formalin-fixed paraffin-embedded primary and metastatic melanoma tumor pairs*. *Pigment Cell Melanoma Research*, 2012. **25**(2): p. 284-6.
123. T.G. Clark, M.J. Bradburn, S.B. Love, and D.G. Altman, *Survival analysis part I: basic concepts and first analyses*. *British Journal of Cancer*, 2003. **89**(2): p. 232-8.

124. M. Gail, K. Krickeberg, J. Samet, A. Tsiatis, and W. Wong, *Statistics for biology and health*. 2007, Springer.
125. M.J. Bradburn, T.G. Clark, S.B. Love, and D.G. Altman, *Survival analysis part II: multivariate data analysis--an introduction to concepts and methods*. British Journal of Cancer, 2003. **89**(3): p. 431-6.
126. C. Wild and G. Seber, *The Wilcoxon rank-sum test*. 2011, Chapter.
127. G.W. Corder and D.I. Foreman, *Nonparametric statistics: A step-by-step approach*. 2014: John Wiley & Sons.
128. S. Boslaugh, *Statistics in a nutshell: A desktop quick reference*. 2012: "O'Reilly Media, Inc."
129. T.M. Therneau and T. Lumley, *Package 'survival'*. 2017, Verze.
130. H. Wickham, *ggplot2: elegant graphics for data analysis*. 2016: Springer.
131. D. Collett, *Modelling survival data in medical research*. 1993: Chapman and Hall/CRC.
132. A. Zhang, *Advanced analysis of gene expression microarray data*. Vol. 1. 2006: World Scientific Publishing Company.
133. R. Xu and D. Wunsch, 2nd, *Survey of clustering algorithms*. IEEE Trans Neural Netw, 2005. **16**(3): p. 645-78.
134. S. Bandyopadhyay and S. Bandyopadhyay, *Analysis of Biological Data: A Soft Computing Approach - Vol. 3*. 2007: World Scientific Publishing Co., Inc. 352.
135. L. Kaufman and P. Rousseeuw, *Clustering by means of medoids*. in 'Y. Dodge (editor) *Statistical Data Analysis based on L1 Norm*', 405-416. 1987, Elsevier/North-Holland.
136. J. Crowley and A. Hoering, *Handbook of statistics in clinical oncology*. 2012: p. 572-574.
137. S. Monti, P. Tamayo, J. Mesirov, and T. Golub, *Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data*. Machine learning, 2003. **52**(1-2): p. 91-118.
138. A.S. Shirkhorshidi, S. Aghabozorgi, and T.Y. Wah, *A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data*. PloS One, 2015. **10**(12): p. e0144059.
139. M.D. Wilkerson and D.N. Hayes, *ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking*. Bioinformatics, 2010. **26**(12): p. 1572-3.
140. J.F. Healey, *Statistics: A tool for social research*. 2014: Cengage Learning.

141. D. Meyer, A. Zeileis, and K. Hornik, *vcd: Visualizing Categorical Data. R package version 1.3-1*. 2013.
142. R. Xu and D.C. Wunsch, 2nd, *Clustering algorithms in biomedical research: a review*. IEEE Reviews in Biomedical Engineering, 2010. **3**: p. 120-54.
143. C. Wiwie, J. Baumbach, and R. Rottger, *Comparing the performance of biomedical clustering methods*. Nature Methods, 2015. **12**: p. 1033.
144. H. Cirenajwis, H. Ekedahl, M. Lauss, K. Harbst, A. Carneiro, J. Enoksson, F. Rosengren, L. Werner-Hartman, T. Tömgren, and A. Kvist, *Molecular stratification of metastatic melanoma using gene expression profiling: Prediction of survival outcome and benefit from molecular targeted therapy*. Oncotarget, 2015. **6**(14): p. 12297.
145. C. Hennig, *fpc: Flexible procedures for clustering. R package version 2.1-5*. 2013.
146. Michael Way. 2018.
147. A.P. Bradley, *The use of the area under the ROC curve in the evaluation of machine learning algorithms*. Pattern Recognition, 1997. **30**(7): p. 1145-1159.
148. K. Hajian-Tilaki, *Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation*. Caspian Journal of Internal Medicine, 2013. **4**(2): p. 627.
149. M. Sachs, *plotROC: Generate Useful ROC Curve Charts for Print and Interactive Use, 2016*. URL <http://sachsmc.github.io/plotROC>. R package version. **2**(1): p. 220.
150. T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer, *ROCR: visualizing classifier performance in R*. Bioinformatics, 2005. **21**(20): p. 3940-1.
151. R. Tibshirani, G. Chu, B. Narasimhan, and J. Li, *samr: SAM: Significance Analysis of Microarrays. R package version 2.0*. 2011.
152. V.G. Tusher, R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response*. Proceedings of the National Academy of Sciences, 2001. **98**(9): p. 5116-5121.
153. G. Chu, J. Li, B. Narasimhan, R. Tibshirani, and V. Tusher, *Significance analysis of microarrays users guide and technical document*. 2001.
154. G. Wu, E. Dawson, A. Duong, R. Haw, and L. Stein, *ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis*. F1000Res, 2014. **3**: p. 146.
155. P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, *Cytoscape: a software environment for*

- integrated models of biomolecular interaction networks*. Genome Research, 2003. **13**(11): p. 2498-504.
156. Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. Journal of the royal statistical society. Series B (Methodological), 1995: p. 289-300.
 157. Y. Senbabaoglu, G. Michailidis, and J.Z. Li, *Critical limitations of consensus clustering in class discovery*. Scientific Reports, 2014. **4**: p. 6207.
 158. S.M. Broekaert, R. Roy, I. Okamoto, J. van den Oord, J. Bauer, C. Garbe, R.L. Barnhill, K.J. Busam, A.J. Cochran, M.G. Cook, D.E. Elder, S.W. McCarthy, M.C. Mihm, D. Schadendorf, R.A. Scolyer, A. Spatz, and B.C. Bastian, *Genetic and morphologic features for melanoma classification*. Pigment Cell Melanoma Research, 2010. **23**(6): p. 763-70.
 159. M. Fallahi-Sichani, N.J. Moerke, M. Niepel, T. Zhang, N.S. Gray, and P.K. Sorger, *Systematic analysis of BRAF V 600E melanomas reveals a role for JNK/c-Jun pathway in adaptive resistance to drug-induced apoptosis*. Molecular Systems Biology, 2015. **11**(3): p. 797.
 160. S. Hu-Lieskovan, L. Robert, B. Homet Moreno, and A. Ribas, *Combining targeted therapy with immunotherapy in BRAF-mutant melanoma: promise and challenges*. Journal of Clinical Oncology, 2014. **32**(21): p. 2248-54.
 161. V. Gray-Schopfer, C. Wellbrock, and R. Marais, *Melanoma biology and new targeted therapy*. Nature, 2007. **445**(7130): p. 851-7.
 162. R. Sullivan, P. LoRusso, S. Boerner, and R. Dummer, *Achievements and challenges of molecular targeted therapy in melanoma*. American Society Clinical Oncology Education Book, 2015. **35**: p. 177-186.
 163. A.H. Shain and B.C. Bastian, *From melanocytes to melanomas*. Nature Reviews Cancer, 2016. **16**(6): p. 345-58.
 164. J. Jaeger, D. Koczan, H.J. Thiesen, S.M. Ibrahim, G. Gross, R. Spang, and M. Kunz, *Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues*. Clinical Cancer Research, 2007. **13**(3): p. 806-15.
 165. A.I. Riker, S.A. Enkemann, O. Fodstad, S. Liu, S. Ren, C. Morris, Y. Xi, P. Howell, B. Metge, R.S. Samant, L.A. Shevde, W. Li, S. Eschrich, A. Daud, J. Ju, and J. Matta, *The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis*. BMC Medical Genomics, 2008. **1**(1): p. 13.
 166. W. Hugo, J.M. Zaretsky, L. Sun, C. Song, B.H. Moreno, S. Hu-Lieskovan, B. Berent-Maoz, J. Pang, B. Chmielowski, G. Cherry, E. Seja, S. Lomeli, X.

- Kong, M.C. Kelley, J.A. Sosman, D.B. Johnson, A. Ribas, and R.S. Lo, *Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma*. *Cell*, 2016. **165**(1): p. 35-44.
167. E.M. Van Allen, D. Miao, B. Schilling, S.A. Shukla, C. Blank, L. Zimmer, A. Sucker, U. Hillen, M.H.G. Foppen, and S.M. Goldinger, *Genomic correlates of response to CTLA4 blockade in metastatic melanoma*. *Science*, 2015: p. aad0095.
168. M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M.M. Hoffman, *Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities*. *Inf Fusion*, 2019. **50**: p. 71-91.
169. H.A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A.B.H. Hassen, L. Thomas, A. Enk, L. Uhlmann, I. Reader study level, and I.I.G. level, *Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists*. *Annals of Oncology*, 2018. **29**(8): p. 1836-1842.
170. L. Shi, G. Campbell, W.D. Jones, F. Campagne, Z. Wen, S.J. Walker, Z. Su, T.M. Chu, F.M. Goodsaid, L. Pusztai, J.D. Shaughnessy, Jr., A. Oberthuer, R.S. Thomas, R.S. Paules, M. Fielden, B. Barlogie, W. Chen, P. Du, M. Fischer, C. Furlanello, B.D. Gallas, X. Ge, D.B. Megherbi, W.F. Symmans, M.D. Wang, J. Zhang, H. Bitter, B. Brors, P.R. Bushel, M. Bylesjo, M. Chen, J. Cheng, J. Cheng, J. Chou, T.S. Davison, M. Delorenzi, Y. Deng, V. Devanarayan, D.J. Dix, J. Dopazo, K.C. Dorff, F. Elloumi, J. Fan, S. Fan, X. Fan, H. Fang, N. Gonzaludo, K.R. Hess, H. Hong, J. Huan, R.A. Irizarry, R. Judson, D. Juraeva, S. Lababidi, C.G. Lambert, L. Li, Y. Li, Z. Li, S.M. Lin, G. Liu, E.K. Lobenhofer, J. Luo, W. Luo, M.N. McCall, Y. Nikolsky, G.A. Pennello, R.G. Perkins, R. Philip, V. Popovici, N.D. Price, F. Qian, A. Scherer, T. Shi, W. Shi, J. Sung, D. Thierry-Mieg, J. Thierry-Mieg, V. Thodima, J. Trygg, L. Vishnuvajjala, S.J. Wang, J. Wu, Y. Wu, Q. Xie, W.A. Yousef, L. Zhang, X. Zhang, S. Zhong, Y. Zhou, S. Zhu, D. Arasappan, W. Bao, A.B. Lucas, F. Berthold, R.J. Brennan, A. Buness, J.G. Catalano, C. Chang, R. Chen, Y. Cheng, J. Cui, W. Czika, F. Demichelis, X. Deng, D. Dosymbekov, R. Eils, Y. Feng, J. Fostel, S. Fulmer-Smentek, J.C. Fuscoe, L. Gatto, W. Ge, D.R. Goldstein, L. Guo, D.N. Halbert, J. Han, S.C. Harris, C. Hatzis, D. Herman, J. Huang, R.V. Jensen, R. Jiang, C.D. Johnson, G. Jurman, Y. Kahlert, S.A. Khuder, M. Kohl, J. Li, L. Li, M. Li, Q.Z. Li, S. Li, Z. Li, J. Liu, Y. Liu, Z. Liu, L. Meng, M. Madera, F. Martinez-Murillo, I. Medina, J. Meehan, K. Miclaus, R.A.

- Moffitt, D. Montaner, P. Mukherjee, G.J. Mulligan, P. Neville, T. Nikolskaya, B. Ning, G.P. Page, J. Parker, R.M. Parry, X. Peng, R.L. Peterson, J.H. Phan, B. Quanz, Y. Ren, S. Riccadonna, A.H. Roter, F.W. Samuelson, M.M. Schumacher, J.D. Shambaugh, Q. Shi, R. Shippy, S. Si, A. Smalter, C. Sotiriou, M. Soukup, F. Staedtler, G. Steiner, T.H. Stokes, Q. Sun, P.Y. Tan, R. Tang, Z. Tezak, B. Thorn, M. Tsyganova, Y. Turpaz, S.C. Vega, R. Visintainer, J. von Frese, C. Wang, E. Wang, J. Wang, W. Wang, F. Westermann, J.C. Willey, M. Woods, S. Wu, N. Xiao, J. Xu, L. Xu, L. Yang, X. Zeng, J. Zhang, L. Zhang, M. Zhang, C. Zhao, R.K. Puri, U. Scherf, W. Tong, R.D. Wolfinger and M. Consortium, *The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models*. Nature Biotechnology, 2010. **28**(8): p. 827-38.
171. L. Breiman, *Random forests*. Machine Learning, 2001. **45**(1): p. 5-32.
172. D. Meyer and F.T. Wien, *Support vector machines*. R News, 2001. **1**(3): p. 23-26.
173. S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, *Handling imbalanced datasets: A review*. GESTS International Transactions on Computer Science and Engineering, 2006. **30**(1): p. 25-36.
174. J. Cohen, *A coefficient of agreement for nominal scales*. Educational and Psychological Measurement, 1960. **20**(1): p. 37-46.
175. A.J. Viera and J.M. Garrett, *Understanding interobserver agreement: the kappa statistic*. Family Medicine, 2005. **37**(5): p. 360-3.
176. J.R. Landis and G.G. Koch, *The measurement of observer agreement for categorical data*. Biometrics, 1977. **33**(1): p. 159-74.
177. A. Liaw and M. Wiener, *Classification and regression by randomForest*. R News, 2002. **2**(3): p. 18-22.
178. G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Vol. 112. 2013: Springer.
179. L. Breiman, *Bagging predictors*. Machine learning, 1996. **24**(2): p. 123-140.
180. M. Kuhn, *Caret package*. Journal of statistical software, 2008. **28**(5): p. 1-26.
181. S. RColorBrewer and M.A. Liaw, *Package 'randomForest'*. 2018.
182. N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, *SMOTE: synthetic minority over-sampling technique*. Journal of Artificial Intelligence Research, 2002. **16**: p. 321-357.

183. A. Estabrooks, T. Jo, and N. Japkowicz, *A multiple resampling method for learning from imbalanced data sets*. Computational Intelligence, 2004. **20**(1): p. 18-36.
184. C. Cortes and V. Vapnik, *Support-vector networks*. Machine Learning, 1995. **20**(3): p. 273-297.
185. B.E. Boser, I.M. Guyon, and V.N. Vapnik. *A training algorithm for optimal margin classifiers*. in *Proceedings of the fifth annual workshop on Computational learning theory*. 1992. ACM.
186. C.-W. Hsu, C.-C. Chang, and C.-J. Lin, *A practical guide to support vector classification*. 2003.
187. T. Eitrich and B. Lang, *Efficient optimization of support vector machine learning parameters for unbalanced datasets*. Journal of computational and applied mathematics, 2006. **196**(2): p. 425-436.
188. E. Byvatov and G. Schneider, *Support vector machine applications in bioinformatics*. Applied Bioinformatics, 2003. **2**(2): p. 67-77.
189. T. Fushiki, *Estimation of prediction error by using K-fold cross-validation*. Statistics and Computing, 2011. **21**(2): p. 137-146.
190. A. Karatzoglou, A. Smola, K. Hornik, and M.A. Karatzoglou, *Package 'kernlab'*. 2018.
191. H. Asri, H. Mousannif, H.A. Moatassime, and T. Noel, *Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis*. Procedia Computer Science, 2016. **83**(Supplement C): p. 1064-1069.
192. S. Gupta, T. Tran, W. Luo, D. Phung, R.L. Kennedy, A. Broad, D. Campbell, D. Kipp, M. Singh, and M. Khasraw, *Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry*. BMJ open, 2014. **4**(3): p. e004007.
193. L. Mervic, *Time course and pattern of metastasis of cutaneous melanoma differ between men and women*. PloS One, 2012. **7**(3): p. e32955.
194. V. Ganganwar, *An overview of classification algorithms for imbalanced datasets*. International Journal of Emerging Technology and Advanced Engineering, 2012. **2**(4): p. 42-47.
195. R. Akbani, S. Kwek, and N. Japkowicz. *Applying support vector machines to imbalanced datasets*. in *European conference on machine learning*. 2004. Springer.
196. J. Gong and H. Kim, *RHSBoost: Improving classification performance in imbalance data*. Computational Statistics & Data Analysis, 2017. **111**: p. 1-13.

197. M. Lauss, M. Donia, K. Harbst, R. Andersen, S. Mitra, F. Rosengren, M. Salim, J. Vallon-Christersson, T. Torngren, A. Kvist, M. Ringner, I.M. Svane, and G. Jonsson, *Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma*. *Nature Communication*, 2017. **8**(1): p. 1738.
198. M.N. Wright and A. Ziegler, *Ranger: a fast implementation of random forests for high dimensional data in C++ and R*. arXiv preprint arXiv:1508.04409, 2015.
199. X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, and M. Muller, *pROC: an open-source package for R and S+ to analyze and compare ROC curves*. *BMC Bioinformatics*, 2011. **12**(1): p. 77.
200. I. Fellows, *wordcloud: Word clouds*. R package version, 2012. **2**: p. 109.
201. E. Neuwirth, *RColorBrewer: ColorBrewer palettes*. R package version 1.1-2. 2014.
202. J.A. Newton-Bishop, J.R. Davies, F. Latheef, J. Randerson-Moor, M. Chan, J. Gascoyne, S. Waseem, S. Haynes, C. O'Donovan, and D.T. Bishop, *25-Hydroxyvitamin D2/D3 levels and factors associated with systemic inflammation and melanoma survival in the Leeds Melanoma Cohort*. *International Journal of Cancer*, 2015. **136**(12): p. 2890-2899.
203. F. Bagherzadeh-Khiabani, A. Ramezankhani, F. Azizi, F. Hadaegh, E.W. Steyerberg, and D. Khalili, *A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results*. *Journal of Clinical Epidemiology*, 2016. **71**: p. 76-85.
204. R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, *Variable selection using random forests*. *Pattern Recognition Letters*, 2010. **31**(14): p. 2225-2236.
205. I. Guyon and A. Elisseeff, *An introduction to variable and feature selection*. *Journal of Machine Learning Research*, 2003. **3**(Mar): p. 1157-1182.
206. A.P. Cogdill, M.C. Andrews, and J.A. Wargo, *Hallmarks of response to immune checkpoint blockade*. *British Journal of Cancer*, 2017. **117**(1): p. 1-7.
207. P. Knipscheer, M. Raschle, A. Smogorzewska, M. Enoiu, T.V. Ho, O.D. Scharer, S.J. Elledge, and J.C. Walter, *The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair*. *Science*, 2009. **326**(5960): p. 1698-701.
208. L.J. Niedernhofer, A.S. Lalai, and J.H. Hoeijmakers, *Fanconi anemia (cross)linked to DNA repair*. *Cell*, 2005. **123**(7): p. 1191-8.
209. G.-L. Moldovan and A.D. D'Andrea, *How the fanconi anemia pathway guards the genome*. *Annual Review of Genetics*, 2009. **43**: p. 223-249.

210. A.D. D'Andrea, *Susceptibility pathways in Fanconi's anemia and breast cancer*. New England Journal of Medicine, 2010. **362**(20): p. 1909-1919.
211. N.G. Howlett, T. Taniguchi, S. Olson, B. Cox, Q. Waisfisz, C. De Die-Smulders, N. Persky, M. Grompe, H. Joenje, G. Pals, H. Ikeda, E.A. Fox, and A.D. D'Andrea, *Biallelic inactivation of BRCA2 in Fanconi anemia*. Science, 2002. **297**(5581): p. 606-9.
212. M. Nepal, R. Che, J. Zhang, C. Ma, and P. Fei, *Fanconi Anemia Signaling and Cancer*. Trends Cancer, 2017. **3**(12): p. 840-856.
213. T. Taniguchi, M. Tischkowitz, N. Ameziane, S.V. Hodgson, C.G. Mathew, H. Joenje, S.C. Mok, and A.D. D'Andrea, *Disruption of the Fanconi anemia–BRCA pathway in cisplatin-sensitive ovarian tumors*. Nature Medicine, 2003. **9**(5): p. 568.
214. M.S. van der Heijden, C.J. Yeo, R.H. Hruban, and S.E. Kern, *Fanconi anemia gene mutations in young-onset pancreatic cancer*. Cancer Research, 2003. **63**(10): p. 2585-8.
215. M.A. Davies, *The role of the PI3K-AKT pathway in melanoma*. Cancer Journal, 2012. **18**(2): p. 142-7.
216. F. Meier, B. Schitteck, S. Busch, C. Garbe, K. Smalley, K. Satyamoorthy, G. Li, and M. Herlyn, *The RAS/RAF/MEK/ERK and PI3K/AKT signaling pathways present molecular targets for the effective treatment of advanced melanoma*. Frontiers in Bioscience, 2005. **10**(2986–3001): p. 2986-3001.
217. T. Sinnberg, K. Lasithiotakis, H. Niessner, B. Schitteck, K.T. Flaherty, D. Kulms, E. Maczey, M. Campos, J. Gogel, C. Garbe, and F. Meier, *Inhibition of PI3K-AKT-mTOR signaling sensitizes melanoma cells to cisplatin and temozolomide*. Journal of Investigative Dermatology, 2009. **129**(6): p. 1500-15.
218. J. Villanueva, A. Vultur, J.T. Lee, R. Somasundaram, M. Fukunaga-Kalabis, A.K. Cipolla, B. Wubbenhorst, X. Xu, P.A. Gimotty, and D. Kee, *Acquired resistance to BRAF inhibitors mediated by a RAF kinase switch in melanoma can be overcome by cotargeting MEK and IGF-1R/PI3K*. Cancer Cell, 2010. **18**(6): p. 683-695.
219. Y. Ye, L. Jin, J.S. Wilmott, W.L. Hu, B. Yosufi, R.F. Thorne, T. Liu, H. Rizos, X.G. Yan, and L. Dong, *PI (4, 5) P2 5-phosphatase A regulates PI3K/Akt signalling and has a tumour suppressive role in human melanoma*. Nature communications, 2013. **4**: p. 1508.
220. A.E. Teschendorff, A. Miremadi, S.E. Pinder, I.O. Ellis, and C. Caldas, *An immune response gene expression module identifies a good prognosis*

- subtype in estrogen receptor negative breast cancer*. *Genome Biology*, 2007. **8**(8): p. R157.
221. M. Neagu, C. Constantin, and S. Zurac, *Immune parameters in the prognosis and therapy monitoring of cutaneous melanoma patients: experience, role, and limitations*. *Biomed Research International*, 2013. **2013**: p. 107940.
222. S. Sivendran, R. Chang, L. Pham, R.G. Phelps, S.T. Harcharik, L.D. Hall, S.G. Bernardo, M.M. Moskalenko, M. Sivendran, Y. Fu, E.H. de Moll, M. Pan, J.Y. Moon, S. Arora, A. Cohain, A. DiFeo, T.C. Feringer, M. Tismenetsky, C.L. Tsui, P.A. Friedlander, M.K. Parides, J. Banchereau, D. Chaussabel, M.G. Lebwohl, J.D. Wolchok, N. Bhardwaj, S.J. Burakoff, W.K. Oh, K. Palucka, M. Merad, E.E. Schadt, and Y.M. Saenger, *Dissection of immune gene networks in primary melanoma tumors critical for antitumor surveillance of patients with stage II-III resectable disease*. *Journal of Investigative Dermatology*, 2014. **134**(8): p. 2202-2211.
223. E. Becht, A. de Reyniès, N.A. Giraldo, C. Pilati, B. Buttard, L. Lacroix, J. Selves, C. Sautès-Fridman, P. Laurent-Puig, and W.-H. Fridman, *Immune and stromal classification of colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy*. *Clinical Cancer Research*, 2016: p. clincanres. 2879.2015.
224. S. Spranger, R. Bao, and T.F. Gajewski, *Melanoma-intrinsic β -catenin signalling prevents anti-tumour immunity*. *Nature*, 2015. **523**(7559): p. 231.
225. L. Marisa, A. de Reyniès, A. Duval, J. Selves, M.P. Gaub, L. Vescovo, M.-C. Etienne-Grimaldi, R. Schiappa, D. Guenot, and M. Ayadi, *Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value*. *PLoS Medicine*, 2013. **10**(5): p. e1001453.
226. P. Bailey, D.K. Chang, K. Nones, A.L. Johns, A.M. Patch, M.C. Gingras, D.K. Miller, A.N. Christ, T.J. Bruxner, M.C. Quinn, C. Nourse, L.C. Murtaugh, I. Harliwong, S. Idrisoglu, S. Manning, E. Nourbakhsh, S. Wani, L. Fink, O. Holmes, V. Chin, M.J. Anderson, S. Kazakoff, C. Leonard, F. Newell, N. Waddell, S. Wood, Q. Xu, P.J. Wilson, N. Cloonan, K.S. Kassahn, D. Taylor, K. Quek, A. Robertson, L. Pantano, L. Mincarelli, L.N. Sanchez, L. Evers, J. Wu, M. Pinese, M.J. Cowley, M.D. Jones, E.K. Colvin, A.M. Nagrial, E.S. Humphrey, L.A. Chantrill, A. Mawson, J. Humphris, A. Chou, M. Pajic, C.J. Scarlett, A.V. Pinho, M. Giry-Laterriere, I. Rومان, J.S. Samra, J.G. Kench, J.A. Lovell, N.D. Merrett, C.W. Toon, K. Epari, N.Q. Nguyen, A. Barbour, N. Zeps, K. Moran-Jones, N.B. Jamieson, J.S. Graham, F. Duthie, K. Oien, J. Hair, R. Grutzmann, A. Maitra, C.A. Iacobuzio-Donahue, C.L. Wolfgang, R.A.

- Morgan, R.T. Lawlor, V. Corbo, C. Bassi, B. Rusev, P. Capelli, R. Salvia, G. Tortora, D. Mukhopadhyay, G.M. Petersen, I. Australian Pancreatic Cancer Genome, D.M. Munzy, W.E. Fisher, S.A. Karim, J.R. Eshleman, R.H. Hruban, C. Pilarsky, J.P. Morton, O.J. Sansom, A. Scarpa, E.A. Musgrove, U.M. Bailey, O. Hofmann, R.L. Sutherland, D.A. Wheeler, A.J. Gill, R.A. Gibbs, J.V. Pearson, N. Waddell, A.V. Biankin and S.M. Grimmond, *Genomic analyses identify molecular subtypes of pancreatic cancer*. *Nature*, 2016. **531**(7592): p. 47-52.
227. T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander, *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring*. *Science*, 1999. **286**(5439): p. 531-7.
228. M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts, and V. Sondak, *Molecular classification of cutaneous malignant melanoma by gene expression profiling*. *Nature*, 2000. **406**(6795): p. 536-40.
229. A.I. Su, J.B. Welsh, L.M. Sapinoso, S.G. Kern, P. Dimitrov, H. Lapp, P.G. Schultz, S.M. Powell, C.A. Moskaluk, and H.F. Frierson, *Molecular classification of human carcinomas by use of gene expression signatures*. *Cancer Research*, 2001. **61**(20): p. 7388-7393.
230. C.H. Chung, J.S. Parker, G. Karaca, J. Wu, W.K. Funkhouser, D. Moore, D. Butterfoss, D. Xiang, A. Zanation, and X. Yin, *Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression*. *Cancer Cell*, 2004. **5**(5): p. 489-500.
231. C. Sotiriou, P. Wirapati, S. Loi, A. Harris, S. Fox, J. Smeds, H. Nordgren, P. Farmer, V. Praz, B. Haibe-Kains, C. Desmedt, D. Larsimont, F. Cardoso, H. Peterse, D. Nuyten, M. Buyse, M.J. Van de Vijver, J. Bergh, M. Piccart, and M. Delorenzi, *Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis*. *Journal of the National Cancer Institute*, 2006. **98**(4): p. 262-72.
232. B. Ryu, D.S. Kim, A.M. Deluca, and R.M. Alani, *Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression*. *PloS One*, 2007. **2**(7): p. e594.

233. E. Tremante, A. Ginebri, E. Lo Monaco, P. Frascione, F. Di Filippo, I. Terrenato, M. Benevolo, M. Mottolose, E. Pescarmona, P. Visca, P.G. Natali, and P. Giacomini, *Melanoma molecular classes and prognosis in the postgenomic era*. *Lancet Oncology*, 2012. **13**(5): p. e205-11.
234. C.G.A.R. Network, *Comprehensive molecular characterization of gastric adenocarcinoma*. *Nature*, 2014. **513**(7517): p. 202.
235. C.G.A.R. Network, *Comprehensive molecular profiling of lung adenocarcinoma*. *Nature*, 2014. **511**(7511): p. 543.
236. C.M. Perou, T. Sorlie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, C.A. Rees, J.R. Pollack, D.T. Ross, H. Johnsen, L.A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S.X. Zhu, P.E. Lonning, A.L. Borresen-Dale, P.O. Brown, and D. Botstein, *Molecular portraits of human breast tumours*. *Nature*, 2000. **406**(6797): p. 747-52.
237. L.J. Van't Veer, H. Dai, M.J. Van De Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. Van Der Kooy, M.J. Marton, and A.T. Witteveen, *Gene expression profiling predicts clinical outcome of breast cancer*. *Nature*, 2002. **415**(6871): p. 530.
238. R.W. Tothill, A.V. Tinker, J. George, R. Brown, S.B. Fox, S. Lade, D.S. Johnson, M.K. Trivett, D. Etemadmoghadam, B. Locandro, N. Traficante, S. Fereday, J.A. Hung, Y.E. Chiew, I. Haviv, G. Australian Ovarian Cancer Study, D. Gertig, A. DeFazio, and D.D. Bowtell, *Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome*. *Clinical Cancer Research*, 2008. **14**(16): p. 5198-208.
239. E.J. Yeoh, M.E. Ross, S.A. Shurtleff, W.K. Williams, D. Patel, R. Mahfouz, F.G. Behm, S.C. Raimondi, M.V. Relling, A. Patel, C. Cheng, D. Campana, D. Wilkins, X. Zhou, J. Li, H. Liu, C.H. Pui, W.E. Evans, C. Naeve, L. Wong, and J.R. Downing, *Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling*. *Cancer Cell*, 2002. **1**(2): p. 133-43.
240. I. Hedenfalk, D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, O.P. Kallioniemi, B. Wilfond, A. Borg, J. Trent, M. Raffeld, Z. Yakhini, A. Ben-Dor, E. Dougherty, J. Kononen, L. Bubendorf, W. Fehrle, S. Pittaluga, S. Gruvberger, N. Loman, O. Johannsson, H. Olsson, and G. Sauter, *Gene-expression profiles in hereditary breast cancer*. *New England Journal of Medicine*, 2001. **344**(8): p. 539-48.
241. M.J. Van De Vijver, Y.D. He, L.J. Van't Veer, H. Dai, A.A. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, and M.J. Marton, *A gene-expression*

- signature as a predictor of survival in breast cancer*. New England Journal of Medicine, 2002. **347**(25): p. 1999-2009.
242. J.C. Chang, E.C. Wooten, A. Tsimelzon, S.G. Hilsenbeck, M.C. Gutierrez, R. Elledge, S. Mohsin, C.K. Osborne, G.C. Chamness, D.C. Allred, and P. O'Connell, *Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer*. The Lancet, 2003. **362**(9381): p. 362-369.
243. Y. Wang, J.G. Klijn, Y. Zhang, A.M. Sieuwerts, M.P. Look, F. Yang, D. Talantov, M. Timmermans, M.E. Meijer-van Gelder, J. Yu, T. Jatkoe, E.M. Berns, D. Atkins, and J.A. Foekens, *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer*. The Lancet, 2005. **365**(9460): p. 671-679.
244. Y. Pawitan, J. Bjohle, L. Amler, A.L. Borg, S. Egyhazi, P. Hall, X. Han, L. Holmberg, F. Huang, S. Klaar, E.T. Liu, L. Miller, H. Nordgren, A. Ploner, K. Sandelin, P.M. Shaw, J. Smeds, L. Skoog, S. Wedren, and J. Bergh, *Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts*. Breast Cancer Research, 2005. **7**(6): p. R953-64.
245. S. Paik, G. Tang, S. Shak, C. Kim, J. Baker, W. Kim, M. Cronin, F.L. Baehner, D. Watson, J. Bryant, J.P. Costantino, C.E. Geyer, Jr., D.L. Wickerham, and N. Wolmark, *Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer*. Journal of Clinical Oncology, 2006. **24**(23): p. 3726-34.
246. S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F.L. Baehner, M.G. Walker, D. Watson, T. Park, W. Hiller, E.R. Fisher, D.L. Wickerham, J. Bryant, and N. Wolmark, *A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer*. New England Journal of Medicine, 2004. **351**(27): p. 2817-26.
247. P. Mumby, S. Lo, J. Norton, J. Smerage, K. Joseph, H. Chew, D. Hayes, and K. Albain, *Prospective multi-center study of the impact of the 21-gene recurrence score assay on patient satisfaction, anxiety and decisional conflict for adjuvant breast cancer treatment selection*. Breast Cancer Research and Treatment, 2007. **106**: p. S73-S74.
248. K.S. Albain, W.E. Barlow, S. Shak, G.N. Hortobagyi, R.B. Livingston, I.T. Yeh, P. Ravdin, R. Bugarini, F.L. Baehner, N.E. Davidson, G.W. Sledge, E.P. Winer, C. Hudis, J.N. Ingle, E.A. Perez, K.I. Pritchard, L. Shepherd, J.R. Gralow, C. Yoshizawa, D.C. Allred, C.K. Osborne, D.F. Hayes, and A. Breast

- Cancer Intergroup of North, *Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial*. *Lancet Oncology*, 2010. **11**(1): p. 55-65.
249. J.S. Reis-Filho and L. Pusztai, *Gene expression profiling in breast cancer: classification, prognostication, and prediction*. *The Lancet*, 2011. **378**(9805): p. 1812-23.
250. J.A. Sparano, R.J. Gray, D.F. Makower, K.I. Pritchard, K.S. Albain, D.F. Hayes, C.E. Geyer, Jr., E.C. Dees, E.A. Perez, J.A. Olson, Jr., J. Zujewski, T. Lively, S.S. Badve, T.J. Saphner, L.I. Wagner, T.J. Whelan, M.J. Ellis, S. Paik, W.C. Wood, P. Ravdin, M.M. Keane, H.L. Gomez Moreno, P.S. Reddy, T.F. Goggins, I.A. Mayer, A.M. Brufsky, D.L. Toppmeyer, V.G. Kaklamani, J.N. Atkins, J.L. Berenberg, and G.W. Sledge, *Prospective Validation of a 21-Gene Expression Assay in Breast Cancer*. *New England Journal of Medicine*, 2015. **373**(21): p. 2005-14.
251. J.A. Sparano, R.J. Gray, D.F. Makower, K.I. Pritchard, K.S. Albain, D.F. Hayes, C.E. Geyer, Jr., E.C. Dees, M.P. Goetz, J.A. Olson, Jr., T. Lively, S.S. Badve, T.J. Saphner, L.I. Wagner, T.J. Whelan, M.J. Ellis, S. Paik, W.C. Wood, P.M. Ravdin, M.M. Keane, H.L. Gomez Moreno, P.S. Reddy, T.F. Goggins, I.A. Mayer, A.M. Brufsky, D.L. Toppmeyer, V.G. Kaklamani, J.L. Berenberg, J. Abrams, and G.W. Sledge, Jr., *Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer*. *New England Journal of Medicine*, 2018. **379**(2): p. 111-121.
252. A.C. Berger, R.S. Davidson, J.K. Poitras, I. Chabra, R. Hope, A. Brackeen, C.E. Johnson, D.J. Maetzold, B. Middlebrook, K.M. Oelschlager, R.W. Cook, F.A. Monzon, and A.R. Miller, *Clinical impact of a 31-gene expression profile test for cutaneous melanoma in 156 prospectively and consecutively tested patients*. *Current Medical Research and Opinion*, 2016. **32**(9): p. 1599-604.
253. L.K. Ferris, A.S. Farberg, B. Middlebrook, C.E. Johnson, N. Lassen, K.M. Oelschlager, D.J. Maetzold, R.W. Cook, D.S. Rigel, and P. Gerami, *Identification of high-risk cutaneous melanoma tumors is improved when combining the online American Joint Committee on Cancer Individualized Melanoma Patient Outcome Prediction Tool with a 31-gene expression profile-based classification*. *Journal of the American Academy of Dermatology*, 2017. **76**(5): p. 818-825. e3.
254. J.S. Zager, B.R. Gastman, S. Leachman, R.C. Gonzalez, M.D. Fleming, L.K. Ferris, J. Ho, A.R. Miller, R.W. Cook, K.R. Covington, K. Meldi-Plasseraud,

- B. Middlebrook, L.H. Kaminester, A. Greisinger, S.I. Estrada, D.M. Pariser, L.D. Cranmer, J.L. Messina, J.T. Vetto, J.D. Wayne, K.A. Delman, D.H. Lawson, and P. Gerami, *Performance of a prognostic 31-gene expression profile in an independent cohort of 523 cutaneous melanoma patients*. BMC Cancer, 2018. **18**(1): p. 130.
255. E.P. Diamandis, *The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem?* BMC Medicine, 2012. **10**(1): p. 87.
256. E. Wang, W.C. Cho, S.C. Wong, and S. Liu, *Disease biomarkers for precision medicine: challenges and future opportunities*. 2017, Elsevier.
257. Cancer research UK 21st September 2018, <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer/diagnosis-and-treatment#ref-1>.
258. Cancer Research UK, <http://www.cancerresearchuk.org/about-cancer/melanoma/survival>. 2016.
259. B. Andreopoulos, A. An, X. Wang, and M. Schroeder, *A roadmap of clustering algorithms: finding a match for a biomedical application*. Briefings in Bioinformatics, 2009. **10**(3): p. 297-314.
260. R. Sloutsky, N. Jimenez, S.J. Swamidass, and K.M. Naegle, *Accounting for noise when clustering biological data*. Briefings in Bioinformatics, 2012. **14**(4): p. 423-436.
261. M. Angelova, P. Charoentong, H. Hackl, M.L. Fischer, R. Snajder, A.M. Krogsdam, M.J. Waldner, G. Bindea, B. Mlecnik, J. Galon, and Z. Trajanoski, *Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy*. Genome Biology, 2015. **16**(1): p. 64.
262. F. Pages, J. Galon, M. Dieu-Nosjean, E. Tartour, C. Sautes-Fridman, and W. Fridman, *Immune infiltration in human tumors: a prognostic factor that should not be ignored*. Oncogene, 2010. **29**(8): p. 1093.
263. N.R. West, S.E. Kost, S.D. Martin, K. Milne, R.J. Deleeuw, B.H. Nelson, and P.H. Watson, *Tumour-infiltrating FOXP3(+) lymphocytes are associated with cytotoxic immune responses and good clinical outcome in oestrogen receptor-negative breast cancer*. British Journal of Cancer, 2013. **108**(1): p. 155-62.
264. J. Goc, C. Germain, T.K.D. Vo-Bourgais, A. Lupo, C. Klein, S. Knockaert, L. de Chaisemartin, H. Ouakrim, E. Becht, and M. Alifano, *Dendritic cells in*

- tumor-associated tertiary lymphoid structures signal a Th1 cytotoxic immune contexture and license the good positive prognostic value of infiltrating CD8+ T cells.* Cancer Research, 2013.
265. A. Ladanyi, *Prognostic and predictive significance of immune cells infiltrating cutaneous melanoma.* Pigment Cell Melanoma Research, 2015. **28**(5): p. 490-500.
266. G. Finak, N. Bertos, F. Pepin, S. Sadekova, M. Souleimanova, H. Zhao, H. Chen, G. Omeroglu, S. Meterissian, A. Omeroglu, M. Hallett, and M. Park, *Stromal gene expression predicts clinical outcome in breast cancer.* Nature Medicine, 2008. **14**(5): p. 518-27.
267. R.R. Ji, S.D. Chasalow, L. Wang, O. Hamid, H. Schmidt, J. Cogswell, S. Alaparthi, D. Berman, M. Jure-Kunkel, N.O. Siemers, J.R. Jackson, and V. Shahabi, *An immune-active tumor microenvironment favors clinical response to ipilimumab.* Cancer Immunology, Immunotherapy, 2012. **61**(7): p. 1019-1031.
268. J.S. Ankeny, B. Labadie, J. Luke, E. Hsueh, J. Messina, and J.S. Zager, *Review of diagnostic, prognostic, and predictive biomarkers in melanoma.* Clinical and Experimental Metastasis, 2018. **35**(5-6): p. 487-493.
269. L.E. Clarke, M.B. Warf, D.D. Flake, A.R. Hartman, S. Tahan, C.R. Shea, P. Gerami, J. Messina, S.R. Florell, and R.J. Wenstrup, *Clinical validation of a gene expression signature that differentiates benign nevi from malignant melanoma.* Journal of Cutaneous Pathology, 2015. **42**(4): p. 244-252.
270. E.C. Minca, R.N. Al-Rohil, M. Wang, P.W. Harms, J.S. Ko, A.M. Collie, I. Kovalyshyn, V.G. Prieto, M.T. Tetzlaff, and S.D. Billings, *Comparison between melanoma gene expression score and fluorescence in situ hybridization for the classification of melanocytic lesions.* Modern Pathology, 2016. **29**(8): p. 832.
271. A. De Rienzo, R.W. Cook, J. Wilkinson, C.E. Gustafson, W. Amin, C.E. Johnson, K.M. Oelschlager, D.J. Maetzold, J.F. Stone, M.D. Feldman, M.J. Becich, B.Y. Yeap, W.G. Richards, and R. Bueno, *Validation of a Gene Expression Test for Mesothelioma Prognosis in Formalin-Fixed Paraffin-Embedded Tissues.* Journal of Molecular Diagnostics, 2017. **19**(1): p. 65-71.
272. A. Mitra, C. Conway, C. Walker, M. Cook, B. Powell, S. Lobo, M. Chan, M. Kissin, G. Layer, and J. Smallwood, *Melanoma sentinel node biopsy and prediction models for relapse and overall survival.* British Journal of Cancer, 2010. **103**(8): p. 1229.

273. R. Díaz-Uriarte and S.A. De Andres, *Gene selection and classification of microarray data using random forest*. BMC Bioinformatics, 2006. **7**(1): p. 3.
274. A. Statnikov, L. Wang, and C.F. Aliferis, *A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification*. BMC Bioinformatics, 2008. **9**(1): p. 319.