Bioinformatic Analysis of Copy Number Alteration Signature Mechanisms in Ovarian Carcinoma

Name: Conor Giles Doran

Student number: 17408582

Principal Investigator: Prof. Donal Brennan

Supervisor: Dr. Bruce Moran

Module name: PATH40220-Research Project: BSc Biomed

BSc. Biomedical Health and Life Sciences

School of Medicine

University College Dublin

Assessment Submission Form

Student Name	Conor Giles Doran
Student Number	17408582
Assessment Title	Bioinformatic Analysis of Copy Number Alteration Signature Mechanisms in Ovarian Carcinoma
Module Title	PATH40220 - Research Project: BSc Biomed
Module Co-ordinator	Prof. Bill Watson
Supervisor	Prof. Donal Brennan and Dr. Bruce Moran
Date Submitted	04/03/2021
OFFICE USE ONLY **Date Received**	
OFFICE USE ONLY *Grade/Mark*	

A SIGNED COPY OF THIS FORM MUST ACCOMPANY ALL SUBMISSIONS FOR ASSESSMENT.

STUDENTS SHOULD KEEP A COPY OF ALL WORK SUBMITTED.

Procedures for Submission and Late Submission

Ensure that you have checked the School's procedures for the submission of assessments.

Note: ***There are penalties for the late submission of assessments. For further information please see the University's*** Policy on Late Submission of Coursework**,** ***(****http://www.ucd.ie/registry/academicsecretariat/pol.htm****)***

Plagiarism: the unacknowledged inclusion of another person's writings or ideas or works, in any formally presented work (including essays, examinations, projects, laboratory reports or presentations). The penalties associated with plagiarism designed to impose sanctions that reflect the seriousness of University's commitment to academic integrity. Ensure that you have read the University's ***Briefing for Students on Academic Integrity and Plagiarism*** and the UCD ***Plagiarism Statement, Plagiarism Policy and Procedures*** (http://www.ucd.ie/registry/academicsecretariat/pol.htm)

Declaration of Authorship

I declare that all material in this assessment is my own work except where there is clear acknowledgement and appropriate reference to the work of others.

Signed: ***Conor Giles Doran*** Date: ***04/03/2021***

1. Summary

Context

High grade serous ovarian (HGSOC) cancer is a complex disease associated with treatment resistance and poor prognosis. Largely driven by DNA copy number alterations (CNAs), it is characterized by an extreme level of chromosomal instability, giving rise to a vast degree of intratumoural heterogeneity among patient samples. CNA signatures have offered a novel form of advanced patient stratification, aiming to identify molecular patterns within the heterogeneity that exists.

Objectives

1) Confirm CNA signature findings from a previously published HGSOC study.

2) Use RNA-sequencing data to generate differential expression (DE) profiles for the CNA signatures.

3) Computationally derive RNA signatures based on CNA signatures from 1.

4) Assess the prognostic potential of the CNA and RNA signature classifications.

Results

Using published methods, seven CNA signatures were derived from 415 HGSOC samples. Of which, six were used to generate DE profiles, revealing the key hallmark pathways and genes underlying HGSOC. CNA s5 and s6 were significantly associated ($p = 0.031$) with the worst and best overall prognosis, respectively. Four RNA signatures were subsequently derived for 311 HGSOC samples, of which r2 and r4 dominated, r2 being associated with the more favourable prognosis ($p = 0.21$).

Conclusions

Although with limitations, CNA signatures have an interesting potential as both future prognostic and diagnostic biomarkers in cancer, providing an advanced level of stratification that might aid in deciphering the underlying complexity of HGSOC.

2. Table of Contents

3. Introduction

3.1 Ovarian Cancer

Ovarian cancer is an umbrella term used to describe numerous molecularly distinct diseases which vary in aetiology but share a mutual location within the female body. Serous ovarian carcinoma, the most common form of ovarian cancer, affects the epithelial layer of the ovaries. Both high and low-grade serous ovarian cancers exist (HGSOC and LGSOC). HGSOC is the most prevalent and is found at advanced stages. Prognosis is poor and HGSOC causes 70 – 80% of ovarian cancer deaths, with survival outcomes failing to improve in the last two decades (Bowtell et al., 2015; Gockley et al., 2017; Vaughan et al., 2011).

HGSOC is a disease characterized by abundant DNA copy number alterations (CNAs), indicative of a high level of chromosomal instability that results in a complex, heterogenetic molecular profile (Bowtell et al., 2015; Vaughan et al., 2011). HGSOC is driven by missense and nonsense TP53 mutations, alongside ~20% of cases in which BRCA1 and/or BRCA2 mutations are involved (Ahmed et al., 2010; The Cancer Genome Atlas Research Network, 2011). DNA repair is often significantly disrupted as a result, with up to 50% of HGSOCs estimated as being homologous recombination deficient (HRD; Mukhopadhyay et al., 2012; The Cancer Genome Atlas Research Network, 2011).

The current standard treatment for ovarian cancer is maximal cytoreductive surgical debulking in conjunction with platinum-based chemotherapy such as carboplatin, paclitaxel and bevacizumab (Bowtell et al., 2015; Centre for Drug Evaluation and Research, 2019; Cortez et al., 2018). Treatment resistance remains a prevalent issue, emerging in 80 - 90% of HGSOC patients with widespread disease (Bowtell et al., 2015). Despite genomic advances, the lack of putatively actionable molecular targets remains a barrier to therapeutic intervention. This is due in part to the complex genomic profile of HGSOC, which is often mistakenly treated as a single disease entity (Bowtell et al., 2015; Ciriello et al., 2013; Gockley et al., 2017; Vaughan

et al., 2011).

For HGSOC outcomes to improve, experts believe that novel bioinformatic stratification methods based on early driver processes and specific histotypes will be required, as well as a greater understanding of resistance development (Bowtell et al., 2015; Vaughan et al., 2011). Informative and reliable genomic biomarkers must therefore be developed that can significantly aid in the early detection, treatment and management of HGSOC patients. A relatively new advance, known as CNA signatures, may be capable of fulfilling these goals. The potential of CNA signatures to predict patient survival outcomes, predict relapse post-treatment, and unearth novel therapeutic targets for developing precision-based medicine is a very current research topic (Alexandrov et al., 2020, 2013; Ciriello et al., 2013; Macintyre et al., 2018; Zou et al., 2018).

3.2 Copy Number Alterations

The concept of CNA signatures is based on the biology of 'copy number'. The structural variation within a cancer genome can take numerous forms, ranging from single nucleotide variants or polymorphisms (SNV/SNPs) to large rearrangements of the chromosome (Redon et al., 2006; Shao et al., 2019). SNVs were previously thought to account for a high proportion of genomic variation, as well as the resulting phenotypic variation (Gibbs et al., 2003; Sachidanandam et al., 2001). However, more recent studies have shown abundant alterations in copy number, demonstrating the substantial contribution of CNAs (Conrad et al., 2006; Iafrate et al., 2004; Levy et al., 2007; Locke et al., 2006; Sebat et al., 2004; Sharp et al., 2005; Tuzun et al., 2005; Zarrei et al., 2015).

The human genome consists of 23 pairs of chromosomes and the normal copy number state is diploid (Levy et al., 2007). The structure of the genome is actively modified in healthy individuals and those with disease (Zarrei et al., 2015). A CNA is defined as being an amplification or deletion of a DNA segment that is ≥ 1kb within the human genome (Iafrate et

al., 2004; Redon et al., 2006; Shao et al., 2019). CNAs affect the diploid state of DNA segments and can range from simple deletions, insertions and duplications to more complex alterations involving multiple chromosomal locations within the genome (Redon et al., 2006). An example of CNAs detected within a melanoma sample is presented in Figure 1. CNAs are thought to arise from errors in homologous recombination, which operates in DNA repair mechanisms such as double-strand break (DSB) repair and break-induced recombination (Hastings et al., 2009). Dysfunctional DNA repair can cause tandem duplications, a known cancer genomic configuration (Menghi et al., 2016). Nonhomologous recombination is also thought to give rise to CNAs through the breakage-fusion-bridge (BFB) cycle, wherein a DSB establishes a repetitive cycle of telomere loss and dicentric chromosome formation, leading to inverted duplications (Hastings et al., 2009; Kinsella and Bafna, 2012; Murnane, 2012). Copy number is also influenced by chromothripsis, which leads to structural rearrangement with an extensive loss of DNA sequence fragments (Korbel and Campbell, 2013).

CNAs drive genomic evolution and can directly lead to disease or increase disease risk by influencing gene expression, causing direct gene dosage disruption as well as indirect positional effects (Feuk et al., 2006; Freeman, 2006; Lupski and Stankiewicz, 2005; Redon et al., 2006; Zhang et al., 2009). CNAs have been established as a key contributing factor of genetic diversity and cause of differences in disease susceptibility for both inter-individual and inter-ethnic populations (Almal and Padh, 2012; Fanciulli et al., 2010; Jakobsson et al., 2008).

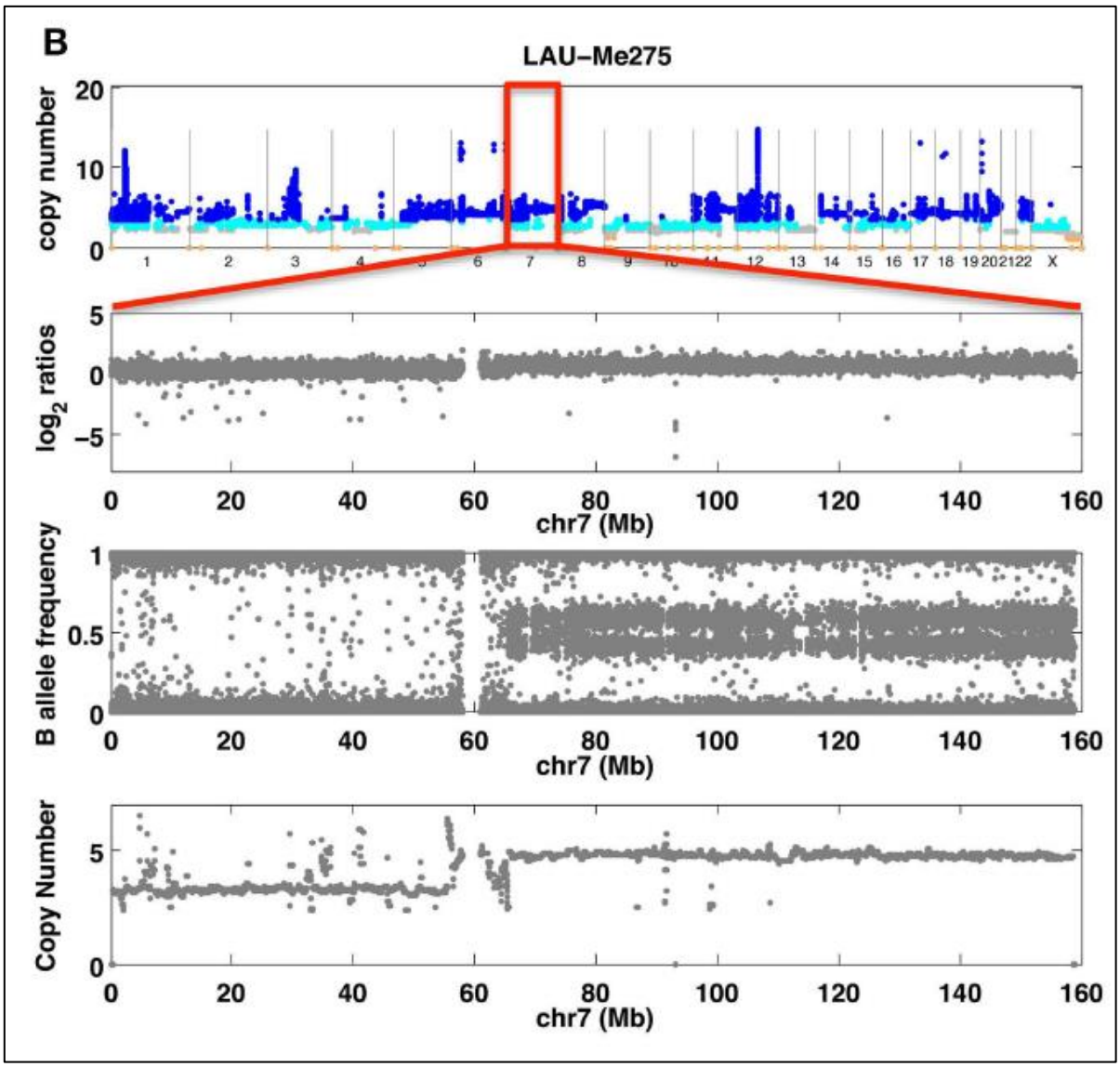


Figure 1 - Example of CNAs detected within a melanoma sample (taken from Valsesia et al., 2013)

The copy number profile of a melanoma sample generated using SNP array. The genome wide copy number state of DNA segments is presented in the top panel. Orange represents <2 copies, grey represents the diploid state, cyan represents 3 copies and dark blue represents >3 copies. Specific analyses of chromosome 7, involving hybridization $\log_2$ ratios, B-allele frequencies, and overall predicted copy number of DNA segments are presented in subsequent panels. The B-allele frequency is an allele specific measurement of copy number whereby 0.5 indicates equal presence of both alleles and 1 or 0 indicates the absence of one allele.

The significant effect of CNAs on transcription has been demonstrated in multiple cancer studies (Dong et al., 2017; Kuzyk et al., 2015; Xu et al., 2017; Yang et al., 2017; Zhao et al., 2014; Zhou et al., 2017). A recent pan-cancer study has shown that CNAs and differential gene expression are highly correlated. This highlights the qualitative association of genetic diversity and its resulting downstream impact in cancer (Shao et al., 2019). Several recent type-specific cancer studies in particular have demonstrated that CNAs are associated with disease prognosis, such as in clear cell renal cell carcinoma and breast cancer (Huang et al., 2017; Kumaran et al., 2017). Studies have stated CNA prognostic power may be subject to modulation by sample purity, an area that requires further study (Hieronymus et al., 2018). It has also been recently demonstrated that copy number analysis of primary tumour oncogenes and tumour suppressors are effective in stratifying patient risk. Single base-pair mutations within the majority of cancer driver genes proved to be non-prognostic, whereas CNAs within these same genes were directly linked with the outcome of a patient (Smith and Sheltzer, 2018). With CNA signatures, the underlying genomic mechanisms involved in tumorigenesis can be derived, and thus signatures might also reveal the prognostic potential of CNAs in the form of a gene-expression level picture of tumour biology. This would serve as a powerful clinical tool (Califf, 2018; Hieronymus et al., 2018; Macintyre et al., 2018; Van Hoeck et al., 2019).

3.3 Copy Number Alteration Signatures

CNA signatures are a form of genomic classification with the potential to change the scope of future prognostic and diagnostic biomarkers. CNA profiles are initially required. Profiles are best determined through whole genome sequencing (WGS) but whole exome sequencing (WES) is often used. Shallow WGS (sWGS) in particular is a cost-effective method for determining CNAs, using only 0.5 – 1x depth of sequencing when the traditional coverage is ~30x (Doherty et al., 2019; Macintyre et al., 2018; Scheinin et al., 2014). This depth of coverage technique can be utilized on fresh and archival DNA samples and is based on the

Illumina® next generation sequencing (NGS) by synthesis protocol. The human reference genome to which sequence reads are aligned is divided into fixed-sized bins that are non-overlapping. Raw copy number is then estimated by calculating the read count of each bin, allowing detection of CNAs (Illumina Inc., 2017; Scheinin et al., 2014). A number of CNA algorithms for analysing sequencing data exist: FACETS, Sequenza, ABSOLUTE, AbsCN-seq, TITAN and FALCON (Bao et al., 2014; Chen et al., 2015; Cibulskis et al., 2012; Ha et al., 2014; Shen and Seshan, 2016).

Using sWGS copy number profiles, a landmark study successfully implemented the concept of genome-wide CNA signatures, demonstrating their functionality in classifying HGSOC (Macintyre et al., 2018). Seven CNA signatures were successfully derived from 253 primary and relapsed HGSOC samples. These signatures were identified by computing genome-wide distributions of six fundamental copy number features. These are: breakpoint count per 10Mb, overall copy number, copy number change point, breakpoint count per chromosome arm, length of oscillating copy number and chromosome segment size. These features were established as known hallmarks of genomic alterations, such as the tandem duplicator phenotype, chromothripsis and BFB events (Korbel and Campbell, 2013; Menghi et al., 2016; Murnane, 2012). The seven signatures revealed key associations with predicted survival and relapse outcomes, as well as the proposed functional mechanisms that drive tumorigenesis (Macintyre et al., 2018).

Building on these important findings, recent CNA signature based studies have been conducted in kidney cancer and chemoradiotherapy-treated head and neck squamous cell carcinoma, revealing key associations with survival and the specific genomic abnormalities underlying the disease (Benstead-Hume et al., 2019; Essers et al., 2020). Additionally, current research has investigated the concept of junction copy number (JCN), wherein copy number, in addition to describing genomic interval dosage, includes possible structural variant junctions

that may also be present in copies of a cancer cell. Hadi et. al aimed to analyse CNAs and rearrangements in combination, rather than being interpreted separately. Three novel complex mutational phenomena were identified, defined as pyrgo, rigma and tyfonas. Each phenom was associated with novel mutational processes and potential somatic selection points, unearthing intricate details within the evolution of cancer (Hadi et al., 2020).

The field of CNA signatures is relatively new and there are a number of important caveats that must be noted. The fitting of signatures is a mathematical practice and is subject to uncertainty. Computational and statistical framework must be improved in order to ensure accurate, reproducible signatures that are biologically relevant and present within the data, preventing spurious associations and false-positive results (Degasperi et al., 2020; Koh et al., 2020). In NGS, the introduction of biases, technical errors and 'noise' remains an issue (Nam et al., 2020; Valsesia et al., 2013). In addition, not all signatures will be clinically relevant or applicable as biomarkers. It is important that signatures are accurately defined and can distinguish distinct mechanisms that are operating within the cancer (Alexandrov et al., 2020; Degasperi et al., 2020; Van Hoeck et al., 2019). Current research also suggests that experimental studies must be carried out in conjunction with computational analysis, in order to validate the accuracy of the predicted biological events, highlighting the necessity of careful consideration and caution when inferring clinical judgment (Alexandrov et al., 2020; Degasperi et al., 2020; Koh et al., 2020). Although limited, a precedent and methodology has been established for further investigating CNA signatures. With this novel form of analysis, an intriguing opportunity is offered for implementation as a powerful genomic stratification tool in cancer.

3.4 Hypothesis and Objectives

This study has the hypothesis that tumour gene expression is divergent between sets of HGSOC patient groups classified based on CNA signatures.

The primary objectives of the study were to:

1. Confirm CNA signature findings from a previously published HGSOC study.
2. Use RNA-seq data to generate differential expression (DE) profiles for the CNA signatures.
3. Computationally derive RNA signatures based on CNA signatures from 1.
4. Assess the prognostic potential of the CNA and RNA signature classifications.

4. Materials and Methods

4.1 Methods Outline

All method components were carried out using the R programming language (version 4.0.2) and the RStudio (version 1.3.1073) integrated development environment. R packages were downloaded from both The Comprehensive R Archive Network (CRAN) and Bioconductor online repositories (Hornik, 2012; Huber et al., 2015; Mora-Cantallops et al., 2020). Data from a total of 587 HGSOC patient samples were used in the study, made publicly available through The Cancer Genome Atlas (TCGA), an online genomics database from the National Cancer Institute and the National Human Genome Research Institute (The Cancer Genome Atlas Research Network, 2011). The methods comprised four sections: CNA signature extrapolation, DE profiling, RNA signature identification and survival analysis. From the 587 TCGA HGSOC samples, there were 415 samples with available CNA data. These 415 samples were used for CNA signature extrapolation. There were 587 samples with available clinical data and 425 samples with available RNA-seq count data. Across the three data types (CNA, clinical and RNA-seq data), there were 311 matching samples used for DE analysis, RNA signature identification and survival analysis. The filtering of samples can be visualized in Appendix 1. No ethical approval was required for this study as the use of TCGA data is in accordance with TCGA Ethics, Law and Policy Group, allowing for effective and fair use of cancer genomic data. All R scripts used to carry out the study are accessible through the GitHub repository (https://github.com/ConorGilesDoran/UCD_BHLS_Thesis).

4.2 CNA Signature Extrapolation

The landmark study from Macintyre et. al successfully derived seven CNA signatures using 253 HGSOC samples, which were validated in 527 independent cases from TCGA and the Pan-Cancer Analysis of Whole Genomes (PCAWG; Macintyre et al., 2018). All computational analysis from Macintyre and colleagues, including R scripts and data, were accessible for public download (https://bitbucket.org/britroc/cnsignatures/src/master/). The published R scripts were inspected in RStudio where the various functions used to process the data and derive signatures were tested. This posed both as a learning exercise for understanding the functionalities of R and the CNA signature extrapolation methods. A strong focus was placed on linking the computational approaches in R with the fundamental biology of cancer genomics and copy number. Due to issues with variable naming within the published R scripts and miscommunication of essential method components, it was not possible to recapitulate the results from raw data as intended. Progression was only possible using previously generated TCGA results.

In summary, the published methods were performed as follows: (i) the genome-wide distribution of six fundamental copy number features (breakpoint count per 10Mb, overall copy number, copy number change point, breakpoint count per chromosome arm, length of oscillating copy number and chromosome segment size) within 415 TCGA HGSOC samples was provided by Macintyre et. al; (ii) using feature results and component parameters provided by Macintyre et. al., a component-by-sample matrix was generated, representing the summed posterior probability of a copy number event belonging to belonging to one of 36 individual components. Algorithmic methods such as non-negative matrix factorization (NMF) and linear combination decomposition (LCD) were used to deconvolute the component-by-sample matrix into a normalized signature-by-sample matrix, representing the 415 TCGA HGSOC samples and the seven CNA signatures by which they have been classified.

4.3 Differential Gene Expression Profiling

Pairwise differential expression (DE) analysis was performed using 311 TCGA HGSOC samples with matching CNA, clinical and RNA-seq data. DE analysis is a well-established method for evaluating the statistical significance of divergence of an individual gene between two specific conditions. In this study, the conditions being compared were those of the CNA signatures, e.g., s1 vs s2. In total, there were then 15 possible comparison combinations, wherein the effect of maximum CNA signature on gene expression was investigated. The 'maximum CNA signature' represented the signature with the greatest posterior probability value of being associated with a sample. Maximum CNA signatures were identified for each sample from the normalized signature-by-sample matrix generated in section 4.2. The histological grade and clinical stage of samples were included as confounding variables in the DE model. RNA-seq reads were previously mapped to NCBI Build 38 of the human transcriptome (GRCh38) using Kallisto (Bray et al., 2016). Gene expression values in TPM (transcripts-per-million) per sample were available from the Sleuth workflow (Pimentel et al., 2016). Ensembl gene IDs were aggregated for genes with multiple external gene name identifiers in order to prevent duplicate entries, and TPM values were converted to log2 format (Yates et al., 2020). Differentially expressed gene (DEG) sets were derived using three DE packages: DESeq2 (v1.30.0), limma (v3.46.0) and edgeR (v3.32.0; Love et al., 2014; Ritchie et al., 2015; Robinson et al., 2010). Packages were used within the RNAseqon (v0.0.2) framework and generated 15 gene sets each, representing the 15 signature comparisons (Moran, 2021). This resulted in 45 DEG sets in total, from which the overlap across the three packages was calculated, allowing DEGs to be determined conservatively to remove potential false positives. Principal component analysis (PCA) was performed to assess the level of variance attributed to each variable within the DE model. Only overlapping genes from edgeR and DESeq2 were used for downstream analysis, as limma failed to generate full DEG sets for

certain signature comparisons. This was unexpected as comparative studies have shown limma to perform well (Corchete et al., 2020; Seyednasrollah et al., 2015). It is possible that errors occurred with the multiple contrasts that were being analysed. Different limma 'method' parameters could have been tested before proceeding with DEseq2 and edgeR results.

To identify the significant biological pathways involved, gene set enrichment analysis (GSEA) was performed using the 15 overlapped DEG sets between DESeq2 and edgeR (4,228 genes; Subramanian et al., 2005). The Broad Institute Molecular Signatures Database (MSigDB-7.0), a collection of 40 hallmark pathway gene sets, was utilized (Liberzon et al., 2015). A normalized enrichment score (NES) for each hallmark pathway was calculated, representing the degree to which the pathway DEGs were overrepresented within a ranked list of each of the 15 CNA signature comparison gene sets. To account for false discovery rates (FDRs), an adjusted p value of ≤ 0.01 was considered a significant result. R packages used for GSEA included fgsea (v1.16.0) and msigdbr (v7.2.1; Dolgalev, 2020; Korotkevich et al., 2021). An example of DEG overlap between packages, and GSEA output can be viewed in Appendix 2.

4.4 RNA Signature Identification

To derive RNA signatures based on the DEG results, single sample gene set enrichment analysis (ssGSEA) was performed (Barbie et al., 2009; Hänzelmann et al., 2013). In ssGSEA, gene expression values are rank-normalized per individual sample. A NES is then generated using the Empirical Cumulative Distribution Functions (ECDF) of the signature enriched genes and all other genes. Unique-to-signature DEG sets were required. Pairwise DE profiling resulted in overlap between DEG sets across the maximum CNA signature comparisons, making it difficult to isolate these unique genes. To achieve this, the tally of DEGs in each hallmark pathway per maximum CNA signature was calculated. The signature with the highest number of DEGs for that particular pathway was then assigned those DEGs. These results

represented DEG sets per maximum CNA signature per significant hallmark pathway involved. RNA signatures were derived using ssGSEA, with the distinct CNA signature gene sets and their associated log2TPMs per sample used as input. R packages used for this analysis included nortest (v1.0-4), GSVA (v1.20.0) and BBmisc (v1.10; Bischl et al., 2017; Gross and Ligges, 2015; Hänzelmann et al., 2013). Default settings were used for each and to account for multiple hypotheses testing, an FDR of < 0.25 was employed.

4.5 Survival Analysis

To assess the prognostic potential of the RNA and CNA signatures, survival analysis was carried out for the maximum RNA and CNA signatures previously identified in sections 4.2 and 4.4, respectively. Clinical data was used for each of the 311 samples, detailing the overall survival outcome and overall survival time recorded for patients. Both univariate and multivariate Cox Proportional-Hazards models were generated using this maximum RNA and CNA signature data. Clinical stage, histological grade and age at initial pathological diagnosis were included as multivariate confounding variables. Kaplan-Meier survival curves were generated for each data set, with a p value of <0.05 considered significant. Survival analysis was performed using the survminer (v0.4.8) and survival (v v3.2-7) packages (Kassambara et al., 2020; Therneau et al., 2020). In addition, PCA was performed through the PCAtools package (v2.2.0) to assess the level of variance within RNA signature data (Blighe et al., 2021).

5. Results

The following are the main findings of the study. Plots were formed using the ggplot2 (v3.3.3) package and the base plotting functionality in R (Wickham et al., 2020). A full set of results, including specific colour-blind friendly figures, are located within the linked GitHub repository (https://github.com/ConorGilesDoran/UCD_BHLS_Thesis).

5.1 CNA Signature Results

Copy number feature data and model component parameters for 415 TCGA HGSOC samples were provided by Macintyre et. al, allowing each sample to be classified by their underlying CNA profile. A total of seven CNA signatures (s1 – s7) were derived and presented as a signature-by-sample matrix, with HGSOC samples classified by a combination of signatures. The 415 CNA signature results were filtered down to 311 for downstream DE analysis (Fig. 2B). Certain samples were clearly defined by one or two CNA signatures, whereas others involved a multitude of signature patterns, highlighting the intratumoural heterogeneity and variability present within HGSOC samples. The difficulty in classifying HGSOC and generating accurate signature classifications is also evident. Despite removing 104 HGSOC samples due to filtering, the order of maximum CNA signature tallies remained the same (Table 1), and the visual pattern present within the filtered matrix (Fig. 2B) also appeared similar to the original (Fig. 2A). The maximum CNA signature represented the signature with the greatest posterior probability value of being associated with a sample. From both sets of results, the maximum CNA signature per sample was identified (Table 1), wherein s2 and s4 dominated. S7 had minimal involvement and was removed from downstream analysis.

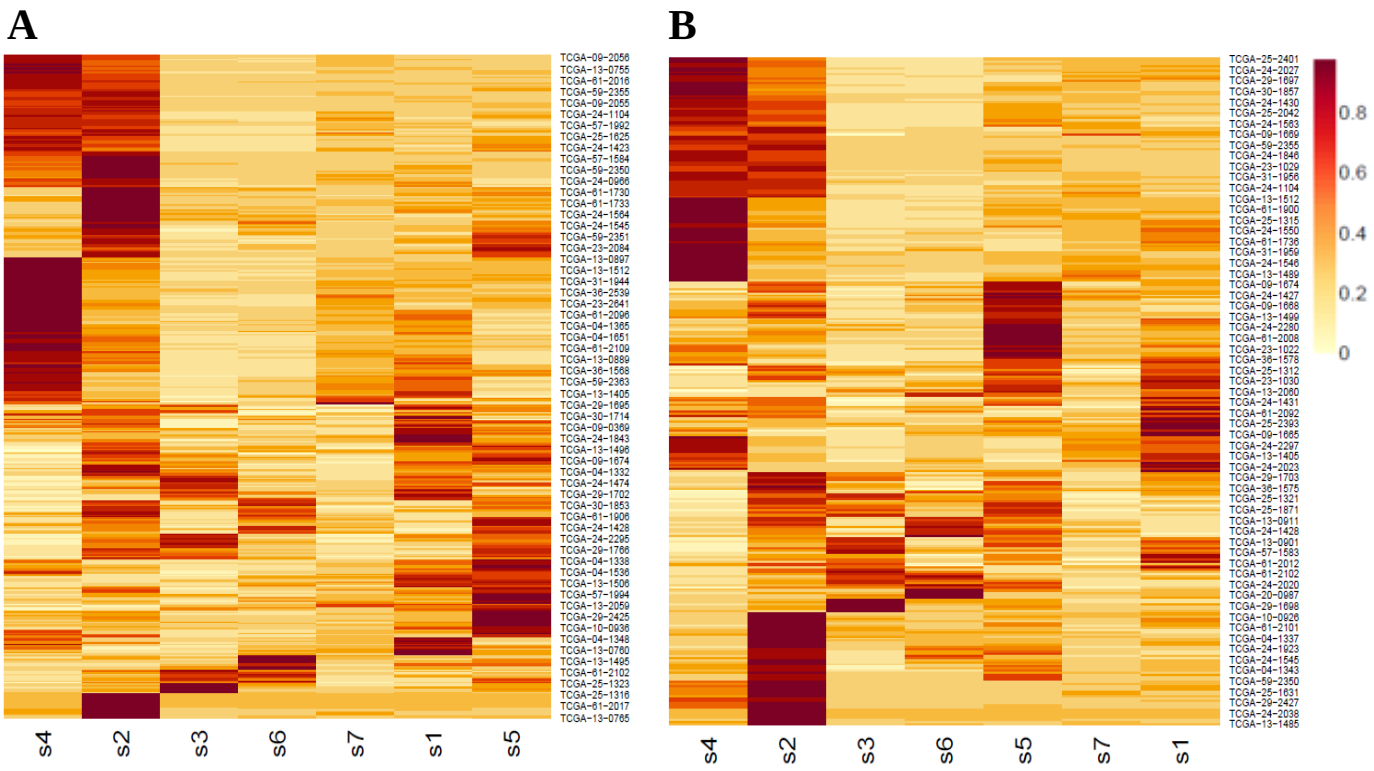


Figure 2 – Normalized signature-by-sample matrices generated using the published CNA signature methods from Macintyre et. al.

(A) Normalized signature-by-sample matrix generated from 415 HGSOC samples. CNA signatures (s1-s7) are displayed on the x axis, while TCGA sample identifiers are shown on the y axis to the right (e.g., TCGA-13-0765). The normalized posterior probability value of each signature being associated with a sample is represented by the extent of colour present within each band of the matrix. The normalized scale of 0-1 has been utilized. Deeper red translates to a higher probability, while a lighter orange demarcates a lower probability. **(B)** Filtered normalized signature-by-sample matrix involving the 311 HGSOC samples to be used for downstream analysis. X and y axes labelling, as well as normalized probability colour scaling, is represented equally to that of part A. CNA s4 and s2 dominated, evident from their dark red bands present within each figure. In the top left of each figure, these dark red bands almost blur into one, making it difficult to separate their distinct associations with a particular sample.

Table 1 – Maximum CNA signature (s) tallies for both the 415 and filtered 311 TCGA HGSOC signature-by-sample datasets.

415 TCGA[a] HGSOC[b] Samples [c]		**311 TCGA HGSOC Samples [d]**	
Max Signature	No. of Samples[e]	Max Signature	No. of Samples
s1	46	s1	35
s2	117	s2	85
s3	30	s3	23
s4	135	s4	104
s5	67	s5	46
s6	19	s6	18
s7	1	s7	0

[a]The Cancer Genome Atlas (TCGA)

[b]High Grade Serous Ovarian Cancer (HGSOC)

[c]Results from the signature-by-sample matrix generated from 415 TCGA HGSOC samples with CNA data available.

[d]Results from the filtered signature-by-sample matrix consisting of 311 TCGA HGSOC samples with matching CNA, clinical and RNA-seq data necessary for downstream analysis.

[e]The tally of samples with a specific maximum signature i.e. the number of samples for which s1-s7 had the highest posterior probability of being associated with.

5.2 Differential Gene Expression Profiles

To computationally analyse the effect of the maximum CNA signatures on differential gene expression, pairwise DE analysis and GSEA were performed using 311 HGSOC samples, allowing for a gene-expression-level profile of each signature. This involved the identification of DEGs within each signature comparison and the subsequent identification of distinct DEGs per individual signature for significant hallmark pathways ($p \leq 0.0020$). The signature enrichment of the 40 hallmark pathways present within the data was derived, revealing the underlying biological mechanisms characteristic of each remaining maximum CNA signature (Fig. 3). CNA signatures significantly represented a combination of biological mechanisms, evident in the fact that 34 of the 40 hallmark pathways were enriched within >3 signatures. As is observed in Figure 3, a number of hallmark pathways, including the complement and coagulation pathways, were present within all six CNA signature gene sets. In contrast, several pathways were only present within a limited number of signatures, such as the reactive oxygen species and apoptosis pathways, which were enriched within two signatures each. Consistent with the maximum CNA signature results of Table 1, s2 and s4 were dominant across all hallmark pathways, whereas s5 and s6 had limited enrichment.

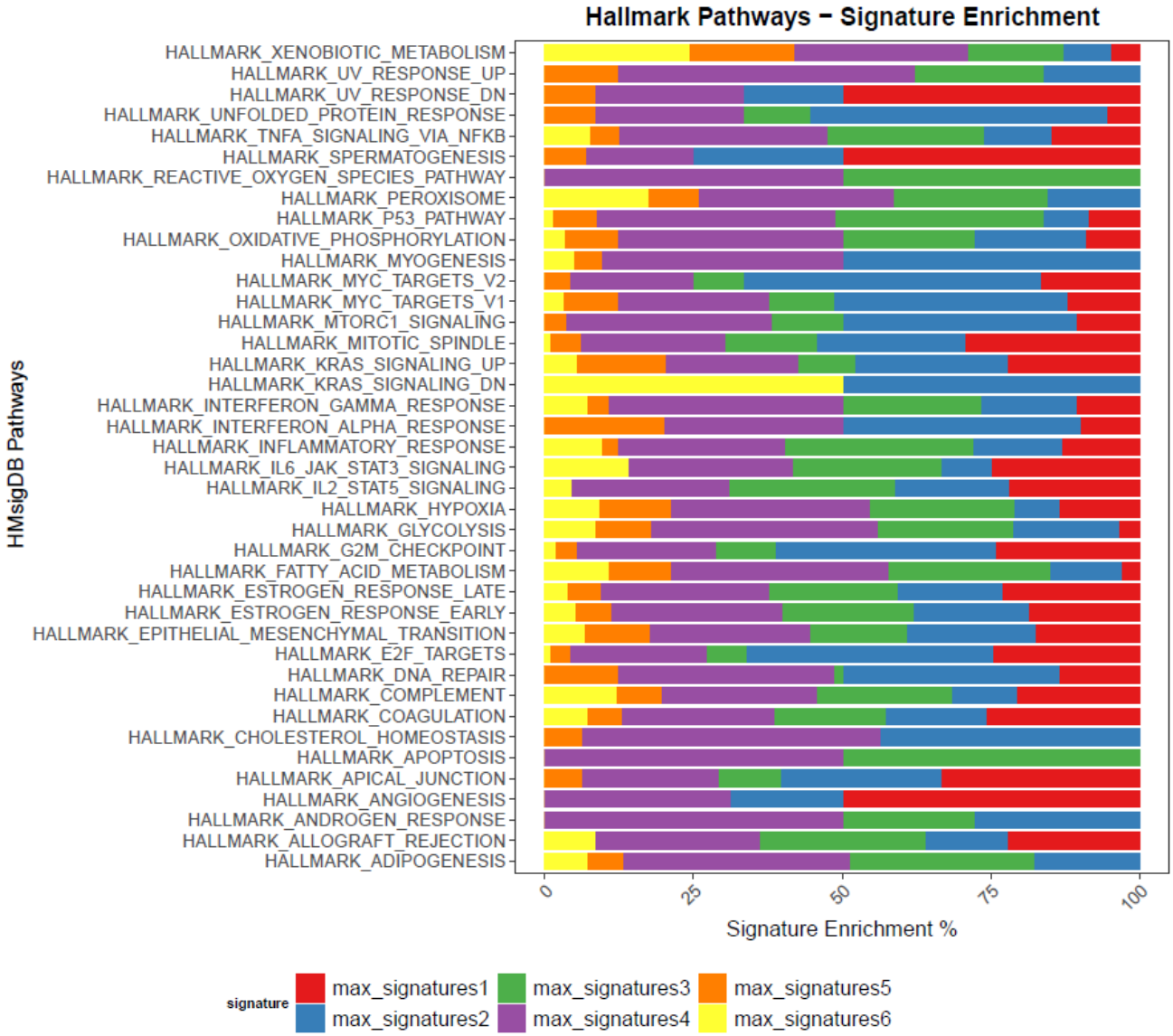


Figure 3 – CNA signature enrichment percentage (%) for 40 hallmark pathways

40 hallmark pathways identified by the Broad Institute MSigDB (7.0) are displayed on the y axis. The x axis ranges from 0-100 and represents the % enrichment of each differentially expressed maximum CNA signature gene set for a specific hallmark pathway. The greater number of CNA signature genes present within the known hallmark pathway gene sets, the higher the enrichment %. Each maximum CNA signature is represented by a distinct colour.

5.3 RNA Signature Results

To investigate whether CNA signatures could be further classified by their DE profiles, RNA signatures were derived. Using the CNA signature enrichment results from section 5.2, the maximum CNA signature per hallmark pathway was identified. A total of four CNA signatures (s1 - s4) were maximally enriched across the 40 hallmark pathways, with no pathways being maximally enriched in s5 or s6. The hallmark pathways maximally enriched by s1 – s4 were recorded and unique gene sets derived for each. The maximum CNA signature involvement per hallmark pathway is broken down in Figure 4 and Table 2. It is evident that s2 and s4 were the CNA signatures with the highest level of pathway involvement, a result that follows the trend of the aforementioned findings. The unique gene sets that were derived for s1 – s4 were subsequently used for ssGSEA, whereby four RNA signatures were derived (r1 –r 4), presented as the RNA signature-by-sample matrix shown in Figure 5. Both r2 and r4 were the most strongly enriched RNA signatures present, with each identified as being the maximum signature for 167 and 144 samples, respectively. The stark contrast in appearances between r2, r4 and r1, r3 (Fig. 5) would be expected, as both r2 and r4 were maximally enriched in a greater number of pathways and involved a larger number of DEGs (Table 2). As mentioned in section 5.1, a blurring of signature boundaries is seen for r2 and r4, as they appear to overlap for particular samples. Furthermore, a similar NES pattern, although to a lesser extent, can be seen for certain samples of r2 and r3 (Fig. 5).

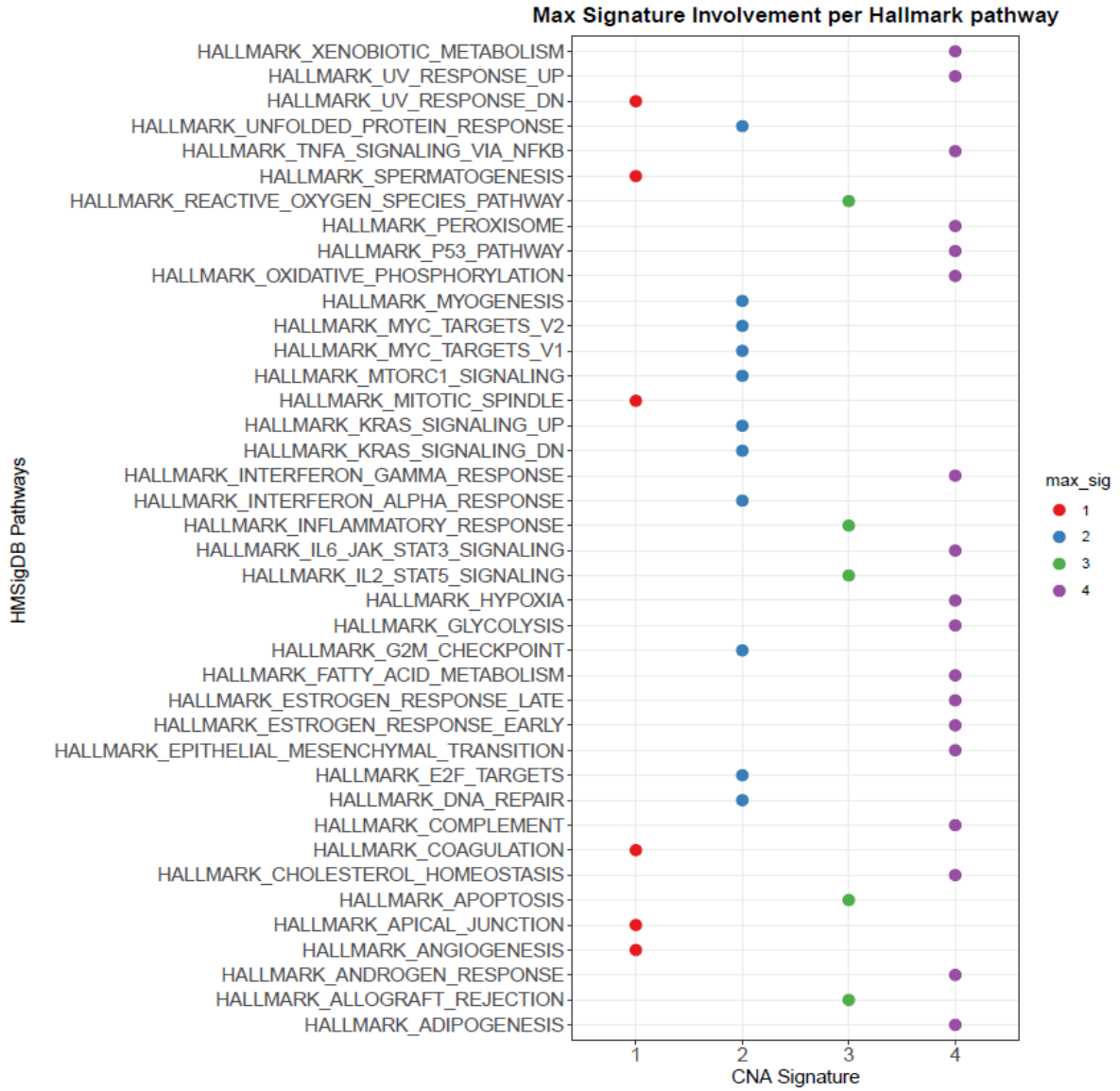


Figure 4 – Maximum signature involvement per hallmark pathway

This plot shows the CNA signature for which each specific hallmark pathway was maximally enriched. The 40 hallmark pathways from the Broad Institute MSigDB (7.0) are displayed on the y axis to the left. Maximum CNA signatures (s1 – s4) are displayed on the x axis. A pathway's maximum CNA signature is represented by an associated coloured point on the plot.

Table 2 – Tally of the maximally enriched pathways and associated DEGs per CNA signature

Maximum CNA signatures	No. of hallmark pathways with maximum enrichment[a]	No. of DEGs involved[b]
s1	6	104
s2	11	363
s3	5	77
s4	18	454
s5	0	-
s6	0	-

[a]The number of hallmark pathways for which each CNA signature was maximally enriched, providing a tally of the results displayed in Figure 3.

[b]The total number of differentially expressed genes (DEGs) that were involved within the maximally enriched hallmark pathways of each signature.

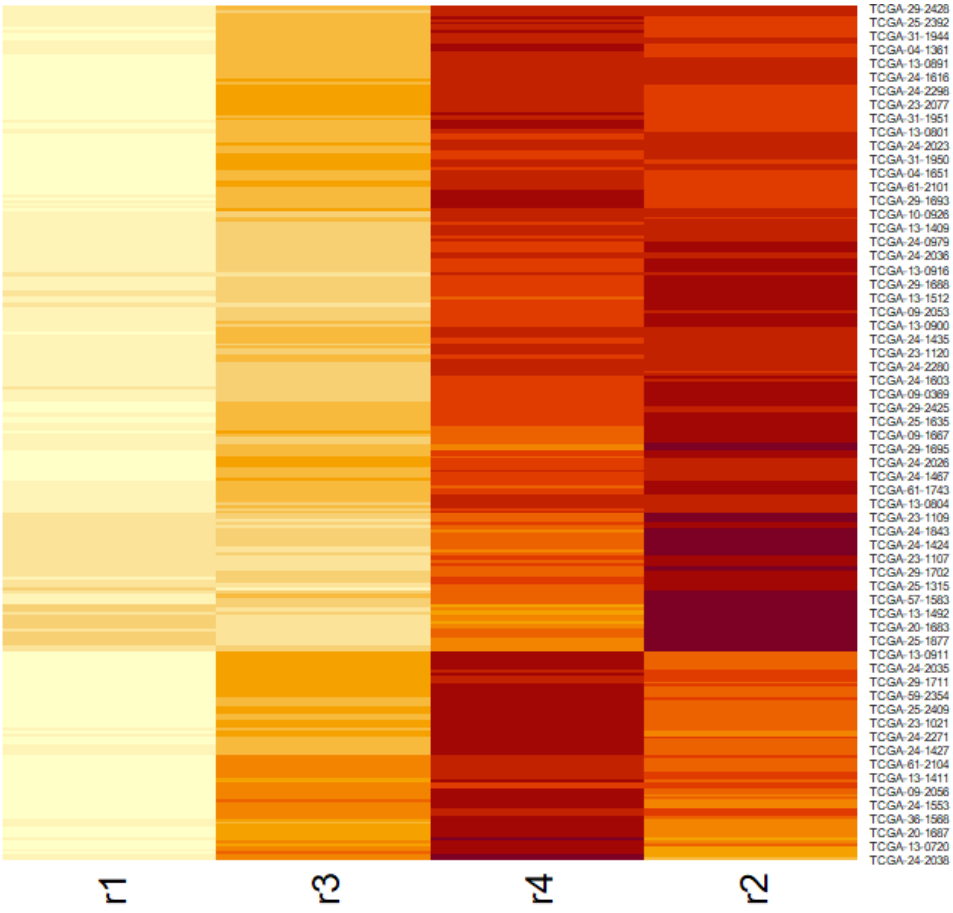


Figure 5 – Normalized RNA signature-by-sample matrix generated by ssGSEA

RNA signatures (r1-r4) are displayed on the x axis, while TCGA sample identifiers are shown on the y axis to the right (e.g., TCGA-13-0720). The NES of signatures for each specific sample is represented by the extent of colour present within each band of the matrix. Deeper red translates to a higher NES, while a lighter orange/yellow translates to a lower NES. Signatures were generated through ssGSEA, with a default false discovery rate (FDR) of < 0.25.

5.4 Survival Results

To assess the clinical translatability of maximum CNA and RNA signatures, their prognostic potential was analysed. Univariate and multivariate Cox Proportional-Hazards regression analysis was performed (Table 3), covering all signature comparisons, allowing for a comprehensive overview of predicted survival outcome across signatures. Kaplan-Meier survival curves were generated for maximum CNA (s1 – s6) and RNA (r2 & r4) signatures, allowing the survival probability of each signature to be visualized over time (Figures 6 and 7). A p-value of < 0.05 was considered a significant result. As observed in Figure 6, maximum CNA s3 and s6 displayed the best survival probability over time, while s5 was associated with the poorest survival outcome. Signatures s4, s2 and s1 displayed more centred survival outcomes, with s2 and s1 curves decreasing in a similar pattern over time. These findings were considered significant ($p < 0.031$). Maximum RNA signatures r2 and r4 displayed similar survival probabilities, with r2 associated with the best survival outcome over time (Fig. 7). However, these results were considered non-significant ($p < 0.21$).

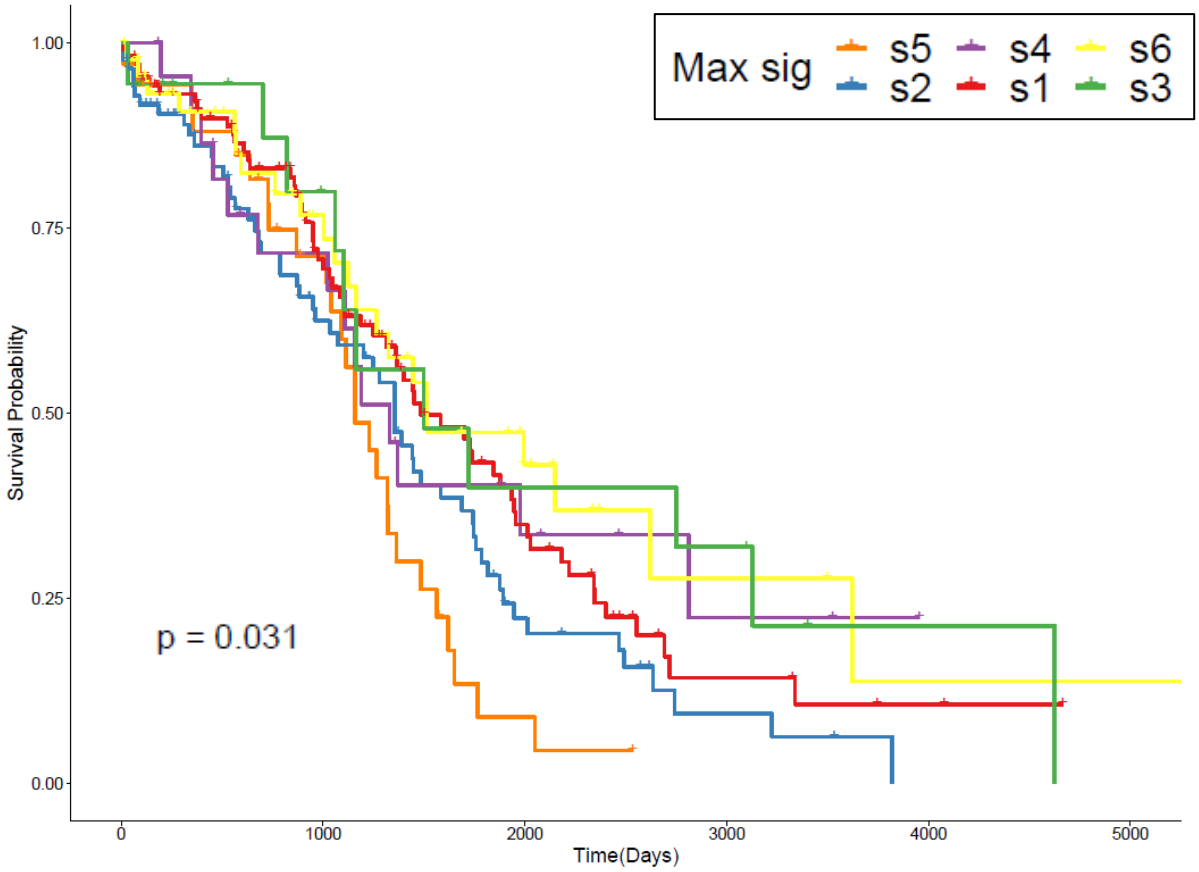


Figure 6 – Kaplan-Meier curves generated for maximum CNA signatures s1 – s6

In the above figure, each coloured curve represents the survival probability of individual maximum CNA signatures over time. Survival probability is displayed on the y axis, ranging from 0 – 1. Time is represented on the x axis by the number of days, ranging from 0 – 5000. The p-value associated with these findings is displayed in the figure as $p = 0.031$, where $p < 0.05$ was considered a significant result.

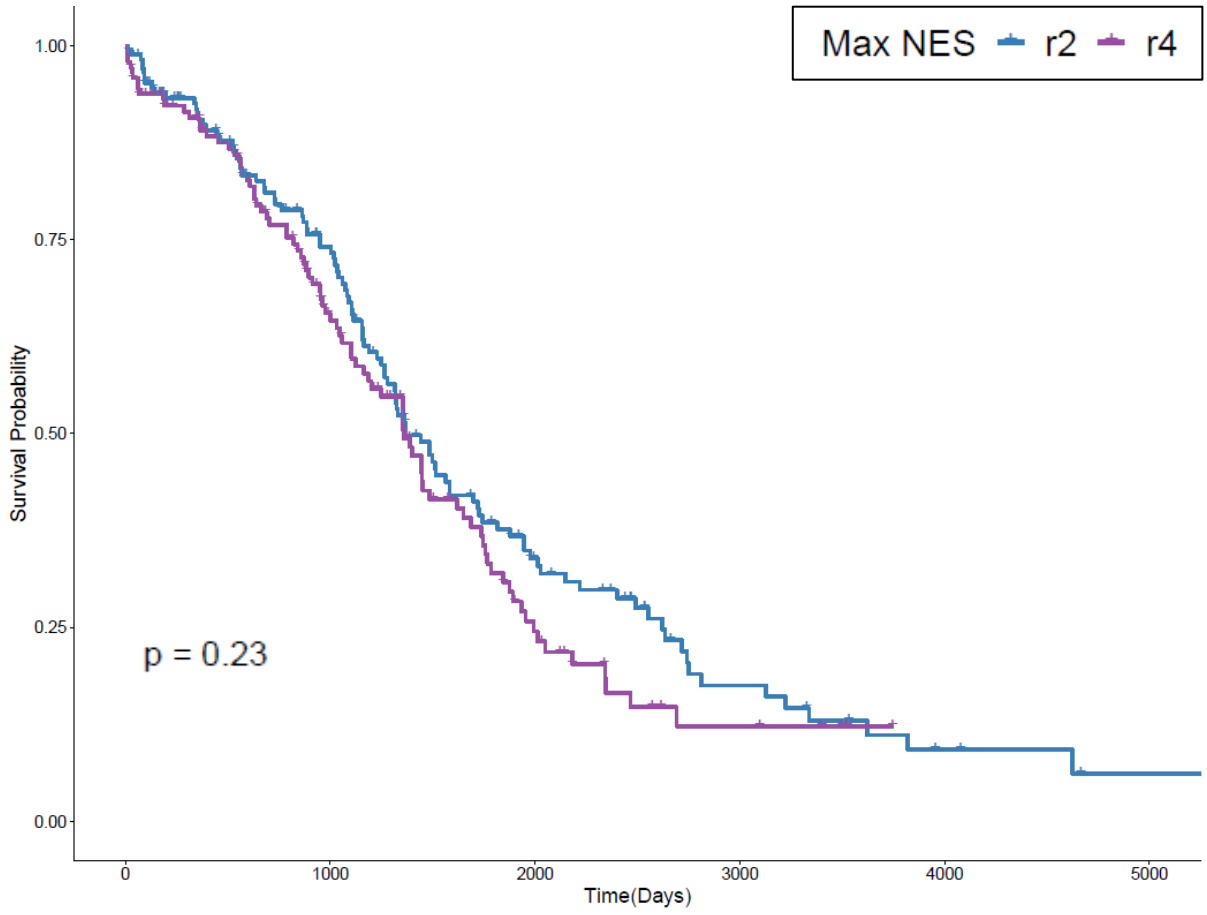


Figure 7 – Kaplan-Meier curves generated for maximum RNA signatures r2 & r4

In the above figure, each curve represents the survival probability of individual maximum RNA signatures over time. Max RNA signatures were identified as the signature representing the highest NES per sample. Survival probability is displayed on the y axis, ranging from 0 – 1. Time is represented on the x axis by the number of days, ranging from 0 – 5000. The p-value associated with these findings is displayed in the figure as $p = 0.23$, where $p < 0.05$ was considered a significant result.

Table 3 – Univariate and multivariate Cox Proportional-Hazards regression models for maximum CNA and RNA signature comparisons.

	Univariate				Multivariate[d]			
Comparison[a]	HR[b]	95% Confidence Interval		*p* value[c]	HR	95% Confidence Interval		*p* value
		Lower	Upper			Lower	Lower	
s1 vs s2	1.33	0.83	2.16	0.239	1.20	0.75	1.98	0.421
s1 vs s3	1.91	0.99	3.68	0.055	1.77	0.89	3.54	0.105
s1 vs s4	1.79	1.12	2.85	**0.014**	1.58	0.98	2.53	0.060
s1 vs s5	2.23	1.25	3.95	**0.006**	2.25	1.23	4.10	**0.008**
s1 vs s6	2.23	1.08	4.58	**0.029**	2.57	1.24	5.34	**0.011**
s2 vs s3	1.44	0.80	2.58	0.225	1.45	0.78	2.69	0.235
s2 vs s4	1.35	0.94	1.94	0.105	1.30	0.89	1.88	0.179
s2 vs s5	1.68	1.03	2.75	**0.039**	1.84	1.10	3.08	**0.020**
s2 vs s6	1.68	0.87	3.23	0.120	2.11	1.08	4.13	**0.030**
s3 vs s4	0.94	0.52	1.68	0.831	0.89	0.48	1.64	0.707
s3 vs s5	1.17	0.60	2.29	0.649	1.27	0.63	2.56	0.509
s3 vs s6	1.17	0.53	2.58	0.700	1.45	0.64	3.27	0.370
s4 vs s5	1.25	0.76	2.03	0.378	1.43	0.85	2.39	0.180
s4 vs s6	1.25	0.65	2.38	0.507	1.63	0.84	3.18	0.151
s5 vs s6	1.00	0.48	2.07	1.000	1.14	0.54	2.42	0.724
r2 vs r4	0.84	0.63	1.12	0.230	0.77	0.57	1.04	0.090

[a]As the Cox Proportional-Hazards regression model required a baseline comparison signature, all possible signature comparison permutations were analysed.

[b]HR = Hazard Ratio

[c]p values - **blue** = significant results of **$p < 0.05$**, **yellow** = significant results of **$p < 0.01$**.

[d]Clinical stage, histological grade and age at initial pathological diagnosis were included as multivariate confounding variables.

6. Discussion

By stratifying patients based on their molecular profile, CNA signatures have offered a possible computational solution to deal with the genomic complexity of HGSOC, aiming to make patient diagnosis and management tailored to a more precise level. In this study, the seven CNA signatures from a previously published HGSOC study were confirmed, wherein the intratumoural heterogeneity among HGSOC samples was evident. By utilizing the published methods of Macintyre et. al, it is evident that computational algorithms are required in order to deconstruct and analyse the intricate genomic profile of a disease like HGSOC. However, the practice of fully implementing the methods of Macintyre et. al posed difficult, with progression made possible only with their findings. Having access to these methods was greatly appreciated, but reproducibility must be ensured, allowing methods to be actively reused and tested by the research community.

From the results presented, one can fail to reject the hypothesis that tumour gene expression is divergent between sets of HGSOC patient groups classified based on CNA signatures. A large number of studies, such as pan-cancer investigations, as well as disease specific analyses, have demonstrated the effect of CNAs on differential gene expression (Dong et al., 2017; Kuzyk et al., 2015; Shao et al., 2019; Zhao et al., 2014). The DE results presented in this study reinforce that knowledge, as the six remaining CNA signatures represented DEGs involved in actual biological pathways. From utilizing these DE profiles, four RNA expression signatures were successfully derived, wherein the sample-specific enrichment of DEGs involved in significant hallmark pathways for maximally enriched CNA signatures was calculated. These results demonstrated the potential for CNA signatures to be further classified based on their RNA expression and emphasized the fact that HGSOC encompasses a multitude of biological mechanisms, giving rise to the complex challenge of stratifying patients. However, only four RNA signatures were derived wherein r2 and r4 dominated, raising

questions as to why other signatures were underrepresented. Pairwise DE analysis resulted in significant overlap possibly masking lowly expressed signatures, limiting RNA signature identification. The problematic variability is likely due to the fundamental nature of CNAs, which drive extensive variation within the cancer genome (Redon et al., 2006; Sebat et al., 2004; Shao et al., 2019; Zarrei et al., 2015). The aim of signature analysis was to investigate if a genome-wide pattern of gene disruption, and hence transcription, could be identified. The CNA signature results highlighted the difficulty in achieving this, with the elaborate genomic interplay underlying the heterogeneity of HGSOC posing many challenges. If repeated, DE comparisons could be made between one signature and all other signatures grouped together, or specific parameters established for gene set derivation. Although warranting further investigation, classifying CNA signatures by RNA expression has demonstrated the potential for HGSOC patient stratification in the future.

A number of interesting comparisons can be made between the findings of this study and the published results (Macintyre et al., 2018). Macintyre et. al identified CNA s1 as representing RAS-MAPK signalling, telomere shortening and BFB events, resulting in a poor prognosis. RNA signature results of this study, formed from the maximum CNA signature expression profiles (Fig. 4), show that r1 was maximally enriched for six hallmark pathways, including the mitotic spindle pathway. This pathway is of particular interest, as breakage-fusion-bridge (BFB) events are known to arise from mechanical and physical stress of mitotic spindle, ultimately leading to significant chromosomal instability (Bohlander et al., 2019; Gisselsson et al., 2000; Kinsella and Bafna, 2012). It is surprising, however, that no hallmark RAS-signalling pathway was maximally enriched for r1, and that CNA s1 was associated with a relatively good prognosis here (Fig. 6). Macintyre et. al stated that CNA s2 is associated with tandem duplications via CDK12 inactivation, leading to a poor survival outcome. Here, RNA signature 2 was maximally enriched for several related hallmarks, including MYC targets V1

and V2, DNA repair and G2/M checkpoint pathways. These findings correlate to the published results, as inactivation of CDK12, a fundamental regulator of the cell cycle, is known to influence the G2/M checkpoint, and that focal tandem duplications from CDK12 inactivation can result in secondary overexpression of c-MYC enhancers (Pilarova et al., 2020). Furthermore, the tandem duplicator phenotype is significantly associated with loss of DNA damage repair (Menghi et al., 2016). Similar to Macintyre et. al., poor prognosis was characteristic of the CNA s2 presented here (Fig.6). CNA s3 of Macintyre et. al was characteristic of BRCA1/2-related HRD. RNA signature 3 was not associated with any pathways related to BRCA1/2 related HRD specifically, but was maximally enriched for immune-related hallmarks, including Il-2/STAT5 signalling, inflammatory response, reactive oxygen species and allograft rejection, suggesting r3 may represent a more immune-focused signature (Heeger and Dinavahi, 2012; Hoppe et al., 2018; Konstantinopoulos et al., 2015; Ross and Cantrell, 2018; Spolski et al., 2018; Yang et al., 2013). Interestingly, it has been suggested that BRCA1/2 related HGSOC tumours display a heightened presence of immune cell infiltrates (McAlpine et al., 2012; Soslow et al., 2012; Strickland et al., 2015). Il-2 is an important cytokine involved in immune cell differentiation and activation-induced cell death, which correlates to the finding that r3 was maximally enriched for the apoptosis hallmark pathway (Fig. 4; Ross and Cantrell, 2018; Spolski et al., 2018). Several studies have demonstrated that an increased level of immune cell infiltrates is also associated with a favourable prognosis, which was similarly observed for CNA s3 presented in this study (Fig. 6; Han et al., 2008; Leffers et al., 2008; Zhang et al., 2003). Macintyre et. al deduced that CNA s4 was linked with whole genome duplication arising from cell cycle dysfunction through PI3K–AKT, TLR cascade and interleukin signalling. RNA signature 4 represented a number of similar pathways, with maximum enrichment observed for IL-6 JAK/STAT3, interferon gamma (IFNγ), p53, and TNFα/ NF-κB hallmark signalling pathways (Fig. 4). Both TNFα/NF-

κB and IL-6 JAK/STAT3 signalling mechanisms interact downstream with the PI3K/AKT pathway, regulating important cellular mechanisms such as survival, proliferation, angiogenesis and inflammation (Ataie-Kachoie et al., 2014; Garbers et al., 2012; Liu et al., 2009; Nidai Ozes et al., 1999; Romashkova and Makarov, 1999; Sansone and Bromberg, 2012; Wegiel et al., 2008). Additionally, PI3K/AKT signalling, acting through MDM2, is an important mediator of p53 regulation (Abraham and O'Neill, 2014). IFNγ is primarily secreted by T cells and natural killer cells and signals through JAK2/JAK1 (Schroder et al., 2004). Macintyre et. al noted the involvement of JAK2 within CNA s4, as well as that of MYC, which is known to interact with p53 (Dang, 2012; Sachdeva et al., 2009; Wierstra and Alves, 2008).

There are several possibilities as to why CNA s5, s6 and s7 were filtered out of downstream analysis. Macintyre et. al concluded that s5 represented chromothriptic-like events. Chromothripsis is a calamitous genomic event that drives extrachromosomal DNA amplification, and is rarely observed in HGSOC (Malhotra et al., 2013; Patch et al., 2015; Shoshani et al., 2020; Zack et al., 2013). However, chromothripsis is known to be associated with poor patient prognosis in other cancers (Hirsch et al., 2013; Magrangeas et al., 2011; Rausch et al., 2012). Consistent with published literature, s5 of this study was significantly associated ($p = 0.031$) with the worst predicted survival outcome of all signatures. The suggested link between s5 and chromothripsis warrants further investigation, and as stated by Macintyre et. al, may highlight its underestimated influence on the ongoing genomic instability of HGSOC. CNA s6 of Macintyre et. al was associated with focal amplifications due to cell cycle dysfunction. Focal CNAs are of limited size and overlap with larger CNAs, making them difficult to isolate. Several software packages have been developed to try and distinguish large and focal CNAs, such as WIFA, CGHcall and GISTIC2.0, but it remains a difficult task, with best practice being a direct comparison with germline DNA matched for patient tumour samples (Krijgsman et al., 2014). The absence of CNA s7 from signature results was due to

sample filtering for DE analysis, preventing comparisons with the published results.

In conclusion, it is evident that CNA signatures have an interesting potential as both future prognostic and diagnostic biomarkers in cancer. With their detailed stratification of HGSOC genomes, they enabled insights into the fundamental mutational mechanisms that make tumorigenesis so prolific. It will be a difficult framework to implement clinically in the near future, but computational genomics will allow for significant advancements to be made long-term. Complex diseases like HGSOC will require complex solutions, and with the wealth of omics data currently being generated, these solutions are needed and there is potential. For clinical validity, the mathematical derivation of CNA signatures must be refined, with the incorporation of larger sample sizes, stringent experimental studies, as well as clinical trials. In current research, the concept of JCN and single-cell CNAs will develop in the coming years (Dago et al., 2014; Navin et al., 2011). The published methods of Macintyre et. al should continue to be tested on other cancers, such as uterine serous carcinoma, which has genomic similarities to HGSOC (Levine et al., 2013). The complex genomic landscape of HGSOC poses arduous challenges for researchers and clinicians, with the development of improved stratification, diagnostics, prognostics, and treatments all required for overcoming this prevalent disease. CNA signatures represent a new form of biomarker that can aid in this process, aiming to decipher the underlying complexity of cancer and improve the lives of those effected by its impact.

7. Acknowledgements

Sincere thanks to Professor Donal Brennan and Doctor Bruce Moran for the opportunity to work on this stimulating project. Their continued guidance, communication and time was greatly appreciated, especially during the ongoing COVID-19 pandemic. During these testing times, meeting virtually with Dr. Bruce Moran semi-weekly was a massive help personally and for which I am very thankful. This project was really interesting to work on and I thank them both for the enjoyable and enriching experience.

Thank you also to my family members for all the brilliant support during my time in UCD and enabling me to pursue my degree.

Finally, I wish to thank Professor William Watson for his guidance and enthusiasm throughout the four years, as well as the BHLS staff for organising and supporting the final year projects during the pandemic.

Word Count: 5998 (excluding title page, table of contents, tables, acknowledgements, references, appendices)

8. References

Abraham, A., O'Neill, E., 2014. PI3K/AKT-mediated regulation of p53 in cancer. Biochemical Society transactions 42, 798–803. https://doi.org/10.1042/BST20140070

Ahmed, A.A., Etemadmoghadam, D., Temple, J., Lynch, A.G., Riad, M., Sharma, R., Stewart, C., Fereday, S., Caldas, C., deFazio, A., Bowtell, D., Brenton, J.D., 2010. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. The Journal of Pathology 221, 49–56. https://doi.org/10.1002/path.2696

Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., Islam, S.M.A., Lopez-Bigas, N., Klimczak, L.J., McPherson, J.R., Morganella, S., Sabarinathan, R., Wheeler, D.A., Mustonen, V., Getz, G., Rozen, S.G., Stratton, M.R., 2020. The repertoire of mutational signatures in human cancer. Nature 578, 94–101. https://doi.org/10.1038/s41586-020-1943-3

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjörd, J.E., Foekens, J.A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jäger, N., Jones, D.T.W., Jones, D., Knappskog, S., Kool, M., Lakhani, S.R., López-Otín, C., Martin, S., Munshi, N.C., Nakamura, H., Northcott, P.A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J.V., Puente, X.S., Raine, K., Ramakrishna, M., Richardson, A.L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T.N., Span, P.N., Teague, J.W., Totoki, Y., Tutt, A.N.J., Valdés-Mas, R., van Buuren, M.M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L.R., Zucman-Rossi, J., Andrew Futreal, P., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S.M., Siebert, R., Campo, E., Shibata, T., Pfister, S.M., Campbell, P.J., Stratton, M.R., 2013. Signatures of mutational processes in human cancer. Nature 500, 415–421. https://doi.org/10.1038/nature12477

Almal, S.H., Padh, H., 2012. Implications of gene copy-number variation in health and diseases. J Hum Genet 57, 6–13. https://doi.org/10.1038/jhg.2011.108

Ataie-Kachoie, P., Pourgholami, M.H., Richardson, D.R., Morris, D.L., 2014. Gene of the month: Interleukin 6 (IL-6). Journal of Clinical Pathology 67, 932–937. https://doi.org/10.1136/jclinpath-2014-202493

Bao, L., Pu, M., Messer, K., 2014. AbsCN-seq: a statistical method to estimate tumor purity, ploidy and absolute copy numbers from next-generation sequencing data. Bioinformatics 30, 1056–1063. https://doi.org/10.1093/bioinformatics/btt759

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E.M., Sos, M.L., Michel, K., Mermel, C., Silver, S.J., Weir, B.A., Reiling, J.H., Sheng, Q., Gupta, P.B., Wadlow, R.C., Le, H., Hoersch, S., Wittner, B.S., Ramaswamy, S., Livingston, D.M., Sabatini, D.M., Meyerson, M., Thomas, R.K., Lander, E.S., Mesirov, J.P., Root, D.E., Gilliland, D.G., Jacks, T., Hahn, W.C., 2009. Systematic RNA interference reveals that oncogenic KRAS -driven cancers require TBK1. Nature 462, 108–112. https://doi.org/10.1038/nature08460

Benstead-Hume, G., Wooller, S.K., Downs, J.A., Pearl, F.M.G., 2019. Defining Signatures of Arm-Wise Copy Number Change and Their Associated Drivers in Kidney Cancers. Int J Mol Sci 20. https://doi.org/10.3390/ijms20225762

Bischl, B., Lang, M., Bossek, J., Horn, D., Richter, J., Surmann, D., 2017. BBmisc: Miscellaneous Helper Functions for B. Bischl.

Blighe, K., Brown, A.-L., Carey, V., Hooiveld, G., Lun, A., 2021. PCAtools: PCAtools: Everything Principal Components Analysis. Bioconductor version: Release (3.12). https://doi.org/10.18129/B9.bioc.PCAtools

Bohlander, S.K., Kakadiya, P.M., Coysh, A., 2019. Chromosome Rearrangements and Translocations, in: Boffetta, P., Hainaut, P. (Eds.), Encyclopedia of Cancer (Third Edition). Academic Press, Oxford, pp. 389–404. https://doi.org/10.1016/B978-0-12-801238-3.65105-X

Bowtell, D.D., Böhm, S., Ahmed, A.A., Aspuria, P.-J., Bast, R.C., Beral, V., Berek, J.S., Birrer, M.J., Blagden, S., Bookman, M.A., Brenton, J.D., Chiappinelli, K.B., Martins, F.C., Coukos, G., Drapkin, R., Edmondson, R., Fotopoulou, C., Gabra, H., Galon, J., Gourley, C., Heong, V., Huntsman, D.G., Iwanicki, M., Karlan, B.Y., Kaye, A., Lengyel, E., Levine, D.A., Lu, K.H., McNeish, I.A., Menon, U., Narod, S.A., Nelson, B.H., Nephew, K.P., Pharoah, P., Powell, D.J., Ramos, P., Romero, I.L., Scott, C.L., Sood, A.K., Stronach, E.A., Balkwill, F.R., 2015. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. Nature Reviews Cancer 15, 668–679. https://doi.org/10.1038/nrc4019

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 34, 525–527. https://doi.org/10.1038/nbt.3519

Califf, R.M., 2018. Biomarker definitions and their applications. Exp Biol Med (Maywood) 243, 213–221. https://doi.org/10.1177/1535370217750088

Centre for Drug Evaluation and Research, F., 2019. FDA approves bevacizumab in combination with chemotherapy for ovarian cancer. FDA.

Chen, H., Bell, J.M., Zavala, N.A., Ji, H.P., Zhang, N.R., 2015. Allele-specific copy number profiling by next-generation DNA sequencing. Nucleic Acids Res 43, e23–e23. https://doi.org/10.1093/nar/gku1252

Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P., Onofrio, R., Winckler, W., Weir, B., Beroukhim, R., Pellman, D., Levine, D., Lander, E., Meyerson, M., Getz, G., 2012. Absolute quantification of somatic DNA alterations in human cancer. Nature biotechnology 30, 413–21. https://doi.org/10.1038/nbt.2203

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., Sander, C., 2013. Emerging landscape of oncogenic signatures across human cancers. Nature Genetics 45, 1127–1133. https://doi.org/10.1038/ng.2762

Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E., Pritchard, J.K., 2006. A high-resolution survey of deletion polymorphism in the human genome. Nature Genetics 38, 75–81. https://doi.org/10.1038/ng1697

Corchete, L.A., Rojas, E.A., Alonso-López, D., De Las Rivas, J., Gutiérrez, N.C., Burguillo, F.J., 2020. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. Scientific Reports 10, 19737. https://doi.org/10.1038/s41598-020-76881-x

Cortez, A.J., Tudrej, P., Kujawa, K.A., Lisowska, K.M., 2018. Advances in ovarian cancer therapy. Cancer Chemother Pharmacol 81, 17–38. https://doi.org/10.1007/s00280-017-3501-8

Dago, A.E., Stepansky, A., Carlsson, A., Luttgen, M., Kendall, J., Baslan, T., Kolatkar, A., Wigler, M., Bethel, K., Gross, M.E., Hicks, J., Kuhn, P., 2014. Rapid Phenotypic and Genomic Change in Response to Therapeutic Pressure in Prostate Cancer Inferred by High Content Analysis of Single Circulating Tumor Cells. PLOS ONE 9, e101777. https://doi.org/10.1371/journal.pone.0101777

Dang, C.V., 2012. MYC on the Path to Cancer. Cell 149, 22–35. https://doi.org/10.1016/j.cell.2012.03.003

Degasperi, A., Amarante, T.D., Czarnecki, J., Shooter, S., Zou, X., Glodzik, D., Morganella, S., Nanda, A.S., Badja, C., Koh, G., Momen, S.E., Georgakopoulos-Soares, I., Dias, J.M.L., Young, J., Memari, Y., Davies, H., Nik-Zainal, S., 2020. A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. Nature Cancer 1, 249–263. https://doi.org/10.1038/s43018-020-0027-5

Doherty, G.J., Petruzzelli, M., Beddowes, E., Ahmad, S.S., Caldas, C., Gilbertson, R.J., 2019. Cancer Treatment in the Genomic Era. Annual Review of Biochemistry 88, 247–280. https://doi.org/10.1146/annurev-biochem-062917-011840

Dolgalev, I., 2020. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format.

Dong, Gaochao, Mao, Q., Yu, D., Zhang, Y., Qiu, M., Dong, Gaoyue, Chen, Q., Xia, W., Wang, J., Xu, L., Jiang, F., 2017. Integrative analysis of copy number and transcriptional expression profiles in esophageal cancer to identify a novel driver gene for therapy. Scientific Reports 7, 42060. https://doi.org/10.1038/srep42060

Essers, P.B.M., Heijden, M. van der, Vossen, D., Roest, R.H. de, Leemans, C.R., Brakenhoff, R.H., Brekel, M.W.M. van den, Bartelink, H., Verheij, M., Vens, C., 2020. Ovarian cancer-derived copy number alterations signatures are prognostic in chemoradiotherapy-treated head and neck squamous cell carcinoma. International Journal of Cancer 147, 1732–1739. https://doi.org/10.1002/ijc.32962

Fanciulli, M., Petretto, E., Aitman, T.J., 2010. Gene copy number variation and common human disease. Clin Genet 77, 201–213. https://doi.org/10.1111/j.1399-0004.2009.01342.x

Feuk, L., Carson, A.R., Scherer, S.W., 2006. Structural variation in the human genome. Nature Reviews Genetics 7, 85–97. https://doi.org/10.1038/nrg1767

Freeman, J.L., 2006. Copy number variation: New insights in genome diversity. Genome Research 16, 949–961. https://doi.org/10.1101/gr.3677206

Garbers, C., Hermanns, H.M., Schaper, F., Müller-Newen, G., Grötzinger, J., Rose-John, S., Scheller, J., 2012. Plasticity and cross-talk of Interleukin 6-type cytokines. Cytokine & Growth Factor Reviews 23, 85–97. https://doi.org/10.1016/j.cytogfr.2012.04.001

Gibbs, R.A., Belmont, J.W., Hardenbol, P., Willis, T.D., Yu, F., Yang, H., Ch'ang, L.-Y., Huang, W., Liu, B., Shen, Y., Tam, P.K.-H., Tsui, L.-C., Waye, M.M.Y., Wong, J.T.-F., Zeng, C., Zhang, Q., Chee, M.S., Galver, L.M., Kruglyak, S., Murray, S.S., Oliphant, A.R., Montpetit, A., Hudson, T.J., Chagnon, F., Ferretti, V., Leboeuf, M., Phillips, M.S., Verner, A., Kwok, P.-Y., Duan, S., Lind, D.L., Miller, R.D., Rice, J.P., Saccone, N.L., Taillon-Miller, P., Xiao, M., Nakamura, Y., Sekine, A., Sorimachi, K., Tanaka, T., Tanaka, Y., Tsunoda, T., Yoshino, E., Bentley, D.R., Deloukas, P., Hunt, S., Powell, D., Altshuler, D., Gabriel, S.B., Zhang, H., Zeng, C., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Aniagwu, T., Marshall, P.A., Matthew, O., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Stein, L.D., Cunningham, F., Kanani, A., Thorisson, G.A., Chakravarti, A.,

Chen, P.E., Cutler, D.J., Kashuk, C.S., Donnelly, P., Marchini, J., McVean, G.A.T., Myers, S.R., Cardon, L.R., Abecasis, G.R., Morris, A., Weir, B.S., Mullikin, J.C., Sherry, S.T., Feolo, M., Altshuler, D., Daly, M.J., Schaffner, S.F., Qiu, R., Kent, A., Dunston, G.M., Kato, K., Niikawa, N., Knoppers, B.M., Foster, M.W., Clayton, E.W., Wang, V.O., Watkin, J., Gibbs, R.A., Belmont, J.W., Sodergren, E., Weinstock, G.M., Wilson, R.K., Fulton, L.L., Rogers, J., Birren, B.W., Han, H., Wang, H., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Todani, K., Fujita, T., Tanaka, S., Holden, A.L., Lai, E.H., Collins, F.S., Brooks, L.D., McEwen, J.E., Guyer, M.S., Jordan, E., Peterson, J.L., Spiegel, J., Sung, L.M., Zacharia, L.F., Kennedy, K., Dunn, M.G., Seabrook, R., Shillito, M., Skene, B., Stewart, J.G., Valle (chair), D.L., Clayton (co-chair), E.W., Jorde (co-chair), L.B., Belmont, J.W., Chakravarti, A., Cho, M.K., Duster, T., Foster, M.W., Jasperse, M., Knoppers, B.M., Kwok, P.-Y., Licinio, J., Long, J.C., Marshall, P.A., Ossorio, P.N., Wang, V.O., Rotimi, C.N., Royal, C.D.M., Spallone, P., Terry, S.F., Lander (chair), E.S., Lai (co-chair), E.H., Nickerson (co-chair), D.A., Abecasis, G.R., Altshuler, D., Bentley, D.R., Boehnke, M., Cardon, L.R., Daly, M.J., Deloukas, P., Douglas, J.A., Gabriel, S.B., Hudson, R.R., Hudson, T.J., Kruglyak, L., Kwok, P.-Y., Nakamura, Y., Nussbaum, R.L., Royal, C.D.M., Schaffner, S.F., Sherry, S.T., Stein, L.D., Tanaka, T., †The International HapMap Consortium, Genotyping centres: Baylor College of Medicine and ParAllele BioScience, Chinese HapMap Consortium, Illumina, McGill University and Génome Québec Innovation Centre, University of California at San Francisco and Washington University, University of Tokyo and RIKEN, Wellcome Trust Sanger Institute, Whitehead Institute/MIT Center for Genome Research, Community engagement/public consultation and sample-collection groups: Beijing Normal University and Beijing Genomics Institute, Health Sciences University of Hokkaido, E.E.I. and S.U., Howard University and University of Ibadan, University of Utah, Analysis Groups: Cold Spring Harbor Laboratory, Johns Hopkins University School of Medicine, University of Oxford, University of Oxford, W.T.C. for H.G., US National Institutes of Health, Ethical, L. and S.I.C.A. of S.S., Genetic Interest Group, Howard University, Kyoto University, Nagasaki University, University of Montréal, University of Oklahoma, Vanderbilt University, Wellcome Trust, SNP Discovery: Baylor College of Medicine, Washington University, Scientific Management: Chinese Academy of Sciences, Chinese Ministry of Science and Technology, Genome Canada, Génome Québec, Japanese Ministry of Education, C., Sports, Science and Technology, The SNP Consortium, Initial Planning Groups: Populations and Ethical, L. and S.I.G., Methods Group, 2003. The International HapMap Project. Nature 426, 789–796. https://doi.org/10.1038/nature02168

Gisselsson, D., Pettersson, L., Höglund, M., Heidenblad, M., Gorunova, L., Wiegant, J., Mertens, F., Cin, P.D., Mitelman, F., Mandahl, N., 2000. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. PNAS 97, 5357–5362. https://doi.org/10.1073/pnas.090013497

Gockley, A., Melamed, A., Bregar, A.J., Clemmer, J.T., Birrer, M., Schorge, J.O., del Carmen, M.G., Rauh-Hain, J.A., 2017. Outcomes of Women With High-Grade and Low-Grade Advanced-Stage Serous Epithelial Ovarian Cancer. Obstetrics & Gynecology 129, 439–447. https://doi.org/10.1097/AOG.0000000000001867

Gross, J., Ligges, U., 2015. nortest: Tests for Normality.

Ha, G., Roth, A., Khattra, J., Ho, J., Yap, D., Prentice, L., Melnyk, N., McPherson, A., Bashashati, A., Laks, E., Biele, J., Ding, J., Le, A., Rosner, J., Shumansky, K., Marra, M., Gilks, C., Huntsman, D., Mcalpine, J., Shah, S., 2014. TITAN: Inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. Genome research 24. https://doi.org/10.1101/gr.180281.114

Hadi, K., Yao, X., Behr, J.M., Deshpande, A., Xanthopoulakis, C., Tian, H., Kudman, S., Rosiene, J., Darmofal, M., DeRose, J., Mortensen, R., Adney, E.M., Shaiber, A., Gajic, Z., Sigouros, M., Eng, K.,

Wala, J.A., Wrzeszczyński, K.O., Arora, K., Shah, M., Emde, A.-K., Felice, V., Frank, M.O., Darnell, R.B., Ghandi, M., Huang, F., Dewhurst, S., Maciejowski, J., de Lange, T., Setton, J., Riaz, N., Reis-Filho, J.S., Powell, S., Knowles, D.A., Reznik, E., Mishra, B., Beroukhim, R., Zody, M.C., Robine, N., Oman, K.M., Sanchez, C.A., Kuhner, M.K., Smith, L.P., Galipeau, P.C., Paulson, T.G., Reid, B.J., Li, X., Wilkes, D., Sboner, A., Mosquera, J.M., Elemento, O., Imielinski, M., 2020. Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. Cell 183, 197-210.e32. https://doi.org/10.1016/j.cell.2020.08.006

Han, L.Y., Fletcher, M.S., Urbauer, D.L., Mueller, P., Landen, C.N., Kamat, A.A., Lin, Y.G., Merritt, W.M., Spannuth, W.A., Deavers, M.T., De Geest, K., Gershenson, D.M., Lutgendorf, S.K., Ferrone, S., Sood, A.K., 2008. HLA Class I Antigen Processing Machinery Component Expression and Intratumoral T-Cell Infiltrate as Independent Prognostic Markers in Ovarian Carcinoma. Clin Cancer Res 14, 3372–3379. https://doi.org/10.1158/1078-0432.CCR-07-4433

Hänzelmann, S., Castelo, R., Guinney, J., 2013. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics 14, 7. https://doi.org/10.1186/1471-2105-14-7

Hastings, P., Lupski, J.R., Rosenberg, S.M., Ira, G., 2009. Mechanisms of change in gene copy number. Nat Rev Genet 10, 551–564. https://doi.org/10.1038/nrg2593

Heeger, P.S., Dinavahi, R., 2012. Transplant immunology for non-immunologist. Mt Sinai J Med 79, 376–387. https://doi.org/10.1002/msj.21314

Hieronymus, H., Murali, R., Tin, A., Yadav, K., Abida, W., Moller, H., Berney, D., Scher, H., Carver, B., Scardino, P., Schultz, N., Taylor, B., Vickers, A., Cuzick, J., Sawyers, C.L., 2018. Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. Elife 7. https://doi.org/10.7554/eLife.37294

Hirsch, D., Kemmerling, R., Davis, S., Camps, J., Meltzer, P.S., Ried, T., Gaiser, T., 2013. Chromothripsis and Focal Copy Number Alterations Determine Poor Outcome in Malignant Melanoma. Cancer Res 73, 1454–1460. https://doi.org/10.1158/0008-5472.CAN-12-0928

Hoppe, M.M., Sundar, R., Tan, D.S.P., Jeyasekharan, A.D., 2018. Biomarkers for Homologous Recombination Deficiency in Cancer. JNCI: Journal of the National Cancer Institute 110, 704–713. https://doi.org/10.1093/jnci/djy085

Hornik, K., 2012. The Comprehensive R Archive Network. WIREs Computational Statistics 4, 394–398. https://doi.org/10.1002/wics.1212

Huang, Y.-S., Liu, W.-B., Han, F., Yang, J.-T., Hao, X.-L., Chen, H.-Q., Jiang, X., Yin, L., Ao, L., Cui, Z.-H., Cao, J., Liu, J.-Y., 2017. Copy number variations and expression of MPDZ are prognostic biomarkers for clear cell renal cell carcinoma. Oncotarget 8, 78713–78725. https://doi.org/10.18632/oncotarget.20220

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K.D., Irizarry, R.A., Lawrence, M., Love, M.I., MacDonald, J., Obenchain, V., Oleś, A.K., Pagès, H., Reyes, A., Shannon, P., Smyth, G.K., Tenenbaum, D., Waldron, L., Morgan, M., 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12, 115–121. https://doi.org/10.1038/nmeth.3252

Iafrate, A., Feuk, L., Rivera, M., Listewnik, M., Donahoe, P., Qi, Y., Scherer, S., Lee, C., 2004. Detection of large-scale variation in the human genome. Nature genetics 36, 949–51. https://doi.org/10.1038/ng1416

Illumina Inc., 2017. An Introduction to Next-Generation Sequencing Technology.

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.-C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., Bras, J.M., Schymick, J.C., Hernandez, D.G., Traynor, B.J., Simon-Sanchez, J., Matarin, M., Britton, A., van de Leemput, J., Rafferty, I., Bucan, M., Cann, H.M., Hardy, J.A., Rosenberg, N.A., Singleton, A.B., 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451, 998–1003. https://doi.org/10.1038/nature06742

Kassambara, A., Kosinski, M., Biecek, P., Fabian, S., 2020. survminer: Drawing Survival Curves using "ggplot2."

Kinsella, M., Bafna, V., 2012. Combinatorics of the Breakage-Fusion-Bridge Mechanism. Journal of Computational Biology 19, 662–678. https://doi.org/10.1089/cmb.2012.0020

Koh, G., Zou, X., Nik-Zainal, S., 2020. Mutational signatures: experimental design and analytical framework. Genome Biology 21, 37. https://doi.org/10.1186/s13059-020-1951-5

Konstantinopoulos, P.A., Ceccaldi, R., Shapiro, G.I., D'Andrea, A.D., 2015. Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. Cancer Discov 5, 1137–1154. https://doi.org/10.1158/2159-8290.CD-15-0714

Korbel, J.O., Campbell, P.J., 2013. Criteria for Inference of Chromothripsis in Cancer Genomes. Cell 152, 1226–1236. https://doi.org/10.1016/j.cell.2013.02.023

Korotkevich, G., Sukhov, V., Budin, N., Sergushichev, A., 2021. Fast Gene Set Enrichment Analysis.

Krijgsman, O., Carvalho, B., Meijer, G.A., Steenbergen, R.D.M., Ylstra, B., 2014. Focal chromosomal copy number aberrations in cancer—Needles in a genome haystack. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1843, 2698–2704. https://doi.org/10.1016/j.bbamcr.2014.08.001

Kumaran, M., Cass, C.E., Graham, K., Mackey, J.R., Hubaux, R., Lam, W., Yasui, Y., Damaraju, S., 2017. Germline copy number variations are associated with breast cancer risk and prognosis. Sci Rep 7, 14621. https://doi.org/10.1038/s41598-017-14799-7

Kuzyk, A., Booth, S., Righolt, C., Mathur, S., Gartner, J., Mai, S., 2015. MYCN overexpression is associated with unbalanced copy number gain, altered nuclear location, and overexpression of chromosome arm 17q genes in neuroblastoma tumors and cell lines. Genes, Chromosomes and Cancer 54, 616–628. https://doi.org/10.1002/gcc.22273

Leffers, N., Gooden, M.J.M., de Jong, R.A., Hoogeboom, B.-N., ten Hoor, K.A., Hollema, H., Boezen, H.M., van der Zee, A.G.J., Daemen, T., Nijman, H.W., 2008. Prognostic significance of tumor-infiltrating T-lymphocytes in primary and metastatic lesions of advanced stage ovarian cancer. Cancer Immunol Immunother 58, 449. https://doi.org/10.1007/s00262-008-0583-5

Levine, D.A., The Cancer Genome Atlas Research Network, Getz, G., Gabriel, S.B., Cibulskis, K., Lander, E., Sivachenko, A., Sougnez, C., Lawrence, M., Kandoth, C., Dooling, D., Fulton, R., Fulton, L., Kalicki-Veizer, J., McLellan, M.D., O'Laughlin, M., Schmidt, H., Wilson, R.K., Ye, K., Ding, L., Mardis, E.R., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y.S.N., Carlsen, R., Carter, C., Chu, A., Chuah, E., Chun, H.-J.E., Dhalla, N., Guin, R., Hirst, C., Holt, R.A., Jones, S.J.M., Lee, D., Li, H.I., Marra, M.A., Mayo, M., Moore, R.A., Mungall, A.J., Plettner, P., Schein, J.E., Sipahimalani, P., Tam, A., Varhol, R.J., Gordon Robertson, A., Cherniack, A.D., Pashtan, I., Saksena, G., Onofrio, R.C., Schumacher, S.E., Tabak, B., Carter, S.L., Hernandez, B., Gentry, J., Salvesen, H.B., Ardlie, K., Getz, G., Winckler, W., Beroukhim, R., Gabriel, S.B., Meyerson, M., Hadjipanayis, A., Lee, S., Mahadeshwar, H.S., Park, P.,

Protopopov, A., Ren, X., Seth, S., Song, X., Tang, J., Xi, R., Yang, Lixing, Zeng, D., Kucherlapati, R., Chin, L., Zhang, J., Todd Auman, J., Balu, S., Bodenheimer, T., Buda, E., Neil Hayes, D., Hoyle, A.P., Jefferys, S.R., Jones, C.D., Meng, S., Mieczkowski, P.A., Mose, L.E., Parker, J.S., Perou, C.M., Roach, J., Shi, Y., Simons, J.V., Soloway, M.G., Tan, D., Topal, M.D., Waring, S., Wu, J., Hoadley, K.A., Baylin, S.B., Bootwalla, M.S., Lai, P.H., Triche Jr, T.J., Van Den Berg, D.J., Weisenberger, D.J., Laird, P.W., Shen, H., Chin, L., Zhang, J., Getz, G., Cho, J., DiCara, D., Frazer, S., Heiman, D., Jing, R., Lin, P., Mallard, W., Stojanov, P., Voet, D., Zhang, H., Zou, L., Noble, M., Lawrence, M., Reynolds, S.M., Shmulevich, I., Arman Aksoy, B., Antipin, Y., Ciriello, G., Dresdner, G., Gao, J., Gross, B., Jacobsen, A., Ladanyi, M., Reva, B., Sander, C., Sinha, R., Onur Sumer, S., Taylor, B.S., Cerami, E., Weinhold, N., Schultz, N., Shen, R., Benz, S., Goldstein, T., Haussler, D., Ng, S., Szeto, C., Stuart, J., Benz, C.C., Yau, C., Zhang, W., Annala, M., Broom, B.M., Casasent, T.D., Ju, Z., Liang, H., Liu, G., Lu, Y., Unruh, A.K., Wakefield, C., Weinstein, J.N., Zhang, N., Liu, Y., Broaddus, R., Akbani, R., Mills, G.B., Adams, C., Barr, T., Black, A.D., Bowen, J., Deardurff, J., Frick, J., Gastier-Foster, J.M., Grossman, T., Harper, H.A., Hart-Kothari, M., Helsel, C., Hobensack, A., Kuck, H., Kneile, K., Leraas, K.M., Lichtenberg, T.M., McAllister, C., Pyatt, R.E., Ramirez, N.C., Tabler, T.R., Vanhoose, N., White, P., Wise, L., Zmuda, E., Barnabas, N., Berry-Green, C., Blanc, V., Boice, L., Button, M., Farkas, A., Green, A., MacKenzie, J., Nicholson, D., Kalloger, S.E., Blake Gilks, C., Karlan, B.Y., Lester, J., Orsulic, S., Borowsky, M., Cadungog, M., Czerwinski, C., Huelsenbeck-Dill, L., Iacocca, M., Petrelli, N., Rabeno, B., Witkin, G., Nemirovich-Danchenko, E., Potapova, O., Rotin, D., Berchuck, A., Birrer, M., DiSaia, P., Monovich, L., Curley, E., Gardner, J., Mallery, D., Penny, R., Dowdy, S.C., Winterhoff, B., Dao, L., Gostout, B., Meuter, A., Teoman, A., Dao, F., Olvera, N., Bogomolniy, F., Garg, K., Soslow, R.A., Levine, D.A., Abramov, M., Bartlett, J.M.S., Kodeeswaran, S., Parfitt, J., Moiseenko, F., Clarke, B.A., Goodman, M.T., Carney, M.E., Matsuno, R.K., Fisher, J., Huang, M., Kimryn Rathmell, W., Thorne, L., Van Le, L., Dhir, R., Edwards, R., Elishaev, E., Zorn, K., Broaddus, R., Goodfellow, P.J., Mutch, D., Schultz, N., Liu, Y., Akbani, R., Cherniack, A.D., Cerami, E., Weinhold, N., Shen, H., Hoadley, K.A., Kahn, A.B., Bell, D.W., Pollock, P.M., Wang, C., A.Wheeler, D., Shinbrot, E., Karlan, B.Y., Berchuck, A., Dowdy, S.C., Winterhoff, B., Goodman, M.T., Gordon Robertson, A., Beroukhim, R., Pashtan, I., Salvesen, H.B., Laird, P.W., Noble, M., Stuart, J., Ding, L., Kandoth, C., Blake Gilks, C., Soslow, R.A., Goodfellow, P.J., Mutch, D., Broaddus, R., Zhang, W., Mills, G.B., Kucherlapati, R., Mardis, E.R., Levine, D.A., Ayala, B., Chu, A.L., Jensen, M.A., Kothiyal, P., Pihl, T.D., Pontius, J., Pot, D.A., Snyder, E.E., Srinivasan, D., Kahn, A.B., Mills Shaw, K.R., Sheth, M., Davidsen, T., Eley Martin L. Ferguson, G., Demchok, J.A., Yang, Liming, Guyer, M.S., Ozenberger, B.A., Sofia, H.J., Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Gordon Robertson, A., Pashtan, I., Shen, R., Benz, C.C., Yau, C., Laird, P.W., Ding, L., Zhang, W., Mills, G.B., Kucherlapati, R., Mardis, E.R., Levine, D.A., 2013. Integrated genomic characterization of endometrial carcinoma. Nature 497, 67.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W.C., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.-H., Frazier, M.E., Scherer, S.W., Strausberg, R.L., Venter, J.C., 2007. The Diploid Genome Sequence of an Individual Human. PLoS Biol 5. https://doi.org/10.1371/journal.pbio.0050254

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P., 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 1, 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Liu, P., Cheng, H., Roberts, T.M., Zhao, J.J., 2009. Targeting the phosphoinositide 3-kinase (PI3K) pathway in cancer. Nat Rev Drug Discov 8, 627–644. https://doi.org/10.1038/nrd2926

Locke, D.P., Sharp, A.J., McCarroll, S.A., McGrath, S.D., Newman, T.L., Cheng, Z., Schwartz, S., Albertson, D.G., Pinkel, D., Altshuler, D.M., Eichler, E.E., 2006. Linkage Disequilibrium and Heritability of Copy-Number Polymorphisms within Duplicated Regions of the Human Genome. Am J Hum Genet 79, 275–290.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lupski, J.R., Stankiewicz, P., 2005. Genomic Disorders: Molecular Mechanisms for Rearrangements and Conveyed Phenotypes. PLOS Genetics 1, e49. https://doi.org/10.1371/journal.pgen.0010049

Macintyre, G., Goranova, T.E., De Silva, D., Ennis, D., Piskorz, A.M., Eldridge, M., Sie, D., Lewsley, L.-A., Hanif, A., Wilson, C., Dowson, S., Glasspool, R.M., Lockley, M., Brockbank, E., Montes, A., Walther, A., Sundar, S., Edmondson, R., Hall, G.D., Clamp, A., Gourley, C., Hall, M., Fotopoulou, C., Gabra, H., Paul, J., Supernat, A., Millan, D., Hoyle, A., Bryson, G., Nourse, C., Mincarelli, L., Sanchez, L.N., Ylstra, B., Jimenez-Linan, M., Moore, L., Hofmann, O., Markowetz, F., McNeish, I.A., Brenton, J.D., 2018. Copy number signatures and mutational processes in ovarian carcinoma. Nature Genetics 50, 1262–1270. https://doi.org/10.1038/s41588-018-0179-8

Magrangeas, F., Avet-Loiseau, H., Munshi, N.C., Minvielle, S., 2011. Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. Blood 118, 675–678. https://doi.org/10.1182/blood-2011-03-344069

Malhotra, A., Lindberg, M., Faust, G.G., Leibowitz, M.L., Clark, R.A., Layer, R.M., Quinlan, A.R., Hall, I.M., 2013. Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. Genome Res 23, 762–776. https://doi.org/10.1101/gr.143677.112

McAlpine, J.N., Porter, H., Köbel, M., Nelson, B.H., Prentice, L.M., Kalloger, S.E., Senz, J., Milne, K., Ding, J., Shah, S.P., Huntsman, D.G., Gilks, C.B., 2012. BRCA1 and BRCA2 mutations correlate with TP53 abnormalities and presence of immune cell infiltrates in ovarian high-grade serous carcinoma. Modern Pathology 25, 740–750. https://doi.org/10.1038/modpathol.2011.211

Menghi, F., Inaki, K., Woo, X., Kumar, P.A., Grzeda, K.R., Malhotra, A., Yadav, V., Kim, H., Marquez, E.J., Ucar, D., Shreckengast, P.T., Wagner, J.P., MacIntyre, G., Murthy Karuturi, K.R., Scully, R., Keck, J., Chuang, J.H., Liu, E.T., 2016. The tandem duplicator phenotype as a distinct genomic configuration in cancer. Proc Natl Acad Sci USA 113, E2373–E2382. https://doi.org/10.1073/pnas.1520010113

Mora-Cantallops, M., Sánchez-Alonso, S., García-Barriocanal, E., 2020. A complex network analysis of the Comprehensive R Archive Network (CRAN) package ecosystem. Journal of Systems and Software 170, 110744. https://doi.org/10.1016/j.jss.2020.110744

Moran, B., 2021. RNAseqon - RNAseq DE Analysis Downstream from Nextflow Pipelines.

Mukhopadhyay, A., Plummer, E.R., Elattar, A., Soohoo, S., Uzir, B., Quinn, J.E., McCluggage, W.G., Maxwell, P., Aneke, H., Curtin, N.J., Edmondson, R.J., 2012. Clinicopathological features of homologous recombination-deficient epithelial ovarian cancers: sensitivity to PARP inhibitors, platinum, and survival. Cancer Res 72, 5675–5682. https://doi.org/10.1158/0008-5472.CAN-12-0324

Murnane, J.P., 2012. Telomere dysfunction and chromosome instability. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, Telomeres and disease 730, 28–36. https://doi.org/10.1016/j.mrfmmm.2011.04.008

Nam, A.S., Chaligne, R., Landau, D.A., 2020. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. Nature Reviews Genetics 1–16. https://doi.org/10.1038/s41576-020-0265-5

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., Muthuswamy, L., Krasnitz, A., McCombie, W.R., Hicks, J., Wigler, M., 2011. Tumour evolution inferred by single-cell sequencing. Nature 472, 90–94. https://doi.org/10.1038/nature09807

Nidai Ozes, O., Mayo, L.D., Gustin, J.A., Pfeffer, S.R., Pfeffer, L.M., Donner, D.B., 1999. NF-κB activation by tumour necrosis factor requires the Akt serine–threonine kinase. Nature 401, 82–85. https://doi.org/10.1038/43466

Patch, A.-M., Christie, E.L., Etemadmoghadam, D., Garsed, D.W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P.J., Kassahn, K.S., Newell, F., Quinn, M.C.J., Kazakoff, S., Quek, K., Wilhelm-Benartzi, C., Curry, E., Leong, H.S., Australian Ovarian Cancer Study Group, Hamilton, A., Mileshkin, L., Au-Yeung, G., Kennedy, C., Hung, J., Chiew, Y.-E., Harnett, P., Friedlander, M., Quinn, M., Pyman, J., Cordner, S., O'Brien, P., Leditschke, J., Young, G., Strachan, K., Waring, P., Azar, W., Mitchell, C., Traficante, N., Hendley, J., Thorne, H., Shackleton, M., Miller, D.K., Arnau, G.M., Tothill, R.W., Holloway, T.P., Semple, T., Harliwong, I., Nourse, C., Nourbakhsh, E., Manning, S., Idrisoglu, S., Bruxner, T.J.C., Christ, A.N., Poudel, B., Holmes, O., Anderson, M., Leonard, C., Lonie, A., Hall, N., Wood, S., Taylor, D.F., Xu, Q., Fink, J.L., Waddell, Nick, Drapkin, R., Stronach, E., Gabra, H., Brown, R., Jewell, A., Nagaraj, S.H., Markham, E., Wilson, P.J., Ellul, J., McNally, O., Doyle, M.A., Vedururu, R., Stewart, C., Lengyel, E., Pearson, J.V., Waddell, Nicola, deFazio, A., Grimmond, S.M., Bowtell, D.D.L., 2015. Whole-genome characterization of chemoresistant ovarian cancer. Nature 521, 489–494. https://doi.org/10.1038/nature14410

Pilarova, K., Herudek, J., Blazek, D., 2020. CDK12: cellular functions and therapeutic potential of versatile player in cancer. NAR Cancer 2. https://doi.org/10.1093/narcan/zcaa003

Pimentel, H., L, B.N., Puente, S., Melsted, P., Pachter, L., 2016. Differential analysis of RNA-Seq incorporating quantification uncertainty. bioRxiv 058164. https://doi.org/10.1101/058164

Rausch, T., Jones, D.T.W., Zapatka, M., Stütz, A.M., Zichner, T., Weischenfeldt, J., Jäger, N., Remke, M., Shih, D., Northcott, P.A., Pfaff, E., Tica, J., Wang, Q., Massimi, L., Witt, H., Bender, S., Pleier, S., Cin, H., Hawkins, C., Beck, C., von Deimling, A., Hans, V., Brors, B., Eils, R., Scheurlen, W., Blake, J., Benes, V., Kulozik, A.E., Witt, O., Martin, D., Zhang, C., Porat, R., Merino, D.M., Wasserman, J., Jabado, N., Fontebasso, A., Bullinger, L., Rücker, F.G., Döhner, K., Döhner, H., Koster, J., Molenaar, J.J., Versteeg, R., Kool, M., Tabori, U., Malkin, D., Korshunov, A., Taylor, M.D., Lichter, P., Pfister, S.M., Korbel, J.O., 2012. Genome Sequencing of Pediatric Medulloblastoma Links Catastrophic DNA Rearrangements with TP53 Mutations. Cell 148, 59–71. https://doi.org/10.1016/j.cell.2011.12.013

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, Junjun, Zerjal, T., Zhang, Jane, Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. Nature 444, 444–454. https://doi.org/10.1038/nature05329

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47. https://doi.org/10.1093/nar/gkv007

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Romashkova, J.A., Makarov, S.S., 1999. NF-κB is a target of AKT in anti-apoptotic PDGF signalling. Nature 401, 86–90. https://doi.org/10.1038/43474

Ross, S.H., Cantrell, D.A., 2018. Signaling and Function of Interleukin-2 in T Lymphocytes. Annu Rev Immunol 36, 411–433. https://doi.org/10.1146/annurev-immunol-042617-053352

Sachdeva, M., Zhu, S., Wu, F., Wu, H., Walia, V., Kumar, S., Elble, R., Watabe, K., Mo, Y.-Y., 2009. p53 represses c-Myc through induction of the tumor suppressor miR-145. PNAS 106, 3207–3212. https://doi.org/10.1073/pnas.0808042106

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.-Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., Altshuler, D., The International SNP Map Working Group, Cold Spring Harbor Laboratories:, National Center for Biotechnology Information:, The Sanger Centre:, Washington University in St. Louis:, Whitehead/MIT Center for Genome Research:, 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409, 928–933. https://doi.org/10.1038/35057149

Sansone, P., Bromberg, J., 2012. Targeting the Interleukin-6/Jak/Stat Pathway in Human Malignancies. J Clin Oncol 30, 1005–1014. https://doi.org/10.1200/JCO.2010.31.8907

Scheinin, I., Sie, D., Bengtsson, H., Van de Wiel, M., Olshen, A., van Thuijl, H., Essen, H.F., Eijk, P., Rustenburg, F., Meijer, G., Reijneveld, J., Wesseling, P., Pinkel, D., Ylstra, B., 2014. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome research 24. https://doi.org/10.1101/gr.175141.114

Schroder, K., Hertzog, P.J., Ravasi, T., Hume, D.A., 2004. Interferon-γ: an overview of signals, mechanisms and functions. Journal of Leukocyte Biology 75, 163–189. https://doi.org/10.1189/jlb.0603252

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., Wigler, M., 2004. Large-Scale Copy Number Polymorphism in the Human Genome. Science 305, 525–528. https://doi.org/10.1126/science.1098918

Seyednasrollah, F., Laiho, A., Elo, L.L., 2015. Comparison of software packages for detecting differential expression in RNA-seq studies. Briefings in Bioinformatics 16, 59–70. https://doi.org/10.1093/bib/bbt086

Shao, X., Lv, N., Liao, J., Long, J., Xue, R., Ai, N., Xu, D., Fan, X., 2019. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. BMC Medical Genetics 20, 175. https://doi.org/10.1186/s12881-019-0909-5

Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R., Oseroff, V.V., Albertson, D.G., Pinkel, D., Eichler, E.E., 2005. Segmental Duplications and Copy-Number Variation in the Human Genome. The American Journal of Human Genetics 77, 78–88. https://doi.org/10.1086/431652

Shen, R., Seshan, V.E., 2016. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res 44, e131–e131. https://doi.org/10.1093/nar/gkw520

Shoshani, O., Brunner, S.F., Yaeger, R., Ly, P., Nechemia-Arbely, Y., Kim, D.H., Fang, R., Castillon, G.A., Yu, M., Li, J.S.Z., Sun, Y., Ellisman, M.H., Ren, B., Campbell, P.J., Cleveland, D.W., 2020. Chromothripsis drives the evolution of gene amplification in cancer. Nature 1–5. https://doi.org/10.1038/s41586-020-03064-z

Smith, J.C., Sheltzer, J.M., 2018. Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. Elife 7. https://doi.org/10.7554/eLife.39217

Soslow, R.A., Han, G., Park, K.J., Garg, K., Olvera, N., Spriggs, D.R., Kauff, N.D., Levine, D.A., 2012. Morphologic patterns associated with BRCA1 and BRCA2 genotype in ovarian carcinoma. Modern Pathology 25, 625–636. https://doi.org/10.1038/modpathol.2011.183

Spolski, R., Li, P., Leonard, W.J., 2018. Biology and regulation of IL-2: from molecular mechanisms to human therapy. Nature Reviews Immunology 18, 648–659. https://doi.org/10.1038/s41577-018-0046-y

Strickland, K., Howitt, B., Rodig, S., Ritterhouse, L., D'Andrea, A., Matulonis, U., Konstantinopoulos, P., 2015. Tumor infiltrating and peritumoral T cells and expression of PD-L1 in BRCA1/2-mutated high grade serous ovarian cancers. Journal of Clinical Oncology 33, 5512–5512. https://doi.org/10.1200/jco.2015.33.15_suppl.5512

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102

The Cancer Genome Atlas Research Network, 2011. Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–615. https://doi.org/10.1038/nature10166

Therneau, T.M., until 2009), T.L. (original S.->R port and R. maintainer, Elizabeth, A., Cynthia, C., 2020. survival: Survival Analysis.

Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., Olson, M.V., Eichler, E.E., 2005. Fine-scale structural variation of the human genome. Nature Genetics 37, 727–732. https://doi.org/10.1038/ng1562

Valsesia, A., Macé, A., Jacquemont, S., Beckmann, J.S., Kutalik, Z., 2013. The Growing Importance of CNVs: New Insights for Detection and Clinical Interpretation. Front Genet 4. https://doi.org/10.3389/fgene.2013.00092

Van Hoeck, A., Tjoonk, N.H., van Boxtel, R., Cuppen, E., 2019. Portrait of a cancer: mutational signature analyses for cancer diagnostics. BMC Cancer 19, 457. https://doi.org/10.1186/s12885-019-5677-2

Vaughan, S., Coward, J.I., Bast, R.C., Berchuck, A., Berek, J.S., Brenton, J.D., Coukos, G., Crum, C.C., Drapkin, R., Etemadmoghadam, D., Friedlander, M., Gabra, H., Kaye, S.B., Lord, C.J., Lengyel, E., Levine, D.A., McNeish, I.A., Menon, U., Mills, G.B., Nephew, K.P., Oza, A.M., Sood, A.K., Stronach, E.A., Walczak, H., Bowtell, D.D., Balkwill, F.R., 2011. Rethinking ovarian cancer: recommendations for improving outcomes. Nature Reviews Cancer 11, 719–725. https://doi.org/10.1038/nrc3144

Wegiel, B., Bjartell, A., Culig, Z., Persson, J.L., 2008. Interleukin-6 activates PI3K/Akt pathway and regulates cyclin A1 to promote prostate cancer cell survival. International Journal of Cancer 122, 1521–1529. https://doi.org/10.1002/ijc.23261

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., RStudio, 2020. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.

Wierstra, I., Alves, J., 2008. The c-myc Promoter: Still MysterY and Challenge, in: Advances in Cancer Research. Academic Press, pp. 113–333. https://doi.org/10.1016/S0065-230X(07)99004-1

Xu, B., Lefringhouse, J., Liu, Z., West, D., Baldwin, L.A., Ou, C., Chen, L., Napier, D., Chaiswing, L., Brewer, L.D., Clair, D.S., Thibault, O., van Nagell, J.R., Zhou, B.P., Drapkin, R., Huang, J.-A., Lu, M.L., Ueland, F.R., Yang, X.H., 2017. Inhibition of the integrin/FAK signaling axis and c-Myc synergistically disrupts ovarian cancer malignancy. Oncogenesis 6, e295–e295. https://doi.org/10.1038/oncsis.2016.86

Yang, L., Wang, Y.-Z., Zhu, H.-H., Chang, Y., Li, L.-D., Chen, W.-M., Long, L.-Y., Zhang, Y.-H., Liu, Y.-R., Lu, J., Qin, Y.-Z., 2017. PRAME Gene Copy Number Variation Is Related to Its Expression in Multiple Myeloma. DNA Cell Biol 36, 1099–1107. https://doi.org/10.1089/dna.2017.3951

Yang, Y., Bazhin, A.V., Werner, J., Karakhanova, S., 2013. Reactive Oxygen Species in the Immune System. International Reviews of Immunology 32, 249–270. https://doi.org/10.3109/08830185.2012.755176

Yates, A.D., Achuthan, P., Akanni, W., Allen, James, Allen, Jamie, Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J.C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., Giron, C.G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., Kay, M., Lavidas, I., Le, T., Lemos, D., Martinez, J.G., Maurel, T., McDowall, M., McMahon, A., Mohanan, S., Moore, B., Nuhn, M., Oheh, D.N., Parker, A., Parton, A., Patricio, M., Sakthivel, M.P., Abdul Salam, A.I., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Sycheva, M., Szuba, M., Taylor, K., Thormann, A., Threadgold, G., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Flint, B., Frankish, A., Hunt, S.E., IIsley, G., Kostadima, M., Langridge, N., Loveland, J.E., Martin, F.J., Morales, J., Mudge, J.M., Muffato, M., Perry, E., Ruffier, M., Trevanion, S.J., Cunningham, F., Howe, K.L., Zerbino, D.R., Flicek, P., 2020. Ensembl 2020. Nucleic Acids Research 48, D682–D688. https://doi.org/10.1093/nar/gkz966

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhang, C.-Z., Wala, J., Mermel, C.H., Sougnez, C., Gabriel, S.B., Hernandez, B., Shen, H., Laird, P.W., Getz, G., Meyerson, M., Beroukhim, R., 2013. Pan-cancer patterns of somatic copy-number alteration. Nat Genet 45, 1134–1140. https://doi.org/10.1038/ng.2760

Zarrei, M., MacDonald, J.R., Merico, D., Scherer, S.W., 2015. A copy number variation map of the human genome. Nature Reviews Genetics 16, 172–183. https://doi.org/10.1038/nrg3871

Zhang, F., Gu, W., Hurles, M.E., Lupski, J.R., 2009. Copy Number Variation in Human Health, Disease, and Evolution. Annu Rev Genomics Hum Genet 10, 451–481. https://doi.org/10.1146/annurev.genom.9.081307.164217

Zhang, L., Conejo-Garcia, J.R., Katsaros, D., Gimotty, P.A., Massobrio, M., Regnani, G., Makrigiannakis, A., Gray, H., Schlienger, K., Liebman, M.N., Rubin, S.C., Coukos, G., 2003. Intratumoral T Cells, Recurrence, and Survival in Epithelial Ovarian Cancer. New England Journal of Medicine 348, 203–213. https://doi.org/10.1056/NEJMoa020177

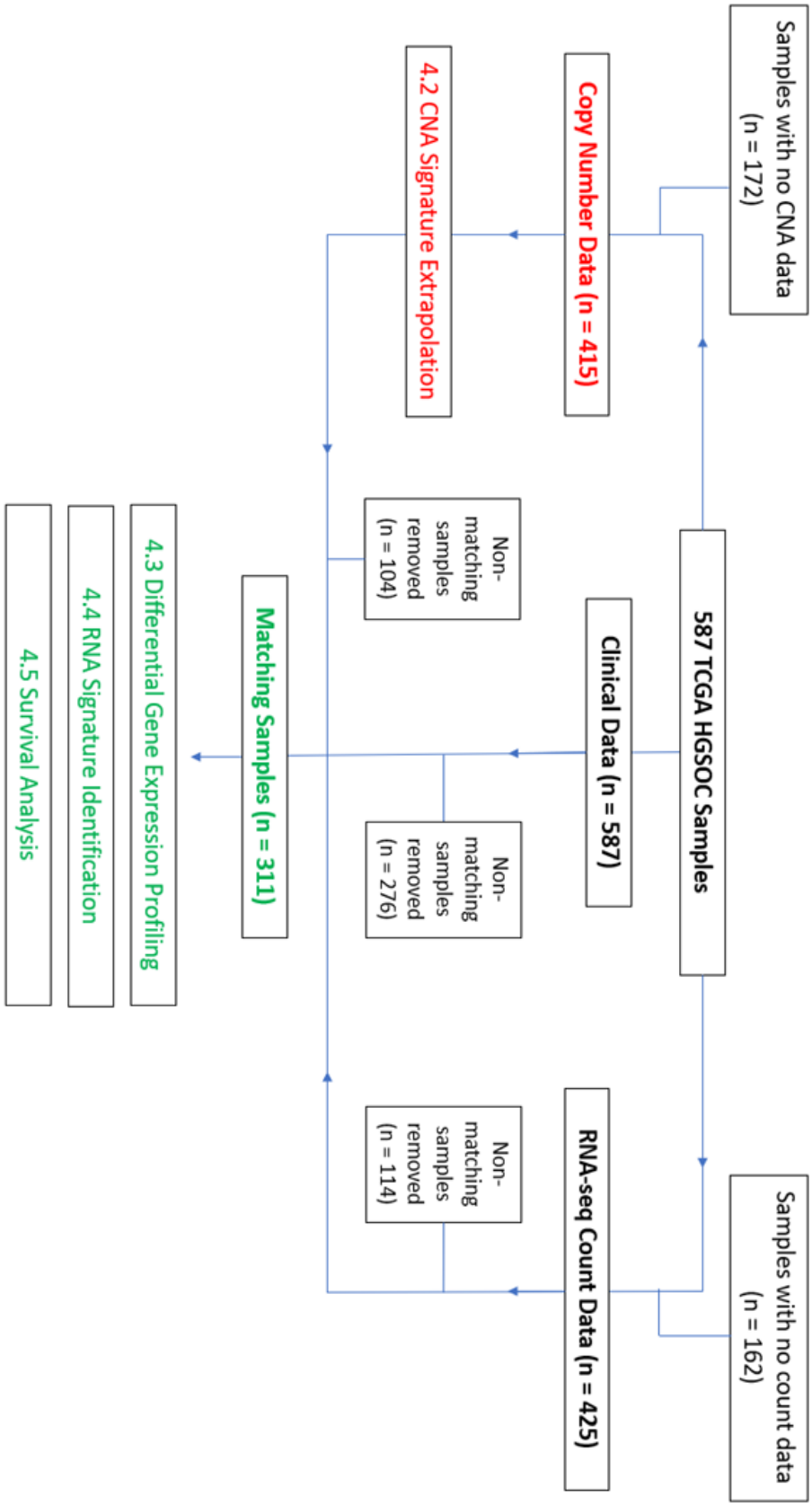
Zhao, N., Wilkerson, M.D., Shah, U., Yin, X., Wang, A., Hayward, M.C., Roberts, P., Lee, C.B., Parsons, A.M., Thorne, L.B., Haithcock, B.E., Grilley-Olson, J.E., Stinchcombe, T.E., Funkhouser, W.K., Wong, K.-K., Sharpless, N.E., Hayes, D.N., 2014. Alterations of LKB1 and KRAS and risk of brain metastasis: comprehensive characterization by mutation analysis, copy number, and gene expression in non-small-cell lung carcinoma. Lung Cancer 86, 255–261. https://doi.org/10.1016/j.lungcan.2014.08.013

Zhou, C., Zhang, W., Chen, W., Yin, Y., Atyah, M., Liu, S., Guo, L., Shi, Y., Ye, Q., Dong, Q., Ren, N., 2017. Integrated Analysis of Copy Number Variations and Gene Expression Profiling in Hepatocellular carcinoma. Sci Rep 7, 10570. https://doi.org/10.1038/s41598-017-11029-y

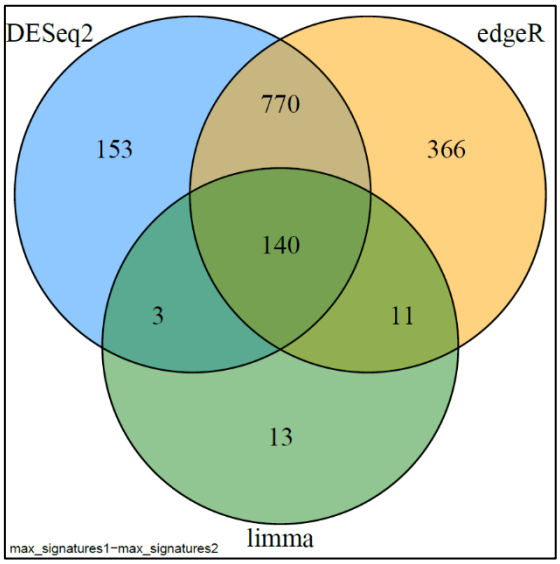
Zou, X., Owusu, M., Harris, R., Jackson, S.P., Loizou, J.I., Nik-Zainal, S., 2018. Validating the concept of mutational signatures with isogenic cell models. Nature Communications 9, 1–16. https://doi.org/10.1038/s41467-018-04052-8

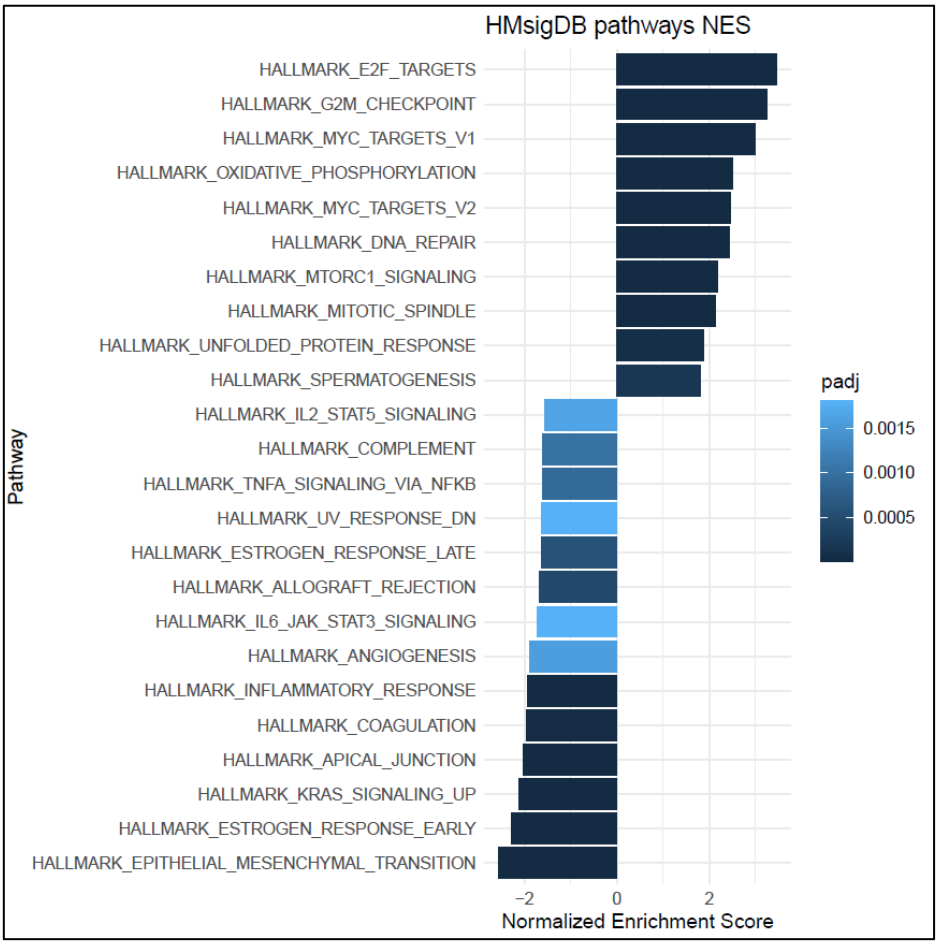
9. Appendices

Appendix 1 – Breakdown of sample filtering for each component of the methods



Appendix 2 – Example of DEG overlap between DESeq2, edgeR and limma, and an example GSEA output (maximum CNA signature comparison s1 vs s2).





Appendix 3 – Abbreviations used within the thesis document

BFB	Breakage-fusion-bridge
CNA	Copy number alteration
CRAN	Comprehensive R Archive Network
DE	Differential expression
DEG	Differentially expressed gene
ECDF	Empirical Cumulative Distribution Functions
FDR	False discovery rate
DSB	Double strand break
JCN	Junction copy number
GSEA	Gene set enrichment analysis
HGSOC	High grade serous ovarian cancer
HR	Hazard ratio
IFNγ	Interferon gamma
LCD	Linear combination decomposition
LGSOC	Low grade serous ovarian cancer
MSigDB	Molecular Signatures Database
NES	Normalized enrichment score
NGS	Next generation sequencing
NMF	Non-negative matrix factorization
PCA	Principal component analysis
r1, r2, etc	RNA signature 1, signature 2
RNA-seq	RNA-sequencing
s1, s2, etc	CNA signature 1, signature 2
ssGSEA	Single sample gene set enrichment analysis
sWGS	Shallow whole genome sequencing
SNV	Single nucleotide variant
SNP	Single nucleotide polymorphism
PCA	Principal Component Analysis
PCAWG	Pan Cancer Analysis of Whole Genomes
TCGA	The Cancer Genome Atlas
WES	Whole exome sequencing
WGS	Whole genome sequencing
TPM	Transcripts per million