Genes Identified via Bayesian Network Analysis of Temporal Dynamic Expression are Prognostic of Survival in Ovarian Cancer

Hannah Currant[1] **Athanasios Vogogias**[2] **V. Anne Smith**[3]

[1]EMBL-EBI , Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, United Kingdom
[2]The Roslin Institute , University of Edinburgh , Easter Bush Campus , Midlothian, EH25 9RG, United Kingdom
[3]School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, United Kingdom

Abstract

The treatment of ovarian cancer still faces several challenges: it is often diagnosed at a late stage, and many patients relapse due to acquisition of resistance to chemotherapeutic agents. We aimed to identify gene sets which can predict prognosis in patients. We used experimentally-obtained temporal gene expression from a mouse xenograft study to identify groups, or clusters, of genes sharing similar differential expression profiles. Dependency of gene expression values on treatment was assessed by using Bayesian network structure learning to identify connections among clusters and with a treatment variable. We hypothesised that gene clusters directly dependent on treatment could be prognostic of patient survival. Prognostic ability of all such network-identified clusters were assessed in an independent clinical dataset. Expression values of genes in some of these clusters allowed the clinical dataset to be separated into groups with significantly different progression-free (time to relapse) survival and thus are prognostic. The prognostic genes identified provide avenues of further research into the development of a clinically viable prognostic tool and into potential therapeutic targets.

1 INTRODUCTION

Ovarian cancer is a major cause of cancer fatality in women, In 2020, it was estimated that 13,940 women in the United States of America died of ovarian cancer [Siegel et al., 2020]. Treatment options include a number of relatively effective chemotherapeutic drugs in combination with surgery. However, lack of early-stage symptoms leads to late diagnosis, thus poor long-term survival.

A major issue in the field is the development of resistance to one of the main treatment regimens. Both Carboplatin, and Carboplatin in combination with Paclitaxel are utilised to treat ovarian cancer. Carboplatin is a platinum based alkylating agent which functions by creating guanine cross-links in the DNA, disrupting replication and preventing cell growth. Resistance to this drug often develops with increased DNA repair and drug inactivation [Apps et al., 2015]. Paclitaxel is a non-platinum based agent that acts by inhibiting microtubules during mitosis and meiosis and thus causing cell cycle arrest and apoptosis [Jordan and Wilson, 2004]. Even when used in combination, the drugs are imperfect. Therefore, another current research focus is elucidating mechanisms underlying drug resistance, as well as resistance prediction and identification of novel drug targets.

Gene expression analysis can identify novel biomarkers and aid prognosis. In particular, network analysis has been used to identify prognostic signatures of ovarian cancer, however, only using static clinical gene expression [Coveney et al., 2015, Yan et al., 2020]. Experimental temporal data provides increased dimensionality and an ability to see subtle changes in expression over time that may lead to novel insight.

Bayesian network analysis offers the ability to connect experimental condition with gene expression measurements and further benefits from its ability to handle "noisy data", common in biological data. Bayesian network analysis of gene expression data has previously been implemented in the study of breast and thyroid cancer amongst others [Gevaert et al., 2006, Polanski et al., 2007].

In keeping with these aims, here we analyse a temporal gene expression data series looking at cancer tumours' response to the two major medication regimens in clinical use, with gene expression sampled at multiple time points after drug treatment. The experiment includes both a platinum-sensitive ovarian cancer cell line, OV1002, and a platinum-resistant ovarian cancer cell line, HOX424. Genes differentially expressed compared to untreated control in each cell line were identified for each treatment regimen. Here

Accepted for 15th Bayesian Modeling Applications Workshop at UAI 2021 (BMAW2021).

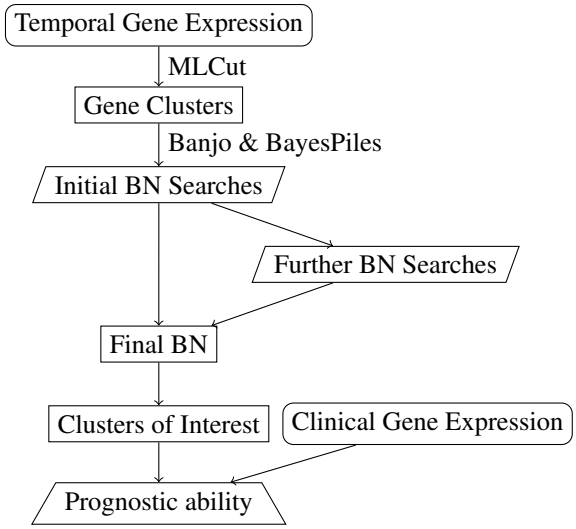

Figure 1: Flowchart of our analysis approach. BN = Bayesian network.

we aim to examine whether we are able to identify prognostic genes by employing Bayesian networks to statistically model dependent relationships between drug treatment and gene expression. To perform this analysis, we utilise two visualisation techniques to cluster the differentially expressed genes and to guide the Bayesian network search. We hypothesise that differentially expressed genes with direct statistical interaction with the treatment condition will be prognostic of survival in ovarian cancer: we thus test our identified genes in an independent clinical cohort. We found a number of gene sets connected to the treatment condition. Of these four had prognostic capabilities in differentiating patients' survival.

2 METHODS

2.1 OVERVIEW

Figure 1 provides an overview of our analysis, described in detail in the following sections. We cluster temporal gene expression profiles using the visualisation tool MLCut [Vogogias et al., 2016]. These clusters are used as data from which to learn a Bayesian network structure using Banjo [Smith et al., 2006, Yu et al., 2004], with initial searches first analysed via the visualisation tool BayesPiles [Vogogias et al., 2018] to guide final search and network selection. From final networks, we identify those clusters directly connected to the treatment node as of interest. Each of these clusters were then analysed for their prognostic ability in an independent clinical dataset.

2.2 EXPERIMENTAL TEMPORAL GENE EXPRESSION

The experimental temporal gene expression from cancer tumours was taken from Koussounadis et al. [2014]. Briefly, this experiment developed mouse xenografts by implanting tumours in mice using either the platinum-sensitive cell line OV1002 or the platinum-resistant cell line HOX424. After a period of tumour growth, the tumours were treated with either Carboplatin, Carboplatin in combination with Paclitaxel, or left untreated as a control. Drug-treated tumour samples were taken at days 1, 2, 4, 7 and 14 after treatment, and control tumour samples were taken before treatment and at days 1, 7 and 14. Gene expression was measured via Illumina bead chips and differential expression determined for each drug treatment/day/cell line combination compared to controls pooled for each cell line [Koussounadis et al., 2014]. We accessed gene expression values via the Gene Expression Omnibus (GEO), accession number GSE49577, and differential gene expression as log fold-change from control via Supplementary Data 1 of the publication.

2.3 CLUSTERING USING MLCUT

For each of the tested cell lines, groups of similar gene profiles were identified using hierarchical clustering of the fold-change differential expression. MLCut [Vogogias et al., 2016] was used to visually guide the selection of clusters. Initially, `TSclust` [Montero and Vilar, 2014] was used to calculate the distance matrix for all pairs of genes, based on the Euclidean distance between the fold-change across multiple time-points. A dendrogram was constructed using average-linkage clustering [R Core Team, 2020]. Similarity levels were scaled to range between 0 and 1, and the output was transformed into a JSON format using the `rjson` package in R [Couture-Beil, 2018]. For each cell line, both the JSON-formatted dendrogram and the original temporal data (in CSV), were visualised in MLCut (Figure 2).

For each cell type and treatment regimen, the same clustering distinctiveness parameter of 0.5 was used throughout and a similarity parameter of 0.5 was used as a beginning point. From there, similarity parameter and cluster selection decisions were made based on visual feedback provided by MLCut, maximising homogeneity of clusters whilst retaining a viable number of clusters to use in Bayesian network analysis.

2.4 BAYESIAN NETWORK STRUCTURE LEARNING

Bayesian networks were recovered using Banjo [Smith et al., 2006, Yu et al., 2004], which uses a score-based search algorithm. Search decisions were guided by BayesPiles [Vogogias et al., 2018], an interactive tool that allows visuali-

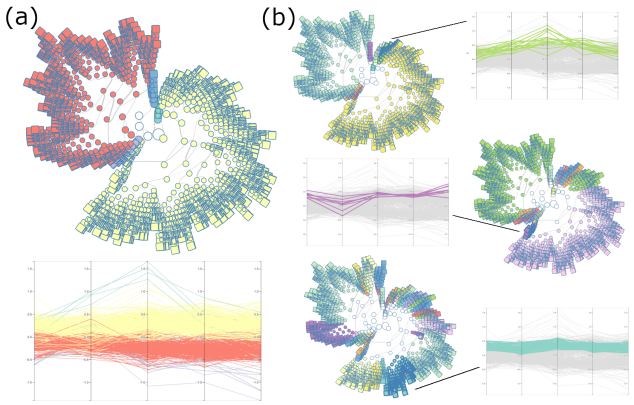

Figure 2: MLCut clustering: example with OV1002 treated with Carboplatin. Initial cut (a) was modified via visual inspection, iteratively selecting clusters. MLCut shows a radial dendrogram (top) and a line plot (bottom) plotting differential gene expression as log fold-change from control. Different clusters are shown in different colours; colours in the dendrogram and line graph are congruent such that a gene's differential expression line is shown as the same colour as its node in the dendrogram. (b) Intermediate cuts showing Cluster 11 (top), Cluster 14 (middle) and, on the final cut, Cluster 18 (bottom). In each intermediate cut, one cluster is highlighted in the dendrogram (thicker outlines) and only those genes' differential expression coloured in the line plot (shown via connecting line).

sation of scores and structures of search results. The original (not differential) gene expression values were used for Bayesian network analysis, including all cell line, treatment, and control samples. Mean gene expression was calculated across all genes in each cluster. Networks were built from these mean values plus a binary variable indicating whether the sample was treatment or control, with one network solution per cell line/treatment combination.

Basic settings shared between all searches were BDe score, quantile discretisation into 3 states, maximum parent count 5, and equivalent sample size 1. Initial searches consisted of two sets of 5 replicate searches visiting 25 million networks, one set using a greedy search with all local moves (makes change that increases the score most), and one using simulated annealing with the default settings. If needed, we ran an additional set of 5 replicate searches using greedy search with random local moves (selects first change encountered that increases the score). Final search configurations were decided based on inspection via BayesPiles as reported in Results. Results could be a single top network, or the consensus network model averaging feature of Banjo. The consensus network includes links with probability ≥ 0.5 based on a set of top-scoring networks weighted by network score.

2.5 EVALUATION OF PROGNOSTIC ABILITY

Those clusters directly connected to the treatment node for each cell line/treatment combination were evaluated for ability to predict survival in an independent clinical sample of ovarian cancer patients. In vivo gene expression and survival data were taken from Tothill et al. [2008]. Briefly, gene expression of tumours from 285 patients was evaluated via Affymetrix gene chips. Survival was recorded as months to relapse (progression-free survival) and death (overall survival), and right-censored with status at final follow-up (e.g., progression-free, relapsed, or deceased) [Tothill et al., 2008]. We accessed gene expression via the Gene Expression Omnibus (GEO), accession number GSE9891, and patient survival data via Supplementary Table 1 of the publication.

For each cluster directly connected to the treatment node, hierarchical clustering was applied to scaled patient expression values (mean $= 0$, standard deviation $= 1$) using Euclidean distance and complete linkage [R Core Team, 2020]. We aimed to split the hierarchy into two groups of patients with most different gene expression by using the highest level cut. However, for four clusters these cuts were extremely uneven (< 6 patients in one group), thus the small groups were discarded and clusters cut consecutively lower until at least 10% of the patients were in the smallest group, using the second cut for three clusters (discarding 1, 2 and 6 patients) and the fourth for another (discarding 3 patients). We compared survival of the two groups of patients defined by each cluster with Kaplan-Meier survival curves using the `survival` package in R [Therneau and Grambsch, 2000]. Statistical probability of the two cluster-defined groups having different progression-free and overall survival was assessed for each cluster, using Bonferonni-corrected significance criteria adjusted for the number of clusters.

3 RESULTS

3.1 CLUSTERING

3.1.1 OV1002 Carboplatin divides into 19 clusters

The similarity value of 0.5 yielded five clusters. The smallest three clusters were selected following visual assessment of the temporal expression. The smaller of the two remaining clusters was further split with a similarity value of 0.606, at which point the smaller of the two clusters was selected. The remaining cluster was split with similarity parameters 0.622 and 0.661, again at which the smaller of the clusters was selected. The remaining cluster was split at similarity = 0.702, at which point both clusters were selected as further splitting did not yield more visually homogeneous clusters. Returning to the largest cluster generated using similarity = 0.5, a similar method was used where by the cluster was

repeatedly split, the smaller of the two clusters selected, and the remaining further split. This was done at similarity values of 0.508, 0.569, 0.591, 0.641, 0.653, 0.677, 0.697, 0.704, 0.74 and a final split at 0.771 where both clusters were selected. In total this yielded 19 clusters.

3.1.2 OV1002 Carboplatin and Paclitaxel divides into 20 clusters

The starting similarity value of 0.5 identified eight clusters of genes. The six smallest clusters were selected. The largest remaining cluster was split with a similarity parameter of 0.52 and the smallest cluster selected. The remaining cluster was further split and the smaller clusters selected at values of 0.552, 0.577, 0.603 and 0.638 at which the remaining cluster was also selected. The second largest cluster identified using similarity = 0.5 using the same methodology was split at similarity values of 0.501, 0.526, 0.529, 0.535, 0.577, 0.593 and 0.66 at which the remaining cluster was also selected. This yielded a total of 20 clusters.

3.1.3 HOX424 Carboplatin divides into 6 clusters

The initial similarity parameter of 0.5 clustered the expression data into six well defined clusters; all were selected.

3.1.4 HOX424 Carboplatin and Paclitaxel divides into 11 clusters

The initial similarity value of 0.5 yielded 7 clusters. The five smallest clusters were selected. The largest remaining cluster was split with a similarity parameter of 0.537 and both clusters selected. The smaller remaining cluster from using a similarity parameter of 0.5 was further split at 0.517, 0.555 and 0.644 at which point both resulting clusters were selected. In total this yielded 11 clusters.

3.2 BAYESPILES-GUIDED NETWORK SEARCH

3.2.1 Networks of OV1002 clusters identify 4 clusters directly linked to treatment

For networks of clusters found from both OV1002 treated with Carboplatin only and with Carboplatin plus Paclitaxel, Bayespiles inspection guided the same search choices. In both cases, initial greedy searches showed clear multiple hills in the search space. Simulated annealing searches had variation across the five runs, but all showed higher scoring networks than the greedy searches. Further runs of greedy search with only random local moves resulted in all different networks with scores between those of the all local moves greedy search and simulated annealing (Figure 3a). Thus, we determined the search space was complex and most suited to

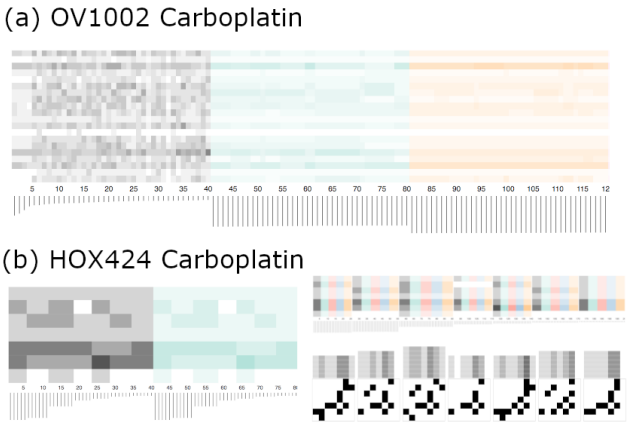

Figure 3: BayesPiles investigations. (a) Greedy searches with all local moves (left, grey) for OV1002 treated with Carboplatin had worse scores than greedy with random local moves (middle, turquoise) and simulated annealing (right, orange). Horizontal axis represents networks sorted by score; shaded area represents summary of out-degree for each node on the vertical axis (darker=higher). Line below horizontal axis represents score (longer=higher). OV1002 Carboplatin and Paclitaxel showed similar results. (b) Two representative of the identical greedy searches for HOX424 treated with Carboplatin (left) and matrix summaries of top structures from all five searches (right). Plots of networks and scores as in (a); matrix summaries show a skeleton view, for consideration of equivalence classes, of all links in the network, with filled boxes indicating a link between nodes on both axes. Shaded bars above matrix show summary of out-degree for each node across horizontal axis.

simulated annealing and model averaging. Additional simulated annealing searches were run to the increased value of 250 million networks searched, and consensus networks calculated from the top 100 networks after pruning for equivalence classes (24-31 networks for OV1002 Carboplatin; 25-43 networks for OV1002 Carboplatin and Paclitaxel). These searches were repeated 10 times. The treatment variable linked in at least 9 of 10 consensus networks to three clusters for OV1002 treated with Carboplatin (Clusters 11, 14 and 18; Figure 4a) and one cluster for OV1002 treated with Carboplatin and Paclitaxel (Cluster 2; Figure 4b).

3.2.2 Networks of HOX424 clusters identify 2 clusters directly linked to treatment

For networks of clusters found from both HOX424 treated with Carbpoplatin only and with Carboplatin plus Paclitaxel, Bayespiles inspection guided the same search choices. In both cases, the intial greedy searches all found the same identical top network (Figure 3b), also found in 2 (Carboplatin plus Paclitaxel) to all 5 (Carboplatin only) of the simulated annealing searches (the other simulated annealing

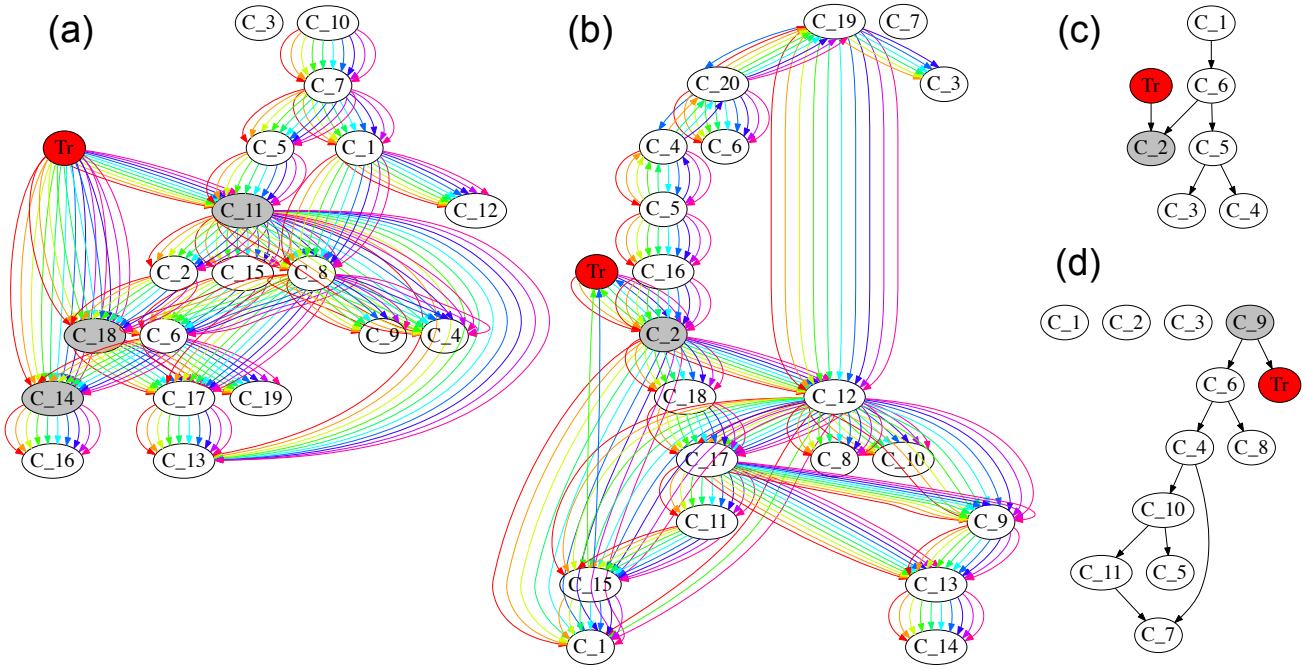

Figure 4: Networks of clusters and Treatment variable for (a) OV1002 treated with Carboplatin, (b) OV1002 treated with Carboplatin and Paciltaxel, (c) HOX424 treated with Carboplatin and (d) HOX424 treated with Carboplatin and Paciltaxel. Tr = Treatment node (red fill); C_# = Cluster #. Clusters linking to Treatment with grey fill. (a) and (b) show consensus networks from 10 searches each in different colours; (c) and (d) show single identical top network from all searches.

searches found lower-scoring networks). This top network was first discovered after less than 1 million networks in the 25 million networks examined. Thus, we determined this single top network was the clear solution. The treatment variable linked to one cluster each for HOX424 treated with Carbpoplatin (Cluster 2; Figure 4c) and HOX424 treated with Carboplatin and Paclitaxel (Cluster 9; Figure 4d).

3.3 PROGNOSTIC ABILITY OF CLUSTERS

3.3.1 Network-identified clusters are prognostic of progression-free survival

Four clusters separated patients by progression-free survival, two significantly against a Bonferonni-adjusted significance criteria of $\alpha < 0.0083$: Cluster 11 identified from OV1002 treated with Carboplatin ($p = 0.00649$; Figure 5a) and Cluster 9 identified from HOX424 treated with Carboplatin and Paclitaxel ($p = 0.00816$ Figure 5b; others: OV1002 Carboplatin Cluster 14 $p = 0.03309$, OV1002 Carboplatin and Paclitaxel Cluster 2 $p = 0.02557$; Figure 5cd).

The remainder of comparisons did not separate patients by progression-free ($p > 0.42$) or overall survival ($p > 0.14$).

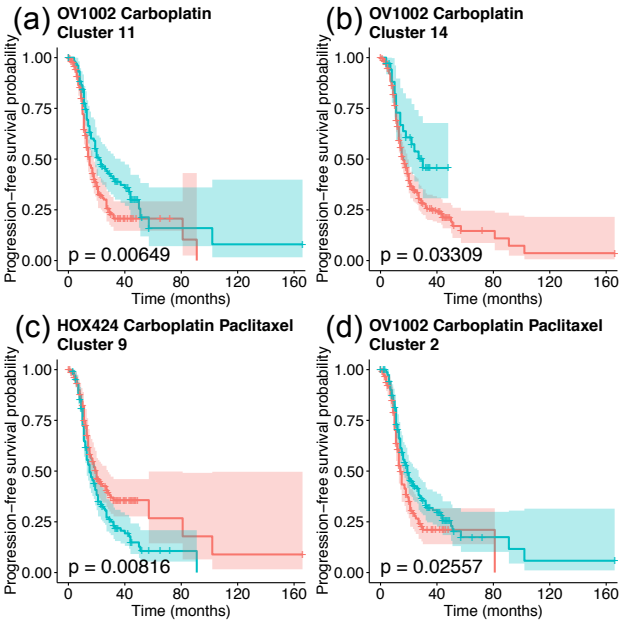

Figure 5: Survival analysis on clinical data. Kaplan-Meier plots for progression-free survival on patients split by network-identified clusters. Curves represent probability of survival at given months for each of the two cluster-defined groups, with shaded regions representing 95% confidence intervals; p-value is statistical comparison between curves.

4 DISCUSSION

This study enquired whether Bayesian network analysis could identify gene clusters in vitro that were able to prognose a clinical dataset in vivo. We used two ovarian cancer cell lines with different platinum sensitivity to look at gene expression following administration of two drug regimens. Bayesian network analysis of clusters of differentially expressed genes successfully identified gene sets directly linked to drug treatment in an experimental study of temporally dynamic gene expression. Of these, four were prognostic – two significantly after correction for multiple testing – of patient survival in an independent dataset.

To come to these results we used visualisation tools to assist in both clustering of differential expression profiles and making decisions for Bayesian network structure learning.

MLCut was used to select clusters of differentially expressed genes in each cell line/treatment. An advantage of this software is that it allows simultaneous visualisation of clustering in the dendrogram space, and in the original data space. This enabled intuitive selection of clusters, and use of different similarity parameters, balancing cluster size with distinctiveness of the cluster expression profiles.

BayesPiles guided the network search. Two datasets (clusters from the OV1002 treatments) had more challenging search spaces than the other two (clusters from the HOX424 treatments). Easy visualisation of test searches in BayesPiles gave confidence to the decisions made. Without such a tool, standard practice would lead to use of one somewhat arbitrarily-determined criteria to guide the heuristic search, e.g., a single top network or model averaging method. However, a single top network would have been inappropriate for the OV1002 clusters as it was clear the search space was variable enough that no single top network had high confidence. Yet model averaging would have been inappropriate for the HOX424 networks as it would have diluted the signal of the consistently-found top network, potenitally losing interesting dependencies. Thus, use of BayesPiles enabled tailoring of heuristic search to each dataset, providing increased confidence in the resulting networks.

The analysis identified six clusters connected to the treatment variable across the different cell lines/treatment conditions. These clusters contained a number of genes with prior associations to cancer, in addition to other biological phenotypes.

OV1002 Carboplatin Cluster 11 contained 17 genes including some with prior associations to cancer including: *C1ORF24* which has shown to be involved in both renal cancer and thyroid cancer [Adachi et al., 2004, Carvalheira et al., 2015]; *PCK2* which is known to be involved in tumour cell survival in lung cancer [Leithner et al., 2015]; *CDC25A*, a known regulator of the cell cycle involved in numerous cancers [Busino et al., 2004]; *PHGDH* which is responsible for an increased risk of oncogenesis in breast cancer [Locasale et al., 2011]; and *LAMP3* which has been identified as a potential prognostic biomarker for cervical cancer, promoting metastasis [Kanao et al., 2005].

OV1002 Carboplatin cluster 14 contained nine genes. Several of these including *MYC*, *SGK1* and *FABP5* are known to be involved in cancer [Dang, 2012, Sang et al., 2021, Carbonetti et al., 2019]. *FABP5* in particular has prior associations to ovarian cancer [Gharpure et al., 2018].

OV1002 Carboplatin cluster 18 contained 59 genes. Functional annotation using DAVID [Huang et al., 2009a,b] showed significant association with nucleoplasm and DNA replication in addition to several DNA replication regulation terms and DNA damage. This matches with the therapeutic action of Carboplatin being to introduce cross-links in the DNA, inducing DNA damage and inhibiting cell-growth [Apps et al., 2015].

OV1002 Carboplatin and Paclitaxel Cluster 2 contained 19 genes, several of which have known involvement in cancer. These include *FAM129A* which has been shown to be involved in both prostate and lung cancer [Pällmann et al., 2019, Zhang et al., 2019], *AURKB* which is involved in the resistance of lung cancer treatment [Bertran-Alamillo et al., 2019], and *PCK2* which was also in the OV1002 Carboplatin Cluster 11 and is associated with tumour cell survival [Leithner et al., 2015]. Functional analysis with DAVID [Huang et al., 2009a,b] found significant associations with cell cycle regulation. This is supported by known involvement of several of the genes in cycle control including *CDC25A* and *E2F2* [Busino et al., 2004, Laresgoiti et al., 2013]. This is in line with Paclitaxel's therapeutic action being disruption of the cell cycle. Additionally several of the genes have known involvement with cell fate such as *NUP210* [D'Angelo et al., 2012] and DNA damage response such as *FOXM1* [Khongkow et al., 2014]. This may reflect the therapeutic action of carboplatin.

HOX424 Carboplatin Cluster 2 contained a single gene, *TRIB3*. This gene has previous associations to cancer [Wang et al., 2013, Lee et al., 2019] and cell survival [Hua et al., 2015]. Interestingly, *TRIB3* has been targeted by a new potential therapeutic agent which could be clinically implemented in combination with Carboplatin and Paclitaxel [López-Plana et al., 2020]. This supports the use of Bayesian network analysis to identify novel therapeutic targets.

HOX424 Carboplatin and Paclitaxel Cluster 9 contained 135 genes. Functional annotation using DAVID showed significant associations with translational initiation. The cluster also contained a number of genes with known involvement in cancer including: *MEN1*, mutations in which cause multiple endocrine neoplasia type 1 which is characterised by tumours in the endocrine glands [Marx et al., 1998]; *CCND1* which has been implicated in endometrial cancer [Moreno-Bueno et al., 2003] and is known to be involved in cell cycle

regulation [Wang et al., 2018]; *PIK3R1* which has shown to have prognostic capabilities in breast cancer [Cizkova et al., 2013]; and *EML4* which has previously been implicated in lung cancer [Soda et al., 2007], amongst others.

It is interesting to note that several of the genes were discovered across the different treatment types. Notably *PCK2* and *CDC25A* were found in the clusters connected to treatment in both the OV1002 treated with Carboplatin, and the OV1002 cell line treated with Carboplatin and Paclitaxel. Neither of the these genes were in the clusters connected with the treatment condition in the HOX424 treatment-resistant cell line. This may suggest that *PCK2* and *CDC25A* are involved in the pathways that lead to drug resistance, and warranting further investigation into their metabolic response to the drug regimens.

Further, it is notable that numerous clusters across the conditions had no direct links to any nodes – either other gene clusters or the treatment node – within the network analysis: OV1002 Carboplatin Cluster 3, OV1002 Carboplatin and Paclitaxel Cluster 7 and HOX424 Carboplatin and Paclitaxel Clusters 1, 2 and 3. These represent a set of genes that currently do not appear to be affected by medication and therefore could indicate previously unused therapy targets.

Of the six clusters identified using the Bayesian networks, four were able to stratify patients by progression-free survival with marginal significance, two with significance following Bonferroni correction. These two were OV1002 Carboplatin Cluster 11 and HOX424 Carboplatin and Paclitaxel Cluster 9. HOX424 Carboplatin and Paclitaxel Cluster 9 was a large cluster, containing 135 genes. It contained a number of known cancer genes, some which have previously been shown to have prognostic capabilities in other cancer types. Comparatively OV1002 Carboplatin Cluster 11 contained only 17 genes, but also contained genes that have previous associations to cancer patient prognosis.

In summary, the work here has demonstrated the utility of Bayesian network analysis when applied to biological data in identification of important ovarian cancer-related genes. Here we used the technique to identify genes with prognostic ability in an independent clinical dataset. This has identified a number of genes, many of which may warrant further study to assess their viability as therapeutic targets. This method could be further applied to other datasets to discover further biomarkers and potential drug targets in other cancers and other diseases.

Author Contributions

VAS and HC conceived the idea. AV and HC performed visually guided clustering; VAS performed BayesPiles analysis, supported by AV; HC and VAS performed remaining analyses. HC provided biological interpretations. All authors wrote the manuscript together.

Acknowledgements

VAS and HC received support from St Andrews University School of Biology.

References

H Adachi, S Majima, S Kon, T Kobayashi, K Kajino, H Mitani, Y Hirayama, H Shiina, M Igawa, and O Hino. Niban gene is commonly expressed in the renal tumors: A new candidate marker for renal carcinogenesis. *Oncogene*, 23: 3495–3500, 2004.

MG Apps, EHY Choi, and NJ Wheate. The state-of-play and future of platinum drugs. *Endocrine-Related Cancer*, 22:R219–R233, 2015.

J Bertran-Alamillo, V Cattan, M Schoumacher, J Codony-Servat, A Giménez-Capitán, F Cantero, M Burbridge, S Rodríguez, C Teixidó, R Roman, J Castellví, S García-Román, C Codony-Servat, S Viteri, AF Cardona, N Karachaliou, R Rosell, and MA Molina-Vila. AURKB as a target in non-small cell lung cancer with acquired resistance to anti-EGFR therapy. *Nature Communications*, 10:1–14, 2019.

L Busino, M Chiesa, GF Draetta, and M Donzelli. Cdc25A phosphatase: Combinatorial phosphorylation, ubiquitylation and proteolysis. *Oncogene*, 23:2050–2056, 2004.

G Carbonetti, T Wilpshaar, J Kroonen, K Studholme, C Converso, S D'Oelsnitz, and M Kaczocha. FABP5 coordinates lipid signaling that promotes prostate cancer metastasis. *Scientific Reports*, 9:18944, 2019.

G Carvalheira, BH Nozima, and JM Cerutti. MicroRNA-106b-mediated down-regulation of C1orf24 expression induces apoptosis and suppresses invasion of thyroid cancer. *Oncotarget*, 6:28357–28370, 2015.

M Cizkova, S Vacher, D Meseure, M Trassard, A Susini, D Mlcuchova, C Callens, E Rouleau, F Spyratos, R Lidereau, and I Bièche. PIK3R1 underexpression is an independent prognostic marker in breast cancer. *BMC Cancer*, 13:1–15, 2013.

A Couture-Beil. *rjson: JSON for R*, 2018. URL `https://CRAN.R-project.org/package=rjson`.

C Coveney, D Boocock, R Rees, S Deen, and G Ball. Data mining of gene arrays for biomarkers of survival in ovarian cancer. *Microarrays*, 4:324–338, 2015.

CV Dang. MYC on the path to cancer. *Cell*, 149:22–35, 2012.

MA D'Angelo, JS Gomez-Cavazos, A Mei, DH Lackner, and MW Hetzer. A change in nuclear pore complex composition regulates cell differentiation. *Developmental Cell*, 22:446–458, 2012.

O Gevaert, F De Smet, D Timmerman, Y Moreau, and B De Moor. Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics*, 22:e184–e190, 2006.

KM Gharpure, S Pradeep, M Sans, R Rupaimoole, C Ivan, SY Wu, E Bayraktar, AS Nagaraja, LS Mangala, X Zhang, M Haemmerle, W Hu, C Rodriguez-Aguayo, M McGuire, CSL Mak, X Chen, MA Tran, A Villar-Prados, GA Pena, R Kondetimmanahalli, R Nini, P Koppula, P Ram, J Liu, G Lopez-Berestein, K Baggerly, LS Eberlin, and AK Sood. FABP4 as a key determinant of metastatic potential of ovarian cancer. *Nature Communications*, 9, 2018.

F Hua, K Li, JJ Yu, XX Lv, J Yan, XW Zhang, W Sun, H Lin, S Shang, F Wang, B Cui, R Mu, B Huang, JD Jiang, and ZW Hu. TRB3 links insulin/IGF to tumour promotion by interacting with p62 and impeding autophagic/proteasomal degradations. *Nature Communications*, 6:7951, 2015.

DW Huang, BT Sherman, and RA Lempicki. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37:1–13, 2009a.

DW Huang, BT Sherman, and RA Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4:44–57, 2009b.

MA Jordan and L Wilson. Microtubules as a target for anticancer drugs. *Nature Reviews Cancer*, 4:253–265, 2004.

H Kanao, T Enomoto, T Kimura, M Fujita, R Nakashima, Y Ueda, Y Ueno, T Miyatake, T Yoshizaki, GS Buzard, A Tanigami, K Yoshino, and Y Murata. Overexpression of LAMP3/TSC403/DC-LAMP promotes metastasis in uterine cervical cancer. *Cancer Research*, 65:8640–8645, 2005.

P Khongkow, U Karunarathna, M Khongkow, C Gong, AR Gomes, E Yagüe, LJ Monteiro, M Kongsema, S Zona, EPS Man, JWH Tsang, RC Coombes, KJ Wu, US Khoo, RH Medema, R Freire, and EWF Lam. FOXM1 targets NBS1 to regulate DNA damage-induced senescence and epirubicin resistance. *Oncogene*, 33:4144–4155, 2014.

A Koussounadis, SP Langdon, DJ Harrison, and VA Smith. Chemotherapy-induced dynamic gene expression changes in vivo are prognostic in ovarian cancer. *British Journal of Cancer*, 110:2975–2984, 2014.

U Laresgoiti, A Apraiz, M Olea, J Mitxelena, M Osinalde, JA Rodriguez, A Fullaondo, and AM Zubiaga. E2F2 and CREB cooperatively regulate transcriptional activity of cell cycle genes. *Nucleic Acids Research*, 41:10185–10198, 2013.

YC Lee, WL Wang, WC Chang, YH Huang, GC Hong, HL Wang, YH Chou, HC Tseng, HT Lee, ST Li, HL Chen, CC Wu, HF Yang, BY Wang, and WW Chang. Tribbles homolog 3 involved in radiation response of triple negative breast cancer cells by regulating Notch1 activation. *Cancers*, 11:127, 2019.

K Leithner, A Hrzenjak, M Trötzmüller, T Moustafa, HC Köfeler, C Wohlkoenig, E Stacher, J Lindenmann, AL Harris, A Olschewski, and H Olschewski. PCK2 activation mediates an adaptive response to glucose depletion in lung cancer. *Oncogene*, 34:1044–1050, 2015.

JW Locasale, AR Grassian, T Melman, CA Lyssiotis, KR Mattaini, AJ Bass, G Heffron, CM Metallo, T Muranen, H Sharfi, AT Sasaki, D Anastasiou, E Mullarky, NI Vokes, M Sasaki, R Beroukhim, G Stephanopoulos, AH Ligon, M Meyerson, AL Richardson, L Chin, G Wagner, JM Asara, JS Brugge, LC Cantley, and MG Vander Heiden. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nature Genetics*, 43:869–874, 2011.

A López-Plana, P Fernández-Nogueira, P Muñoz-Guardiola, S Solé-Sánchez, E Megías-Roda, H Pérez-Montoyo, P Jauregui, M Yeste-Velasco, M Gómez-Ferreria, T Erazo, E Ametller, L Recalde-Percaz, N Moragas-Garcia, A Noguera-Castells, M Mancino, T Morán, E Nadal, J Alfón, C Domènech, P Gascon, J M Lizcano, G Fuster, and P Bragado. The novel proautophagy anticancer drug ABTL0812 potentiates chemotherapy in adenocarcinoma and squamous nonsmall cell lung cancer. *International Journal of Cancer*, 147:1163–1179, 2020.

S Marx, AM Spiegel, MC Skarulis, JL Doppman, FS Collins, and LA Liotta. Multiple endocrine neoplasia type 1: Clinical and genetic topics. *Annals of Internal Medicine*, 129:484–494, 1998.

P Montero and J Vilar. Tsclust: An R package for time series clustering. *Journal of Statistical Software*, 62:1–43, 2014.

G Moreno-Bueno, S Rodríguez-Perales, C Sánchez-Estévez, D Hardisson, D Sarrió, J Prat, JC Cigudosa, X Matias-Guiu, and J Palacios. Cyclin D1 gene (CCND1) mutations in endometrial cancer. *Oncogene*, 22:6115–6118, 2003.

N Pällmann, M Livgård, M Tesikova, H Zeynep Nenseth, E Akkus, J Sikkeland, Y Jin, D Koc, OF Kuzu, M Pradhan, HE Danielsen, N Kahraman, HM Mokhlis, B Ozpolat, PP Banerjee, A Uren, L Fazli, PS Rennie, Y Jin, and F Saatcioglu. Regulation of the unfolded protein response through ATF4 and FAM129A in prostate cancer. *Oncogene*, 38:6301–6318, 2019.

A Polanski, J Polanska, M Jarzab, M Wiench, and B Jarzab. Application of Bayesian networks for inferring cause-effect relations from gene expression profiles of cancer

versus normal cells. *Mathematical Biosciences*, 209:528–546, 2007.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

Y Sang, P Kong, S Zhang, L Zhang, Y Cao, X Duan, T Sun, Z Tao, and W Liu. SGK1 in human cancer: Emerging roles and mechanisms. *Frontiers in Oncology*, 10:2987, 2021.

RL Siegel, KD Miller, and A Jemal. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70:7–30, 2020.

VA Smith, J Yu, TV Smulders, AJ Hartemink, and ED Jarvis. Computational inference of neural information flow networks. *PLoS Computational Biology*, 2:1436–1449, 2006.

M Soda, Young L Choi, M Enomoto, S Takada, Y Yamashita, S Ishikawa, S-i Fujiwara, H Watanabe, K Kurashina, H Hatanaka, M Bando, S Ohno, Y Ishikawa, H Aburatani, T Niki, Y Sohara, Y Sugiyama, and H Mano. Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature*, 448:561–566, 2007.

TM Therneau and PM Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer, 2000.

RW Tothill, AV Tinker, J George, R Brown, SB Fox, S Lade, DS Johnson, MK Trivett, D Etemadmoghadam, B Locandro, N Traficante, S Fereday, JA Hung, Y Eng Chiew, I Haviv, D Gertig, A Defazio, and DDL Bowtell. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical Cancer Research*, 14:5198–5208, 2008.

A Vogogias, J Kennedy, D Archambault, VA Smith, and H Currant. MLCut: Exploring multi-level cuts in dendrograms for biological data. *Proceedings of Computer Graphics and Visual Computing, CGVC*, 2016.

A Vogogias, J Kennedy, D Archambault, B Bach, VA Smith, and H Currant. Bayespiles: Visualisation support for bayesian network structure learning. *ACM Transactions on Intelligent Systems and Technology*, 10:5, 2018.

J Wang, JS Park, Y Wei, M Rajurkar, JL. Cotton, Q Fan, BC Lewis, H Ji, and J Mao. TRIB2 acts downstream of Wnt/TCF in liver cancer cells to regulate YAP and C/EBPα function. *Molecular Cell*, 51:211–225, 2013.

Q Wang, G He, M Hou, L Chen, S Chen, A Xu, and Y Fu. Cell cycle regulation by alternative polyadenylation of CCND1. *Scientific Reports*, 8:6824, 2018.

S Yan, J Fang, Y Chen, Y Xie, S Zhang, X Zhu, and F Fang. Comprehensive analysis of prognostic gene signatures based on immune infiltration of ovarian cancer. *BMC Cancer*, 20:1–17, 2020.

J Yu, VA Smith, PP Wang, AJ Hartemink, and ED Jarvis. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, 20:3594–3603, 2004.

N Zhang, X-M Zhou, F-F Yang, Q Zhang, Y Miao, and G Hou. FAM129A promotes invasion and proliferation by activating FAK signaling pathway in non-small cell lung cancer. *International Journal of Clinical and Experimental Pathology*, 12:893–900, 2019.